

Recursive Polynomial Minimum Mean Square Error Estimation With Applications To Orbit Determination

Simone Servadio¹ and Renato Zanetti²
The University of Texas at Austin, Austin, Texas 78751

This paper presents a systematic generalization of the linear update structure associated with the extended Kalman filter for high order polynomial estimation of non-linear dynamical systems. The minimum mean-square error criterion is used as the cost function to determine the optimal polynomial update during the estimation process. The high order series representation is implemented effectively using differential algebra techniques. Numerical examples show that the proposed algorithm, named **High Order Differential Algebra Kalman Filter**, provides superior robustness and/or mean-square error performance compared to linear estimators under the condition considered.

I. Nomenclature

σ_r	Position Standard Deviation
σ_v	Velocity Standard Deviation
$d\mathbf{X}^-$	Predicted State Deviation Vector
$d\mathbf{X}^+$	Updated State Deviation Vector
$d\mathbf{X}^{++}$	Zero-Mean Updated State Deviation Vector
$d\mathbf{Y}$	Measurement Deviation Vector

¹ Ph.D. Student, Aerospace Engineering and Engineering Mechanics

² Assistant Professor, Aerospace Engineering and Engineering Mechanics

$d\mathbf{Y}^{[2]}$	Kronecker Square of the Measurement Deviation Vector
$d\mathcal{Y}$	Augmented Measurement Deviation Vector
$\mathbf{f}[\cdot]$	Process Model
$\mathbf{h}[\cdot]$	Measurement Function
\mathbf{K}	Kalman Gain
$\mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]}$	State Covariance
$\mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[i]}$	State i -th Central Moment
$\mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[i]-}$	Predicted State i -th Central Moment
$\mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[i]+}$	Updated State i -th Central Moment
$\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}^{[2]-}$	Measurement Covariance
$\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}d\mathbf{Y}}^{[3]-}$	Measurement 3^{rd} Central Moment
$\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}d\mathbf{Y}d\mathbf{Y}}^{[4]-}$	Measurement 4^{rd} Central Moment
$\mathbf{P}_{d\mathcal{Y}d\mathcal{Y}}^{[2]-}$	Augmented Measurement Covariance
$\mathbf{Q}^{[2]}$	Covariance of the Process Noise
$\mathbf{Q}^{[i]}$	Process Noise i -th Central Moment
\mathbf{r}	Spacecraft Position Vector
$\mathbf{R}^{[2]}$	Covariance of the Measurement Noise
$\mathbf{R}^{[i]}$	Measurement Noise i -th Central Moment
\mathbf{v}	Spacecraft Velocity Vector
\mathbf{V}	Process Noise Vector
\mathbf{W}	Measurement Noise Vector
$\hat{\mathbf{x}}$	Estimated State
$\hat{\mathbf{x}}^-$	Predicted State
$\hat{\mathbf{x}}^+$	Updated State
\mathbf{X}	State Vector of the System
\mathbf{y}	Numerical Value of the Measurements
$\hat{\mathbf{y}}$	Predicted Measurement Mean
\mathbf{Y}	Measurement Vector

II. Introduction

The optimal solution to the sequential stochastic Minimum Mean Square Error (MMSE) estimation problem is well known: the optimal estimate is the conditional mean. The optimal solution is obtained from the conditional probability density function which is calculated recursively with Bayes's rule. For the linear/Gaussian case, the conditional distribution remains Gaussian at all times and hence fully described by its mean and covariance matrix that can be calculated using Kalman's technique [1, 2]. For practical nonlinear/non-Gaussian problems, such as orbit determination [3], the analytical solution is usually not available in closed form.

The most widely used nonlinear estimator is the extended Kalman filter (EKF) [4]. The EKF linearizes the estimation error around the most current estimate and applies the Kalman filter equations to this linearized system. It has been shown, however, that nonlinearities of the orbit determination problem can make the linearization assumption insufficient to represent the actual uncertainty [5]. An alternative to linearization around the mean is stochastic linearization in the so-called unscented Kalman filter (UKF) [6, 7]. The UKF is typically more robust than the EKF as it is able to better handle the effects of nonlinearities in the dynamics and in the measurements. The UKF is a linear estimator, i.e. the estimate is a linear function of the current measurement.

In some cases, the uncertainty associated with orbital mechanics can be propagated analytically [8]; however these analytical solutions usually do not include perturbations other than J_2 , nor process noise or measurement updates. Park and Scheeres [9] use state transition tensors (STT) to propagate mean and higher order central moments through arbitrary nonlinear dynamics. They subsequently expand their work to create higher order Kalman filters able to handle process noise and measurement updates [10], however their filter is a linear estimator, i.e. the state estimate is a linear function of the current measurement. In Ref. [10] Park and Scheeres only update the mean and covariance when a measurement becomes available, neglecting the contribution the measurement update has on the higher order central moments. Majji, Turner, and Junkins [11], on the other hand, introduce a tensorial mechanization that expands the work by Park and Scheeres to include

measurement updates in all the higher order central moments. Valli *et al.* [12] efficiently replicate these results using differential algebra (DA) techniques.

The EKF, UKF, and the higher order filter proposed by Majji *et al.* utilize a linear update, while it is known that a nonlinear update provides better performance in the nonlinear/non-Gaussian case. The globally optimal update is the conditional mean, which is given by some nonlinear function of the measurement; closed-form calculation of this nonlinear function is usually not tractable. One approach to approximate the optimal nonlinear update is breaking the estimation error distribution in many smaller Gaussian components such that each is small enough to satisfy the linearization assumption of an EKF; this is the basis of the Gaussian sum filter [13, 14]. Another approach is to approximate the nonlinear function with a Taylor series [4]. Truncating this series to first order produces the EKF. Generally, the higher the order of the Taylor series the better the performance of the filter. Truncating the Taylor series to order N requires knowledge of the estimation error's central moments up to order $2N$. For example the EKF truncates to first order and requires knowledge of the covariance. A second order filter requires knowledge of central moments up to order four. To avoid carrying third and fourth order central moments, the Gaussian second order filter (GSOF) [15] approximates them assuming the distribution is Gaussian and constructs them from the covariance matrix.

Calculating the higher order gains is no trivial matter, although it is an operation that is required only once. The GSOF avoids this operation all together by producing a linear update based on a second order approximation of the posterior estimation error. To avoid these calculations, De Santis *et al.* [16] use an augmented state to obtain a nonlinear update but preserving the linear update structure. Their original work focuses on linear but non-Gaussian systems, and on approximations of the optimal non-linear update as either quadratic [16] or polynomial [17]. While maintaining the exact same structure as the traditional Kalman filter, this approach takes the nonlinearity of the measurement update into consideration by implicitly accounting for high order moments in the estimation process. The augmented state is comprised of the high order raw moments. In the presence of non-linear dynamics and measurements, these nonlinear functions are approximated via the Carleman approximation [18], which requires the evaluation of Kronecker powers of the state

and outputs of the systems. Every smooth map can be Carleman-approximated by a composition of homomorphic maps. In a Kalman-filter-type implementation, both the Kronecker powers of the state and their covariance matrices need to be carried.

This work introduces a novel nonlinear recursive filter, named "Higher Order Differential Algebra Kalman Filter" (HODAKF), that performs a polynomial update of arbitrary order N and carries an arbitrary number of central moments $M \geq 2N$. The polynomial-update coefficients are chosen to minimize the mean square estimation error and require knowledge of high order moments of the distribution. The mean and central moments are propagated following the work of Park and Scheeres [9]. The polynomial update is obtained using an augmented measurement approach; however, unlike Ref. [18], the Carleman approximation is not made and hence the state vector is not augmented and all the central moments are updated independently. The algorithm is numerically tested in an orbit determination example.

III. Background

This work utilizes a differential algebra (DA) computer tool for the implementation of the proposed nonlinear system. DA operates on Taylor polynomials with operations such as sums and product of functions, as well as scalar products with real numbers; leading to the so-called truncated power series algebra [19] [20]. Function operations such as composition, inversion (when it exists), and explicit solutions of nonlinear systems of equations, and other common elementary functions are also available [21] [22]. In addition to these algebraic operations, the DA framework is endowed with differentiation and integration operators, thus completing the requisites of a differential algebra structure.

The Taylor series coefficients are obtained up to an arbitrary and specified order c from a library of known Taylor series for elementary functions; no numerical derivatives, e.g. finite differences, are used. For a more detailed explanation of Differential Algebra, refer to Refs. [23–25], and for a deeper analysis of the computer environment used to implement the examples of this paper Refs. [26, 27].

This work utilizes the Kronecker product \otimes and the function $\text{vect}(\mathbf{M})$ that vectorizes matrix \mathbf{M} by placing all its columns on top of each other. Throughout the work, random vectors are denoted

by an upper case, while their outcome are denoted with lower case letters. For example, \mathbf{Y} is the observed random vector and \mathbf{y} is the outcome, i.e. the actual measured value.

Out of different possible stochastic estimators (maximum likelihood, maximum a posteriori) this work concentrates on Minimum Mean Square Error Estimation (MMSE) whose optimal solution is the conditional mean [28], the conditional distribution and the MMSE solution can be computed recursively when the measurement and process noise are white. Generally, the conditional covariance of the state given the measurements and the total covariance of the estimation error are different. For jointly Gaussian random vectors \mathbf{X} and \mathbf{Y} however, the two coincide, which means that the covariance of \mathbf{X} conditioned on \mathbf{Y} is not a function of \mathbf{Y} , that is to say, it is the same regardless of the outcome \mathbf{y} of \mathbf{Y} .

Computing the full nonlinear MMSE estimator (the conditional mean) is usually a computationally heavy endeavor, and it is fairly common to settle for the Linear Minimum Mean Square Error Estimation (LMMSE) given by

$$\hat{\mathbf{X}} = \mathbb{E}_{\mathbf{X}} \{\mathbf{X}\} - \Sigma_{\mathbf{X}\mathbf{Y}} \Sigma_{\mathbf{Y}\mathbf{Y}}^{-1} (\mathbf{Y} - \mathbb{E}_{\mathbf{Y}} \{\mathbf{Y}\}) \quad (1)$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator, $\Sigma_{\mathbf{X}\mathbf{Y}}$ is the joint covariance matrix of \mathbf{X} and \mathbf{Y} , and $\Sigma_{\mathbf{Y}\mathbf{Y}}$ is the covariance matrix of \mathbf{Y} . Generally speaking, the LMMSE estimate cannot be calculated recursively. However, for linear systems (either Gaussian or not) the LMMSE is the Kalman filter and it is recursive. Once again, the key assumption is that all noises are white and independent from all other distributions.

IV. Polynomial MMSE Estimation

It is possible to expand the concept of Linear MMSE to higher order polynomials [17]. Any polynomial function satisfies the conditions of the orthogonality principle, namely closure under addition and scalar multiplication, hence it is possible to calculate an optimal polynomial update using the orthogonality principle [29].

For LMMSE, knowledge of the first two moments (mean and covariance) is needed. In general, for a polynomial update of order p , knowledge of the first $2p$ moments is needed. It is possible to use the Kronecker product \otimes to rewrite polynomial updates as linear ones, e.g. for a quadratic update

($p = 2$), we have that

$$\mathbf{g}(\mathbf{Y}) = \mathbf{a} + \mathbf{K}^{aug} \begin{bmatrix} \mathbf{Y} \\ \mathbf{Y} \otimes \mathbf{Y} \end{bmatrix} = \mathbf{a} + \mathbf{K}^{aug} \mathbf{Y}^{aug} \quad (2)$$

where \mathbf{Y}^{aug} is an augmented measurement vector that includes both \mathbf{Y} and its square, this approach reduces the quadratic update to a Kalman-filter-like update

$$\hat{\mathbf{X}} = \mathbb{E}_{\mathbf{X}} \{\mathbf{X}\} + \Sigma_{\mathbf{X}\mathbf{Y}^{aug}} \Sigma_{\mathbf{Y}^{aug}\mathbf{Y}^{aug}}^{-1} (\mathbf{Y}^{aug} - \mathbb{E}_{\mathbf{Y}} \{\mathbf{Y}^{aug}\}) \quad (3)$$

unfortunately, for non-scalar measurements, the elements of the augmented measurement \mathbf{Y}^{aug} are not unique, hence $\Sigma_{\mathbf{Y}^{aug}\mathbf{Y}^{aug}}$ is not invertible. This is easily alleviated by removing the duplicate elements from \mathbf{Y}^{aug} . Even when the dynamics and the measurement \mathbf{Y} are linear, the system ceases to be linear ($\mathbf{Y} \otimes \mathbf{Y}$ is quadratic in \mathbf{X}), therefore it also ceases to be recursive. By applying the equations recursively nevertheless, the result is a suboptimal estimate that is not the true quadratic MMSE estimate.

For any vector \mathbf{v} , define $\mathbf{v}^{[2]} = \mathbf{v} \otimes \mathbf{v}$. In the case of linear measurements

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W} \quad (4)$$

Ref. [16] proposes to reinstate the linearity by also augmenting the state vector

$$\mathbf{X}^{aug} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X} \otimes \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}^{[2]} \end{bmatrix} \quad (5)$$

hence

$$\mathbf{Y}^{aug} = \begin{bmatrix} \mathbf{H} & \mathbf{O} \\ \mathbf{O} & \mathbf{H} \otimes \mathbf{H} \end{bmatrix} \mathbf{X}^{aug} + \begin{bmatrix} \mathbf{W} \\ (\mathbf{H}\mathbf{X}) \otimes \mathbf{W} + \mathbf{W} \otimes (\mathbf{H}\mathbf{X}) + \mathbf{W}^{[2]} \end{bmatrix} = \mathbf{H}^{aug} \mathbf{X}^{aug} + \mathbf{W}^{aug} \quad (6)$$

and

$$\mathbb{E}_{\mathbf{Y}} \{\mathbf{Y}^{aug}\} = \mathbf{H}^{aug} \mathbb{E}_{\mathbf{X}} \{\mathbf{X}^{aug}\} + \begin{bmatrix} \mathbf{0} \\ \text{vect}(\mathbf{R}) \end{bmatrix} \quad (7)$$

$$\Sigma_{\mathbf{X}^{aug}\mathbf{Y}^{aug}} = \mathbf{P}^{aug} (\mathbf{H}^{aug})^T \quad (8)$$

$$\Sigma_{\mathbf{Y}^{aug}\mathbf{Y}^{aug}} = \mathbf{H}^{aug} \mathbf{P}^{aug} (\mathbf{H}^{aug})^T + \mathbf{R}^{aug} \quad (9)$$

$$\mathbf{P}^{aug} = \begin{bmatrix} \Sigma_{\mathbf{X}\mathbf{X}} & \Sigma_{\mathbf{X}\mathbf{X}^{[2]}} \\ \Sigma_{\mathbf{X}^{[2]}\mathbf{X}} & \Sigma_{\mathbf{X}^{[2]}\mathbf{X}^{[2]}} \end{bmatrix} \quad (10)$$

where

$$\mathbf{W}^{aug} = \begin{bmatrix} \mathbf{W} \\ \mathbf{W}^{[2]} \end{bmatrix} \quad (11)$$

$$\mathbf{R}^{aug} = \begin{bmatrix} \mathbf{R} & \mathbf{R}_{12}^{aug} \\ \mathbf{R}_{21}^{aug} & \mathbf{R}_{22}^{aug} \end{bmatrix} \quad (12)$$

$$\mathbf{R}_{21}^{aug} = (\mathbf{H} \otimes \mathbf{I}) (\mathbb{E}\{\mathbf{X}\} \otimes \mathbf{R}) + (\mathbf{I} \otimes \mathbf{H}) (\mathbf{R} \otimes \mathbb{E}\{\mathbf{X}\}) + \mathbb{E}\{\mathbf{W}^{[2]} \mathbf{W}^T\} \quad (13)$$

$$\mathbf{R}_{12}^{aug} = (\mathbf{R}_{21}^{aug})^T \quad (14)$$

$$\begin{aligned} \mathbf{R}_{22}^{aug} &= (\mathbf{H} \otimes \mathbf{I}) (\mathbb{E}\{\mathbf{X}\mathbf{X}^T\} \otimes \mathbf{R})(\mathbf{H} \otimes \mathbf{I})^T + (\mathbf{I} \otimes \mathbf{H}) (\mathbf{R} \otimes \mathbb{E}\{\mathbf{X}\mathbf{X}^T\})(\mathbf{I} \otimes \mathbf{H})^T \\ &\quad + \mathbb{E}\{\mathbf{W}^{[2]}(\mathbf{W}^{[2]})^T\} - \text{vect}(\mathbf{R}) \text{vect}(\mathbf{R})^T \end{aligned} \quad (15)$$

To simplify the above equations, Ref. [16] assumes \mathbf{X} to be zero mean. Therefore the algorithm carries a propagated mean of \mathbf{X} , together with an augmented estimate $\hat{\mathbf{X}}$ of the deviation from the mean and its augmented covariance matrix. The following simplifications occur

$$\mathbf{R}_{21}^{aug} = \mathbb{E}\{\mathbf{W}^{[2]} \mathbf{W}^T\} \quad (16)$$

$$\begin{aligned} \mathbf{R}_{22}^{aug} &= (\mathbf{H} \otimes \mathbf{I}) (\boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \otimes \mathbf{R})(\mathbf{H} \otimes \mathbf{I})^T + (\mathbf{I} \otimes \mathbf{H}) (\mathbf{R} \otimes \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}})(\mathbf{I} \otimes \mathbf{H})^T + \boldsymbol{\Sigma}_{\mathbf{W}^{[2]} \mathbf{W}^{[2]}} \\ &= \mathbf{H} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \mathbf{H}^T \otimes \mathbf{R} + \mathbf{R} \otimes \mathbf{H} \boldsymbol{\Sigma}_{\mathbf{X}\mathbf{X}} \mathbf{H}^T + \boldsymbol{\Sigma}_{\mathbf{W}^{[2]} \mathbf{W}^{[2]}} \end{aligned} \quad (17)$$

This quadratic update is not recursive because the standard Kalman filter assumptions are not met as the “ \mathbf{R} ” matrix is a function of the “ \mathbf{P} ” matrix. Furthermore, in the linear case the recursive LMMSE uses $\mathbb{E}\{\mathbf{X}\}$ only at initialization, after that the mean of \mathbf{X} is replaced by $\hat{\mathbf{x}}$, the LMMSE estimate, which is not the mean. It is therefore not clear that $\mathbb{E}\{\mathbf{X}\} = \mathbf{0}$ should result in a simplification of Eq. (13) at any step other than the very first. In computing the measurement covariance with Eq. (15), Ref. [16] certainly replaces $\mathbb{E}\{\mathbf{X}\mathbf{X}^T\}$ with the recursively updated covariance of the LMMSE; not the un-updated covariance of \mathbf{X} . It therefore seems that Eq. (13) and Eq. (15) are treated differently in that $\mathbb{E}\{\mathbf{X}\}$ is always the mean of the true state, unchanged by the successive measurements being incorporated, while $\mathbb{E}\{\mathbf{X}\mathbf{X}^T\}$ is the spread around the true state which is updated and reduced as more measurements become available. Given these observations, we propose a different solution to this same problem.

A. Suboptimal Recursive LMMSE

The optimal MMSE estimate is the conditional mean of $\mathbf{X}|\mathbf{Y}_k\mathbf{Y}_{k-1}\dots\mathbf{Y}_1$; in its recursive formulation we start from the PDF of $\mathbf{X}|\mathbf{Y}_{k-1}\dots\mathbf{Y}_1$ and apply Bayes' rule to obtain $\mathbf{X}|\mathbf{Y}_k\mathbf{Y}_{k-1}\dots\mathbf{Y}_1$. A similar recursion holds exactly for LMMSE estimation of linear systems. It is common, however, to use a similar approach in recursive LMMSE estimation for nonlinear problems (unscented Kalman filter [6, 7], quadrature Kalman filter [30], cubature Kalman filter [31], etc). This approach is a suboptimal filter, because the LMMSE is not recursive in the presence of nonlinear measurements and/or nonlinear dynamics. To discuss this approach, we will once again assume a static estimation problem; the conclusions drawn do not change when state dynamics are present in-between measurements.

Let $\hat{\mathbf{x}}_0 = \mathbb{E}_{\mathbf{X}}\{\mathbf{X}\}$. After the first measurement is incorporated, we have

$$\hat{\mathbf{X}}_1 = \hat{\mathbf{x}}_0 + \Sigma_{\mathbf{X}\mathbf{Y}_1}\Sigma_{\mathbf{Y}_1\mathbf{Y}_1}^{-1}(\mathbf{Y}_1 - \mathbb{E}_{\mathbf{Y}_1}\{\mathbf{Y}_1\}) \quad (18)$$

$$\mathbf{E}_1 = \mathbf{X} - \hat{\mathbf{X}}_1 \quad (19)$$

$$\mathbf{P}_1 = \mathbb{E}_{\mathbf{E}_1}\{\mathbf{E}_1\mathbf{E}_1^T\} = \mathbb{E}_{\mathbf{X}\mathbf{Y}_1}\{(\mathbf{X} - \hat{\mathbf{X}}_1)(\mathbf{X} - \hat{\mathbf{X}}_1)^T\} = \Sigma_{\mathbf{X}\mathbf{X}} - \Sigma_{\mathbf{X}\mathbf{Y}_1}\Sigma_{\mathbf{Y}_1\mathbf{Y}_1}^{-1}\Sigma_{\mathbf{Y}_1\mathbf{X}}^T \quad (20)$$

where the linear estimator $\hat{\mathbf{X}}_1$ is a random vector and its outcome (the estimate) is

$$\hat{\mathbf{x}}_1 = \mathbb{E}_{\mathbf{X}}\{\mathbf{X}\} + \Sigma_{\mathbf{X}\mathbf{Y}_1}\Sigma_{\mathbf{Y}_1\mathbf{Y}_1}^{-1}(\mathbf{y}_1 - \mathbb{E}_{\mathbf{Y}_1}\{\mathbf{Y}_1\}) \quad (21)$$

where \mathbf{y}_1 is the actual measurement, i.e. the numerical value or the outcome of the random vector \mathbf{Y}_1 as read from the sensor.

The recursion is obtained using the same LMMSE equations to process the second measurement \mathbf{Y}_2 by creating a new prior \mathbf{X}_1 centered at $\hat{\mathbf{x}}_1$

$$\mathbf{X}_1 = \hat{\mathbf{x}}_1 + \mathbf{E}_1 \quad (22)$$

Notice that \mathbf{X}_1 is neither \mathbf{X} nor, in general, $\mathbf{X}|\mathbf{Y}_1$. The approach taken is to create a recursive algorithm by replacing the random vector \mathbf{X} with one that includes the knowledge gained from the linear inclusion of \mathbf{Y}_1 . However, $\mathbf{X}_1 \neq \mathbf{X}|\mathbf{Y}_1$ (with the notable exception of the linear Gaussian case). In this approach the mean of \mathbf{X}_1 , i.e. $\hat{\mathbf{x}}_1$, is taken as deterministic, while in reality $\hat{\mathbf{x}}_1$ is a function of the outcome of random vector \mathbf{Y}_1 and hence stochastic. The spread of \mathbf{X}_1 , i.e. \mathbf{E}_1 , on

the other hand, includes the randomness of \mathbf{Y}_1 .

$$\mathbb{E}_{\mathbf{X}_1} \{\mathbf{X}_1\} = \hat{\mathbf{x}}_1 \quad (23)$$

$$\Sigma_{\mathbf{X}_1\mathbf{X}_1} = \Sigma_{\mathbf{E}_1\mathbf{E}_1} = \mathbf{P}_1 = \Sigma_{\mathbf{X}\mathbf{X}} - \Sigma_{\mathbf{X}\mathbf{Y}_1} \Sigma_{\mathbf{Y}_1\mathbf{Y}_1}^{-1} \Sigma_{\mathbf{X}\mathbf{Y}_1}^T \quad (24)$$

$$\hat{\mathbf{X}}_2 = \hat{\mathbf{x}}_1 + \Sigma_{\mathbf{X}_1\mathbf{Y}_2} \Sigma_{\mathbf{Y}_2\mathbf{Y}_2}^{-1} (\mathbf{Y}_2 - \mathbb{E}_{\mathbf{Y}_2} \{\mathbf{Y}_2\}) \quad (25)$$

$$\mathbf{E}_2 = \mathbf{X} - \hat{\mathbf{X}}_2 \quad (26)$$

$$\mathbf{P}_2 = \mathbb{E}_{\mathbf{E}_2} \{\mathbf{E}_2 \mathbf{E}_2^T\} = \mathbb{E}_{\mathbf{X}\mathbf{Y}_1\mathbf{Y}_2} \{(\mathbf{X} - \hat{\mathbf{X}}_2)(\mathbf{X} - \hat{\mathbf{X}}_2)^T\} = \Sigma_{\mathbf{X}_1\mathbf{X}_1} - \Sigma_{\mathbf{X}_1\mathbf{Y}_2} \Sigma_{\mathbf{Y}_2\mathbf{Y}_2}^{-1} \Sigma_{\mathbf{X}_1\mathbf{Y}_2}^T \quad (27)$$

and so on for subsequent measurements.

Assuming all means and covariance matrices can be calculated exactly (which is usually not possible for nonlinear systems), this approach produces a consistent estimator that reduces the mean-squared estimation error (\mathbf{P}_k) with each subsequent measurement incorporated, but it performs worse, in a MMSE sense, than processing all measurements at once with the optimal LMMSE. This concept is exemplified in the following section.

For a general distribution and an arbitrary nonlinear measurement, it is usually impossible to exactly calculate all mean and covariance matrices in the above equations. To overcome this difficulty, different linear estimators approximate them in different ways. The extended Kalman filter (EKF) uses a first order Taylor series expansion centered at the current estimate while the gaussian second order filter approximates the same Taylor series to second order and calculates third and fourth-order central moments of \mathbf{X}_1 as if the distribution was Gaussian. The unscented Kalman filter (UKF) approximates the moments with a set of deterministic regression points and belongs to the family of estimators that use statistical linear regression. Other linear filters in the same family as the UKF are the central difference filter [32], the first and second order divided difference filters [33], the quadrature Kalman filter [30] and the cubature Kalman filter [31].

B. Example: Recursive LMMSE of non-linear measurement

Define two independent Gaussian random variables as $Z \sim \mathcal{N}(m_z, \sigma_{zz}^2)$ and $W \sim \mathcal{N}(0, \sigma_{ww}^2)$. The notation $\mathcal{N}(m, \sigma^2)$ indicates a Gaussian random variable with mean m and variance σ^2 . The random variable to be estimated is $X = Z + W$, therefore X is Gaussian with mean $m_x = m_z$ and

variance $\sigma_{xx}^2 = \sigma_{zz}^2 + \sigma_{ww}^2$. Suppose the observed random variable is $Y = (X - W)^3$, therefore the measurement is nonlinear and W is interpreted as measurement noise. Notice that $Y = Z^3$ therefore

$$m_y = \mathbb{E}\{Z^3\} = m_z^3 + 3m_z \sigma_{zz}^2 \quad (28)$$

$$\begin{aligned} \sigma_{yy}^2 &= \mathbb{E}\{Y^2\} - m_y^2 = \mathbb{E}\{Z^6\} - (m_z^3 + 3m_z \sigma_{zz}^2)^2 \\ &= 9m_z^4 \sigma_{zz}^2 + 36m_z^2 \sigma_{zz}^4 + 15\sigma_{zz}^6 \end{aligned} \quad (29)$$

$$\begin{aligned} \sigma_{xy}^2 &= \mathbb{E}\{XY\} - m_x m_y = \mathbb{E}\{Z^4 + Z^3 W\} - m_z(m_z^3 + 3m_z \sigma_{zz}^2) \\ &= 3m_z^2 \sigma_{zz}^2 + 3\sigma_{zz}^4. \end{aligned} \quad (30)$$

the estimation error variance is

$$\begin{aligned} \sigma_{ee}^2 &= \mathbb{E}_{xy}\{(x - \hat{x})^2\} = \sigma_{xx}^2 - \frac{\sigma_{xy}^4}{\sigma_{yy}^2} \\ &= \frac{18m_z^2 \sigma_{zz}^2 + 6\sigma_{zz}^4}{9m_z^4 + 36m_z^2 \sigma_{zz}^2 + 15\sigma_{zz}^4} \sigma_{zz}^2 + \sigma_{ww}^2. \end{aligned} \quad (31)$$

Let $m_z = 0$, $\sigma_{zz}^2 = 1$, $\sigma_{ww}^2 = 1$ and make the measurement a vector:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} (X - W)^3 \\ X^3 + V \end{bmatrix} = \begin{bmatrix} Z^3 \\ Z^3 + 3Z^2 w + 3ZW^2 + W^3 + V \end{bmatrix} \quad (32)$$

where $V \sim \mathcal{N}(0, 1)$ is independent from X and from W .

$$\mathbb{E}\{\mathbf{Y}\} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (33)$$

$$\mathbf{P}_{\mathbf{Y}\mathbf{Y}} = \begin{bmatrix} 15 & 24 \\ 24 & 121 \end{bmatrix} \quad (34)$$

$$\mathbf{P}_{X\mathbf{Y}} = \begin{bmatrix} 3 & 12 \end{bmatrix} \quad (35)$$

The LMMSE is

$$\hat{X} = m_x + \mathbf{P}_{X\mathbf{Y}} \mathbf{P}_{\mathbf{Y}\mathbf{Y}}^{-1} (\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\}) = \begin{bmatrix} 75 & 108 \end{bmatrix} \frac{\mathbf{Y}}{1239} \quad (36)$$

the estimation error variance is

$$\sigma_{ee}^2 = \sigma_{xx}^2 - \sigma_{\hat{x}\hat{x}}^2 = 2 - \frac{1521}{1239} = 0.7724 \quad (37)$$

If we process only Y_1

$$\hat{X}_1 = \frac{3}{15}Y_1 = \frac{1}{5}Y_1 \quad (38)$$

$$\hat{x}_1 = \frac{1}{5}y_1 \quad (39)$$

$$E_1 = X - \hat{X}_1 = Z + W - \frac{1}{5}Z^3 \quad (40)$$

$$\sigma_{e_1 e_1}^2 = \sigma_{xx}^2 - \sigma_{\hat{x}_1 \hat{x}_1}^2 = 2 - \frac{15}{25} = \frac{7}{5} \quad (41)$$

$$\mathbb{E}\{E_1^2\} = \sigma_{e_1 e_1}^2 \quad (42)$$

$$\mathbb{E}\{E_1^3\} = \mathbb{E}\{E_1^5\} = 0 \quad (43)$$

$$\mathbb{E}\{E_1^4\} = \frac{1024}{125} \quad (44)$$

$$\mathbb{E}\{E_1^6\} = \frac{263397}{625} \quad (45)$$

we now process Y_2 , but we start from a prior distributed as $X_1 = \hat{x}_1 + E_1$. We take $\mathbb{E}\{X_1\} = \hat{x}_1$, i.e. the mean is a function of the measurement outcome y_1 and central moments to coincide with the central moments of E_1 , indeed they are a function of both random variables X and Y_1 .

$$\begin{aligned} \hat{y}_2 &= \mathbb{E}\{X_1^3 + V\} = \mathbb{E}\{(\hat{x}_1 + E_1)^3 + V\} = \hat{x}_1^3 + 3\hat{x}_1^2 \mathbb{E}\{E_1\} + 3\hat{x}_1 \mathbb{E}\{E_1^2\} + \mathbb{E}\{E_1^3\} + \mathbb{E}\{V\} \\ &= \hat{x}_1^3 + \frac{21}{5}\hat{x}_1 \end{aligned} \quad (46)$$

$$\begin{aligned} \sigma_{x_1 y_2}^2 &= \mathbb{E}\{X_1^4 + VX_1\} - \hat{x}_1 \hat{y}_2 = \mathbb{E}\{\hat{x}_1^4 + 4\hat{x}_1^3 e_1 + 6\hat{x}_1^2 e_1^2 + 4\hat{x}_1 e_1^3 + e_1^4 + V\hat{x}_1 + Ve_1\} - (\hat{x}_1^4 + \frac{21}{5}\hat{x}_1^2) \\ &= 6\hat{x}_1^2 \mathbb{E}\{e_1^2\} + \mathbb{E}\{e_1^4\} - \frac{21}{5}\hat{x}_1^2 = \frac{42}{5}\hat{x}_1^2 + \frac{1024}{125} - \frac{21}{5}\hat{x}_1^2 = \frac{21}{5}\hat{x}_1^2 + \frac{1024}{125} \end{aligned} \quad (47)$$

$$\begin{aligned} \sigma_{y_2 y_2}^2 &= \mathbb{E}\{X_1^6 + 2VX_1^3 + V^2\} - \hat{y}_2^2 \\ &= \mathbb{E}\{\hat{x}_1^6 + 6\hat{x}_1^5 e_1 + 15\hat{x}_1^4 e_1^2 + 20\hat{x}_1^3 e_1^3 + 15\hat{x}_1^2 e_1^4 + 6\hat{x}_1 e_1^5 + e_1^6 + V^2\} - \hat{x}_1^6 - \frac{42}{5}\hat{x}_1^4 - \frac{21^2}{25}\hat{x}_1^2 \\ &= 15\hat{x}_1^4 \mathbb{E}\{e_1^2\} + 15\hat{x}_1^2 \mathbb{E}\{e_1^4\} + \mathbb{E}\{e_1^6\} + 1 - \frac{42}{5}\hat{x}_1^4 - \frac{21^2}{25}\hat{x}_1^2 = \frac{63}{5}\hat{x}_1^4 + \frac{2631}{25}\hat{x}_1^2 + \frac{263397}{625} + 1 \end{aligned} \quad (48)$$

and

$$\hat{X}_2 = \hat{x}_1 + \frac{\sigma_{x_1 y_2}^2}{\sigma_{y_2 y_2}^2}(Y_2 - \hat{y}_2) \quad (49)$$

$$\hat{x}_2 = \hat{x}_1 + \frac{\sigma_{x_1 y_2}^2}{\sigma_{y_2 y_2}^2}(y_2 - \hat{y}_2) \quad (50)$$

The predicted estimation error variance is

$$\sigma_{e_2 e_2}^2 = \sigma_{x_1 x_1}^2 - \frac{(\sigma_{x_1 y_2}^2)^2}{\sigma_{y_2 y_2}^2} \quad (51)$$

the variance $\sigma_{e_2 e_2}^2$ is a function of \hat{x}_1 and hence of y_1 ; which is typical of nonlinear systems in which the evolution of the uncertainty is a function of the measurements' outcome history. For example, in the EKF, the measurement Jacobians are evaluated at the estimated state which is a function of the measurement history. The performance index is the mean-squared error $\mathbb{E}_{XY_1Y_2} \{E_2^2\}$; we can calculate its value performing Monte Carlo analysis and dispersing all random variables

$$E_2 = X - \hat{X}_2 = X - \hat{X}_1 - \frac{\sigma_{x_1 y_2}^2}{\sigma_{y_2 y_2}^2} (Y_2 - \hat{X}_1^3 + \frac{21}{5} \hat{X}_1) \quad (52)$$

$$= X - \hat{X}_2 = X - \frac{1}{5} Y_1 - \frac{\sigma_{x_1 y_2}^2}{\sigma_{y_2 y_2}^2} (Y_2 - \frac{1}{125} Y_1^3 + \frac{21}{25} Y_1) \quad (53)$$

where again $\sigma_{x_1 y_2}^2$ and $\sigma_{y_2 y_2}^2$ are also functions of Y_1 . From the Monte Carlo analysis we obtain that

$$\sigma_{e_2 e_2}^2 = \mathbb{E}_{XY_1Y_2} \{E_2^2\} = 1.085 \quad (54)$$

which is clearly suboptimal since is greater than the optimal LMMSE value of 0.7724 from Eq. (37).

V. Nonlinear Updates with Polynomial Residuals

While the update procedure proposed in this paper can be expanded to polynomials of all orders, the quadratic update is shown. Let's rewrite the quadratic update as

$$\mathbf{g}(\mathbf{Y}) = \mathbf{a} + \mathbf{B}\mathbf{Y} + \mathbf{C}\mathbf{Y} \otimes \mathbf{Y} \quad (55)$$

where it is understood that the redundant components of $\mathbf{Y} \otimes \mathbf{Y}$ are eliminated. Without any loss of generality we can redefine \mathbf{a} , \mathbf{B} , and \mathbf{C} by adding and subtracting constants in order to obtain a different, but equivalent, family of quadratic estimators

$$\mathbf{g}(\mathbf{Y}) = \mathbf{a} + \mathbb{E}\{\mathbf{X}\} + \mathbf{B}(\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\}) + \mathbf{C}[(\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\}) \otimes (\mathbf{Y} - \mathbb{E}\{\mathbf{Y}\})] \quad (56)$$

the quantity $d\mathbf{Y} = \mathbf{Y} - \mathbb{E}\{\mathbf{Y}\}$ is usually referred to as the measurement residual. Similarly, we define state deviation from the mean as $d\mathbf{X} = \mathbf{X} - \mathbb{E}\{\mathbf{X}\}$. We know that the optimal values of the estimator's coefficients (denoted with an asterisk) satisfy the orthogonality principle

$$\mathbb{E} \left\{ \left(d\mathbf{X} - \mathbf{a}^* - \mathbf{B}^* d\mathbf{Y} - \mathbf{C}^* (d\mathbf{Y} \otimes d\mathbf{Y}) \right) \left(\mathbf{a} + \mathbf{B}d\mathbf{Y} + \mathbf{C} (d\mathbf{Y} \otimes d\mathbf{Y}) \right)^T \right\} = \mathbf{O}, \quad \forall \mathbf{a}, \mathbf{B}, \mathbf{C} \quad (57)$$

therefore the optimal coefficients are found solving the following linear system

$$\mathbf{a}^* + \mathbf{C}^* \mathbb{E} \{ (d\mathbf{Y} \otimes d\mathbf{Y}) \} = \mathbf{0} \quad (58)$$

$$\mathbf{B}^* \mathbf{P}_{d\mathbf{Y}d\mathbf{Y}} + \mathbf{C}^* \mathbf{P}_{d\mathbf{Y}^{[2]}d\mathbf{Y}} = \mathbf{P}_{d\mathbf{X}d\mathbf{Y}} \quad (59)$$

$$\mathbf{a}^* \mathbf{p}^T + \mathbf{B}^* \mathbf{P}_{d\mathbf{Y}d\mathbf{Y}^{[2]}} + \mathbf{C}^* \mathbf{P}_{d\mathbf{Y}^{[2]}d\mathbf{Y}^{[2]}} = \mathbf{P}_{d\mathbf{X}d\mathbf{Y}^{[2]}} \quad (60)$$

with the following definition for \mathbf{p}

$$\mathbf{p} = \text{vect}(\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}) \quad (61)$$

where the redundant terms of $(\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}})$ are removed. This results in

$$\mathbf{a}^* = -\mathbf{C}^* \mathbf{p} \quad (62)$$

$$\begin{bmatrix} \mathbf{B}^* & \mathbf{C}^* \end{bmatrix} \begin{bmatrix} \mathbf{P}_{d\mathbf{Y}d\mathbf{Y}} & \mathbf{P}_{d\mathbf{Y}^{[2]}d\mathbf{Y}} \\ \mathbf{P}_{d\mathbf{Y}d\mathbf{Y}^{[2]}} & \mathbf{P}_{d\mathbf{Y}^{[2]}d\mathbf{Y}^{[2]}} - \mathbf{p}\mathbf{p}^T \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{d\mathbf{X}d\mathbf{Y}} & \mathbf{P}_{d\mathbf{X}d\mathbf{Y}^{[2]}} \end{bmatrix} \quad (63)$$

The optimal estimator is

$$\hat{\mathbf{X}} = \mathbb{E} \{ \mathbf{X} \} + \mathbf{B}^* d\mathbf{Y} + \mathbf{C}^* d\mathbf{Y}^{[2]} - \mathbf{C}^* \mathbf{p} \quad (64)$$

and the posterior estimation error is

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} = \mathbf{X} - \mathbf{g}(\mathbf{Y}) = \mathbf{X} - \left(\mathbb{E} \{ \mathbf{X} \} + \mathbf{B}^* d\mathbf{Y} + \mathbf{C}^* d\mathbf{Y}^{[2]} - \mathbf{C}^* \mathbf{p} \right) \quad (65)$$

$$= d\mathbf{X} - \mathbf{B}^* d\mathbf{Y} - \mathbf{C}^* (d\mathbf{Y}^{[2]} - \mathbf{p}) \quad (66)$$

From the update functions shown above, we build the full filtering algorithm. For our high-order filter, we take the same approach commonly used by linear estimators, i.e. we will apply the measurements recursively, hence obtaining a consistent, but suboptimal, estimator (refer to the discussions and the example in prior sections).

VI. High Order Differential Algebra Kalman Filter - HODAKF

The new filtering technique is now proposed in details. In the following pages, the High Order Differential Algebra Kalman Filter (HODAKF) is presented carefully, specifying how the innovative algorithm works in the DA framework, with polynomial approximations.

Consider a dynamic system where the state evolves according to a discrete-time nonlinear state transition equation and where the only information about the system is a set of measurements, related to the state vector, acquired at discrete times.

$$\mathbf{X}(k+1) = \mathbf{f}[\mathbf{X}(k)] + \mathbf{V}(k) \quad (67)$$

$$\mathbf{Y}(k+1) = \mathbf{h}[\mathbf{X}(k+1)] + \mathbf{W}(k+1) \quad (68)$$

Where \mathbf{f} is the process model, $\mathbf{X}(k)$ is the n -dimensional state at time-step k , $\mathbf{Y}(k+1)$ is the m -dimensional vector of the actual measurement at time-step $k+1$, and \mathbf{h} is the measurement function. The process noise \mathbf{V} and the measurement noise \mathbf{W} are potentially non-Gaussian zero-mean random sequences which satisfy the conditions $\forall i, j > 0$:

$$\mathbb{E}\{\mathbf{V}(i)\} = \mathbb{E}\{\mathbf{W}(i)\} = 0 \quad (69)$$

$$\mathbb{E}\{\mathbf{V}(i)\mathbf{V}^T(j)\} = \mathbf{Q}^{[2]}(i)\delta_{ij} \quad (70)$$

$$\mathbb{E}\{\mathbf{W}(i)\mathbf{W}^T(j)\} = \mathbf{R}^{[2]}(i)\delta_{ij} \quad (71)$$

$$\mathbb{E}\{\mathbf{V}(i)\mathbf{W}^T(j)\} = 0 \quad (72)$$

Differential Algebra (DA) expresses quantities such as the state vector and the measurement vector in their Taylor series expansion up to any desired order. Therefore, the propagation function is applied directly on each polynomial (one polynomial per component) of the state. The same holds true for the measurement function.

$$\mathbf{X}_i(k+1) = \mathbf{f}_i[\hat{\mathbf{x}}(k|k)] + \sum_{r=1}^c \frac{1}{r!} \sum \frac{\partial^r \mathbf{f}_i[\mathbf{X}(k|k)]}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_r}} \delta \mathbf{X}_1^{\gamma_1}(k) \dots \delta \mathbf{X}_n^{\gamma_r}(k) + \mathbf{V}_i(k) \quad (73)$$

$$\mathbf{Y}_j(k+1) = \mathbf{h}_j[\mathbf{f}[\hat{\mathbf{x}}(k+1|k)]] + \sum_{r=1}^c \frac{1}{r!} \sum \frac{\partial^r \mathbf{h}_j[\mathbf{X}(k+1)]}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_r}} \delta \mathbf{X}_1^{\gamma_1}(k) \dots \delta \mathbf{X}_n^{\gamma_r}(k) + \mathbf{W}_j(k) \quad (74)$$

Where c is the arbitrary order of the expansion; the second summation is over all permutations of $\gamma_i \in \{1, \dots, n\}$ with $i \in \{1, \dots, r\}$. Hence, the summation of $\mathbf{f}_i[\mathbf{X}(k|k)]$ includes the higher-order partials of the solution flow, which maps the deviations from time k to time $k+1$ and, in an analogous way, the summation of $\mathbf{h}_j[\mathbf{X}(k+1)]$ includes the higher-order partials of the measurement function. These functions are obtained by either integrating in the DA framework the equations of motion or evaluating the measurement equation with DA. Therefore, each function needs to

be c -times differentiable: $\mathbf{f}, \mathbf{h} \in \mathcal{C}^c$. Our approach differs from how prior linear DA estimators handled noise (both process and measurement) [34]. The influence of the noise is included directly on the polynomial integration and evaluation rather than being added after the nonlinear function evaluation: it gives the advantage of easily considering all cross-terms between estimation error and noise. The noises are initialized as DA variables and included in an augmented state vector, leading to the computation of polynomials with coefficients also accounting for the noise and the state-noise mixed terms. Equations (73) and (74) assume the following form:

$$\begin{aligned} \mathbf{X}_i(k+1) &= \mathbf{f}_i[\hat{\mathbf{x}}(k|k)] + \\ &+ \sum_{r=1}^c \frac{1}{r!} \sum \frac{\partial^r (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_r} \partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_r}} \delta \mathbf{X}_1^{\gamma_1}(k) \dots \delta \mathbf{X}_n^{\gamma_r}(k) \delta \mathbf{V}_1^{\gamma_1}(k) \dots \delta \mathbf{V}_n^{\gamma_r}(k) + \mathbf{V}_i(k) \end{aligned} \quad (75)$$

$$\begin{aligned} \mathbf{Y}_j(k+1) &= \mathbf{h}_j[\mathbf{f}[\hat{\mathbf{x}}(k+1|k)]] + \\ &+ \sum_{r=1}^c \frac{1}{r!} \sum \frac{\partial^r (\mathbf{h}_j[\mathbf{X}(k+1)] + \mathbf{W}_j(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_r} \partial \mathbf{W}_1^{\gamma_1} \dots \partial \mathbf{W}_m^{\gamma_r}} \delta \mathbf{X}_1^{\gamma_1}(k) \dots \delta \mathbf{X}_n^{\gamma_r}(k) \delta \mathbf{W}_1^{\gamma_1}(k) \dots \delta \mathbf{W}_m^{\gamma_r}(k) + \mathbf{W}_j(k) \end{aligned} \quad (76)$$

As a consequence, the filter works with an augmented state vector comprising a total of $2n + m$ DA variables; n associated with the actual state vector, n with the process noise, and m with the measurement noise.

The presence of process noise and measurement noise is considered when computing the expected value of the polynomials; high-order central moments of the noise contribute to the transformed variable. The c -th order central moments of the noises are evaluated with the integral definition

$$\mathbf{Q}_{ij\dots q}^{[c]} = \int \mathbf{V}_i \mathbf{V}_j \dots \mathbf{V}_q p_{\mathbf{V}}(\mathbf{V}_i, \mathbf{V}_j, \dots, \mathbf{V}_q) d\mathbf{x} \quad (77)$$

$$\mathbf{R}_{ij\dots q}^{[c]} = \int \mathbf{W}_i \mathbf{W}_j \dots \mathbf{W}_q p_{\mathbf{W}}(\mathbf{W}_i, \mathbf{W}_j, \dots, \mathbf{W}_q) d\mathbf{x} \quad (78)$$

where $p_{\mathbf{V}}()$ and $p_{\mathbf{W}}()$ are the probability density function of the random processes.

A. Prediction

Starting from the knowledge of the state estimate $\hat{\mathbf{x}}(k)$ and its central moments up to a selected order c ; $\mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]}$, $\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}}^{[3]}$, $\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}}^{[4]}$, \dots , $\mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[c]}$, the prediction part of the Kalman filter is described in this section.

It is now possible to predict the state mean using Equation (75) for each state component:

$$\hat{\mathbf{x}}_i^-(k+1) = \mathbb{E} \{ \mathbf{f}_i[\mathbf{X}(k)] + \mathbf{V}_i(k) \} = \quad (79)$$

$$= \mathbf{f}_i[\hat{\mathbf{x}}(k|k)] + \sum_{r=1}^c \frac{1}{r!} \sum \frac{\partial^r (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_r} \partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_r}} \\ \mathbb{E} \{ \delta \mathbf{X}_1^{\gamma_1}(k) \dots \delta \mathbf{X}_n^{\gamma_r}(k) \delta \mathbf{V}_1^{\gamma_1}(k) \dots \delta \mathbf{V}_n^{\gamma_r}(k) \} \quad (80)$$

$$= \mathbf{f}_i[\hat{\mathbf{x}}(k|k)] + \sum_{\rho=1}^n \sum_{\sigma=1}^n \frac{1}{2!} \sum \frac{\partial^2 (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_2}} \mathbf{P}_{\rho\sigma}^{[2]} + \\ + \sum_{\rho=1}^n \sum_{\sigma=1}^n \frac{1}{2!} \sum \frac{\partial^2 (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_2}} \mathbf{Q}_{\rho\sigma}^{[2]} + \\ + \sum_{\rho=1}^n \sum_{\sigma=1}^n \sum_{\tau=1}^n \frac{1}{3!} \sum \frac{\partial^3 (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_3}} \mathbf{P}_{\rho\sigma\tau}^{[3]} + \\ + \sum_{\rho=1}^n \sum_{\sigma=1}^n \sum_{\tau=1}^n \frac{1}{3!} \sum \frac{\partial^3 (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_3}} \mathbf{Q}_{\rho\sigma\tau}^{[3]} + \\ + \sum_{\rho=1}^n \sum_{\sigma=1}^n \sum_{\tau=1}^n \sum_{v=1}^n \frac{1}{4!} \sum \frac{\partial^4 (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_4}} \mathbf{P}_{\rho\sigma\tau v}^{[4]} + \\ + \sum_{\rho=1}^n \sum_{\sigma=1}^n \sum_{\tau=1}^n \sum_{v=1}^n \frac{1}{4!} \sum \frac{\partial^4 (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_4}} \mathbf{Q}_{\rho\sigma\tau v}^{[4]} + \\ + \sum_{\rho=1}^n \sum_{\sigma=1}^n \sum_{\tau=1}^n \sum_{v=1}^n \frac{1}{4!} \sum \frac{\partial^4 (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_2} \partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_2}} \left(\mathbf{P}^{[2]} \mathbf{Q}^{[2]} \right)_{\rho\sigma\tau v} + \\ \vdots \\ + \sum_{\rho=1}^n \dots \sum_{\omega=1}^n \frac{1}{c!} \sum \frac{\partial^c (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_c}} \mathbf{P}_{\rho\dots\omega}^{[c]} + \\ + \sum_{\rho=1}^n \dots \sum_{\omega=1}^n \frac{1}{c!} \sum \frac{\partial^c (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_c}} \mathbf{Q}_{\rho\dots\omega}^{[c]} + \\ + \sum_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_{\rho=1}^n \dots \sum_{\omega=1}^n \frac{1}{c!} \sum \frac{\partial^c (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i(k))}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_\alpha} \partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_\beta}} \left(\mathbf{P}^{[\boldsymbol{\alpha}]} \mathbf{Q}^{[\boldsymbol{\beta}]} \right)_{\rho\dots\omega} \quad (81)$$

$$= \text{eval}(\mathbf{f}_i[\mathbf{X}(k)] + \mathbf{V}_i(k), \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]}(k), \mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}}^{[3]}(k), \mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}}^{[4]}(k), \dots \\ \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[c]}(k), \mathbf{Q}^{[2]}(k), \mathbf{Q}^{[3]}(k), \mathbf{Q}^{[4]}(k), \dots, \mathbf{Q}^{[c]}(k)) \quad (82)$$

where in Equation (81), the last summation in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ considers all the possible combinations such that $\boldsymbol{\alpha} + \boldsymbol{\beta} = c$, with $\boldsymbol{\alpha}, \boldsymbol{\beta} \neq 1$. From now on, for the sake of brevity, the expected value

of polynomial evaluations will be denoted as $\text{eval}()$, as defined in Equation (82). The first input to $\text{eval}()$ is the polynomial function to be evaluated at the specific values given by the subsequent inputs. In the evaluation of the predicted mean, the contribution of the noise is equal to zero, since the propagation equation is linear with respect to the noise: $\mathbf{f}_i[\mathbf{X}(k)] + \mathbf{V}_i(k)$ and the noise is assumed to have zero mean. Therefore, Equation (82) can be expressed as:

$$\hat{\mathbf{x}}_i^-(k+1) = \text{eval}(\mathbf{f}_i[\mathbf{X}(k)] + \mathbf{V}_i(k), \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]}(k), \mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}}^{[3]}(k), \mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}}^{[4]}(k), \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[c]}(k)) \quad (83)$$

Calculation of high-order moments, on the other hand, are strongly affected by the influence of the noise and the $\text{eval}()$ function defined above correctly evaluates the expected value so long as the input central moments are provided up to the requested order (Equation (90)).

The predicted state covariance is evaluated considering the deviation of the propagated state polynomials with respect to the estimated mean.

$$\mathbf{P}_{\mathbf{X}\mathbf{X},ij}^{[2]-}(k+1) = \mathbb{E} \{ [\mathbf{f}_i[\mathbf{X}(k)] - \hat{\mathbf{x}}_i^- + \mathbf{V}_i(k)] [\mathbf{f}_j[\mathbf{X}(k)] - \hat{\mathbf{x}}_j^- + \mathbf{V}_j(k)] \} \quad (84)$$

$$= \sum_{r=1}^c \sum_{s=1}^c \frac{1}{r!s!} \sum \frac{\partial^r (\mathbf{f}_i[\mathbf{X}(k|k)] + \mathbf{V}_i)}{\partial \mathbf{X}_1^{\gamma_1} \dots \partial \mathbf{X}_n^{\gamma_r} \partial \mathbf{V}_1^{\gamma_1} \dots \partial \mathbf{V}_n^{\gamma_r}} \sum \frac{\partial^s (\mathbf{f}_j[\mathbf{X}(k|k)] + \mathbf{V}_j)}{\partial \mathbf{X}_1^{\xi_1} \dots \partial \mathbf{X}_n^{\xi_s} \partial \mathbf{V}_1^{\xi_1} \dots \partial \mathbf{V}_n^{\xi_s}} \cdot \mathbb{E} \left\{ \delta \mathbf{X}_1^{\gamma_1}(k) \dots \delta \mathbf{X}_n^{\gamma_r}(k) \delta \mathbf{V}_1^{\gamma_1}(k) \dots \delta \mathbf{V}_n^{\gamma_r}(k) \delta \mathbf{X}_1^{\xi_1}(k) \dots \delta \mathbf{X}_n^{\xi_s}(k) \delta \mathbf{V}_1^{\xi_1}(k) \dots \delta \mathbf{V}_n^{\xi_s}(k) \right\} - \delta \hat{\mathbf{x}}_i^- \delta \hat{\mathbf{x}}_j^- \quad (85)$$

Where $\xi_i \in \{1, \dots, n\}$ and $\delta \hat{\mathbf{x}}_i^- = \mathbf{f}_i[\hat{\mathbf{x}}(k|k)] - \hat{\mathbf{x}}_i^-$. Defining the state deviation vector such as

$$d\mathbf{X}_i^- = \mathbf{f}_i[\mathbf{X}(k)] - \hat{\mathbf{x}}_i^- + \mathbf{V}_i(k) \quad (86)$$

equation (85) can be then written using the $\text{eval}()$ function

$$\mathbf{P}_{\mathbf{X}\mathbf{X},ij}^{[2]-} = \text{eval}(d\mathbf{X}_i^- d\mathbf{X}_j^-, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[2c]}, \mathbf{Q}^{[2]}) \quad (87)$$

where it must be noted that, since the polynomial to be evaluated is the product of two polynomials, the order of the function to be evaluated is $2c$. The evaluation of the covariance requires a multiplication between two c -th order polynomials, therefore the expansion order of the series doubles. The pursuit of the filter is to estimate correctly up to the second moment, thus the precise evaluation of moments up to order $2c$ must be implemented. The prediction of these moments is performed

through to the eval() function;

$$\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X},ijk}^{[3]-} = \text{eval}(d\mathbf{X}_i^-, d\mathbf{X}_j^-, d\mathbf{X}_k^-, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[3c]}, \mathbf{Q}^{[2]}, \mathbf{Q}^{[3]}) \quad (88)$$

$$\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X},ijkl}^{[4]-} = \text{eval}(d\mathbf{X}_i^-, d\mathbf{X}_j^-, d\mathbf{X}_k^-, d\mathbf{X}_l^-, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[4c]}, \mathbf{Q}^{[2]}, \mathbf{Q}^{[3]}, \mathbf{Q}^{[4]}) \quad (89)$$

⋮

$$\mathbf{P}_{\mathbf{X}\dots\mathbf{X},i\dots q}^{[2c]-} = \text{eval}(d\mathbf{X}_i^-, \dots, d\mathbf{X}_q^-, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[2c^2]}, \mathbf{Q}^{[2]}, \dots, \mathbf{Q}^{[2c]}) \quad (90)$$

notice that to calculate the central moment of \mathbf{X}^- of order q , it is necessary to input the central moments of \mathbf{X} up to order qc . As a consequence, it is only possible to calculate all the moments exactly when $c = 1$. Notice also that the following identities hold: $\mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]-} = \mathbf{P}_{d\mathbf{X}d\mathbf{X}}^{[2]-}$, $\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}}^{[3]-} = \mathbf{P}_{d\mathbf{X}d\mathbf{X}d\mathbf{X}}^{[3]-}$, $\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}}^{[4]-} = \mathbf{P}_{d\mathbf{X}d\mathbf{X}d\mathbf{X}d\mathbf{X}}^{[4]-}$, \dots

It is clear that, for $c > 1$, the analytic evaluation of central moments of the transformed variable requires knowledge of all central moments without limit of the original variable. In this work, we only evaluate exactly the first $2c$ central moments, while the required central moments of order higher than that are approximated. The approximation varies depending whether the moment is of odd or even order. An odd moment of order between $(2c + 1)$ and $2c^2$ is approximated such that its only non-zero terms are those where all indexes are the same, e.g. for a fifth order central moment $\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X},iiii}^{[5]-}$. Even-order moments, on the other hand, are approximated with non-zero terms corresponding to all permutation in which all indexes appear an even number of times, e.g. for a sixth-order moment $\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X},iijjkk}^{[6]-}$.

The state of the system and the central moments up to order $2c$ are propagated forward in time with the eval() function. The next step of the algorithm is to re-define the polynomial inside the DA framework by re-centering the Taylor series expansion around the predicted mean $\hat{\mathbf{x}}^-$. Through this shift, the time propagation step is completed and the predicted moments $\mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[i]-}$ are the central moments of the propagated state. The next step is to incorporate the measurement, which is done by defining the following prior distribution of the state

$$\mathbf{X}(k+1) = \hat{\mathbf{x}}^-(k+1) + d\mathbf{X}^-(k+1) \quad (91)$$

where the central moments of \mathbf{X} and $d\mathbf{X}^-$ coincide and are given by $\mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]-}$, $\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}}^{[3]-}$, $\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}}^{[4]-}$, \dots and the mean of $\mathbf{X}(k+1)$ coincides with $\hat{\mathbf{x}}^-(k+1)$.

B. Update

The time prediction step is followed by the measurement update. We start by evaluating the predicted measurement mean and its covariance.

$$\hat{\mathbf{y}}_i = \text{eval}(\mathbf{h}_i[\mathbf{X}(k+1)] + \mathbf{W}_i(k+1), \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[c]-}) \quad (92)$$

$$d\mathbf{Y}_i = \mathbf{h}_i[\mathbf{X}(k+1)] - \hat{\mathbf{y}}_i + \mathbf{W}_i(k+1) \quad (93)$$

where in Equation (92) the influence of the noise moments is null since they appear linearly.

As previously seen, the quadratic update requires augmenting the measurements with their second power by using the Kronecker product. By working with deviations it is possible to avoid the computation of the predicted mean-square of the measurements. The quadratic deviation vector is given by

$$d\mathbf{Y}^{[2]} = d\mathbf{Y} \otimes d\mathbf{Y} \quad (94)$$

Defining the augmented measurement deviation vector as

$$d\mathcal{Y} = \begin{bmatrix} d\mathbf{Y} \\ d\mathbf{Y}^{[2]} - \text{vect}(\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}^{[2]-}) \end{bmatrix} \quad (95)$$

the augmented measurement covariance matrix can be calculated block-wise:

$$\mathbf{P}_{d\mathcal{Y}d\mathcal{Y}} = \begin{bmatrix} \mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}^{[2]-} & \mathbf{P}_{d\mathbf{Y}d\mathbf{Y}^{[2]}}^{[2]-} \\ \mathbf{P}_{d\mathbf{Y}^{[2]}d\mathbf{Y}}^{[2]-} & \mathbf{P}_{d\mathbf{Y}^{[2]}d\mathbf{Y}^{[2]}}^{[2]-} \end{bmatrix} \quad (96)$$

where all the redundant components from the Kronecker product are removed such that $\mathbf{P}_{d\mathcal{Y}d\mathcal{Y}}$ is not singular. In Equation (95), the presence of $\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}^{[2]-}$ keeps the mean of $d\mathcal{Y}$ null. To obtain the three distinct blocks we start from:

$$\mathbf{P}_{d\mathbf{Y}d\mathbf{Y},ij}^{[2]-} = \text{eval}(d\mathbf{Y}_i d\mathbf{Y}_j, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[2c]-}, \mathbf{R}^{[2]}) \quad (97)$$

$$\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}d\mathbf{Y},ijk}^{[3]-} = \text{eval}(d\mathbf{Y}_i d\mathbf{Y}_j d\mathbf{Y}_k, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[3c]-}, \mathbf{R}^{[2]}, \mathbf{R}^{[3]}) \quad (98)$$

$$\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}d\mathbf{Y}d\mathbf{Y},ijkl}^{[4]-} = \text{eval}(d\mathbf{Y}_i d\mathbf{Y}_j d\mathbf{Y}_k d\mathbf{Y}_l, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[4c]-}, \mathbf{R}^{[2]}, \mathbf{R}^{[3]}, \mathbf{R}^{[4]}) \quad (99)$$

Since the quadratic update requires working with the covariance of the square of the measurements, the following quantity is needed

$$\mathbf{P}_{d\mathbf{Y}^{[2]}d\mathbf{Y}^{[2]}}^{[2]-} = \mathbf{P}_{d\mathbf{Y}d\mathbf{Y}d\mathbf{Y}d\mathbf{Y}}^{[4]-} - \text{vect}(\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}^{[2]-}) \text{vect}(\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}^{[2]-})^T \quad (100)$$

Working with deviations simplifies the expression of the off-diagonal block of the covariance matrix of $d\mathbf{Y}^{[2]}$

$$\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}^{[2]}}^{[2]-} = (\mathbf{P}_{d\mathbf{Y}^{[2]}d\mathbf{Y}}^{[2]-})^T = \mathbf{P}_{d\mathbf{Y}d\mathbf{Y}d\mathbf{Y}}^{[3]-} \quad (101)$$

with the understanding the the 3D array $\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}d\mathbf{Y}}^{[3]-}$ is re-written as a matrix.

The cross-covariance between the state with the square of the measurement is evaluated as a third order moment because of the equality $\mathbf{P}_{\mathbf{X}d\mathbf{Y}^{[2]}}^{[2]-} = \mathbf{P}_{\mathbf{X}d\mathbf{Y}d\mathbf{Y}}^{[3]-}$.

$$\mathbf{P}_{\mathbf{X}d\mathbf{Y},ij}^{[2]-} = \text{eval}(d\mathbf{X}_i^-, d\mathbf{Y}_j, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[2c]-}) \quad (102)$$

$$\mathbf{P}_{\mathbf{X}d\mathbf{Y}d\mathbf{Y},ijk}^{[3]-} = \text{eval}(d\mathbf{X}_i^-, d\mathbf{Y}_j, d\mathbf{Y}_k, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[3c]-}) \quad (103)$$

There are no joint contributions to the central moments from process and measurement noises because the two are assumed independent from one another. The augmented state-measurement cross covariance matrix is computed block-wise.

$$\mathbf{P}_{\mathbf{X}d\mathbf{Y}} = \begin{bmatrix} \mathbf{P}_{\mathbf{X}d\mathbf{Y}}^{[2]-} & \mathbf{P}_{\mathbf{X}d\mathbf{Y}^{[2]}}^{[2]-} \end{bmatrix} \quad (104)$$

Where, again, $\mathbf{P}_{\mathbf{X}d\mathbf{Y}}^{[2]-} = \mathbf{P}_{d\mathbf{X}d\mathbf{Y}}^{[2]-}$ and $\mathbf{P}_{\mathbf{X}d\mathbf{Y}^{[2]}}^{[2]-} = \mathbf{P}_{d\mathbf{X}d\mathbf{Y}^{[2]}}^{[2]-}$. The redundant components from the Kronecker product are again removed to be consistent with the dimensions of $\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}$.

From these quantities it is now possible to solve Eq. (63) to obtain the Kalman gain associated with the augmented measurement, $d\mathcal{Y}$.

$$\mathbf{K} = \mathbf{P}_{\mathbf{x}d\mathcal{Y}} \mathbf{P}_{d\mathcal{Y}d\mathcal{Y}}^{-1} \quad (105)$$

The residual is defined as

$$d\mathbf{y} = \mathbf{y} - \hat{\mathbf{y}} \quad (106)$$

where \mathbf{y} is the actual numerical value of the measurement read from the sensor (the outcome of random vector \mathbf{Y}). It is now possible to evaluate the optimal state estimate,

$$\hat{\mathbf{x}}^+ = \hat{\mathbf{x}}^- + \mathbf{K} \begin{bmatrix} d\mathbf{y} \\ d\mathbf{y}^{[2]} - \text{vect}(\mathbf{P}_{d\mathbf{Y}d\mathbf{Y}}^{[2]-}) \end{bmatrix} \quad (107)$$

and the updated estimation error,

$$d\mathbf{X}^+ = d\mathbf{X}^- - \mathbf{K}d\mathcal{Y} \quad (108)$$

Since the update is quadratic, the updated state deviation vector is no longer a polynomial of order c , as $d\mathbf{X}^-$, but of order $2c$. Since \mathbf{K} is deterministic, from equation (153) it follows that the mean of $d\mathbf{X}^+$ is zero. Roundoff errors may result in the mean of $d\mathbf{X}^+$ being slightly different from zero; this inaccuracy can be easily removed by subtracting the calculated mean from the polynomial.

The updated central moments are calculated with `eval()` and the new state deviation vector $d\mathbf{X}^+$. Notice that in the evaluation of the covariance matrix, by squaring the polynomial $d\mathbf{X}^+$, the order of the Taylor series expansion increases to 4 times the initial filter order c . Consequently, the computation of the updated version of moment $2c$ -th requires, as input, state moments up to order $4c^2$ -th; however, moments of order greater than $2c^2$ -th are neglected.

$$\mathbf{P}_{\mathbf{X}\mathbf{X},ij}^{[2]^+} = \text{eval}(d\mathbf{X}_i^+ \ d\mathbf{X}_j^+, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]^-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[4c]^-}, \mathbf{R}^{[2]}, \mathbf{R}^{[3]}, \mathbf{R}^{[4]}) \quad (109)$$

$$\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X},ijk}^{[3]^+} = \text{eval}(d\mathbf{X}_i^+ \ d\mathbf{X}_j^+ \ d\mathbf{X}_k^+, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]^-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[6c]^-}, \mathbf{R}^{[2]}, \mathbf{R}^{[3]}, \mathbf{R}^{[4]}, \mathbf{R}^{[5]}, \mathbf{R}^{[6]}) \quad (110)$$

$$\mathbf{P}_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X},ijkl}^{[4]^+} = \text{eval}(d\mathbf{X}_i^+ \ d\mathbf{X}_j^+ \ d\mathbf{X}_k^+ \ d\mathbf{X}_l^+, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]^-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[8c]^-}, \mathbf{R}^{[2]}, \dots, \mathbf{R}^{[8]}) \quad (111)$$

⋮

$$\mathbf{P}_{\mathbf{X}\dots\mathbf{X},i\dots q}^{[2c]^+} = \text{eval}(d\mathbf{X}_i^+ \ \dots \ d\mathbf{X}_q^+, \mathbf{P}_{\mathbf{X}\mathbf{X}}^{[2]^-}, \dots, \mathbf{P}_{\mathbf{X}\dots\mathbf{X}}^{[2c^2]^-}, \mathbf{R}^{[2]}, \dots, \mathbf{R}^{[2c]}) \quad (112)$$

As done in the propagation portion of the filter, only central moments up to order $2c^2$ are calculated exactly with `eval()`. The remaining moments from order $(2c+1)$ -th to order $2c^2$ -th are approximated in an analogous way as described in the prediction step.

VII. Numerical Examples

For the remainder of this paper, the high order filters are denoted as ‘‘HODAKF’’ followed by two numbers: the first number indicates the order of the filter (the order of the Taylor series at which functions are approximated); the second number indicates the order of the update (1 for linear update and 2 for quadratic update). When a statement or a conclusion is valid for any order, the number is left generic and replaced by ‘‘N’’ (e.g. HODAKF-N-2 stands for a quadratic update and any Taylor series truncation order).

A. Simple Non-Gaussian Noise Example

The performance of the proposed methodology is first assessed with a linear problem with non-Gaussian process and measurement noise. The chosen example is the one proposed by De Santis and Germani [16], with a simple linear discrete-time scalar system:

$$x_{k+1} = ax_k + f_k \quad \bar{x} = 0 \quad (113)$$

$$y_k = cx_k + g_k \quad (114)$$

with $a = 0.6$, $c = 0.8$. The noises f_k and g_k are independent zero-mean random sequences with distribution described in Table 1. This example has been used as a test base to have a quick

f_k	-1	3	9		g_k	1	-3	-9
$P(f_k)$	15/18	2/18	1/18		$P(g_k)$	15/18	2/18	1/18

Table 1 Random sequences distribution

visualization of the advantages of a quadratic update. The HODAKF-N-N is compared with the classical Kalman filter. In a linear example such this one, the selection of the order of the filter has no influence of the results because the system is linear and therefore perfectly modeled. Therefore, the HODAKF-N-1 reduces to the Kalman filter, but with the added capability of estimating higher order central moments. The initial condition is set with a perfect knowledge of the starting point, i.e. null initial error and null moments. A single run for the example is displayed in Figure 1, the figure shows $n = 50$ steps. Figure 1 shoes that HODAKF-N-2 (blue line) follows better the true state of the system (black line) under the influence of this non-Gaussian noise than the Kalman filter (red line). This improvement is especially marked when a noise peak occurs. HODAKF-N-2 has a significantly smaller error standard deviation, as shown in Figure 2.

Table 2 shows the accuracy and the consistency of the estimated covariance against the ensemble statistics from a Monte Carlo analysis of 5000 runs. The table shows that for this linear/non-Gaussian problem a quadratic estimator, HODAKF-N-2, achieves a 40% reduction of the estimation error covariance. Additionally, Table 2 shows that HODAKF-N-N is able to estimate central moments other than the covariance, up to order $2c$ (3^{rd} and 4^{th} in this example). Due to the linear dynamics/measurement, all linear update filters (EKF, UKF, HODAKF-N-1) reduce to the classical

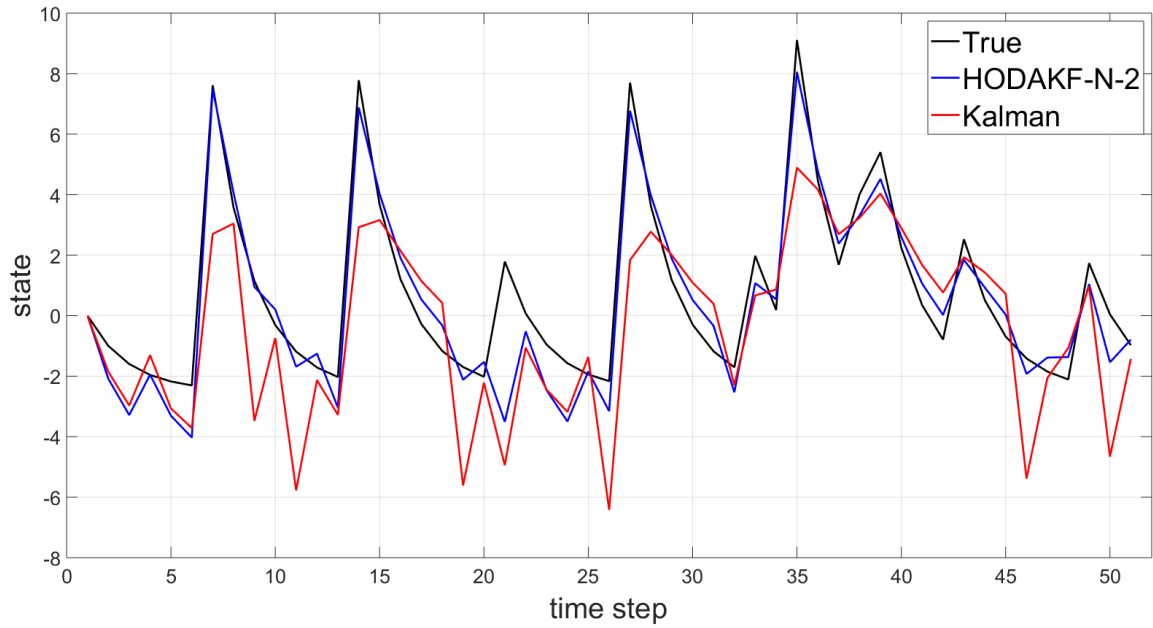


Fig. 1 HODAKF-N-2 vs Kalman Filter

Kalman Filter.

	$\sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}$	$\sqrt{P_{\mathbf{X}\mathbf{X}}^{[2]}}$	$\sqrt[3]{\frac{\sum(x_i - \bar{x})^3}{N}}$	$\sqrt[3]{P_{\mathbf{X}\mathbf{X}\mathbf{X}}^{[3]}}$	$\sqrt[4]{\frac{\sum(x_i - \bar{x})^4}{N}}$	$\sqrt[4]{P_{\mathbf{X}\mathbf{X}\mathbf{X}\mathbf{X}}^{[4]}}$
KALMAN	2.0924	2.0968	2.4712	0 ^a	3.2101	2.7595 ^a
HODAKF-N-1	2.0924	2.0968	2.4712	2.4768	3.2101	3.2161
HODAKF-N-2	1.2681	1.2728	1.9096	1.9144	2.7277	2.7510

^a Approximated from the variance assuming a Gaussian distribution

Table 2 Comparison of 5000 Monte Carlo runs vs. Predicted Central Moments.

Figure 2 shows that HODAKF-N-2 settles to a standard deviation steady state level which is lower when compared to the linear update (Kalman filter and HODAKF-N-1). The consistency of the estimators is shown through Monte Carlo analysis. The continuous lines depict the filter's prediction of the estimation error standard deviations, while the dashed lines are the estimation error standard deviations calculated from the the 5000 samples of the Monte Carlo analysis. Overlap between the two indicates consistency.

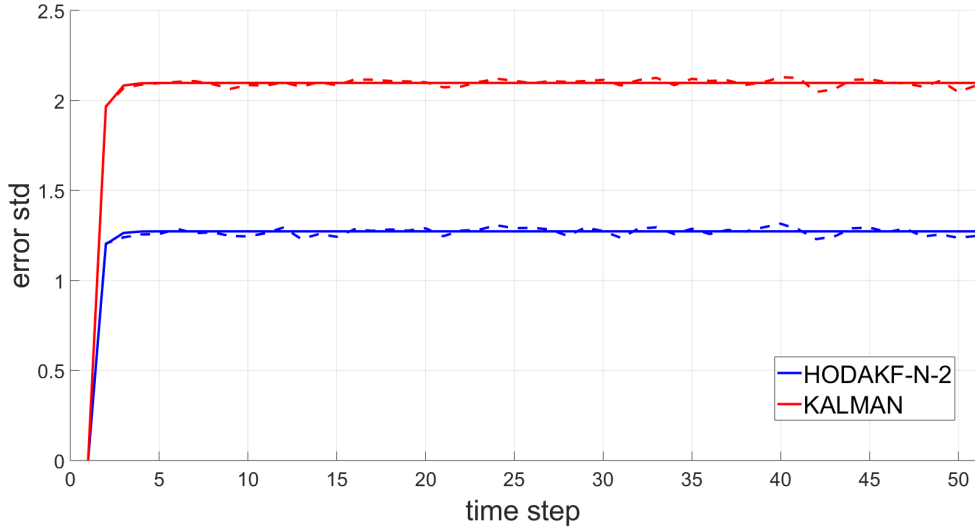


Fig. 2 Error Standard Deviation: Kalman and HODAKF-N-1 (red) vs. HODAKF-N-2 (blue). Continuous lines are the filters' predictions while dashed lines are the Monte Carlo ensemble values.

B. Orbit Determination Example

In this example the performance of the filter is applied to the orbit determination problem. The equations of motion governing the system are these associated to the Keplerian dynamics, where \mathbf{r} is the position vector of the spacecraft and μ is the Earth gravitational parameter.

$$\ddot{\mathbf{r}} = -\frac{\mu}{r^3}\mathbf{r} \quad (115)$$

The initial conditions and uncertainty values used in this example follow those from Refs. [12, 34]. The problem is normalized to be non-dimensional with length scaled by the orbit semi-major axis, $a = 8788$ km, and time scaled by the orbital period $\sqrt{\frac{a^3}{\mu}}$. The true initial condition for the simulation is the following:

$$\mathbf{x}_0 = \begin{pmatrix} \mathbf{r}_0 \\ \mathbf{v}_0 \end{pmatrix} = \begin{pmatrix} -0.68787 \\ -0.39713 \\ 0.28448 \\ -0.51330 \\ 0.98266 \\ 0.37611 \end{pmatrix} \quad (116)$$

For the purpose of this example, range and bearing angles are taken with respect of the center of the planet:

$$y_1 = r + \eta_1 \quad (117)$$

$$y_2 = \arctan2(x_2, x_1) + \eta_2 \quad (118)$$

$$y_3 = \arcsin\left(\frac{x_3}{r}\right) + \eta_3 \quad (119)$$

where η_i , $i = 1, 2, 3$, is the measurement noise, assumed to be Gaussian and $\arctan2$ is the two argument arctangent function that returns an angle in $(-\pi \pi]$. The standard deviation of the error is assumed to be 0.1 m in range and 0.1 arcsec for the angles. The first $4c$ central moments of the moment noise are (where c is the selected filter order)

$$\mathbf{R}_{ij}^{[2]} = \begin{cases} \sigma_{\nu,ij}^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (120)$$

$$\mathbf{R}_{ijk}^{[3]} = 0 \quad \forall i, j, k \quad (121)$$

$$\mathbf{R}_{ijkl}^{[4]} = \begin{cases} 3\sigma_{\nu,ijkl}^4 & \text{if } i = j = k \\ \sigma_{\nu,ij}^2 \sigma_{\nu,kl}^2 & \text{if } i = j \wedge k = l \\ 0 & \text{otherwise} \end{cases} \quad (122)$$

$$\mathbf{R}_{ijklm}^{[5]} = 0 \quad \forall i, j, k, l, m \quad (123)$$

⋮

and so on.

The initial error covariance matrix is assumed diagonal, the value of the standard deviation for the position vector components is 0.01, while the standard deviation for the velocity vector components is 10^{-4} . Thus,

$$\sigma_r = 10^{-2}a \quad (124)$$

$$\sigma_v = 10^{-4} \sqrt{\frac{\mu}{a}} \quad (125)$$

Each Monte Carlo run is initialized with the same true initial state given by Equation (116), while

the initial filter's estimate is dispersed such that the initial estimation error is Gaussian with zero mean and standard deviation given by Equations (124) and (125).

Since the initial uncertainty is assumed Gaussian, the high order central moments are evaluated with the same formulas shown in Equations (120) - (123), and associating σ_r to indexes 1, 2, 3 and σ_v to indexes 4, 5, 6. For the computer implementation in DACE [27, 34, 35] (Differential Algebra Core Engine) all high order moments are expressed using their vectorization representation, that is to say, rather than expressing the c -th order moment of the state as a c -dimensional array they are represented as vectors. The table shows that, except for the first two indexes, each additional odd

term	row position	column position	order c
$\mathbf{P}_{ijklp}^{[5]}$	$in^2 + kn + l$	$jn + p$	odd
$\mathbf{P}_{ijklpq}^{[6]}$	$jn^2 + ln + p$	$in^2 + kn + q$	even
$\mathbf{P}_{ijth\dots pq}^{[c]}$	$in^{(c-1)/2} + tn^{(c-1)/2-1} \dots + p$	$jn^{(c-1)/2-1} + hn^{(c-1)/2-2} + \dots + q$	odd
$\mathbf{P}_{ijth\dots pq}^{[c]}$	$jn^{c/2} + hn^{c/2-1} + \dots n + p$	$in^{c/2} + tn^{c/2-1} \dots + q$	even

Table 3 High order moments representation

dimension adds a column of matrices while each even dimension adds a matrix of column vectors.

The approximated central moments, (those of order $2c + 1$ up to $2c^2$), are stored as column vectors. Odd moments are represented as an n -dimensional vector, while even central moments of order p (with $2c < p \leq 2c^2$) are represented as $n^{p/2}$ -dimensional vectors given by the diagonal of the corresponding matrices calculated using the indexes from Table 3.

The filter performance is shown in Figure 3 for HODAKF-2-1 and Figure 4 for HODAKF-2-2. The left columns represent the position error in the 3 components, while the right columns are the velocity errors. The error is evaluated as the difference between the estimated state and the true state. The two figures show the results for a duration of two orbits, with a total of 24 equally spaced observations per orbit. The initial uncertainty is beyond the scale depicted in the figure, which shows the convergence of the error and the error covariance (shown with $\pm 3\sigma$ values). A Monte Carlo analysis with 100 runs is performed and it can be noted that the filter's predicted covariance (blue continuous line) is consistent with the sample covariance from Monte Carlo (blue dashed line). Moreover, the Figures depict the ability of HODAKF-2-N to correctly estimate high order

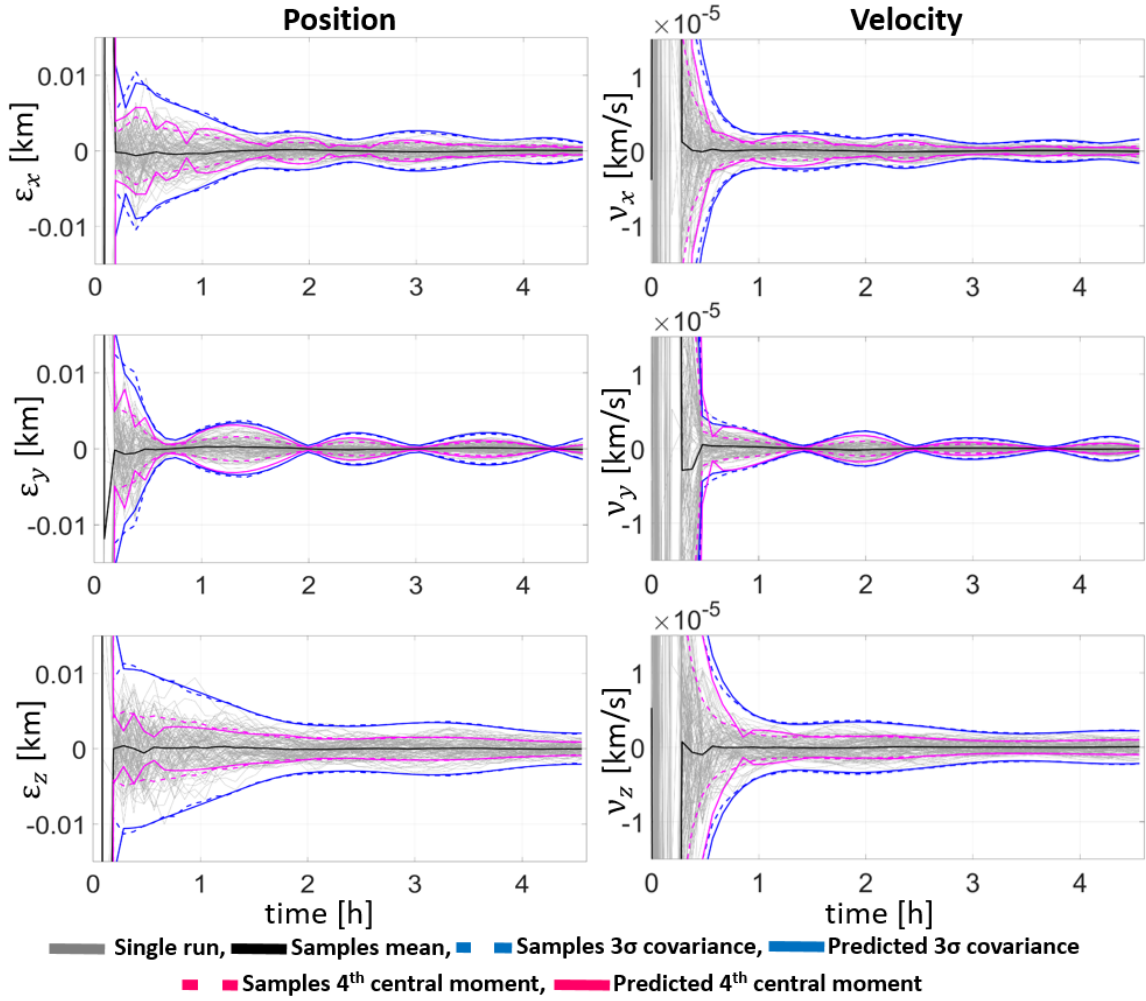


Fig. 3 Simulation results for HODAKF-2-1 with $n = 100$ samples.

moments. In each graph, the continuous magenta line represents the estimated 4^{th} central moment expressed as $\sqrt[4]{P_{XXXX}^{[4]}}$ of the corresponding state component, while the dashed magenta line is the 4^{th} central moment evaluated from the Monte Carlo analysis, as a mean among all the single runs: $\sqrt[4]{\frac{\sum (x_i - \bar{x})^4}{N}}$. It can be noted that HODAKF-2-2 has a more consistent prediction of the 4th central moment with respect to HODAKF-2-1 since the dashed magenta lines and the continue ones perfectly overlap. Figure 5 shows the EKF results under the same simulation conditions. The estimated covariance is three orders of magnitude smaller with respect to the sample one: the EKF is not consistent and the filter diverges. Figure 6 shows the simulation results when the UKF is used. The filter converges and the estimated covariance is consistent with the sample covariance. Figure 7 shows the simulation results when the DAHO-2 filter is used [12].

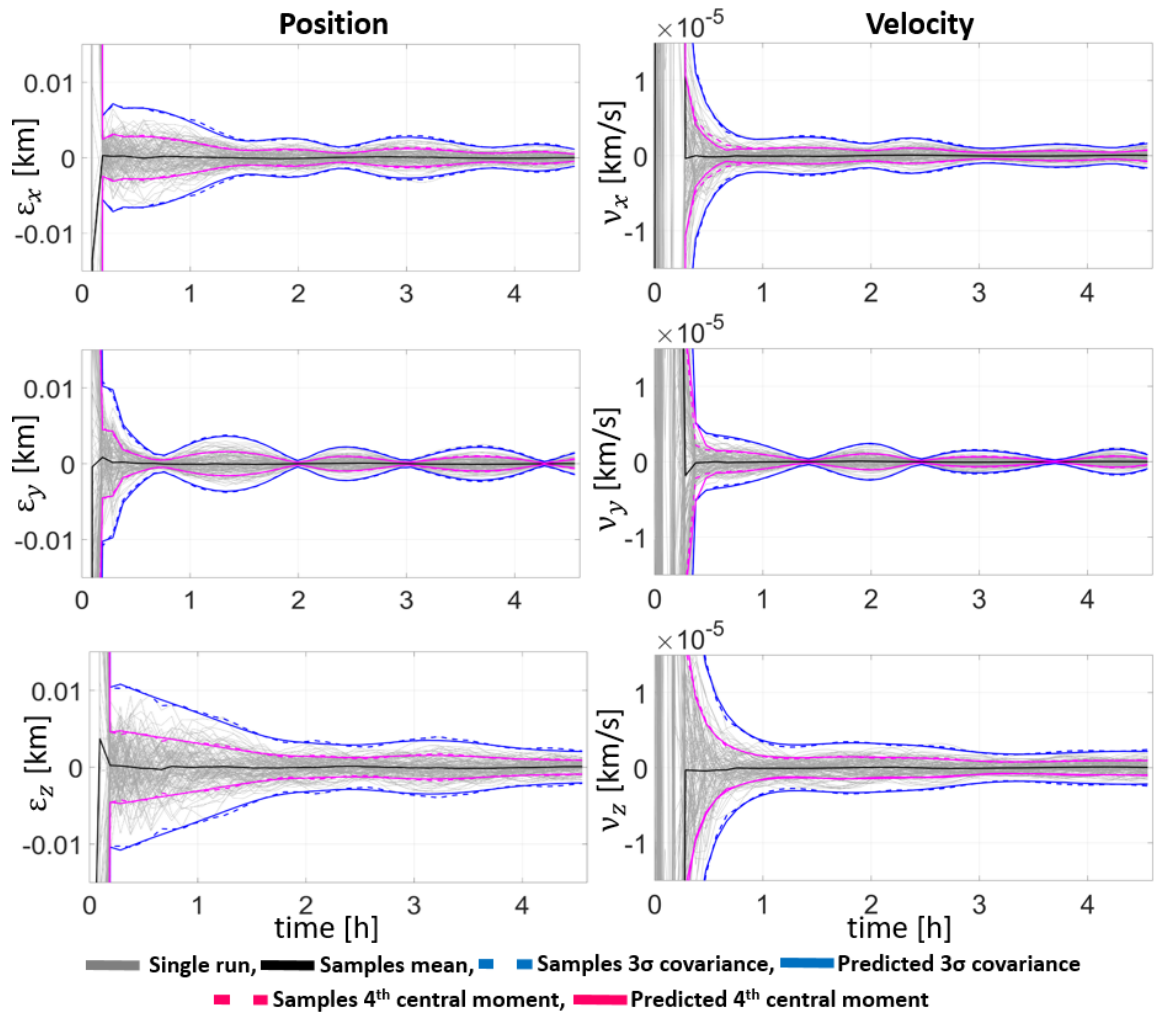


Fig. 4 Simulation results for HODAKF-2-2 with $n = 100$ samples.

Figure 8 summarizes the performance of all filters with their velocity standard deviations calculated from the diagonal terms of the velocity covariance matrices as:

$$\sigma_v = \sqrt{\sigma_{vx}^2 + \sigma_{vy}^2 + \sigma_{vz}^2} \quad (126)$$

Time zero in the figures is apogee, which explains the initial increase in the velocity standard deviation. The overall trace of the covariance matrix is monotonically decreasing until steady state is reached. The red line in Figure 8 is HODAKF-2-1. This line represents the covariance behavior of a linear update filter when second order terms of measurement and dynamics are taken into consideration. HODAKF-2-1 has the same accuracy level of Valli's DAHO-2, as the two lines overlap. The only difference between the two is that HODAKF-2-1 is also able to estimate higher

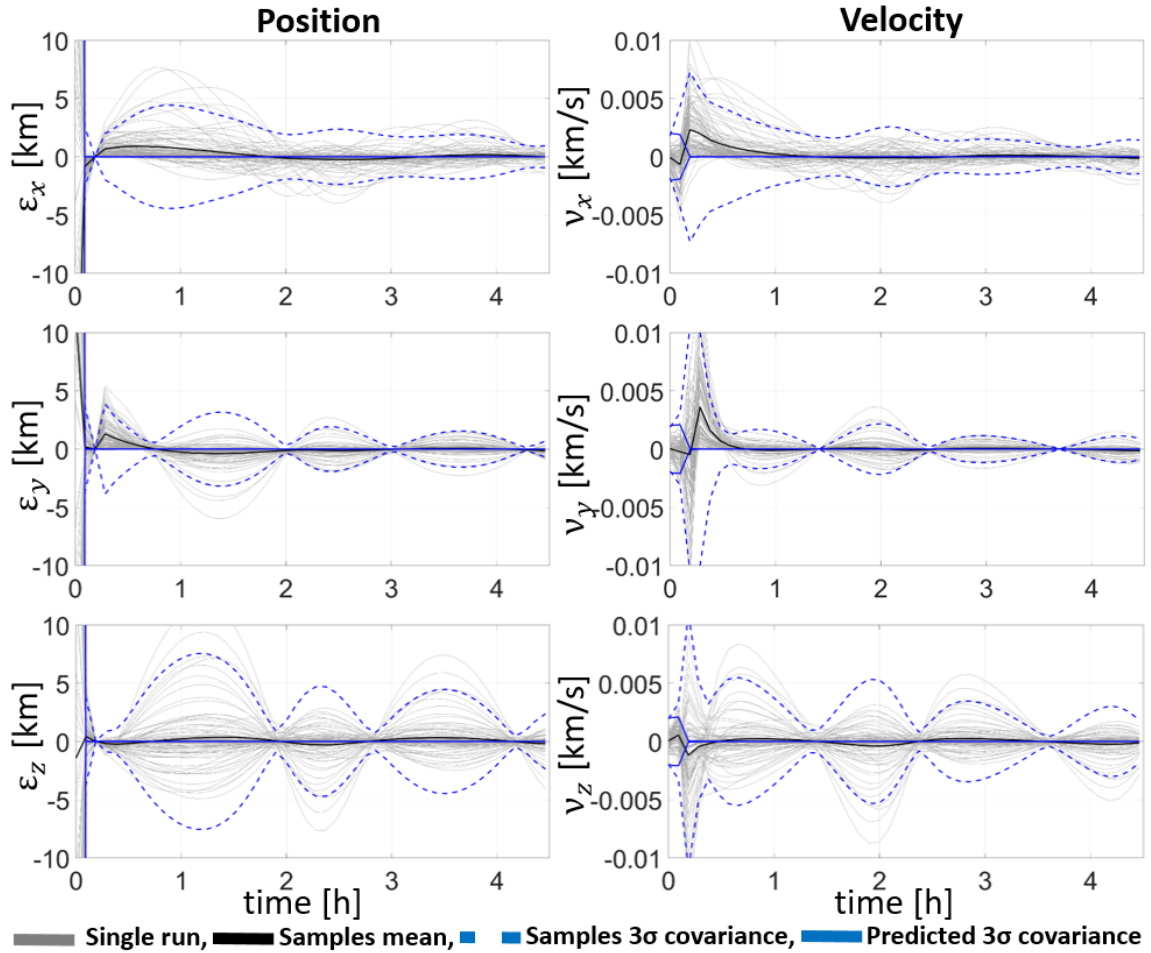


Fig. 5 Simulation results for EKF with $n = 100$ samples.

order central moments. The green line represents HODAKF-1-1 and DAHO-1 which are nothing but the Extended Kalman Filter. While it seems the EKF performs better, what is shown is the predicted error by the filter, since the EKF was shown to diverge, the actual EKF error is considerably larger than all other filters. The UKF is shown in magenta, it too is a linear filter that accounts for the linearities in the measurement and dynamics. In this particular example, the UKF is shown to converge faster than the other linear filters HODAKF-2-1 and DAHO-2. Finally, the black line shows the improvement in convergence using HODAKF-2-2 which poses a quadratic update. During the initial steps, the filter is able to decrease the state uncertainty faster than the linear filters. Through the quadratic update, the measurement information is better fused with the state prediction to produce superior performance. After convergence, once the uncertainty is low, the performance of the linear update matches that of the quadratic update. This fact is further

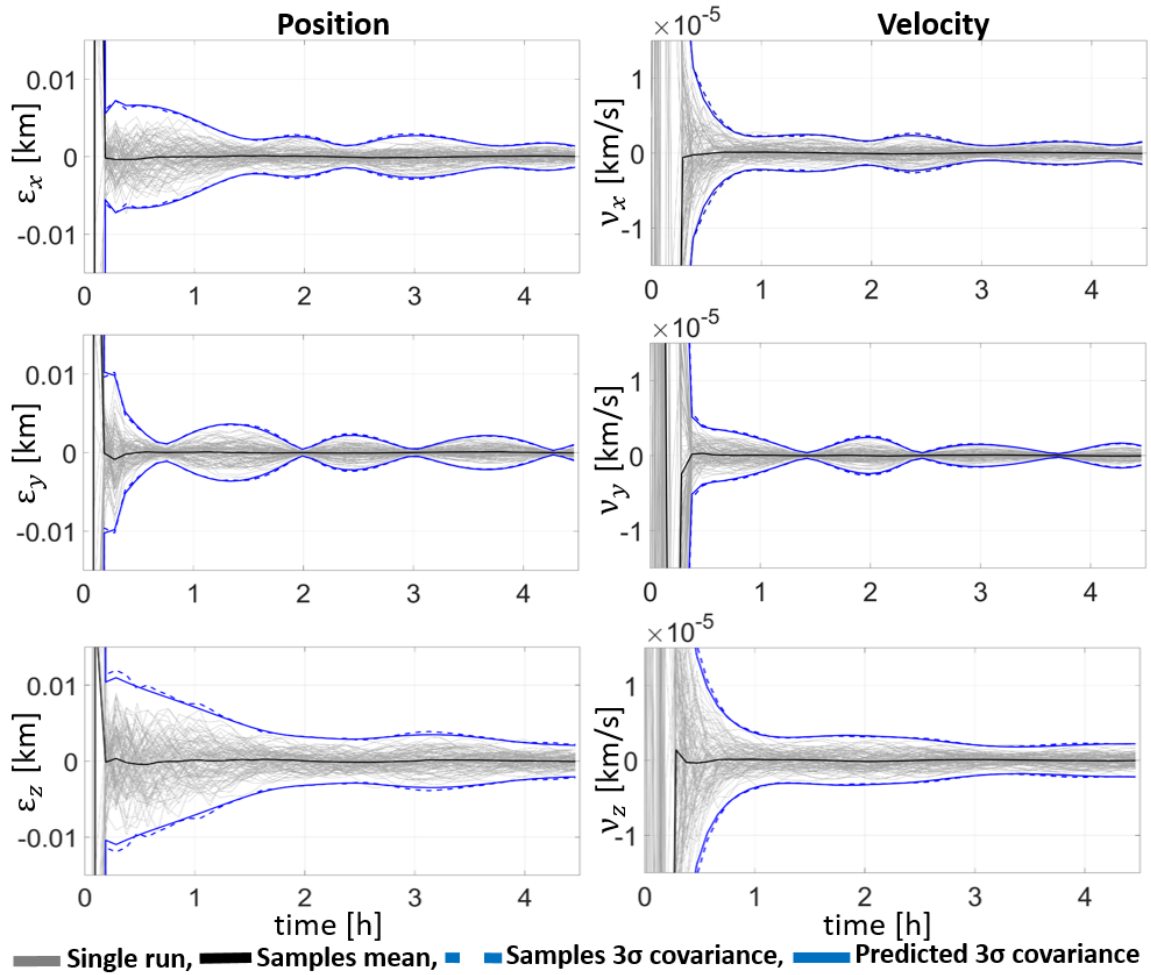


Fig. 6 Simulation results for UKF with $n = 100$ samples.

exacerbated in the next example where only angular measurements are used. Finally, a last example is presented where the steady state performance of the linear estimators is considerably worse than that of the quadratic estimator.

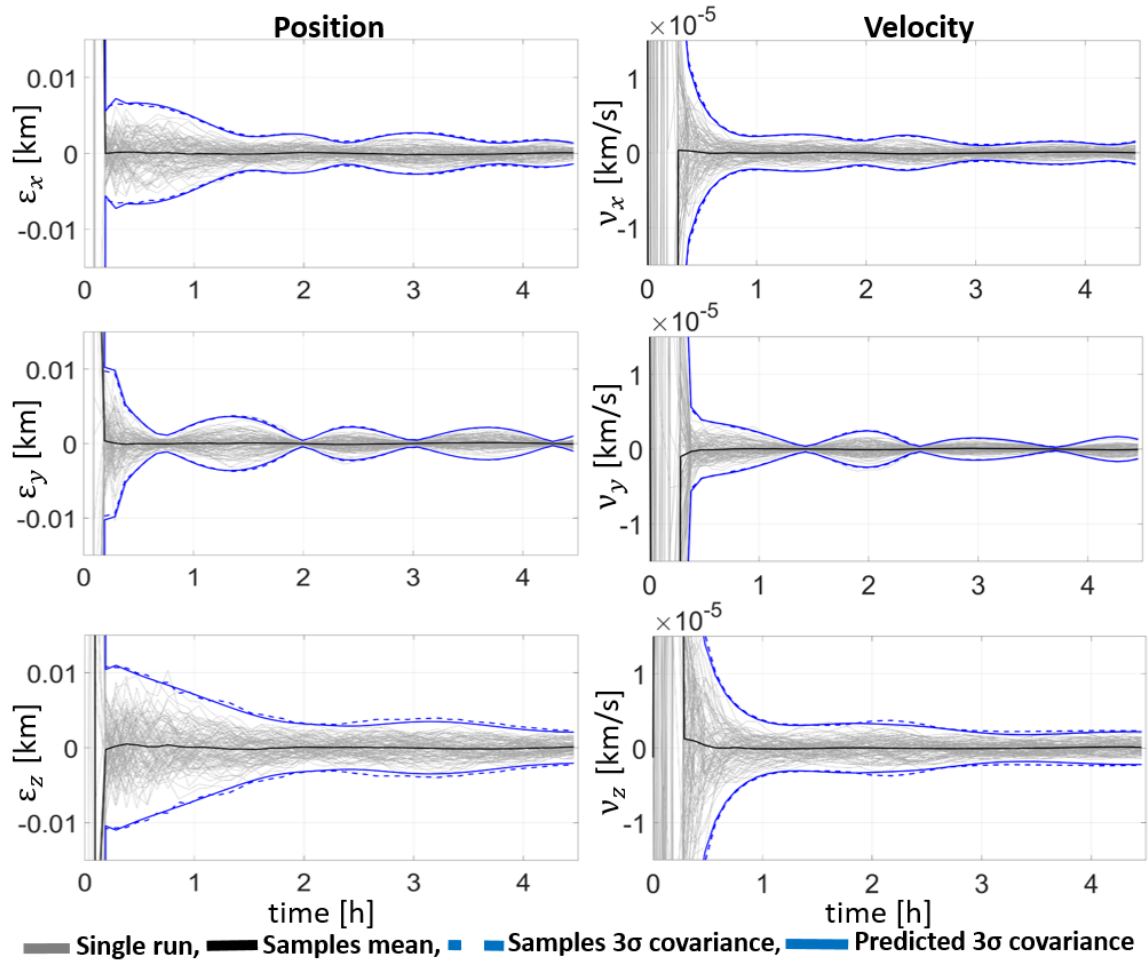


Fig. 7 Simulation results for DAHO-2 with $n = 100$ samples.

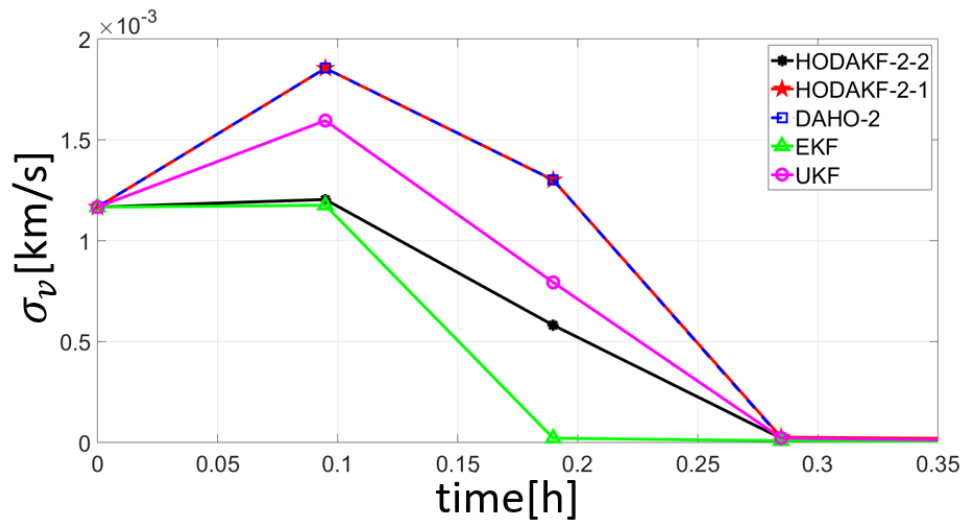


Fig. 8 Covariance comparison among different filters.

C. Angles-Only Orbit Determination

The orbit determination problem is now analyzed when the sensors provide only angular measurements and no range. A Monte Carlo analysis (100 runs) is performed with the four filters mentioned in the previous section. Figure 9 summarizes on the left the position error calculated as

$$\sigma_r = \sqrt{\sigma_{rx}^2 + \sigma_{ry}^2 + \sigma_{rz}^2} \quad (127)$$

and on the right the velocity error calculated with Equation (126). The errors are shown on a logarithmic scale for one orbit with 24 observations. Two different standard deviations for each filter are reported: the continuous lines represent the predicted standard deviation of the filter at the end of each update step, while the dashed lines represent the effective standard deviation of the Monte Carlo runs, calculated among all the samples at each time step. The EKF, red lines, diverges again, and the effective error level of the simulation settles two orders of magnitude over the predicted one, proving again the inconsistency of the filter. By looking at the green lines it can be noted how the UKF is also too confident on its own estimate: the predicted lines remain below the effective ones all the way until the very end. Both HODAKF-2-2 and DAHO-2 however, provide consistent estimates by matching the estimated and sample covariances. The two filters behaves similarly and, after a short transient, consistency is assessed. As before, HODAKF-2-2 converges faster than DAHO-2 and is also able to estimate the skewness and kurtosis of the error's distribution. In the next example we show a situation in which the quadratic update is not only superior in convergence rate, but also in steady-state performance.

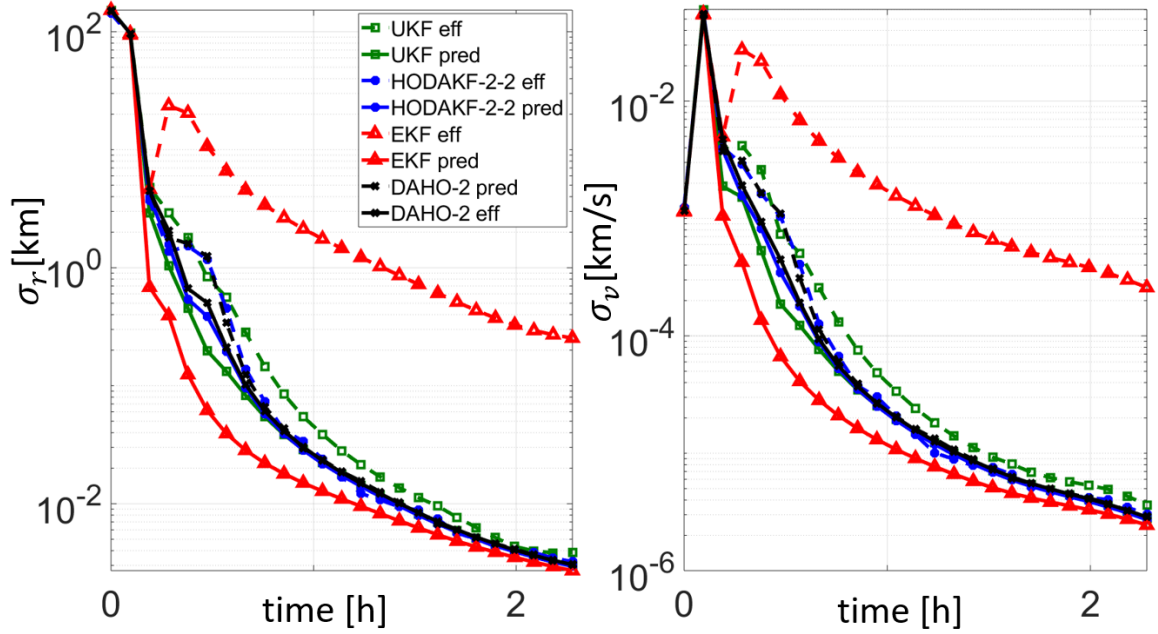


Fig. 9 Position and velocity covariance comparison among different filters. Angles measurements only.

D. Non-Symmetric Initial Conditions

Starting the state estimation problem with a Gaussian initial distribution uncertainties helps filters that rely (directly or indirectly) on the Gaussian assumption of the distribution, such as the UKF. However, the actual shape of the probability density in an orbit determination problem often resembles more a lens than a Gaussian bell. In this example the initial distribution of the state is that depicted by Figure 10.

Figure 10 shows the distribution of independent samples in position (left) and velocity (right) of the spacecraft being tracked at the initial simulation time. DAHO-2 and UKF approximate the initial pdf as Gaussian, and are initialized with the mean and covariance of the samples. HODAKF-2-2 is initialized with the additional knowledge of the third and fourth central moment of the distribution, which are calculated from samples.

The performance of the filters is shown in Figure 12 for HODAKF-2-2 and in Figure 11 for the UKF. The Monte Carlo simulation for DAHO-2 behaves similarly to the UKF, thus its graph has been omitted. By comparing the two figures, it can be noted how the UKF is not able to correctly

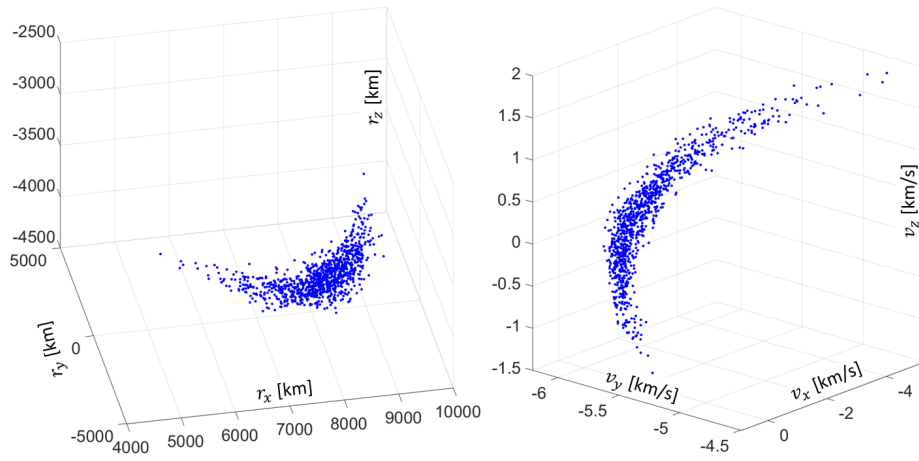


Fig. 10 Skewed initial distribution for position and velocity.

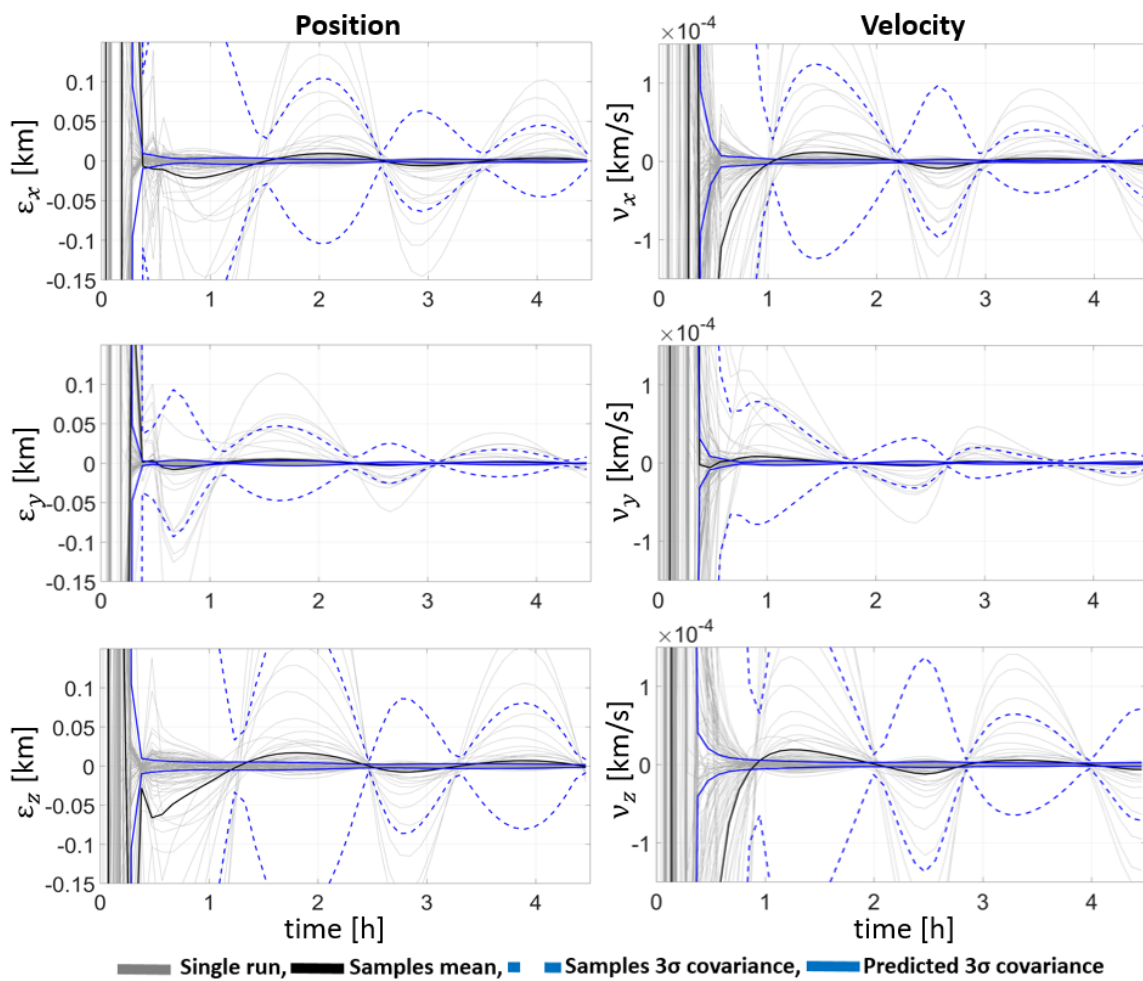


Fig. 11 Simulation results for UKF with $n = 100$ samples with a skewed initial condition.

predict the uncertainties of the state. By relying on the Gaussian assumption the UKF diverges and

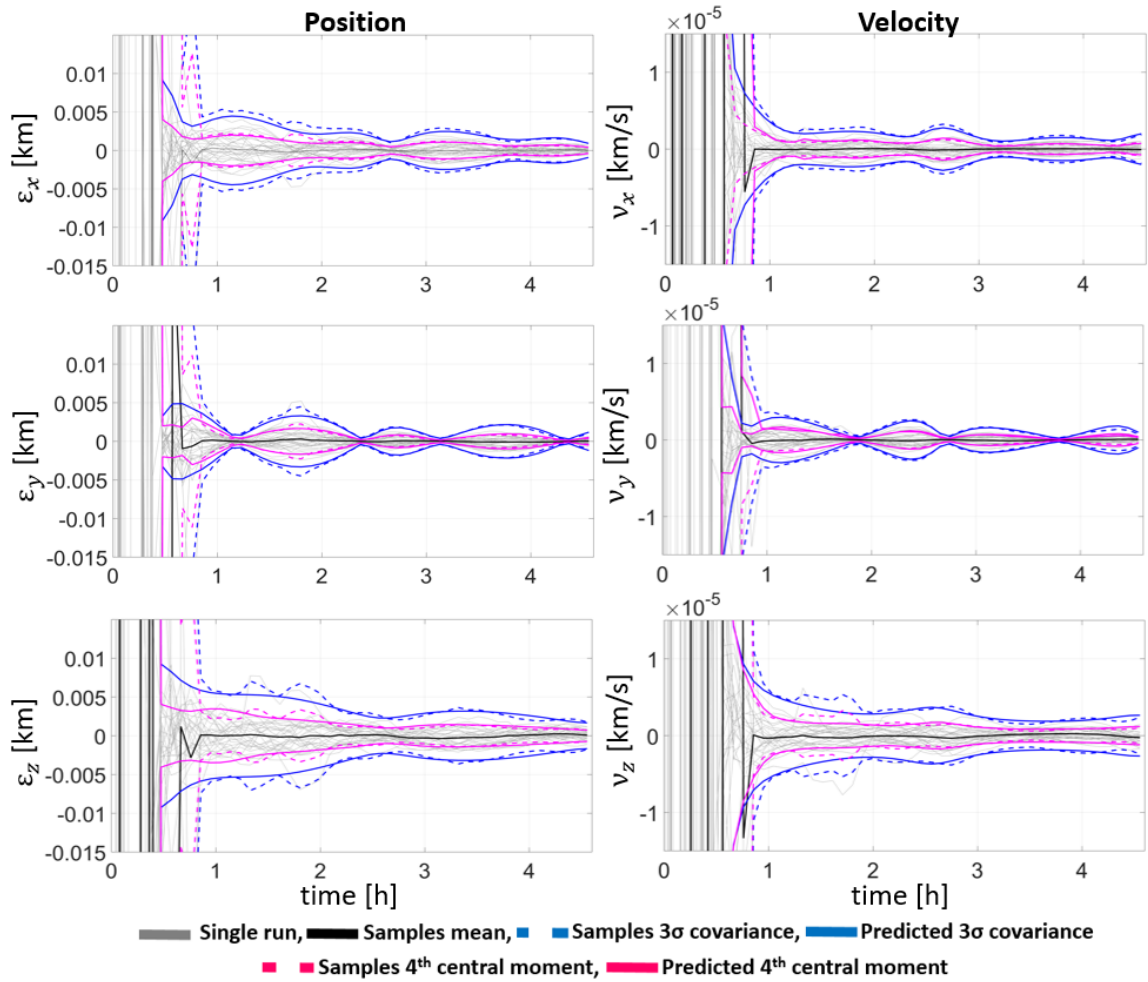


Fig. 12 Simulation results for HODAKF-2-2 with $n = 100$ samples with a skewed initial condition.

the effective error levels of the states are one order of magnitude bigger with respect to HODAKF-2-2. HODAKF-2-2 has more knowledge of the distribution and is able to use it to produce a better estimated. After few steps of transient behavior, the HODAKF-2-2 settles to steady state values and is able to predict its covariance (blue lines) and kurtosis (magenta lines).

Figure 13 compares the performance of the filters. Figure 13 shows the effective standard deviation, dashed lines, calculated from the samples runs, and the predicted standard deviation, continuous lines, estimated from the filters both for position (left) and velocity (right). Both the DAHO-2 and the UKF diverge and the effective covariance sets one order of magnitude over the predicted one: the two filters behave similarly and fail to achieve consistency. HODAKF-2-2, on the other hand, provides a more precise estimate and the filter converges. Through knowledge of

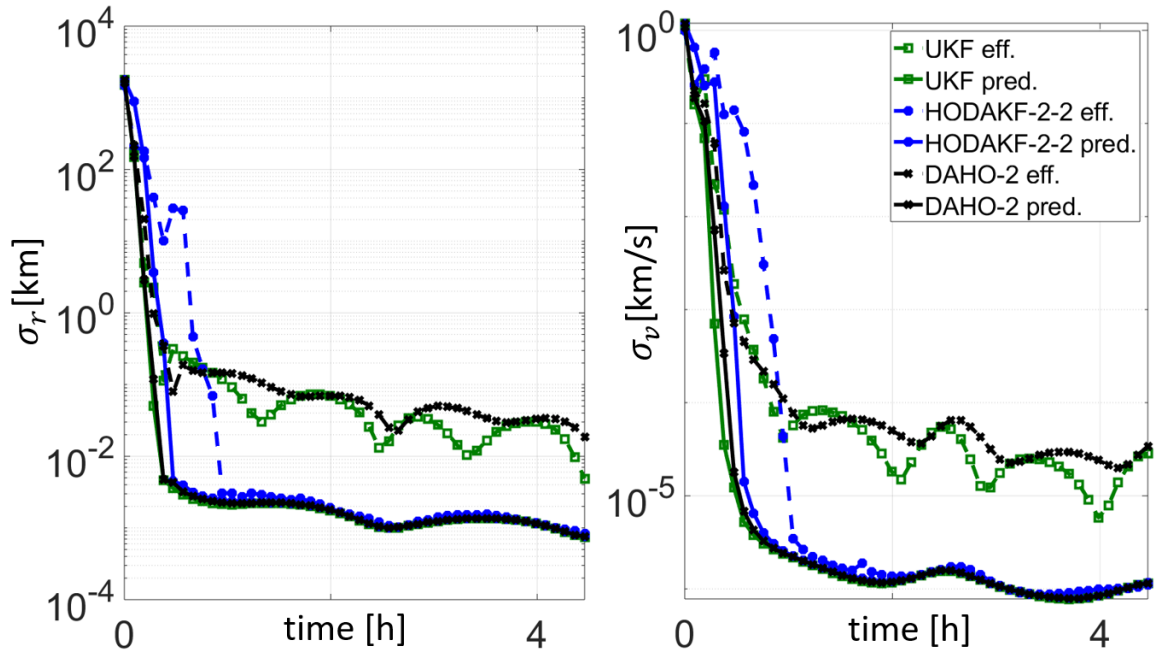


Fig. 13 Position and velocity covariance comparison among different filters. Skewed initial conditions.

higher order moments, HODAKF can perform a higher order update that improves the accuracy of the filter, making it robust to asymmetric pdfs. It should be noticed, however, that the extra performance comes at the cost of additional computations.

VIII. Conclusions

The high order differential algebra Kalman filter has been presented. The classic linear Kalman update is replaced by a polynomial update up to arbitrary order. Higher order Taylor series expansions of measurement and dynamic functions provide robustness against nonlinearities, as also done by other algorithms that improve over the classic extended Kalman filter (EKF) by directly or indirectly retaining higher-than-linear terms of the Taylor series. This work also proposes to perform the measurement update as a polynomial function, rather than as a linear function of the measurement. A polynomial update is shown to perform better than a linear one in terms of mean square error (MSE), especially in terms of convergence rate when the filter starts from high initial estimation errors. The high order approximation is also shown to provide more consistent covariance prediction over linear methods such as the EKF and the unscented Kalman filter (UKF) in highly

nonlinear, highly non-Gaussian systems.

The minimum MSE solution is theoretically known (the expectation of the state conditioned on the measurement) and is typically a nonlinear function of the measurement. Several simulated examples show that the proposed algorithm is able to outperform linear estimators both in terms of convergence and, especially, in terms of MSE performance. Simulated examples show that the proposed technique is able to handle system nonlinearities shown to make algorithms such the EKF and UKF diverge or, at a minimum, be inconsistent. The numerical examples also show that a nonlinear update results in a faster convergence rate.

IX. 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] Tapley, B. D., Schutz, B. E., and Born, G. H., *Statistical Orbit Determination*, Elsevier Academic Press, 2004, ISBN: 9780080541730.
- [4] Gelb, A., editor, *Applied Optimal Estimation*, The MIT press, Cambridge, MA, 1974, ISBN:9780262200271.
- [5] Junkins, J. and Singla, P., "How Nonlinear is it? A tutorial on Nonlinearity of Orbit and Attitude Dynamics," *Journal of the Astronautical Sciences*, Vol. 52, No. 1-2, 2004, pp. 7–60.
- [6] 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.
- [7] 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.

- [8] Fujimoto, K., Scheeres, D. J., and Alfriend, K. T., “Analytical Nonlinear Propagation of Uncertainty in the Two-Body Problem,” *Journal of Guidance, Control, and Dynamics*, Vol. 35, No. 2, 2012, pp. 497–509, doi: 10.2514/1.54385.
- [9] Park, R. S. and Scheeres, D. J., “Nonlinear mapping of Gaussian statistics: theory and applications to spacecraft trajectory design,” *Journal of Guidance, Control, and Dynamics*, Vol. 29, No. 6, 2006, pp. 1367–1375, doi: 10.2514/1.20177.
- [10] Park, R. S. and Scheeres, D. J., “Nonlinear Semi Analytic Methods for Trajectory Estimation,” *Journal of Guidance, Control, and Dynamics*, Vol. 30, No. 6, 2007, pp. 1668–1676, doi: 10.2514/1.29106.
- [11] Majji, M., Junkins, J., and Turner, J., “A High Order Method for Estimation of Dynamic Systems,” *The Journal of the Astronautical Sciences*, Vol. 56, No. 3, 2008, pp. 401–440, doi: 10.1007/BF03256560.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] Germani, A., Manes, C., and Palumbo, P., “Polynomial Extended Kalman Filter,” *IEEE Transactions on Automatic Control*, Vol. 50, No. 12, December 2005, pp. 2059–2064.
- [19] Berz, M., “The new method of TPSA algebra for the description of beam dynamics to high orders,” *Technical Report AT-6: ATN-86-16, Los Alamos National Laboratory*, 1986.
- [20] Berz, M., “The method of power series tracking for the mathematical description of beam dynamics,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. 258, No. 3, 1987, pp. 431–436.

- [21] Berz, M., “High-order computation and normal form analysis of repetitive systems,” *AIP Conference Proceedings*, Vol. 249, AIP, 1992, pp. 456–489.
- [22] Berz, M., “Differential Algebraic Techniques, Entry in Handbook of Accelerator Physics and Engineering,” M. Tinger and A. Chao (Eds.), World Scientific, 1999.
- [23] Lizia, P. D., *Robust Space Trajectory and Space System Design using Differential Algebra*, Phd thesis in aerospace engineering, XX Ciclo, March 2008.
- [24] Valli, M., Armellin, R., Di Lizia, P., and Lavagna, M., “Nonlinear mapping of uncertainties in celestial mechanics,” *Journal of Guidance, Control, and Dynamics*, 2013, doi: 10.2514/1.58068.
- [25] 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.
- [26] 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.
- [27] Rasotto, M., Morselli, A., Wittig, A., Massari, M., Di Lizia, P., Armellin, R., Valles, C., and Ortega, G., “Differential algebra space toolbox for nonlinear uncertainty propagation in space dynamics,” 2016.
- [28] Bar-Shalom, Y., Li, X. R., and Kirubarajan, T., *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*, Wiley, 2001.
- [29] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1st ed., 1965. Page 219.
- [30] 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.
- [31] Arasaratnam, I. and Haykin, S., “Cubature Kalman Filters,” *IEEE Transactions on Automatic Control*, Vol. 54, No. 6, June 2009, pp. 1254 – 1269, doi: 10.1109/TAC.2009.2019800.
- [32] 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.
- [33] 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.
- [34] Di Lizia, P., Massari, M., and Cavenago, F., “Assessment of onboard DA state estimation for spacecraft relative navigation,” 2017, Final report, ESA, Contract Nr. 400117860, GSP Ident: 16/R03, <https://>

[//www.esa.int/gsp/ACT/doc/ARI/ARI%20Study%20Report/ACT-RPT-MAD-ARI-16-5201-Assessment_of_onboard_DA_state_estimation_for_spacecraft_relative_navigation.pdf](http://www.esa.int/gsp/ACT/doc/ARI/ARI%20Study%20Report/ACT-RPT-MAD-ARI-16-5201-Assessment_of_onboard_DA_state_estimation_for_spacecraft_relative_navigation.pdf) .

- [35] 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.
- [36] 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.