AUTOMATIC DOMAIN PRUNING FILTER FOR OPTICAL OBSERVATIONS ON SHORT ARCS

G. Principe[(1)], L. Pirovano[(2)], R. Armellin[(3)], P. Di Lizia[(4)], and H.G. Lewis[(5)]

[(1)]*Astronautics Research Group, University of Southampton, UK, Email: gp1a14@soton.ac.uk*
[(2)]*Departamento de Matemáticas y Computación, Universidad de La Rioja, Spain, Email: laura.pirovano@unirioja.es*
[(3)]*Surrey Space Centre - University of Surrey, UK, Email: r.armellin@surrey.ac.uk*
[(4)]*Politecnico di Milano, Italy, Email: pierluigi.dilizia@polimi.it*
[(5)]*Astronautics Research Group, University of Southampton, UK, Email: H.G.Lewis@soton.ac.uk*

ABSTRACT

The uncertainty region associated with short arcs is typically large, making the initialization of orbit estimators a challenging task. In this work we propose a method to reduce the size of the uncertainty region using automatic domain pruning. The initial orbit and its confidence region are obtained by using a differential algebra-based initial orbit determination algorithm and a least squares algorithm. New measurements are used to reduce the size of the confidence region by retaining only those portions of the domain in which the likelihood is above a certain threshold.

Key words: initial orbit determinaion; differential algebra; uncertainty region.

1. INTRODUCTION

The estimated number of resident space objects (RSOs) larger than 1 cm is around $700,000$ objects [25, 34]. And they all must be seen on a regular basis and very accurately to perform orbit determination (OD). In the OD process it is customary to distinguish between initial orbit determination (IOD) and accurate orbit determination (AOD). The former is used for the computation of six orbital elements from six observations with no a priori knowledge of the spacecraft orbit. IOD provides the first estimate of the OD process and allows us to know which portions of the sky should be monitored to obtain additional observations and refine the OD process. The AOD, on the other hand, is used to improve a priori orbital elements from a large set of tracking data [24]. Thus, identifying observations belonging to the same object (identification problem [23]) will be one of the main challenges of OD. One requirement to perform reliable data association is to have realistic uncertainties for initial orbit solutions, which could also be used to initialize Bayesian estimators for orbit refinement [27].

OD of small objects is even more difficult, as observations can be characterised by long observational gaps due to observability constraints. It is thus important to be able to perform accurate OD with a single passage of the object above an observing station, when a short arc is observed. When an object is detected for the first time on a short arc, it is highly probable (depending on the observation strategy) that more than six observations are taken. The definition of short arc depends on the type of orbiting object, as well as the sensor used to observe it. For instance, at the telescope Fabra ROA Montsec (TFRM) facility, a short arc for optical observation of objects in geostationary Earth orbit (GEO) consists of 5 to 8 observations separated by 2 minutes. Thus, more observations than those required for IOD are available, but their distribution along the orbit is not the typical one of AOD (i.e. observations spread over several orbital revolutions). Indeed, due to the short interval between each detection, only partial information about the curvature of the orbit can be inferred: this means that these observations lead to large uncertainty about the orbit of the object.

One approach to deal with short arcs is based on attributables and admissible regions [22]. An attributable, is a 4-dimensional vector containing the information available from a short arc. In the case of optical observations, an optical attributable is made of two angles (e.g. right ascension and declination) and their angular rates that are computed exploiting the marginal observations of the short arc, $\mathcal{A} = \left(\alpha, \delta, \dot{\alpha}, \dot{\delta}\right)$. Independently from the number of observations acquired for a newly detected object, only four quantities are retained in the attributable. As a result, the orbit is undetermined in the range ρ, range-rate $\dot{\rho}$ space. The admissible region lies in the $2D$ plane generated by the two degrees of freedom of the attributable and its boundaries are defined by some physical constraints such as the orbital energy and minimum/maximum distance from the Earth. Recently, a new feature of the admissible region has been proposed, by endowing it with statistical properties [11, 15, 36]. This resulted in the advantage of allowing for the inclusion of uncertainties in observations, measurements and timing.

Once an estimate of the object orbit has been computed, it is possible to know when to re-observe the object and update the previous estimate when new observations are acquired. Sequential estimators, or filters, update orbit knowledge without re-processing previous data: they use the statistical information from the previous processing and combines it with new measurements. The classical estimator used in orbit determination is the extended Kalman filter (EKF), which assumes a Gaussian distribution. In contrast, the particle filter (PF) is a completely nonlinear Bayesian estimator that can approximate any probability density function by using samples. The PF proves to outperform the EKF in terms of accuracy when the system is highly nonlinear. The price of its accuracy is a considerable computational effort. Indeed, particle filters suffer from the need to propagate a large number of samples.

In this work, we focus on the OD of objects observed on a short arc. Instead of using attributables and admissible regions to deal with short arcs, we attempt to solve a least squares (LS) problem, in which all the observations belonging to a same tracklet are used. The process is started with an initial guess provided by an IOD solver that uses the observations at the ends and in the middle of the observation arc. This strategy was chosen to exploit the full length of the short arc and obtain information about the curvature of the orbit. Then an arbitrary order LS solver is implemented to fully take advantage of differential algebra (DA) techniques [4, 7, 8]. The confidence region computed using the above-mentioned approach can be used to initialize a PF, which can cope with highly nonlinear systems (e.g. an orbiting body) and process sequential measurements yielding subsequent estimates of the state vector. However, the confidence region for short arcs is typically large and nonlinearities play a key role in its accurate representation. Thus, the number of particles required to describe this region is large and using a PF is particularly challenging. In this work we implement a technique to reduce such domains. The uncertainty region is propagated forward to the next available observation, using the automatic domain splitting (ADS) technique to keep the error of the propagated set below a prescribed threshold on the entire domain. As a result, a set of boxes in the phase space are available at the time of a new measurement. The evaluation of the likelihood function on each of the propagated boxes is used to prune away those boxes in which the probability of finding the object is below a given threshold. This procedure can be repeated until the uncertainty region is reduced to a dimension that allows for the use of a PF without a prohibitive computational cost.

Summarising, the proposed orbit estimator is made up of the following blocks:

1. **IOD algorithm** with DA, used for the computation of six orbital elements from six observations with no a priori knowledge of the spacecraft orbit;
2. **LS algorithm** with DA, that exploits a larger set of tracking data to improve the orbital elements computed by the IOD algorithm;
3. **ADS technique**, used to keep the error of the Taylor expansions, computed by DA techniques, below a certain threshold;
4. **Domain pruning algorithm**, which reduces the size of the confidence region computed by the LS algorithm. In so doing, the computational cost of the sequential estimator is alleviated;
5. **PF**, which processes one measurement at a time and yields subsequent estimates of the orbital parameters at the time of each measurement.

This work deals with blocks $1, 2$ and 4. The remainder of the paper is organized as follows. At first, IOD and LS methods are described, taking advantage of DA techniques. The classical computation of the confidence region is also presented. Then, the PF is introduced and the domain pruning is used to reduce the size of the initial confidence region, according to newly-received measurements. Finally, the performances of the proposed approach are assessed.

2. IOD

OD refers to the use of a set of techniques for estimating the orbits of objects and is typically divided into two phases. When the number of observations is equal to the number of unknowns, a nonlinear system of equations needs to be solved. This problem is known as IOD. When many more observations are taken over an orbit arc of adequate length, AOD can be performed [1]. This section presents an IOD algorithm that exploits DA. Before explaining the IOD algorithm, the DA framework is introduced.

2.1. The DA framework

DA supplies the tools to compute the derivatives of functions within a computer environment. More specifically, by substituting the classical implementation of real algebra with the implementation of a new algebra of Taylor polynomials, any function f of n variables is expanded into its Taylor polynomial up to an arbitrary order k with limited computational effort. In addition to basic algebraic operations, operations for differentiation and integration can be easily introduced in the algebra, thus finalizing the definition of the differential algebra structure of DA [5, 6]. In the remainder we indicate with $\delta\boldsymbol{x}$ the DA variables and with $\mathcal{T}_{\mathrm{f}}^{\mathrm{k}}(\delta\boldsymbol{x})$ the k^{th} order approximation of f.

2.2. DA solution of the initial orbit determination algorithm

The IOD algorithm takes as input the observation of an object and gives as output the truncated power series of the object state at the central time of the observation. To do so, an initial estimate of the object position at t_1, t_2 and t_3 is obtained through Gauss' algorithm in double precision [10]. The observatory state is assumed to be known, while the direction cosines are found though the observation angles:

$$\hat{\boldsymbol{\rho}}_{\mathrm{i}} = \begin{bmatrix} \cos\delta_{\mathrm{i}} \cos\alpha_{\mathrm{i}} \\ \cos\delta_{\mathrm{i}} \sin\alpha_{\mathrm{i}} \\ \sin\delta_{\mathrm{i}} \end{bmatrix} \tag{1}$$

where $i = 1,\ 2,\ 3$ refers to the observation instances.

At this point estimates for the position vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ are available in double precision. Now, the velocities have to be computed. Lambert's algorithm takes as input two position vectors and the Δt between them and gives as output the velocity vectors. This means that by computing Lambert's algorithm twice (from t_1 to t_2 and from t_2 to t_3) one should be able to retrieve the three state vectors. However, Gauss' algorithm does not ensure that the three estimated vectors define one unique orbit, thus it does not ensure that the two velocity vectors found at t_2 ($\boldsymbol{v}_2^-$ and $\boldsymbol{v}_2^+$) coincide, as can be seen in Figure 1.

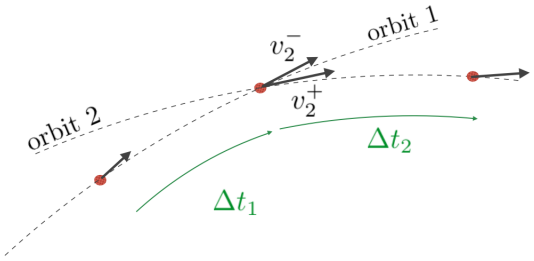


Figure 1. Output of Lambert algorithm taking as input the output from Gauss's algorithm

To fix this problem and obtain the expansion of the state, Lambert's algorithm will be used twice: the first time to find the $\delta\boldsymbol{\rho} = (\delta\rho_1,\ \delta\rho_2,\ \delta\rho_3)$ necessary to ensure that $\boldsymbol{v}_2^- = \boldsymbol{v}_2^+$ and the second one to expand the corrected solution $(\boldsymbol{r}_2,\ \boldsymbol{v}_2)$ with respect to the observation angle variations $\delta\boldsymbol{\alpha}$, $\delta\boldsymbol{\delta}$. The first step allows us to improve the estimation of the ranges made in Gauss' algorithm by forcing the three estimates to be part of the same orbit, while the last step allows us to analyze the variations in the state vectors due to variations in the observations just by means of function evaluations. Indeed, although the observations are our known values, they are not free from errors: sensor accuracy, timing accuracy and observer state knowledge all influence the observation, hence it is important to analyze the neighborhood of the solution.

For the first usage of Lambert's algorithm, the values $\rho_1,\ \rho_2,\ \rho_3$ are initialized as DA variables. Equation 2 shows the mathematical definition:

$$\begin{cases} \rho_1 = \rho_{1,Gauss} + \delta\rho_1 \\ \rho_2 = \rho_{2,Gauss} + \delta\rho_2 \\ \rho_3 = \rho_{3,Gauss} + \delta\rho_3 \end{cases} \tag{2}$$

In this way, the outputs of Lambert's algorithm are the velocity functions depending on variations of the slant ranges. In particular:

$$\boldsymbol{v}_2^- = \boldsymbol{v}_2^-\left(\delta\rho_1,\ \delta\rho_2\right) \tag{3}$$

$$\boldsymbol{v}_2^+ = \boldsymbol{v}_2^+\left(\delta\rho_2,\ \delta\rho_3\right) \tag{4}$$

With the goal of solving the discontinuity in t_2, the $\Delta\boldsymbol{v}$ between the left and right velocities is calculated:

$$\Delta\boldsymbol{v} = \boldsymbol{v}_2^+ - \boldsymbol{v}_2^- = \Delta\boldsymbol{v}\left(\delta\rho_1,\ \delta\rho_2,\ \delta\rho_3\right) \tag{5}$$

By forcing $\Delta\boldsymbol{v} = \boldsymbol{0}$ one wants to find the $\delta\boldsymbol{\rho}$ necessary to obtain it. Newton's method for DA [2] is used here. Thus, from

$$\Delta\boldsymbol{v}\left(\rho_1,\ \rho_2,\ \rho_3;\ \delta\rho_1,\ \delta\rho_2,\ \delta\rho_3\right) = \boldsymbol{0} \tag{6}$$

one obtains:

$$\begin{cases} \rho_{1,L1} = \rho_{1,Gauss} + \Delta\rho_1 \\ \rho_{2,L1} = \rho_{2,Gauss} + \Delta\rho_2 \\ \rho_{3,L1} = \rho_{3,Gauss} + \Delta\rho_3 \end{cases} \tag{7}$$

such that

$$\Delta\boldsymbol{v}\left(\rho_{1,L1},\ \rho_{2,L1},\ \rho_{3,L1};\ 0,\ 0,\ 0\right) = \boldsymbol{0} \tag{8}$$

At the end of this step, one has obtained the states of the object that satisfy the constraint of pertaining to a unique orbit. However, the solution is expanded with respect to the slant ranges. Lambert's algorithm is used again, initializing the observations as DA variables. The resulting solution will thus be expanded with respect to the observations. The non-constant parts of the angle polynomials are scaled by the precision of the observation $\sigma_{P,i}$, so that by evaluating the final solution within the interval $[-1, 1]$, one can find the 3σ interval solution depending on the accuracy. Equation 9 shows the definition:

$$\begin{cases} \alpha_1 = \hat{\alpha}_1 + \sigma_{\mathrm{P},1}\,\delta\alpha_1 \\ \alpha_2 = \hat{\alpha}_2 + \sigma_{\mathrm{P},2}\,\delta\alpha_2 \\ \alpha_3 = \hat{\alpha}_3 + \sigma_{\mathrm{P},3}\,\delta\alpha_3 \\ \delta_1 = \hat{\delta}_1 + \sigma_{\mathrm{P},1}\,\delta\delta_1 \\ \delta_2 = \hat{\delta}_2 + \sigma_{\mathrm{P},2}\,\delta\delta_2 \\ \delta_3 = \hat{\delta}_3 + \sigma_{\mathrm{P},3}\,\delta\delta_3 \end{cases} \tag{9}$$

At this point $\boldsymbol{\rho}$ is the output of the first application of Lambert's algorithm in double precision: $\boldsymbol{\rho} = \boldsymbol{\rho}_{L1}$. The mathematical grounds for this method are the first order Newton:

$$\Delta\boldsymbol{v}\left(\boldsymbol{\rho}\right) = \boldsymbol{0} \ \Rightarrow\ \boldsymbol{\rho}_{i+1} = \boldsymbol{\rho}_i - J^{-1}_{\Delta\mathbf{v}(\boldsymbol{\rho}_0)}\,\Delta\boldsymbol{v}(\boldsymbol{\rho}_i) \tag{10}$$

Here, the assumption is made that the Jacobian does not change in the loop. The iteration is carried out until $i =$ MaxOrder, thus until the highest order of the DA variable is reached. This procedure delivers $\mathcal{T}_{\boldsymbol{\rho}}^{\mathrm{k}}(\delta\boldsymbol{\alpha}, \delta\boldsymbol{\delta})$, which expresses how the slant ranges change as a function of the uncertainty of the observations.

Once the ranges are found depending on the observations, the position vectors are obtained with

$$\boldsymbol{r}_{\mathrm{i}} = \boldsymbol{R}_{\mathrm{i}} + \rho_{\mathrm{i}}\,\hat{\boldsymbol{\rho}}_{\mathrm{i}}, \qquad \text{where } i = 1, 2, 3$$

while the velocities can be calculated with one last Lambert's procedure.

An important outcome of this method is that one not only obtains the point solution (used as initial estimate of the LS, Sec. 3.1), but one can also easily calculate its 3σ variation by means of functions evaluations.

3. LEAST SQUARES

In this section, the basic formulation of OD as a nonlinear LS problem is summarised. First, we introduce the LS problem and classical methods employed for its solution and then we describe the uncertainty of the result as confidence ellipsoids.

3.1. Classic Formulation

The tracking of RSOs requires the solution of the OD problem. The goal is to determine the orbit of an object given some noisy observations. The orbit is described in terms of an n-dimensional state vector $\boldsymbol{x}$ at a reference epoch t_0. The state vector can be expressed in different ways: for example, as a position-velocity vector $(\boldsymbol{r}, \boldsymbol{v})$ in the Earth-centered inertial (ECI) reference frame, or in modified equinoctial elements (MEE) [33].

The standard approach to the OD problem is based on the LS method, devised by Gauss [16]. Starting from a tentative value $\boldsymbol{x} = \boldsymbol{x}(t_0)$, the observations are calculated at each observation epoch. Let $\boldsymbol{y}$ be an m-dimensional vector that contains the computed observations, $\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x})$, with m being the number of measurements. The nonlinear function $\boldsymbol{h}$ includes

- The propagation from t_0 to observation epochs;
- The conversion of the state vector into computed observations.

Due to sensor noise, the observations $\boldsymbol{y}$ differ from the actual ones, $\boldsymbol{y}_{\mathrm{obs}}$ (a vector with the same dimension as $\boldsymbol{y}$): the differences are called *residuals*.

Let $\boldsymbol{\xi} = \boldsymbol{y}_{\mathrm{obs}} - \boldsymbol{y}$ be the m-dimensional vector containing the residuals. The LS solution is the state $\boldsymbol{x}^*$ that minimizes the target function

$$J(\boldsymbol{x}) = \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{\xi}(\boldsymbol{x}) \tag{11}$$

To find the minimum, stationary points of $J(\boldsymbol{x})$ need to be found, i.e.

$$\boldsymbol{x}_{\mathrm{st}} : \frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}_{\mathrm{st}}) = \boldsymbol{0}. \tag{12}$$

Two main difficulties arise when solving Eq. (**??**) [20]:

- It represents a system of nonlinear equations, generally without an explicit solution. An iterative method is thus needed;
- $\boldsymbol{x}_{\mathrm{st}}$ can be a minimum, maximum, as well as a saddle. Thus, to ensure that $\boldsymbol{x}_{\mathrm{st}}$ is a minimum, it is required that the Hessian of the target function in the stationary point, $H = \frac{\partial^2 J}{\partial \boldsymbol{x}^2}\boldsymbol{x}_{\mathrm{st}}$, is definite positive.

A system of nonlinear equations can be solved by using Newton's method. The method will converge when an appropriate choice for the initial estimate is used to start the process. This is generally provided by the solution of the IOD problem described in Sec 2. At a generic step i, the solution $\boldsymbol{x}_{\mathrm{i}}$ is available (or the IOD solution, when $i = 1$). The gradient of $J(\boldsymbol{x})$ can be expanded to first order around $\boldsymbol{x}_{\mathrm{i}}$.

$$\frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}) \approx \frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}_{\mathrm{i}}) + \frac{\partial^2 J}{\partial \boldsymbol{x}^2}(\boldsymbol{x}_{\mathrm{i}})(\boldsymbol{x} - \boldsymbol{x}_{\mathrm{i}}) \tag{13}$$

Then, the solution at the subsequent step is found by imposing (13) to be $\boldsymbol{0}$:

$$\boldsymbol{0} = \frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}_{\mathrm{i}}) + \frac{\partial^2 J}{\partial \boldsymbol{x}^2}(\boldsymbol{x}_{\mathrm{i}})(\boldsymbol{x}_{\mathrm{i+1}} - \boldsymbol{x}_{\mathrm{i}}), \tag{14}$$

Let F be a m x n matrix composed of the partial derivatives of the residuals with respect to the state vector components,

$$F = \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{x}}(\boldsymbol{x}_{\mathrm{i}}); \tag{15}$$

S be a three-index array of shape m x n x n, defined as

$$S = \frac{\partial^2 \boldsymbol{\xi}}{\partial \boldsymbol{x}^2}(\boldsymbol{x}_{\mathrm{i}}); \tag{16}$$

and C the n x n normal matrix,

$$C = F^{\mathrm{T}}F + \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{i}})S. \tag{17}$$

From the solution $\boldsymbol{x}_{\mathrm{i}}$, it is possible to compute the residuals $\boldsymbol{\xi}(\boldsymbol{x}_{\mathrm{i}})$ and then

$$\frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}_{\mathrm{i}}) = 2\boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{i}})F \tag{18}$$

$$\frac{\partial^2 J}{\partial \boldsymbol{x}^2}(\boldsymbol{x}_{\mathrm{i}}) = H = 2(F^{\mathrm{T}}F + \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{x}_{\mathrm{i}})S) = 2C \tag{19}$$

For methods used to correctly compute the design matrix F see [32]. In conclusion, from Eq. (14)-(17), the solution of the iterative method is

$$\boldsymbol{x}_{\mathrm{i}+1} = \boldsymbol{x}_{\mathrm{i}} - C^{-1}F^{\mathrm{T}}\boldsymbol{\xi} \tag{20}$$

Due to practical problems in computing the second derivatives in matrix S, the full Newton's method is typically avoided for the OD problem [19]. The quantity $\boldsymbol{\xi}^{\mathrm{T}}S$ in Eq. (17) is often neglected, indeed when the residuals are small, so is this term. This leads to the so called differential correction technique [20], a variant of Newton's method. However, as the differential correction does not compute the full Hessian of the cost function, it is impossible to ensure that the matrix is positive definite.

In the LS formulation the residuals are often weighted to take into account sensors with different accuracies. Let ξ'_{i} be the true residual and ξ_{i} the normalized one,

$$\xi_{\mathrm{i}} = \frac{\xi'_{\mathrm{i}}}{\sigma_{\mathrm{i}}},$$

with σ_{i} being the standard deviation. Assuming differential correction approximation and weighted residuals, Eq. (17) becomes

$$C = \frac{1}{\sigma_{\mathrm{i}}^2}F^{\mathrm{T}}F$$

Thus, a uniform weight does not matter in the solution, although it matters when describing the confidence region of the solution (i.e. uniform weights appear in the normal matrix that describes the uncertainty, while they disappear in Eq. 20) [20, 32].

3.2. Confidence Region

The solution of the LS $\boldsymbol{x}^*$ is the value of the state vector that minimizes the cost function, $J(\boldsymbol{x}^*) = J^*$. However, $\boldsymbol{x}^*$ does not represent the true orbit, which lies within a confidence region. In order to outline this region, let us consider the value of the target function J in a neighbourhood of $\boldsymbol{x}^*$,

$$J(\boldsymbol{x}) = J^* + \delta J(\boldsymbol{x}) \tag{21}$$

where $\delta J(\boldsymbol{x})$ is called the penalty. Throughout the paper δJ will be used to indicate the functional expression of the penalty, whereas ΔJ will be used when a numerical value is assigned to the penalty. The confidence region of the solution, Z, is defined as the region where δJ is less than the control value K^2. Thus, in terms of the penalty function,

$$Z(K) = \{\boldsymbol{x} \in \mathbb{R}^{\mathrm{n}} : \delta J(\boldsymbol{x}) \leq K^2\}. \tag{22}$$

An expression for $\delta J(\boldsymbol{x})$ can be obtained by expanding $J(\boldsymbol{x})$ around $\boldsymbol{x}^*$ and neglecting terms of order ≥ 3 for $\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^*$:

$$\begin{aligned} J(\boldsymbol{x}) &\approx J^* + \frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}^*)^{\mathrm{T}}\delta\boldsymbol{x} + \frac{1}{2}\delta\boldsymbol{x}^{\mathrm{T}}\frac{\partial^2 J}{\partial \boldsymbol{x}^2}(\boldsymbol{x}^*)\delta\boldsymbol{x} \\ &\approx J^* + \frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}^*)^{\mathrm{T}}\delta\boldsymbol{x} + \frac{1}{2}\delta\boldsymbol{x}^{\mathrm{T}}H\delta\boldsymbol{x}, \end{aligned} \tag{23}$$

where $H = \frac{\partial^2 J}{\partial \boldsymbol{x}^2}(\boldsymbol{x}^*)$. As $\frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x}^*) = \mathbf{0}$, it follows

$$J(\boldsymbol{x}) \approx J^* + \frac{1}{2}\delta\boldsymbol{x}^{\mathrm{T}}H\delta\boldsymbol{x}. \tag{24}$$

By equating Eq. (21) and (24), the confidence region definition becomes

$$\delta J(\boldsymbol{x}) \approx \frac{1}{2}\delta\boldsymbol{x}^{\mathrm{T}}H\delta\boldsymbol{x} = \delta\boldsymbol{x}^{\mathrm{T}}C\delta\boldsymbol{x} \leq K^2. \tag{25}$$

Equation (25) represents the classical expression of the confidence region, with the Taylor expansion of the function $J(\boldsymbol{x})$ limited to 2$^{\mathrm{nd}}$ order. This expression can be further made explicit. According to the Eigen Decomposition Theorem, the normal matrix can be written as

$$C = VC_{\mathrm{d}}V^{-1} = VC_{\mathrm{d}}V^{\mathrm{T}}$$

where V is an orthogonal square matrix whose columns are the eigenvectors of C, and C_{d} is a diagonal matrix which contains the eigenvalues of C [14]. Since V is orthogonal, $V^{-1} = V^{\mathrm{T}}$. Eq. (25) becomes

$$\delta J(\boldsymbol{x}) = \delta\boldsymbol{x}^{\mathrm{T}}VC_{\mathrm{d}}V^{-1}\delta\boldsymbol{x} \tag{26}$$

By applying the transformation

$$\delta\tilde{\boldsymbol{x}} = V^{\mathrm{T}}\delta\boldsymbol{x},$$

the expression of the confidence region will be

$$\delta J(\boldsymbol{x}) = \delta\tilde{\boldsymbol{x}}^{\mathrm{T}}C_{\mathrm{d}}\delta\tilde{\boldsymbol{x}} \leq K^2 \tag{27}$$

In addition, because C is positive definite, C_{d} can be written as

$$C_{\mathrm{d}} = \begin{bmatrix} \frac{1}{\gamma_1^2} & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & \frac{1}{\gamma_n^2} \end{bmatrix},$$

and Eq. (27) becomes

$$\sum_{i=1}^{n}\frac{\delta\tilde{x}_{\mathrm{i}}^2}{\gamma_{\mathrm{i}}^2} \leq K^2 \tag{28}$$

The covariance matrix is $P = C^{-1}$, leading to

$$P_{\mathrm{d}} = C_{\mathrm{d}}^{-1} = \begin{bmatrix} \gamma_1^2 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & \gamma_n^2 \end{bmatrix}, \tag{29}$$

where $P = VP_{\mathrm{d}}V^{-1}$. In conclusion, as the penalty has a quadratic form, the confidence region is described by an ellipsoid whose axes can be determined by the eigen decomposition of the normal matrix (or its inverse, the covariance matrix).

A probabilistic interpretation can be given to the LS method. If the error of each observation is an independent random variable with normal distribution and zero mean, the solution of a LS problem is a random variable with a multivariate gaussian probability density function (pdf) $p_{\mathrm{x}}(\boldsymbol{x})$. In particular, $\boldsymbol{x}^*$ is the mean of the Gaussian distribution and the covariance matrix is the inverse of the normal matrix $P = C^{-1}$ [16]. Then, the LS solution can be statistically described by

$$p_{\mathrm{x}}(\boldsymbol{x}) = \frac{\sqrt{|C|}}{(2\pi)^{\mathrm{n}/2}} e^{-\frac{1}{2}\delta\boldsymbol{x}^{\mathrm{T}} C \delta\boldsymbol{x}} = \frac{\sqrt{|C|}}{(2\pi)^{\mathrm{n}/2}} e^{-\frac{1}{2}\delta J(\boldsymbol{x}^*)}. \tag{30}$$

where $|C|$ is the determinant of C. The contour levels of the penalty function are ellipsoids of equal probability and the confidence level can be obtained by properly selecting the value of the control value K^2. According to the F-test method [28], the value of K^2 for a confidence level of $100(1-\alpha)\%$ can be estimated by

$$\delta J(\boldsymbol{x}) \leq \frac{n}{m-n} J^* F^{\alpha}_{n,m-n} = K^2, \tag{31}$$

in which $F^{\alpha}_{n,m-n}$ is the upper α percentage point of the F-distribution.

3.3. Differential Algebra Least Squares

In this section we present a novel methodology to solve LS problems by applying DA techniques, obtaining a high order iterative procedure.

We start by describing a general algorithm to find the solution of a system of nonlinear equations in the DA framework. It is Newton's method for DA introduced in Sec. 2.1. The aim is to solve $\boldsymbol{g}(\boldsymbol{x}) = \mathbf{0}$. At a general step i:

1. Given the solution $\boldsymbol{x}_{\mathrm{i}}$ (from the previous step, or from the initial guess when $i = 1$), initialize the components of the state vector $\boldsymbol{x}_{\mathrm{i}}$ as k-th order DA variables:
$$\boldsymbol{x}_{\mathrm{i}} = \boldsymbol{x}_{\mathrm{i}} + \delta\boldsymbol{x}_{\mathrm{i}};$$
2. The evaluation of $\boldsymbol{g}$ in the DA framework delivers its k-th order Taylor expansion around $\boldsymbol{x}_{\mathrm{i}}$, $\boldsymbol{\mathcal{T}}^{\mathrm{k}}_{\mathrm{g}}(\delta\boldsymbol{x}_{\mathrm{i}})$. Thus, $\boldsymbol{g}$ will be the sum of a constant part $\boldsymbol{g}_{\mathrm{i}}$ (given by the solution of the last step, $\boldsymbol{g}(\boldsymbol{x}_{\mathrm{i}}) = \boldsymbol{g}_{\mathrm{i}}$) and an origin-preserving Taylor polynomial $\delta\boldsymbol{g}$, function of the DA variables $\delta\boldsymbol{x}_{\mathrm{i}}$
$$\boldsymbol{g}(\boldsymbol{x}) \approx \boldsymbol{\mathcal{T}}^{\mathrm{k}}_{\mathrm{g}}(\delta\boldsymbol{x}_{\mathrm{i}}) = \boldsymbol{g}(\boldsymbol{x}_{\mathrm{i}}) + \delta\boldsymbol{g}(\delta\boldsymbol{x}_{\mathrm{i}}) = \boldsymbol{g}_{\mathrm{i}} + \boldsymbol{\mathcal{T}}^{\mathrm{k}}_{\delta\boldsymbol{g}}(\delta\boldsymbol{x}_{\mathrm{i}}) \tag{32}$$
The following direct map is available
$$\delta\boldsymbol{g} \approx \boldsymbol{\mathcal{T}}^{\mathrm{k}}_{\delta\boldsymbol{g}}(\delta\boldsymbol{x}_{\mathrm{i}}); \tag{33}$$
3. Invert the map of Eq. (33), obtaining
$$\delta\boldsymbol{x}_{\mathrm{i}} \approx \boldsymbol{\mathcal{T}}^{\mathrm{k}}_{\delta\boldsymbol{x}_{\mathrm{i}}}(\delta\boldsymbol{g}); \tag{34}$$
4. Evaluate the inverse map in $-\boldsymbol{g}_{\mathrm{i}}$ to compute the correction $\Delta\boldsymbol{x}_{\mathrm{i}}$
$$\Delta\boldsymbol{x}_{\mathrm{i}} = \boldsymbol{\mathcal{T}}^{\mathrm{k}}_{\delta\boldsymbol{x}_{\mathrm{i}}}(-\boldsymbol{g}_{\mathrm{i}}) \tag{35}$$
$$\boldsymbol{x}_{\mathrm{i+1}} = \boldsymbol{x}_{\mathrm{i}} + \Delta\boldsymbol{x}_{\mathrm{i}} \tag{36}$$
5. Iterate until a convergence criterion is met or the maximum number of iterations is reached.

After convergence, the algorithm provides the solution of the set of nonlinear equations as well as the Taylor expansion of the function $\boldsymbol{g}(\boldsymbol{x})$ at high order around the solution $\boldsymbol{x}^*$.

The solution of the LS problem requires the stationary point of the cost function $J(\boldsymbol{x})$ to be found. If in the previous algorithm we set $\boldsymbol{g}(\boldsymbol{x}) = \frac{\partial J}{\partial \boldsymbol{x}}(\boldsymbol{x})$, then an arbitrary order solver of the LS problem is obtained, i.e. the differential algebra least squares (DALS) solver. The DALS solver has two main advantages with respect to the classical differential correction:

- As the objective function $J(\boldsymbol{x})$ is expanded up to an arbitrary order, we have the correct (full) expression of the Hessian matrix H. This can be used to check whether the stationary point is actually a minimum;
- The polynomial representation of the objective function $J(\boldsymbol{x})$ allows us to analytically represent it in a neighbourhood of the minimum. This feature enables the nonlinear representation of the solution confidence region [26].

As for all the iterative procedures, a convergence criterion needs to be defined. In our implementation we used two convergence criteria: the first one is based on the size of the correction $\Delta\boldsymbol{x}$, while the second one is based on the variation of the target function J. Thus, the iterative process is terminated when one of the two following conditions is met:

$$\begin{aligned} &\|\Delta\boldsymbol{x}\|_{\infty} \leq \epsilon_{\mathrm{x}} \\ &\Delta J \leq \epsilon_{\mathrm{J}} \end{aligned} \tag{37}$$

where ϵ_{x} and ϵ_{J} are established tolerances.

Although the algorithm presented in this section works at arbitrary order, the inclusion of terms above the second did not improve the convergence while significantly increasing the execution time. Thus, a second order DALS solver was used in this work. In contrast, the high-order representation of the confidence region proved to be more accurate than the classical 2^{nd}-order description, as shown in [26].

4. DOMAIN PRUNING FILTER

4.1. Sequential Estimator

The IOD and LS described in the previous sections are batch estimators, in that they obtain or improve an epoch

state estimate by processing a whole set of observations in each run. Sequential estimators, also called filters, aim to address a different problem: they process one measurement at a time and yield subsequent estimates of the state vector at the time of each measurement [24]. Consider we have observations up to time t_k, $\boldsymbol{Y}_k = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k\}$, then the goal is to infer the states $\boldsymbol{X}_k = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$. For this purpose, we are interested in computing the distribution $p(\boldsymbol{X}_k \mid \boldsymbol{Y}_k)$. Thus, filtering techniques make use of pdfs over possible values of the state vector, taking advantage of Bayes's theorem: previous knowledge about a phenomenon, the prior pdf, is updated to the posterior pdf $p(\boldsymbol{X}_k \mid \boldsymbol{Y}_k)$ when new information (i.e. new measurements) is available. The Bayes filter is the most general application of this theorem [12]. Because we are mainly interested in estimating the state vector at the time of the last measurement t_k without changing the estimates at previous time steps, we focus on the so-called filtering distributions $p(\boldsymbol{x}_k \mid \boldsymbol{Y}_k)$. Assuming Markov chain, we have [12]:

Prediction Step (Chapman-Kolmogorov equation)

$$p(\boldsymbol{x}_k \mid \boldsymbol{Y}_{k-1}) = \int p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1} \mid \boldsymbol{Y}_{k-1}) d\boldsymbol{x}_{k-1} \tag{38}$$

Updating Step

$$p(\boldsymbol{x}_k \mid \boldsymbol{Y}_k) = \frac{p(\boldsymbol{y}_k \mid \boldsymbol{x}_k) p(\boldsymbol{x}_k \mid \boldsymbol{Y}_{k-1})}{p(\boldsymbol{y}_k \mid \boldsymbol{Y}_{k-1})} \tag{39}$$

The previous procedure is called the Bayes filter algorithm. The initial distribution $p(\boldsymbol{x}_0 \mid \boldsymbol{Y}_0)$ is necessary to initialize the algorithm, with $\boldsymbol{Y}_0$ being the set of no measurements. That is, $p(\boldsymbol{x}_0 \mid \boldsymbol{Y}_0) = p(\boldsymbol{x}_0)$. $p(\boldsymbol{x}_0)$ can be computed by the LS technique (Eq. (30)).

Eqs. (38)-(39) only represent a conceptual solution in that, in general, they are intractable and cannot be determined analytically. This is the so-called *curse of dimensionality*, the phenomenon whereby virtually all integration methods deteriorate rapidly in performance as the dimension of the integral rises [13]. Thus, simplified variants are used (Fig. 2):

- **Gaussian filters**, which assume that both prior and posterior pdfs are represented by multivariate normal distributions. Once this assumption has been made, different representations are possible. The *moments representation* represents the Gaussian distribution by its mean and covariance, i.e. the first and second moment of a probability distribution. It is worth noting that all other moments are zero for normal distributions. An alternative representation is the *canonical representation*, also called *natural representation*, in which the Gaussian distribution is represented by an information matrix and an information vector [30];
- **Non-parametric filters**, which do not assume a fixed functional expression of the pdfs. Distributions do not have an analytical form, but are approximated by a discrete number of regions (*histogram filter*) or samples (*PF*).

Long-time propagation can make OD problems strongly nonlinear and non-Gaussian. Thus, the posterior pdf may be multi-peaked, heavily-tailed, or skewed [9]: the Kalman filter (KF) and all its variants may fail to provide an accurate description of the uncertainty of the state. This scenario is likely to occur when small objects are tracked and thus there are long gaps between observations. In this cases, the PF is a better solution. All involved pdfs are represented by particles, rather than a fixed functional form. Thus, the PF is able to capture all the moments of these pdfs. From particles it is also convenient to compute statistics such as the mode and the median, to obtain a full description of the distribution.

The goal of the PF is to obtain a set of samples (particles) approximately distributed as the posterior pdf $p(\boldsymbol{x}_\mathrm{k} \mid \boldsymbol{Y}_\mathrm{k})$ [17]. The PF is based on Monte Carlo methods, in which a probability distribution is represented by a set of random samples concentrated in regions of high probability density [9]. The accuracy of the representation increases by increasing the number of particles. When the number of samples tends towards infinity, the discrete representation becomes equivalent to the usual functional description of the posterior pdf.

Perfect Monte Carlo sampling assumes that samples are drawn directly from the posterior distribution. This is however seldom possible and therefore one requires a sampling technique, known as importance sampling (IS). Samples are drawn from an importance function $q(\boldsymbol{x}_\mathrm{k} \mid \boldsymbol{Y}_\mathrm{k})$, which approximates the posterior $p(\boldsymbol{x}_\mathrm{k} \mid \boldsymbol{Y}_\mathrm{k})$. The samples are then weighted by the normalized importance weights $w^{(\mathrm{i})}$, such that

$$\begin{aligned} w^{(\mathrm{i})} &\propto \frac{p(\boldsymbol{x}_\mathrm{k}^{(\mathrm{i})} \mid \boldsymbol{Y}_\mathrm{k})}{q(\boldsymbol{x}_\mathrm{k}^{(\mathrm{i})} \mid \boldsymbol{Y}_\mathrm{k})} \\ \sum_{\mathrm{i}=1}^{\mathrm{N}} w^{(\mathrm{i})} &= 1 \end{aligned} \tag{40}$$

with N being the number of samples. In order to apply the IS technique, the following is done:

- Draw samples $\bar{\boldsymbol{x}}_\mathrm{k} = \{\bar{\boldsymbol{x}}_\mathrm{k}^{(1)}, \ldots, \bar{\boldsymbol{x}}_\mathrm{k}^{(\mathrm{N})}\}$ from $q(\boldsymbol{x}_\mathrm{k} \mid \boldsymbol{Y}_\mathrm{k})$
- Create a discrete distribution over $\bar{\boldsymbol{x}}_\mathrm{k}$, placing mass $w^{(\mathrm{i})}$ on $\bar{\boldsymbol{x}}_\mathrm{k}^{(\mathrm{i})}$.
- Draw samples $\boldsymbol{x}^*$ from this distribution: these samples are approximately distributed according to $p(\boldsymbol{x}_\mathrm{k} \mid \boldsymbol{Y}_\mathrm{k})$

A formal proof can be found in [29]. The samples drawn from the importance function and their importance weights $\{\bar{\boldsymbol{x}}^{(\mathrm{i})}, w^{(\mathrm{i})}\}$ form the components of the IS [9].

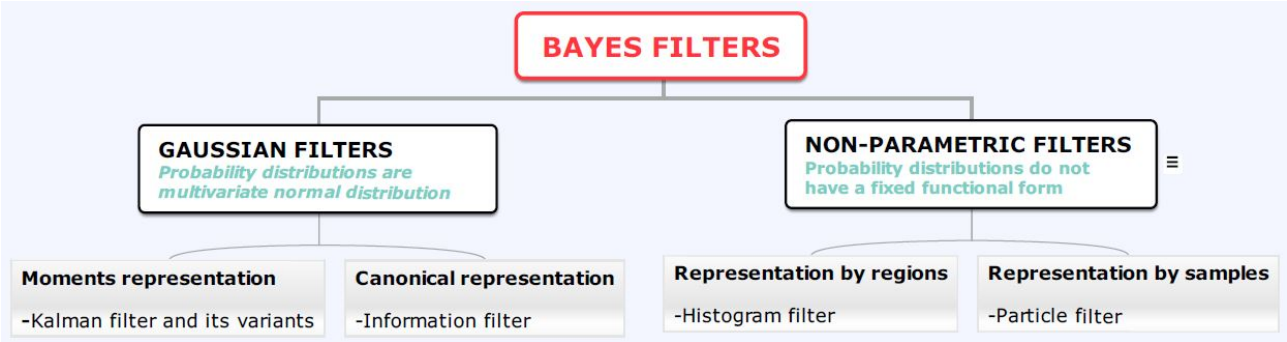


Figure 2. Overview on the types of approximations used to implement a Bayesian filters.

For the IS technique to be accurate, the importance distribution should be as similar as possible to the posterior distribution, so that an adequate number of samples from q falls in high probability regions of p.

The PF uses a recursive form of the IS method, the sequential importance sampling (SIS) [9]. This sampling technique assumes that the importance function is sequentially updated according to

$$q(\boldsymbol{x}_{\mathrm{k}} \mid \boldsymbol{Y}_{\mathrm{k}}) = q(\boldsymbol{x}_{\mathrm{k}} \mid \boldsymbol{x}_{\mathrm{k\text{-}1}}, \boldsymbol{Y}_{\mathrm{k}}) q(\boldsymbol{x}_{\mathrm{k\text{-}1}} \mid \boldsymbol{Y}_{\mathrm{k\text{-}1}}) \tag{41}$$

with $q(\boldsymbol{x}_0 \mid \boldsymbol{Y}_0) = q(\boldsymbol{x}_0) = p(\boldsymbol{x}_0)$ assumed available. Consequently

$$w_{\mathrm{k}}^{(i)} \propto \frac{p(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{Y}_{\mathrm{k}})}{q(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{Y}_{\mathrm{k}})} \propto \frac{p(\boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})} \mid \boldsymbol{Y}_{\mathrm{k\text{-}1}})}{q(\boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})} \mid \boldsymbol{Y}_{\mathrm{k\text{-}1}})} \frac{p(\boldsymbol{y}_{\mathrm{k}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}) p(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})})}{q(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})}, \boldsymbol{Y}_{\mathrm{k}})}$$

Hence, the importance weights are evaluated using [9]

$$w_{\mathrm{k}}^{(\mathrm{i})} \propto w_{\mathrm{k}}^{(\mathrm{i})} \frac{p(\boldsymbol{y}_{\mathrm{k}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}) p(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})})}{q(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})}, \boldsymbol{Y}_{\mathrm{k}})} \tag{42}$$

The most basic implementation of the PF makes use of the SIS. Because the importance function does not need to depend on the observations [3], a convenient choice for $q(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})}, \boldsymbol{Y}_{\mathrm{k}})$ is the transition prior,

$$q(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})}, \boldsymbol{Y}_{\mathrm{k}}) = p(\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} \mid \boldsymbol{x}_{\mathrm{k\text{-}1}}^{(\mathrm{i})})$$

Consequently, a simplified expression of the importance weights is obtained

$$w_{\mathrm{k}}^{(\mathrm{i})} \propto w_{\mathrm{k\text{-}1}}^{(\mathrm{i})} p(\boldsymbol{y}_{\mathrm{k}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}) \tag{43}$$

The PF implemented using Eq. (43) allows for a straightforward interpretation: it updates a sample from the prior (importance distribution) to a sample from the posterior through the likelihood function $p(\boldsymbol{y}_{\mathrm{k}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})})$. This is Bayes' theorem in terms of samples.

The drawback of using the transition prior is that, when the overlap between the prior distribution and the likelihood is small (see Fig. 3), most of the samples will be assigned negligible weights. This can happen frequently, because the prior is generally broader than the likelihood, due to increase in the uncertainty during the propagation compared to the accuracy of the measurements [31]. Thus, if only few samples are used in the approximation of the posterior distribution, the PF results are inaccurate. This phenomenon is known as degeneracy of the filter [3, 18]. The degeneracy problem can be counteracted by means of different approaches, such as the sequential importance resampling (SIR) [3] or usage of population-based optimizers [31, 37, 38]. The bottleneck of the PF is its computational cost, because propagation of a large number of particles is very time-consuming. To solve this problem, DA techniques [4] may be applied to the PF: particle propagation is substituted by polynomial evaluation, and the computational cost is considerably reduced. Reducing the large initial uncertainty region is of key importance, in order to alleviate the drawback of the PF. In this way, samples can be generated only in a portion of the uncertainty region, in particular where the value of $p(\boldsymbol{y}_{\mathrm{k}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})})$ is high. In so doing, the degeneracy problem is less likely to occur.

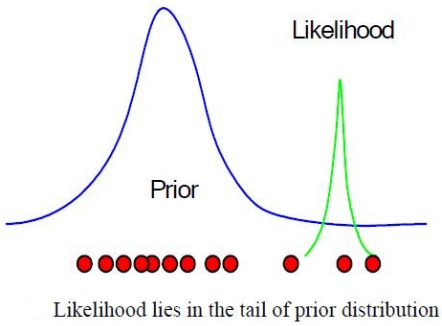


Figure 3. Failed scenario of particle filter sampling [31]

4.2. Domain Pruning

An initial pdf $p(\boldsymbol{x}_0)$ is necessary to initialize the PF. If the filter runs after a batch process (IOD and LS described in Secs. 2-3.3), $p(\boldsymbol{x}_0)$ will depend on the penalty δJ

(Eq. (30)). Thus, samples from $p(\boldsymbol{x}_0)$ can be obtained by sampling the confidence region described in section 3.2.

Although the confidence region associated with the LS solution is in general a n-dimensional region, in many cases of practical interest this region is stretched along one or two directions, called the weak directions. The weak directions are defined as the predominant directions of uncertainty in an orbit determination problem [21], i.e. the directions along which the penalty δJ is less sensitive to variations of the state vector. When the confidence region is highly stretched in two directions, it can be approximated as a $2D$ set. Sampling can be performed in the plane described by the two aforementioned directions, thus drastically reducing the number of samples needed. Let γ_1^2 and γ_2^2 be the two maximum eigenvalues of the covariance matrix $P = C^{-1}$, and $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ be the eigenvectors corresponding to γ_1^2 and γ_2^2. Samples $\tilde{\boldsymbol{x}}_0^{(\mathrm{i})}$ can be drawn in the plane $\boldsymbol{v}_1 - \boldsymbol{v}_2$ and then transformed into samples belonging to the full-dimension space $\boldsymbol{x}_0^{(\mathrm{i})}$,

$$\boldsymbol{x}_0^{(\mathrm{i})} = V\tilde{\boldsymbol{x}}_0^{(\mathrm{i})}, \tag{44}$$

with V being the matrix whose columns are the eigenvectors of P.

The initial uncertainty can be further reduced in accordance with new measurements. At time t_k new measurements $\boldsymbol{y}_{\mathrm{k}}^{\mathrm{obs}}$ are acquired. Let $\boldsymbol{y}_{\mathrm{k}}^{\mathrm{obs}}$ be an $m-$dimensional vector. Each sample $\boldsymbol{x}_0^{(\mathrm{i})}$ is then propagated to t_k, $\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})} = \boldsymbol{f}(\boldsymbol{x}_0^{(\mathrm{i})})$, with $\boldsymbol{f}$ being the function representing the dynamics. Assume independent observations $y_{\mathrm{k,j}}^{\mathrm{obs}}$, with $j = 1, \dots m$, and Gaussian distribution. The likelihood of each sample $p(\boldsymbol{y}_{\mathrm{k}}^{\mathrm{obs}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})})$ is given by

$$\begin{aligned} p(\boldsymbol{y}_{\mathrm{k}}^{\mathrm{obs}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}) &= \prod_{\mathrm{j}=1}^{\mathrm{m}} p(y_{\mathrm{k,j}}^{\mathrm{obs}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}) \\ p(y_{\mathrm{k,j}}^{\mathrm{obs}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}) &= \frac{\exp\left(-\frac{1}{2}\left[\frac{y_{\mathrm{k,j}}^{\mathrm{obs}} - y_{\mathrm{k,j}}^{\mathrm{com}}}{\sigma_{\mathrm{j}}}\right]^2\right)}{\sigma_j\sqrt{2\pi}} \end{aligned} \tag{45}$$

$\boldsymbol{y}_{\mathrm{k}}^{\mathrm{obs}}$ are actual observations, while $\boldsymbol{y}_{\mathrm{k}}^{\mathrm{com}}$ are computed observations, according to a particular value of the state vector, a particle $\boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}$. Then,

$$p(\boldsymbol{y}_{\mathrm{k}}^{\mathrm{obs}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}) = \frac{\exp\left(-\frac{1}{2}\sum_{\mathrm{j}=1}^{\mathrm{m}}\left[\frac{\boldsymbol{y}_{\mathrm{k}}^{\mathrm{obs}}(j) - \boldsymbol{y}_{\mathrm{k}}^{\mathrm{com}}(j)}{\sigma_{\mathrm{j}}}\right]^2\right)}{(\sqrt{2\pi})^{\mathrm{m}}\prod_{\mathrm{j}=1}^{\mathrm{m}}\sum_{\mathrm{j}}} \tag{46}$$

Thus, at time t_{k} we can define a scalar function

$$J_{\mathrm{k}} = \sum_{\mathrm{j}=1}^{\mathrm{m}}\left[\frac{y_{\mathrm{k,j}}^{\mathrm{obs}}(j) - y_{\mathrm{k,j}}^{\mathrm{com}}(j)}{\sigma_{\mathrm{j}}}\right]^2 \tag{47}$$

$$p(\boldsymbol{y}_{\mathrm{k}}^{\mathrm{obs}} \mid \boldsymbol{x}_{\mathrm{k}}^{(\mathrm{i})}) = \frac{\exp\left(-\frac{1}{2}J_{\mathrm{k}}\right)}{(\sqrt{2\pi})^{\mathrm{m}}\prod_{\mathrm{j}=1}^{\mathrm{m}}\sigma_{\mathrm{j}}} \tag{48}$$

Particles with small values of J_{k} have great values of the likelihood. Thus, these particles have significant weights (see Eq. (43)). DA techniques [4] can be used to approximately map samples $\tilde{\boldsymbol{x}}_0^{(\mathrm{i})}$ into J_{k}, by using a Taylor polynomial up to an arbitrary order. However, the accuracy of the approximation tends to decrease drastically when the initial confidence region is large and/or the propagation time is long, due to high nonlinearity of the dynamics [35]. Thus, a single polynomial expansion may be not sufficient to accurately represent the entire confidence region. Indeed, the validity of the expression depends on the truncation error of the polynomial approximation: if the truncation error in the domain of interest is above a required threshold, the polynomial cannot fully represent the domain. Thus, the initial region needs to be split into subdomains over which the Taylor expansion has the desired accuracy. This technique is called the ADS [35]. After applying the ADS, the resulting $2D$ confidence region is divided into boxes, each containing some samples $\tilde{\boldsymbol{x}}_0^{(\mathrm{i})}$. A threshold is used to cut away boxes where J_{k} is too big. This threshold can be computed by using statistical properties. Eq. (47) shows that J_{k} is the sum of the squares of m independent standard normal random variables: it is then distributed as a Chi-squared with m degrees of freedom. Thus, depending on the confidence level we want, the desired threshold can be chosen. The domain pruning procedure is beneficial to the PF for two reasons:

- A smaller confidence region can be described by fewer samples and the resulting computational cost of the PF is reduced;
- Sampling in a region of the initial uncertainty where the likelihood has high value prevents the degeneracy of the filter, because many particles have significant weights (Eq. (43)).

Summarising, when an object is observed on a short arc the IOD and LS algorithms find a solution to the OD problem. Nevertheless, the uncertainty region is very large and the computational cost of the PF may be prohibitive. The domain pruning algorithm approximates the uncertainty region as a $2D$ set and then reduces the region when new observations are acquired. As a result, the size of the uncertainty region reduces such that the PF can be used with a limited computational cost.

5. TEST CASES

Three different orbits were used as test cases: a low earth orbit (LEO), a GEO and a geostationary transfer orbit (GTO). For each object optical observations (i.e. right ascension and declination) were simulated. For objects in GEO and GTO the observation strategy of the TFRM

was simulated, while for the object in LEO the observation strategy of the Defence Science and Technology Laboratory (DSTL) was used. Finally, it is worth mentioning that Kepler's dynamics are considered throughout this section.

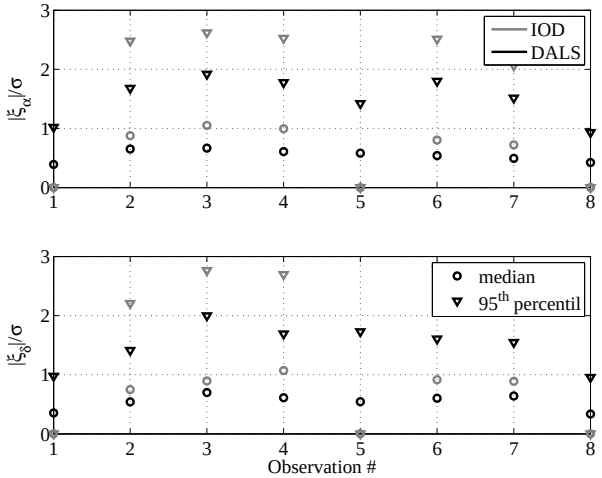
5.1. IOD and DALS Convergence Properties

Two scenarios of batch processes were simulated, with 8 or 15 observations. For each test case, 100 simulations were run in which synthetic observations were generated by adding Gaussian noise with zero mean to the ideal observations. Observations of objects in GEO (NORAD Catalog number 26824) and GTO (NORAD Catalog number 23238) were taken every 2 minutes with standard deviation $\sigma = 0.5$ arcsec, while observations of the object in LEO (NORAD Catalog number 04784) were taken every 7 seconds with standard deviation $\sigma = 5$ arcsec. When 8 observations were used, the observational arcs were 2.08 deg for the GTO, 2.82 deg for the LEO and 3.51 deg for the GEO; when 15 observations were considered, the arc lengths were 4.10 deg for the GTO, 5.65 deg for the LEO and 7.02 deg for the GEO.

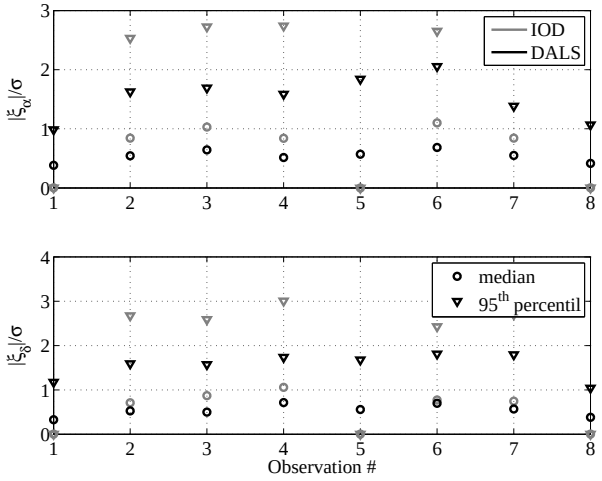
The DALS solver was used to estimate the orbit at the center of the observation window (at observation $\#5$ for the 8-observation scenario and $\#8$ for the 15-observation one) as this was found to maximize the algorithm performances and robustness. The tolerances ϵ_x and ϵ_J were chosen such that convergence was reached when one the following conditions was met

$$||\Delta \boldsymbol{x}||_\infty \leq \begin{cases} 1 \text{ m} & \text{for position} \\ 1 \text{ mm/s} & \text{for velocity} \end{cases} \qquad \Delta J \leq m \left(\frac{\sigma}{100}\right)^2.$$

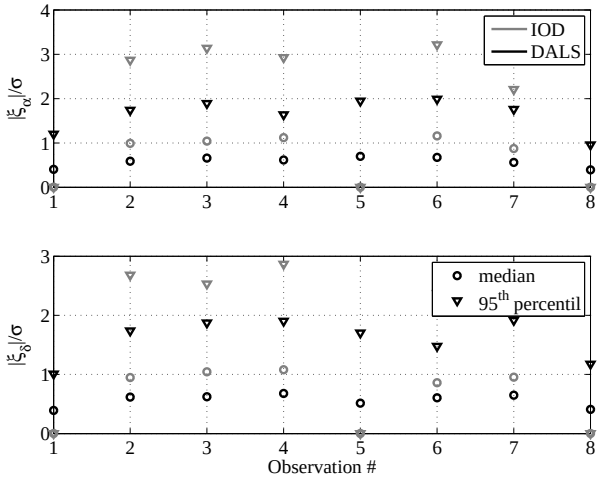
The DALS solver converged for 99.72% of the tests, taking on average only 2 iterations. Note that algorithm convergence does not provide any information on solution quality, e.g. convergence to a local minimum. In Tab. 1 the median of the absolute error with respect to the reference orbit in position (km) and velocity (km/s) is reported for the four test cases and different numbers of observations. On average the estimation errors of the DALS solution were lower than those associated with the IOD solution, showing that the inclusion of all the individual observations can improve the orbit estimation even for short arcs. In addition, the improvement in accuracy of the LS was greater when longer observational arcs were considered. Finally, as one may expect, the median of the errors decreased with the number of observations, i.e. the orbit estimation became more accurate for longer observational arcs. To analyse the solution accuracy without resorting to the true solution, which is unknown in a real scenario, the absolute values of the residuals scaled by the measurements σ were analyzed. The maximum of the median of the absolute values was found for each test case among 100 simulations. These values are reported in Tab. 2, which shows that the errors are compatible with measurement statistics.



(a) Object in LEO



(b) Object in GEO



(c) Object in GTO

Figure 4. Statistics of the absolute value of the dimensionless residuals on α and δ, for IOD and DALS solutions.

Table 1. Median of the absolute value of the error (with respect to the true solution) in position (km) and velocity (km/s), for IOD and DALS solutions

		Number of observations			
		8		15	
		IOD	DALS	IOD	DALS
LEO	Pos	5.8e+01	4.5e+01	1.5e+01	6.1e+00
	Vel	2.6e-01	1.9e-01	7.4e-02	3.1e-02
GEO	Pos	5.8e+02	4.5e+02	1.1e+02	6.9e+01
	Vel	4.4e-02	3.3e-02	8.4e-03	5.4e-03
GTO	Pos	4.8e+02	3.2e+02	9.7e+01	6.6e+01
	Vel	2.7e-02	1.9e-02	5.9e-03	3.4e-03

Table 2. Maximum of the median of the absolute values of the normalized residuals.

		Number of observations	
		8	15
LEO	$\xi_\alpha\ [1/\sigma]$	6.661e-01	8.085e-01
	$\xi_\delta\ [1/\sigma]$	6.980e-01	6.922e-01
GEO	$\xi_\alpha\ [1/\sigma]$	6.853e-01	7.066e-01
	$\xi_\delta\ [1/\sigma]$	7.117e-01	7.402e-01
GTO	$\xi_\alpha\ [1/\sigma]$	6.955e-01	7.250e-01
	$\xi_\delta\ [1/\sigma]$	6.790e-01	7.762e-01

Fig. 4 reports the results of the simulations when 8 observations were considered. The statistics of the absolute value of the dimensionless residuals are plotted and compared with the IOD solutions. Firstly, it is worth noting that the residuals of the IOD solutions were close to zero at the 1-st, 5-th, and 8-th observations, i.e. those used for the IOD, because IOD solutions are deterministic and exactly reproduce the available observations. On the other hand, the residuals significantly increased at other observation epochs. On average the residuals were much smaller and more uniformly distributed when LS solutions were analyzed. From this analysis it can be concluded that the LS solution improved the orbit provided by the IOD even when only few measurements distributed on short arcs were available.

5.2. Domain pruning results

The domain pruning was tested with an object in GEO (NORAD Catalog number 25126) and an object in GTO (NORAD Catalog number 25542). Observations were simulated by using Gaussian noise with zero mean and standard deviation $\sigma = 0.33$ arcsec. The initial uncertainty was obtained by applying the LS method to a batch of 15 observations 2 minutes apart. New observations of the object in GEO were acquired after $3, 24, 60$ hours, while new observations of the object in GTO were acquired after $1.5, 19, 72$ hours.

Table 3. Two objects in GEO, with similar orbital parameters, observed in the same observational window.

NORAD catalog number		36830	37816
Epoch	JED	2457402.69	2457402.69
a	km	42166.068	42166.258
e	-	0.000305	0.000424
i	deg	0.086	0.141
Ω	deg	111.495	58.467
ω	deg	185.691	214.801
M	deg	238.839	262.493

The confidence region was computed to ensure a confidence level of 99.9 percent. However, the two greatest eigenvalues of the covariance matrix γ_1^2, γ_2^2 were found to be at least 2 orders of magnitude larger than the others. Thus, the confidence region was approximated as a 2-D set $\pm\frac{K}{\gamma_i}\boldsymbol{v}_\mathrm{i}$, with $i = 1, 2$, as described in Sec. 4.2. This set was used as initial domain.

Figs. 5-6 display the sequential pruning of the initial domain, when new observations were acquired after a batch process of 15 observations, for objects in GEO and GTO respectively. A confidence level of 99.9% was chosen, leading to a threshold of $T = 13.816$ for the function J_k. When new observations were acquired, the initial domain was propagated from t_0 to t_k. Then, boxes in which the minimum value of J_k was greater than T were pruned away. Fig. 5 illustrates that for the object in GEO 87.5% of the domain was pruned away after 3 hours, 95.3125% after 24 hours and 99.61% after 60 hours. In contrast, Fig. 6 shows that for the object in GTO 75% of the domain was pruned away after 1.5 hours, 96.875% after 19 hours and 99.805% after 72 hours. Pruning proved to be more effective when time separation was longer. Finally, Fig. 5-6 illustrate that the reduced domain contained the true solution of the OD problem.

In a real-world scenario there is no guarantee that observations are of the same object, that is, they are correlated. In Fig. 7 the pruning algorithm is run with correlated and uncorrelated observations of objects in GEO. The uncertainty region of an object (NORAD Catalog number 36830) was computed by applying the LS algorithm to a batch of 15 observations 2 minutes apart. Then, the same object was re-observed after 3 hours and the pruning algorithm was applied. At the same time also observations of a different object (NORAD Catalog number 37816) but with similar orbital parameters were acquired and the pruning algorithm applied with these observations. The orbital parameters of the two objects are shown in Tab. 3. In the latter scenario the high value of J_k due to uncorrelated observations led to the whole initial domain being discarded. This result shows that the pruning procedure has potential for establishing whether observations are correlated and removing outlier observations.

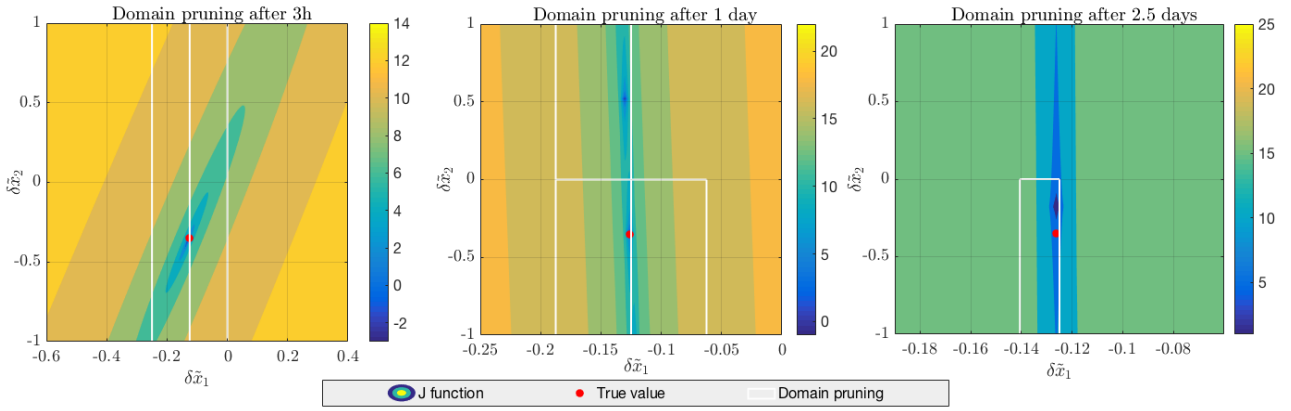


Figure 5. Sequential pruning of the domain for an object in GEO. The 2-D domain is defined by eigenvectors v_1 and v_2 associated with the two largest eigenvalues of the covariance matrix γ_1^2 and γ_2^2. The axes are scaled accordingly to γ_1^2 and γ_2^2. The colour map shows the value of J_k in logarithmic scale.

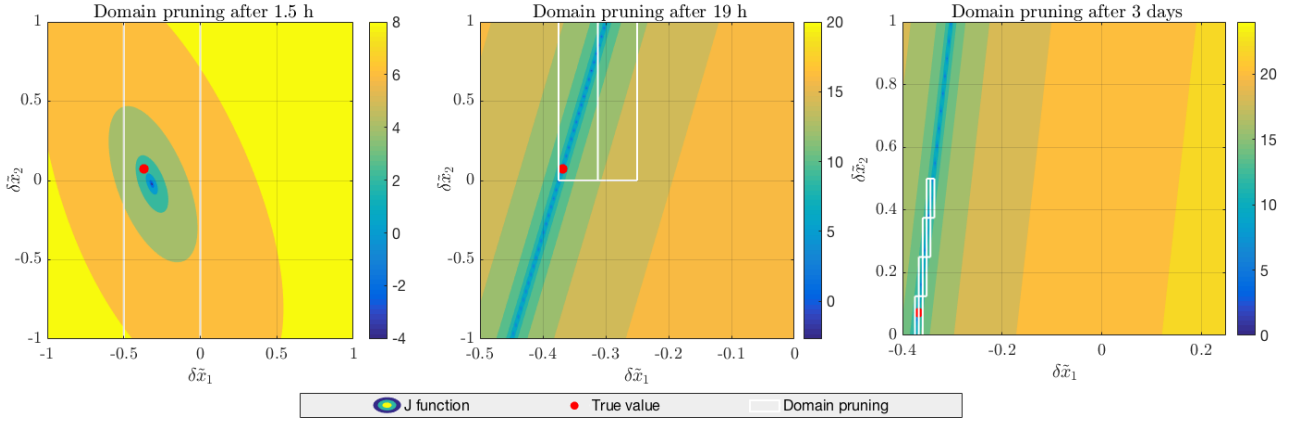


Figure 6. Sequential pruning of the domain for an object in GTO. The 2-D domain is defined by eigenvectors v_1 and v_2 associated with the two largest eigenvalues of the covariance matrix γ_1^2 and γ_2^2. The axes are scaled accordingly to γ_1^2 and γ_2^2. The colour map shows the value of J_k in logarithmic scale.

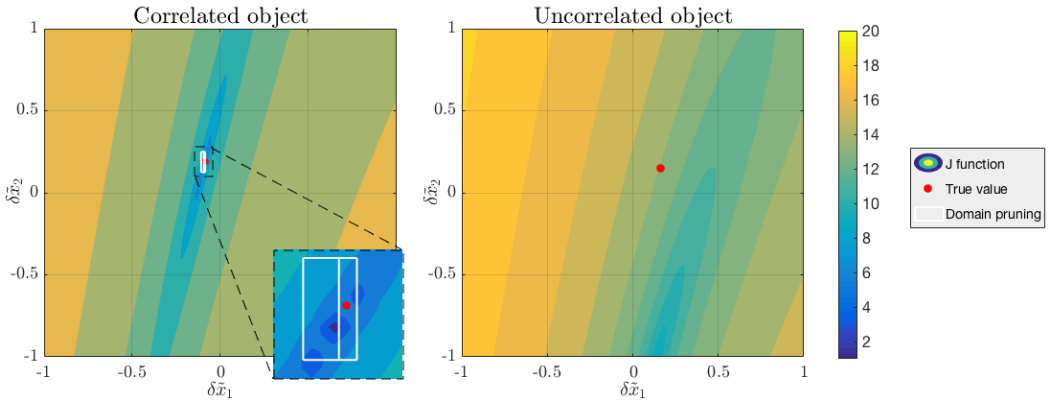


Figure 7. Comparison between the pruning due to correlated and not correlated observations acquired after 3 hours for an object in GEO.

6. CONCLUSIONS

In this work we focused our investigation on the OD problem when a batch of optical observations was taken on short arcs, i.e. less than 10 deg. IOD and DALS algorithms were used to estimate the orbit and the uncertainty region. The observational conditions led to large uncertainty regions. Thus, new observations were used to reduce the size of the region, pruning away portions of the domain where the likelihood was below an established threshold.

The main findings of this work can be summarised as

- The formulation of a least squares problem and its solution via the DALS on average improved the solution made available by IOD;
- The confidence region was found to be highly stretched in two directions and thus it can be approximated as a 2-D set;
- Pruning due to new observations was more effective when time separation was longer;
- Uncorrelated observations led to the removal of the whole initial domain: the algorithm can be potentially used to identify uncorrelated observations.

In our future work we intend to use the resulting uncertainty region to initialize a PF.

ACKNOWLEDGMENTS

Gennaro Principe was funded by an EPSRC Doctoral Training Grant awarded by the Faculty of Engineering and the Environment of UoS. R. Armellin acknowledges the support received by the Marie Sklodowska-Curie grant 627111 (HOPT - Merging Lie perturbation theory and Taylor Differential algebra to address space debris challenges). This work has been partially funded by the Spanish Finance and Competitiveness Ministry under Project ESP2014-57071-R. Laura Pirovano acknowledges the EOARD contract number FA9550-15-1-0244 "Initial Orbit Determination Based on Propagation of Admissible Region With Differential Algebra".

REFERENCES

1. Armellin, R., Di Lizia, P. & Zanetti, R. Dealing With Uncertainties in Initial Orbit Determination. *Aas 15-734,* 1–19 (2015).
2. Armellin, R., Di Lizia, P. & Lavagna, M. High-order expansion of the solution of preliminary orbit determination problem. *Celestial Mechanics and Dynamical Astronomy* **112,** 331–352. ISSN: 09232958 (2012).
3. Bamann, C. & Hugentobler, U. *A particle filter for orbit determination of space debris based on mono- and multi-static laser ranging* English. in *Advances in the Astronautical Sciences* **158** (2016), 2735–2748.
4. Berz, M. *Modern Map Methods in Particle Beam Physics* (Academic Press, 1999).
5. Berz, M. The method of power series tracking for the mathematical description of beam dynamics. *Nuclear Instruments and Methods* **AA258,** 431 (1987).
6. Berz, M. *The new method of TPSA algebra for the description of beam dynamics to high orders* tech. rep. (Los Alamos National Laboratory, 1986).
7. Berz, M. *The new method of TPSA algebra for the description of beam dynamics to high orders. Los Alamos National Laboratory* tech. rep. (Technical, Report AT-6: ATN-86-16, 1986).
8. 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* **258,** 431–436 (1987).
9. Cheng, Y. & Crassidis, J. *Particle filtering for sequential spacecraft attitude estimation* in *AIAA Guidance, Navigation, and Control Conference and Exhibit* (2004), 5337.
10. Curtis, H. *Orbital Mechanics: For Engineering Students* (Elsevier Science, 2015).
11. DeMars, K. J. & Jah, M. K. Probabilistic Initial Orbit Determination Using Gaussian Mixture Models. *Journal of Guidance, Control, and Dynamics* **36,** 1324–1335. ISSN: 0731-5090 (2013).
12. Doucet, A., De Freitas, N. & Gordon, N. *Sequential Monte Carlo methods in practice* (Springer, 2001).
13. Evans, M. & Swartz, T. *Approximating integrals via Monte Carlo and deterministic methods* (OUP Oxford, 2000).
14. Franklin, J. N. *Matrix theory* (Courier Corporation, 2012).
15. Fujimoto, K. New Methods in Optical Track Association and Uncertainty Mapping of Earth-Orbiting Objects (2013).
16. Gauss, C. *Theoria motus corporum coelestium in sectionis conicis solem ambientum* (1809).
17. Gordon, N. J., Salmond, D. J. & Smith, A. F. M. *Novel approach to nonlinear/non-Gaussian Bayesian state estimation* in *IEE Proceedings F (Radar and Signal Processing)* **140** (1993), 107–113.
18. Mashiku, A., Garrison, J. & Carpenter, J. Statistical Orbit Determination using the Particle Filter for incorporating Non-Gaussian Uncertainties. *AIAA Astrodynamics Specialist ...* 1–12 (2012).

19. Milani, A. The Asteroid Identification Problem: I. Recovery of Lost Asteroids. *Icarus* **137,** 269–292 (1999).
20. Milani, A. & Gronchi, G. *Theory of Orbit Determination* English (Cambridge University Press, 2010).
21. Milani, A., Sansaturio, M. E., Tommei, G., Arratia, O. & Chesley, S. R. Multiple solutions for asteroid orbits: computational procedure and applications. *Astronomy & Astrophysics* **431,** 729–746 (2005).
22. Milani, A., Gronchi, G., de' Michieli Vitturi, M. & Knezevic, Z. Orbit determination with very short arcs. I admissible regions. *Celestial Mechanics and Dynamical Astronomy* **90,** 59–87 (2004).
23. Milani, A., Gronchi, G., Knezevic, Z., Sansaturio, M. E. & Arratia, O. Orbit determination with very short arcs. II. Identifications. *Icarus* **179,** 350–374. ISSN: 00191035 (2005).
24. Montenbruck, O. & Gill, E. *Satellite orbits: models, methods, and applications* (Springer, 2000).
25. Pechkis, D. L., Pacheco, N. S. & Botting, T. W. Statistical approach to the operational testing of space fence. *IEEE Aerospace and Electronic Systems Magazine* **31,** 30–39 (2016).
26. Principe, G, Armellin, R & Lewis, H. *Confidence Region of Least Squares Solution for Single-Arc Observations* in *Advanced Maui Optical and Space Surveillance Technologies Conference* (2016).
27. Schumacher, P. W., Sabol, C., Higginson, C. C. & Alfriend, K. T. Uncertain Lambert Problem. *Journal of Guidance, Control, and Dynamics* **38,** 1573–1584. ISSN: 0731-5090 (2015).
28. Seber, A. F. & Wild, C. J. *Nonlinear Regression* (John Wiley & Sons, 2003).
29. Smith, A. F. M. & Gelfand, A. E. Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician* **46,** 84–88 (1992).
30. Thrun, S., Burgard, W. & Fox, D. *Probabilistic robotics* (MIT press, 2005).
31. Tong, G., Fang, Z. & Xu, X. A Particle Swarm Optimized Particle Filter for Nonlinear System State Estimation. *2006 IEEE International Conference on Evolutionary Computation* **1,** 438–442 (2006).
32. Vallado, D. *Fundamentals of Astrodynamics and Applications* English (Space Technology Library, 2013).
33. Walker, M. J. H., Ireland, B. & Joyce, O. A set modified equinoctial orbit elements. *Celestial Mechanics and Dynamical Astronomy* **36,** 409–419 (1985).
34. Wilden, H. *et al. GESTRA - A phased-array based surveillance and tracking radar for space situational awareness* in *Phased Array Systems and Technology (PAST), 2016 IEEE International Symposium on* (2016), 1–5.
35. Wittig, A. *et al.* Propagation of large uncertainty sets in orbital dynamics by automatic domain splitting. *Celestial Mechanics and Dynamical Astronomy* **122,** 239–261 (2015).
36. Worthy III, J. L. & Holzinger, M. J. Incorporating Uncertainty in Admissible Regions for Uncorrelated Detections. *Journal of Guidance, Control, and Dynamics* **38,** 1673–1689. ISSN: 0731-5090 (2015).
37. Zhang, X., Hu, W., Maybank, S., Li, X. & Zhu, M. *Sequential particle swarm optimization for visual tracking* in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (2008), 1–8.
38. Zhang, X., Hu, W. & Maybank, S. *A smarter particle filter* in *Asian Conference on Computer Vision* (2009), 236–246.