

Robust Bayesian Particle Filter for Space Object Tracking Under Severe Uncertainty

Cristian Greco* and Massimiliano Vasile[†]
University of Strathclyde, Glasgow, Scotland, G1 1XJ, United Kingdom

This paper introduces a robust Bayesian particle filter that can handle epistemic uncertainty in the measurements, dynamics, and initial conditions. The robust filter returns robust bounds on the output quantity of interest, rather than a crisp value. Particles are generated with an importance sampling technique and propagated only one time during the estimation process. The proposal distribution is constructed by running a parallel Unscented Kalman Filter to drive particles in regions of high expected likelihood and achieve a high effective sample size. The bounds are then computed by an inexpensive tuning of the importance weights via numerical optimisation. A Branch & Bound algorithm over simplexes with a Lipschitz bounding function is employed to achieve guaranteed convergence to the lower and upper bounds in a finite number of steps. The filter is applied to the robust computation of the collision probability of SENTINEL 2B with a FENGYUN 1C debris in different operational instances, all characterised by a mix of aleatory and epistemic uncertainty on initial conditions and observation likelihoods.

Nomenclature

b	=	mixture weight
C, c	=	computational complexities
\mathbf{D}, \mathbf{d}	=	static parameters' random variable and its realisation
$\mathbb{E}, \underline{\mathbb{E}}, \overline{\mathbb{E}}$	=	expectation, lower and upper expectations
$\mathcal{F}, \tilde{\mathcal{F}}$	=	set of propagated states and its truncated polynomial representation
f	=	equations of motion
h	=	observation model
I_C	=	collision's indicator function
K	=	branching steps
\mathcal{L}	=	list of simplexes
L	=	Lipschitz constant
L_k	=	lower bound of minimum within simplexes

Presented as conference paper AAS 19-845 at the 2019 AAS/AIAA Astrodynamics Specialist Conference, Portland, ME, August 11-15, 2019.

*PhD Candidate, Aerospace Centre of Excellence, Department of Mechanical and Aerospace Engineering, James Weir Building, University of Strathclyde, c.greco@strath.ac.uk.

[†]Professor, Aerospace Centre of Excellence, Department of Mechanical and Aerospace Engineering, James Weir Building, University of Strathclyde, massimiliano.vasile@strath.ac.uk (Corresponding author). Member AIAA.

l	=	number of samples' sets
M	=	number of observations
N	=	number of particles
\mathcal{N}	=	normal distribution
n_d, n_x, n_y, n_λ	=	uncertain variables' dimensionality
n_{eff}	=	effective sample size
\mathcal{P}	=	set of probability density functions
p	=	probability density function
S	=	sphere
\mathcal{S}	=	simplex
T	=	wall time's regression model, s
t	=	time, s
$t_{\alpha, l-1}$	=	t distribution's $1 - \alpha$ two-sided critical value
U_k	=	upper bound of minimum within simplexes
w, \hat{w}	=	unnormalised and normalised importance weights
\mathbf{X}, \mathbf{x}	=	dynamical state random variable and its realisation
\mathbf{Y}, \mathbf{y}	=	observation random variable and its realisation
\mathbf{Z}, \mathbf{z}	=	generic random variable and its realisation
α	=	coefficient of polynomial representation
β	=	regression coefficients
δ_{DCA}	=	collision threshold, m
ε	=	sensor noise
$\hat{\theta}, \underline{\hat{\theta}}, \bar{\hat{\theta}}$	=	expectation estimator, lower and upper estimators
λ	=	epistemic parameter
$\underline{\lambda}$	=	argument of the expectation estimator's minimum
μ	=	distribution mean
μ_χ	=	estimator's mean
ξ	=	joint state and parameters' realisations
π	=	proposal distribution
Σ	=	distribution covariance
σ	=	simplex diameter
σ_χ^{1H}	=	estimator's standard deviation
ϕ	=	auxiliary function
χ	=	collection of particles
Ψ	=	polynomial basis
Ω_λ	=	epistemic domain

I. Introduction

Standard Bayesian state estimation techniques for space object tracking and collision avoidance require the definition of a single precise probability measure, either explicitly or implicitly [1]. Precise probability distributions are used

to model uncertainty both in the prior knowledge of state and in the observations. Consequently, the correctness of the posterior distribution is dependent on the quality of priors and likelihood [2].

The majority of methods for state estimation relies on the use of parametric distributions and simplifying assumptions to keep the analysis computationally tractable [3]. As an example, the Kalman Filter is based on the assumption (or approximation for nonlinear variants) of Gaussian distributions describing the state prior, the dynamical process noise, and the observation likelihood. Different methods exist to deal with more complex distributions, such as Particle Filter (PF) [4] or Gaussian mixture models [5]. However, they still rely on the availability of precise and complete information on the probability distributions associated with prior states and measurements. The performance of these estimation techniques depends on the underlying assumptions on the nature of the uncertainty they can handle.

However, in many real-world applications, it may be difficult to quantify uncertainty with precise probability distributions since it would require a perfect knowledge of all the factors which concur to the definition of such uncertainty and an abundance of data (e.g. full knowledge of the exact dynamics, full sensor characterisation, complete information on the source and quality of the measurements, etc.).

A proper treatment of the uncertainty affecting observations and dynamics is of primary importance when predicting collisions and conjunctions between operational satellites and space debris, in particular if a limited number of, possibly, low-quality observations are available or some information is missing. The majority of collision avoidance approaches employs Conjunction Data Message (CDM) as input to quantify the probability of conjunction. However, a standard CDM presents only an estimation of the covariance of the state at the Time of Closest Approach (TCA) without any information on the full uncertainty distribution [6]. Although several studies have focused on covariance realism to quantify and improve the covariance [7], they often produce a more vacuous estimate rather than a more accurate one [8]. A solely larger covariance at TCA, although more conservative, may produce diluted collision probability and therefore an underestimated risk [9].

In this context, the H-infinity filter [1, 10] may be a viable alternative as it does not require information on the noise statistics. However, this filter minimises the worst-case estimate error of a linear combination of the state only, whereas the collision probability is a highly nonlinear function of such a state. Furthermore, the symmetric positive definite weight matrices in the H-infinity cost function need to be precisely set, thus not allowing one to express epistemic uncertainty on their values. Recent works have investigated outer probability measures, and particularly possibility functions, to model partial uncertainty specifications for tracking resident space objects [11], collision assessment [12], and a generalisation of admissible region analysis [13].

General estimation methods to handle generalised uncertainty models have also been recently developed. Specifically, estimation approaches under p-box uncertainty [14], set-valued robust Kalman filters [15–17], convex polytopes to model epistemic sets [18], coherent lower previsions using closed convex sets of probabilities as model of imprecision [19] have been investigated. However, most of such methods in literature employ only closed convex sets of probability

distributions as epistemic models.

In order to overcome these limitations, this work presents a novel method for state estimation of dynamical systems under an imprecise specification of the probability distributions modelling the problem uncertainty [20]. The proposed state estimation formulation encompasses both aleatory and epistemic models for the uncertain components. The prior, likelihood, and transition distributions are set-valued rather than precisely specified. Specifically, set-valued hyperparameters are employed to parameterise the epistemic sets, making this model more general than the closed convex sets found in most methods in literature. Bayes inference is used to compute a posterior distribution leading to an infinite number of posteriors compatible with this set-valued specification.

An efficient particle filter is proposed to compute robust bounds encompassing all the posterior expectations on a given quantity of interest, The result is an interval-valued estimation of all likely values of the quantity of interest. This interval-valued estimation replaces the traditional single point estimate, which would be sensitive to the choice of prior, transition, and likelihood. The new filtering scheme can also be used to study how the filter output is affected by changes in input parameters of pre-defined families of distributions. We use here the term *particle filter* because the whole probability distribution is propagated through the dynamics by sampling, as it is commonly done in particle filters.

The paper will introduce one estimator for the expectation, based on sequential importance sampling, and one for the bounds together with an analysis of their properties. The proposed filter first performs a precomputation step with a standard PF, and then it carries out the bound computation by inexpensively tuning the precomputed particles weights. The PF uses a proposal constructed by a parallel Unscented Kalman Filter (UKF) to preserve the effective sample size and, therefore, enhance the estimator accuracy. The bounds are computed using a Branch & Bound (B&B) solver to ensure convergence to the global optimum. Theoretical results are presented to prove convergence in a finite number of steps to the required accuracy. In addition, confidence intervals are presented to quantify the accuracy of the bounds.

The new robust filter is tested on a case study in which a piece of debris from FENGYUN 1C is on a potential collision trajectory with the operational satellite SENTINEL 2B. Both aleatory and epistemic uncertainties are considered in the prior and likelihood. Four specific instances are analysed: a collision instance where no measurements are acquired; a collision instance where new measurements are acquired to refine the Probability of Collision (PoC) bounds; a non-collision one again with new measurements; a collision instance where future measurements are simulated rather than acquired to quantify their effect on the PoC estimates. The results of this approach are particularly relevant for space surveillance, where a crisp probability threshold is normally used to issue potential collision warnings.

The remainder of the paper is structured as follows. Section II presents the robust state estimation formulation, the solution approach with the B&B approach, and an analysis of the estimators' complexity. Next, Section III describes the satellite collision scenario and the conjunction probability estimation. Finally, Section IV closes the paper with some concluding remarks and future developments.

II. Robust Filtering

The state estimation problem addressed in this paper is a generalisation of the mixed continuous-discrete one [21]: the system state evolves according to a time-continuous differential equation, whereas indirect noisy observations are collected at discrete instants of time. Let $\mathbf{x} \in \mathbb{R}^{n_x}$ be the state of a dynamical system evolving according to a nonlinear differential equation

$$\dot{\mathbf{x}} = f(t, \mathbf{x}, \mathbf{d}), \quad (1)$$

where $\mathbf{d} \in \mathbb{R}^{n_d}$ are static parameters of the dynamical model. Let the true initial state \mathbf{x}_0 and the true value of the parameters \mathbf{d} be unknown realisations of the random variables \mathbf{X}_0 and \mathbf{D} respectively. These random variables together with Eq. (1) induce the state at a later time t_k to be a random variable itself \mathbf{X}_k . Moreover, the system is observed through indirect noisy measurements. The measurements $\mathbf{y}_k \in \mathbb{R}^{n_y}$, taken at time instances t_k for $k = 1, \dots, M$ with $t_k < t_{k+1}$, is a known realisation of a random variable \mathbf{Y}_k governed by a stochastic measurement model

$$\mathbf{y}_k = h(t_k, \mathbf{x}_k, \boldsymbol{\varepsilon}_k), \quad (2)$$

where $\boldsymbol{\varepsilon}_k$ is a stochastic term modelling sensor noise.

A. Imprecise Formulation

In the case of epistemic uncertainty, the probability distributions cannot be assumed to be known precisely, but they are only specified within an imprecise set \mathcal{P} . Within these sets of distributions, no judgement is made about their relative likeliness. This definition allows one to model more faithfully uncertain scenarios in which information is too scarce to specify a single distribution.

In this work, we will focus on parametric distributions with epistemic uncertainty on their parameters $\boldsymbol{\lambda} \in \mathbb{R}^{n_\lambda}$. Thus, for a generic random variable \mathbf{Z} , the imprecise set is defined

$$\mathcal{P}_{\mathbf{Z}} = \{p(\mathbf{z}; \boldsymbol{\lambda}) \mid \boldsymbol{\lambda} \in \Omega_{\boldsymbol{\lambda}}\}, \quad (3)$$

where the epistemic parameter domain $\Omega_{\boldsymbol{\lambda}}$ is a compact subset of \mathbb{R}^{n_λ} . The pure aleatory structure of classic state estimation problems is a special case when $\Omega_{\boldsymbol{\lambda}}$ is a singleton, that is, when information is sufficient to identify a single distribution.

In this generalised scenario, the probabilistic continuous-discrete estimation problem can be written to account for

the new epistemic component as

$$\begin{cases} \mathbf{X}_0 \sim p(\mathbf{x}_0; \boldsymbol{\lambda}_0) & \in \mathcal{P}_{\mathbf{X}_0} & (4a) \\ \mathbf{X}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}; \boldsymbol{\lambda}_x) & \in \mathcal{P}_{\mathbf{X}_k | \mathbf{X}_{k-1}} & (4b) \\ \mathbf{Y}_k \sim p(\mathbf{y}_k | \mathbf{x}_k; \boldsymbol{\lambda}_y) & \in \mathcal{P}_{\mathbf{Y}_k | \mathbf{X}_k} \text{ for } k = 1, \dots, M, & (4c) \end{cases}$$

where $p(\mathbf{x}_0; \boldsymbol{\lambda}_0)$ is the initial density function, $p(\mathbf{x}_k | \mathbf{x}_{k-1}; \boldsymbol{\lambda}_x)$ is the transition density function, and $p(\mathbf{y}_k | \mathbf{x}_k; \boldsymbol{\lambda}_y)$ is the observation likelihood. $\boldsymbol{\lambda}_0 \in \Omega_{\boldsymbol{\lambda}_0}$, $\boldsymbol{\lambda}_x \in \Omega_{\boldsymbol{\lambda}_x}$ and $\boldsymbol{\lambda}_y \in \Omega_{\boldsymbol{\lambda}_y}$ are the epistemic parameters for the initial, transition, and likelihood distributions respectively. Hereafter, $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_x, \boldsymbol{\lambda}_y]$ is used to indicate the collection of the three epistemic parameters, and $\Omega_{\boldsymbol{\lambda}}$ the collection of their respective sets. Such uncertainty specification induces an infinite number of posterior distributions $p(\mathbf{x}_k | \mathbf{y}_{1:k}; \boldsymbol{\lambda})$ parameterised in $\boldsymbol{\lambda}$, which represent the complete solution of the robust estimation process.

In state estimation, the interest usually lies in a specific quantity of interest connected to the posterior rather than the full posterior itself, e.g. expected value, covariance, collision probability. Therefore, state estimation algorithms primarily focus on the computation of the expectation of a generic function $\phi : \mathbb{R}^{n_x(k+1)} \rightarrow \mathbb{R}$ of the uncertain state given the received measurements as

$$\mathbb{E}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}; \boldsymbol{\lambda}] = \int \phi(\mathbf{x}_{0:k}) p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}; \boldsymbol{\lambda}) d\mathbf{x}_{0:k}. \quad (5)$$

The value of such expectation depends on the epistemic parameter of the posterior, which is set-valued. Hence, the outcome of the robust estimation process is an interval, rather than a crisp value, whose bounds are the lower and upper values of the expectation of the quantity of interest conditional on the observations. The lower and upper expectations are defined as

$$\underline{\mathbb{E}}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}] = \min_{\boldsymbol{\lambda} \in \Omega_{\boldsymbol{\lambda}}} \mathbb{E}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}; \boldsymbol{\lambda}] \quad (6a)$$

$$\overline{\mathbb{E}}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}] = \max_{\boldsymbol{\lambda} \in \Omega_{\boldsymbol{\lambda}}} \mathbb{E}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}; \boldsymbol{\lambda}], \quad (6b)$$

that is the minimum and maximum of the expectation.

These bounds could be computed by a direct optimisation over the parameterised imprecise sets where each objective function evaluation requires an independent precise filtering run for each candidate epistemic parameter $\boldsymbol{\lambda}$ to evaluate. In this case, however, the optimisation would be inefficient because the expectation would be used as a black-box cost function, and therefore the computations performed during the filter call run would be discarded each time. Furthermore, if bounds on the expectation should be updated as new observations arrive, the optimisation would need to be rerun completely.

Hence, we propose a Robust Particle Filter (RPF) to efficiently compute bounds on the expectation of a given quantity of interest by exploiting a precomputation of the particles. Given the sequential nature of this filter, when a new observation is available, the bounds can be updated by still exploiting the precomputations performed at the previous step. The filter is characterised by two fundamental stages: the precomputation step; the bound computation one. These two fundamental steps will be discussed in greater detail in the next two sections.

B. Expectation Estimator

If no specific assumption or parameterisation is imposed on the posterior distribution, the expectation has no closed-form solution, and numerical techniques are required. In this paper we employ a precomputed Sequential Importance Sampling (pSIS) to construct an estimator for the expectation $\mathbb{E}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}; \boldsymbol{\lambda}]$.

1. Precomputed Sequential Importance Sampling

One of the main difficulties in computing the expectation (5) is that in most applications it is impractical to draw samples directly from the posterior distribution $p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}; \boldsymbol{\lambda})$, although it is possible to evaluate its density. To overcome this challenge, the integral in Eq. (5) is written by using a proposal distribution as

$$\mathbb{E}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}; \boldsymbol{\lambda}] = \int \phi(\mathbf{x}_{0:k}) \frac{p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}; \boldsymbol{\lambda})}{\pi(\mathbf{x}_{0:k} | \mathbf{y}_{1:k})} \pi(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}) d\mathbf{x}_{0:k} = \int \phi(\mathbf{x}_{0:k}) w(\mathbf{x}_{0:k}) \pi(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}) d\mathbf{x}_{0:k}, \quad (7)$$

under the condition that π , the so-called *proposal* distribution, has a larger support than $p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}; \boldsymbol{\lambda})$. The function $w(\mathbf{x}_{0:k})$ is the ratio between the target distribution and the proposal one functioning as a weight measuring their deviations. The proposal distribution should be chosen such that it is simple to draw samples from it.

Furthermore, in order to exploit the sequential nature of tracking problem (4), the posterior can be decomposed as

$$\begin{aligned} p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}; \boldsymbol{\lambda}) &\propto p(\mathbf{y}_k | \mathbf{x}_{0:k}, \mathbf{y}_{1:k-1}; \boldsymbol{\lambda}) p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k-1}; \boldsymbol{\lambda}) \\ &= p(\mathbf{y}_k | \mathbf{x}_k; \boldsymbol{\lambda}) p(\mathbf{x}_k | \mathbf{x}_{0:k-1}, \mathbf{y}_{1:k-1}; \boldsymbol{\lambda}) p(\mathbf{x}_{0:k-1} | \mathbf{y}_{1:k-1}; \boldsymbol{\lambda}) \\ &= p(\mathbf{y}_k | \mathbf{x}_k; \boldsymbol{\lambda}_y) p(\mathbf{x}_k | \mathbf{x}_{k-1}; \boldsymbol{\lambda}_x) p(\mathbf{x}_{0:k-1} | \mathbf{y}_{1:k-1}; \boldsymbol{\lambda}), \end{aligned} \quad (8)$$

with initial condition $p(\mathbf{x}_{0:0} | \mathbf{y}_{1:0}; \boldsymbol{\lambda}) = p(\mathbf{x}_0; \boldsymbol{\lambda}_0)$. The first relationship stems from unnormalised Bayes' inference, where the normalisation constant $p(\mathbf{y}_k | \mathbf{y}_{1:k-1})$ is not computed as too expensive to evaluate. The second one results from the conditional independence of the observation \mathbf{y}_k given the state \mathbf{x}_k , and from writing the joint distribution $p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k-1}; \boldsymbol{\lambda})$ in terms of the conditional one. The last relation comes from the Markov property of the dynamical system, that is, conditional independence on the previous history of \mathbf{x}_k once \mathbf{x}_{k-1} is given. Also, the proposal

distribution should be chosen as

$$\pi(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}) = \pi(\mathbf{x}_k | \mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}) \pi(\mathbf{x}_{0:k-1} | \mathbf{y}_{1:k-1}), \quad (9)$$

with initial condition $\pi(\mathbf{x}_{0:0} | \mathbf{y}_{1:0}) = \pi(\mathbf{x}_0)$. With a proposal in this form, the samples can be drawn sequentially, that is, when the samples $\mathcal{X}_k = \{\mathbf{x}_k^{(1)}, \dots, \mathbf{x}_k^{(N)}\}$ at time t_k need to be drawn, the samples until t_{k-1} need not to be drawn again. The new collection of samples is formed as $\mathcal{X}_{0:k} = \{\mathcal{X}_{0:k-1}, \mathcal{X}_k\}$. Hence, the estimator is

$$\hat{\theta}(\mathcal{X}_{0:k}; \boldsymbol{\lambda}) = \sum_{i=1}^N \hat{w}_k^{(i)}(\boldsymbol{\lambda}) \phi(\mathbf{x}_{0:k}^{(i)}), \quad (10)$$

where samples $\mathbf{x}_k^{(i)}$ are drawn from $\pi(\mathbf{x}_k | \mathbf{x}_{0:k-1}, \mathbf{y}_{1:k})$ and $\mathbf{x}_{0:k} = \{\mathbf{x}_{0:k-1}, \mathbf{x}_k\}$ thanks to the sequential choice of the proposal (9), and $w_k^{(i)}(\boldsymbol{\lambda})$ are the *importance* weights, computed in a sequential fashion as

$$w_k^{(i)}(\boldsymbol{\lambda}) = \frac{p(\mathbf{y}_k | \mathbf{x}_k^{(i)}; \boldsymbol{\lambda}_y) p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}; \boldsymbol{\lambda}_x)}{\pi(\mathbf{x}_k^{(i)} | \mathbf{x}_{0:k-1}, \mathbf{y}_{1:k})} \hat{w}_{k-1}^{(i)}(\boldsymbol{\lambda}) \quad (11a)$$

$$\hat{w}_k^{(i)}(\boldsymbol{\lambda}) = \frac{w_k^{(i)}(\boldsymbol{\lambda})}{\sum_{j=1}^N w_k^{(j)}(\boldsymbol{\lambda})}, \quad (11b)$$

with $\hat{w}_0^{(i)}(\boldsymbol{\lambda}_0) = p(\mathbf{x}_0^{(i)}; \boldsymbol{\lambda}_0) / \pi(\mathbf{x}_0^{(i)}) / N$. The weight update rule in Eq. (11a) does not include the posterior normalisation constant, as it has been neglected in Eq. (8), therefore Eq. (11b) is the weight self-normalisation that ensures the estimator consistency, that is, it converges in probability to the true value with increasing sample size [22].

Such estimator brings an advantageous decoupling between the weights, dependent on the epistemic parameter, and the particle sampling, independent of the epistemic parameter. Therefore, once a set of particles has been drawn from a fixed proposal and propagated, the optimisations in Eq. (6) can be carried out by simply tuning the importance weights $w_k^{(i)}(\boldsymbol{\lambda})$. Hence, all the quantities which are independent of the epistemic parameter are precomputed before the optimisation. Thus, once the proposal distributions have been chosen, the particles, the proposal density evaluations and the function ϕ evaluations can be computed offline. The precomputations are outlined in Algorithm 1. A graphical representation is sketched in Fig. 1. First, the initial samples are drawn from the initial proposal distribution $\pi(x_0)$ and the initial densities are computed by evaluating the proposal (Line 1 of Algorithm 1 and Fig. 1(a)). Then, for each of the M observations, the particles are drawn from the transition proposal and the densities evaluated for such samples (Line 2 and Fig. 1(b)). The particles represent a discrete approximation of the true distribution (Fig. 1(c)). Finally, once the whole collection of particles $\mathcal{X}_{0:M}$ has been generated, the function of interest ϕ is evaluated, and the resulting values saved (Line 3).

Once this precomputation step is complete, the pSIS estimator is evaluated for a given epistemic parameter $\boldsymbol{\lambda}$ as

Algorithm 1 Precomputation for sequential importance sampling estimator.

Given:

- the estimation problem (4)
- the proposal distributions $\pi(\mathbf{x}_0)$ and $\pi(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_{1:k})$

1: Draw N particles from the initial distribution $\pi(\mathbf{x}_0)$ and evaluate density

$$\mathbf{x}_0^{(i)} \sim \pi(\mathbf{x}_0), \pi_0^{(i)} = \pi(\mathbf{x}_0^{(i)})$$

for $k = 1 : M$ **do**

2: Draw N particles from the proposal transition and evaluate density

$$\mathbf{x}_k^{(i)} \sim \pi(\mathbf{x}_k | \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_{1:k}), \pi_k^{(i)} = \pi(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_{1:k})$$

end for

3: Evaluate function ϕ for each particle path realisation

$$\phi^{(i)} = \phi(\mathbf{x}_{0:M}^{(i)})$$

Save collection of particles $\mathcal{X}_{0:M}$, densities $\pi_{0:M}^{(i)}$, and function evaluations $\phi^{(i)}$.

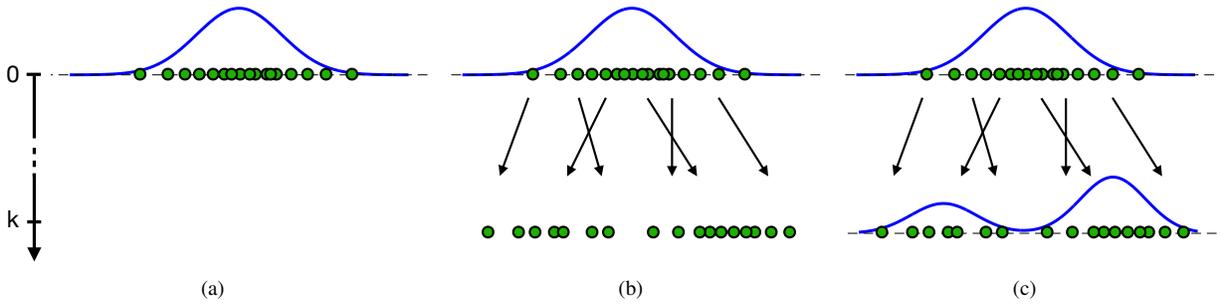


Fig. 1 Representation of precomputation step of sequential importance sampling estimator: a) initial samples, b) propagated samples, c) propagated distribution approximation.

described in Algorithm 2. Its graphical representation is shown in Fig. 2. The pSIS is evaluated by computing the new

Algorithm 2 Algorithmic scheme for precomputed sequential importance sampling estimator.

Given:

- the estimation problem (4)
- the precomputed quantities $\mathcal{X}_{0:M}$, $\pi_{0:M}^{(i)}$ and $\phi^{(i)}$
- epistemic parameter $\lambda = [\lambda_0, \lambda_x, \lambda_y]$ to evaluate

1: Set the weights of the N *precomputed* initial samples \mathcal{X}_0 and normalise them

$$w_0^{(i)} = p(\mathbf{x}_0^{(i)}; \lambda_0) / \pi_0^{(i)} / N$$

for $k = 1 : M$ **do**

2: Update the weights of the *precomputed* samples \mathcal{X}_k and self-normalise them

$$w_k^{(i)} = \hat{w}_{k-1}^{(i)} p(\mathbf{y}_k | \mathbf{x}_k^{(i)}; \lambda_y) p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}; \lambda_x) / \pi_k^{(i)}, \hat{w}_k^{(i)} = w_k^{(i)} / \sum_{j=1}^N w_k^{(j)}$$

end for

3: Evaluate the estimator with the updated weights $\hat{w}_k^{(i)}$

$$\hat{\theta}(\mathcal{X}_{0:M}) = \sum_{i=1}^N \hat{w}_M^{(i)} \phi^{(i)}$$

weights $\hat{w}_k^{(i)}(\lambda)$ for a given λ . By starting from the precomputed initial samples (Fig. 2(a)), the initial distribution with

epistemic parameter λ_0 is targeted by adjusting the weights (Line 1 of Algorithm 2 and Fig. 2(b)) as the ratio between the distribution (in green) and the initial proposal (in blue) normalised by the number of particles. For each observation instance, the precomputed samples are used to save new propagations (Fig. 2(c)). Then the weights are updated (Line 2 and Fig. 2(d)) according to Eq. (11) with the new transition and likelihood (represented in dashed dark yellow curve). Finally, the estimator is computed as a weighted sum of the updated weights and the precomputed function evaluations (Line 3).

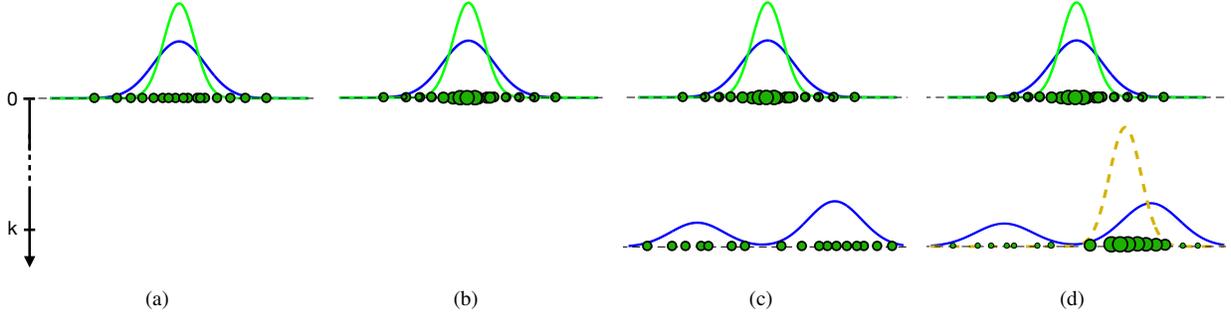


Fig. 2 Representation of precomputed sequential importance sampling estimator: a) initial samples, b) adjusted weights, c) precomputed samples at next time step, d) updated weights after conditioning with the likelihood (dashed curve).

The derivatives of such estimator can be computed analytically. Their explicit representation is reported in Appendix A.

2. Proposal selection

The self-normalised pSIS estimator is asymptotically unbiased, but for a small sample size the bias could be significant [23]. Even when enough particles are employed, the sequential estimator often suffers from the degeneracy problem, that is, when most of the particles have zero weights [3], which causes the estimate to rely on a small *effective* sample. The main cause of degeneracy is the choice of a non-optimal (in terms of bias) proposal distribution, which is one that differs significantly from the target posterior. This problem is exacerbated in the epistemic scenario because there may not be a single proposal distribution which keeps the bias low for all the posteriors resulting from the epistemic set $\lambda \in \Omega_\lambda$.

The effective sample size is computed in percentage as $n_{\text{eff}} = 1 / \sum_{i=1}^N w_k^{(i)2}$. The condition of effective size below a minimum threshold is often used to trigger resampling in particle filtering [24]. However, resampling is not possible in the constructed precomputed estimator as the particles are drawn before the optimisation. Thus, one would need to check the effective sample size at the end of the estimation process and loop back to update the proposal.

An alternative is to use a strategy analogous to the one implemented in Unscented Particle Filter (UPF) [25]. The main idea is to construct an approximated posterior by introducing information on the expected measurement through

an Unscented Kalman Filter (here in Line 2 of Algorithm 1). Starting from a set of particles generated with a given proposal, one would apply a UKF to each particle and do conditioning with the likelihood of the expected measurements. The main computational advantage of a UKF-based proposal in this method is that the sigma points can be propagated inexpensively through the surrogate polynomial expansion approximating the dynamical flow (see Section II.B.4). The posterior of this process is a new proposal that includes information about the last observation resulting in more particles distributed in regions of higher expected likelihood, a crucial step to ensure a satisfactory effective sample size even without resampling.

In the case in which measurements are affected by epistemic uncertainty, this approach presents two challenges: 1) approximate the likelihood with a normal distribution such that the UKF can be applied; 2) construct a posterior which takes into account that the statistical moments of the likelihood are set-valued. We address these two challenges with the following solutions. First, we approximate the generic likelihood with a normal

$$p(\mathbf{y}_k | \mathbf{x}_k; \boldsymbol{\lambda}_y^{(j)}) \approx \mathcal{N}(\mathbf{y}_k | \mathbf{x}_k; \boldsymbol{\mu}_y^{(j)}, \boldsymbol{\Sigma}_y^{(j)}) \quad (12)$$

where $\boldsymbol{\mu}_y^{(j)} = \boldsymbol{\mu}_y^{(j)}(\boldsymbol{\lambda}_y^{(j)})$ and $\boldsymbol{\Sigma}_y^{(j)} = \boldsymbol{\Sigma}_y^{(j)}(\boldsymbol{\lambda}_y^{(j)})$ are the parameterised mean and covariance. In the case of distributions with undefined or infinite mean or covariance, e.g. Lévy or Cauchy distributions, the location and scale parameters could be used as substitutes for the mean and covariance. The second challenge is addressed by constructing the proposal as a convex combination of UKF posteriors, each resulting from a different likelihood in the epistemic set

$$\pi(\mathbf{x}_k | \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_{1:k}) = \sum_j^{N_\pi} b_j \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k^{(i,j)}, \boldsymbol{\Sigma}_k^{(i,j)}), \quad (13)$$

where $\sum_j b_j = 1$.

The overall approach is described in Algorithm 3. Right arrows represent input to output relationships for the corresponding line in the algorithm. The initial proposal is constructed as a convex combination of multiple distributions in the initial epistemic set (Line 1). N particles are drawn from the proposal (Line 2), and for each of them the mean and covariance of the parallel UKFs are initialised (Line 3). For each observation instance, the UKF distributions are propagated forward in time with the Unscented Transform (UT) [26] (Line 4). Then, N_π likelihoods are selected within their epistemic set according to a low-discrepancy sequence, and approximated with normal distributions (Line 5) according to the rule discussed above (see Eq. (12)). For each normal likelihood, the UKF Bayes' update is performed to compute the posterior mean and covariance (Line 6). From here, the proposal is constructed as the collection of UKF posteriors (Line 7). The weights b_j , which sum to one, could be used to express the belief on the relative likelihood of each epistemic parameter.

With this approach, the proposal is constructed to cover the support of each density in the epistemic set. Furthermore,

Algorithm 3 Construction of proposal distributions.

Given the estimation problem (4)

- 1: Set initial proposal as convex combination of distributions in epistemic set
 $\pi(\mathbf{x}_0) = \sum_j^{N_\pi} p(\mathbf{x}_0; \boldsymbol{\lambda}_0^{(j)})/N_\pi$ with $\boldsymbol{\lambda}_0^{(j)} \in \Omega_{\boldsymbol{\lambda}_0}$
- 2: Draw N particles from the initial proposal $\pi(\mathbf{x}_0)$
 $\mathbf{x}_0^{(i)} \sim \pi(\mathbf{x}_0)$
- 3: For each particle, initialise UKF mean as particle value and covariance as sample covariance
 $\boldsymbol{\mu}_0^{(i)} = \mathbf{x}_0^{(i)}, \boldsymbol{\Sigma}_0^{(i)} = \sum_j (\mathbf{x}_0^{(j)} - \boldsymbol{\mu}_0^{(i)})(\mathbf{x}_0^{(j)} - \boldsymbol{\mu}_0^{(i)})^T / (N - 1)$
for $k = 1 : M$ **do**
 - 4: Propagate UKF mean and covariance with UT
 $\boldsymbol{\mu}_{k-1}^{(i)}, \boldsymbol{\Sigma}_{k-1}^{(i)} \rightarrow \boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)}$
for $j = 1 : N_\pi$ **do**
 - 5: Select an epistemic parameter $\boldsymbol{\lambda}_y^{(j)} \in \Omega_{\boldsymbol{\lambda}_y}$ and approximate the corresponding likelihood
 $p(\mathbf{y}_k | \mathbf{x}_k; \boldsymbol{\lambda}_y^{(j)}) \rightarrow \mathcal{N}(\mathbf{y}_k | \mathbf{x}_k; \boldsymbol{\mu}_y^{(j)}, \boldsymbol{\Sigma}_y^{(j)})$
 - 6: Update UKF mean and covariance with Bayes' update
 $\boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)} \rightarrow \boldsymbol{\mu}_k^{(i,j)}, \boldsymbol{\Sigma}_k^{(i,j)}$**end for**
 - 7: Construct proposal with UKF posteriors
 $\pi(\mathbf{x}_k | \mathbf{x}_{k-1}^{(i)}, \mathbf{y}_{1:k}) = \sum_j^{N_\pi} b_j \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k^{(i,j)}, \boldsymbol{\Sigma}_k^{(i,j)})$**end for**

the UKF-based proposal concentrates more particles in high-likelihood areas because it explicitly incorporates information on the observations, thus reducing the risk of sample degeneracy. This behaviour is particularly valuable in space applications characterised by multi-dimensional spaces, highly peaked likelihoods (accurate observations), and little overlap between the predicted distribution and the likelihood.

3. Algorithmic Complexity

In this section, the algorithmic complexity of evaluating the estimator (10) is assessed. Specifically, the complexities of the precomputation step and the pSIS estimator are analysed separately to pinpoint the main saving of precomputation.

Let c_{x_0} be the complexity of drawing one particle from $\pi(\mathbf{x}_0)$, c_x the complexity to sample the proposal $\pi(\mathbf{x}_k | \mathbf{x}_{0:k-1}, \mathbf{y}_{1:k})$, c_π the complexity of evaluating each term of the proposal (9) (assumed equal complexity), c_p the complexity of evaluating each of the distributions in problem (4) (assumed equal complexity), and c_ϕ the complexity of evaluating the function ϕ .

The complexity of the precomputation step can be assessed by looking at the complexity of each step in Algorithm 1:

- 1) $N(c_{x_0} + c_\pi)$ as N particles are drawn from and evaluated with the initial proposal;
- 2) $N(c_x + c_\pi)$ since N particles are drawn from and evaluated with the transition proposal;
- 3) Nc_ϕ because there are N evaluations of the function ϕ .

Step 2) is performed M times because of the for-loop over the number of observation instances. Therefore, the computational complexity of the precomputation step of the pSIS estimator is

$$C_{pre}(M, N) = N(c_{x_0} + c_\pi + c_\phi) + MN(c_x + c_\pi) = \mathcal{O}(N + MN), \quad (14)$$

which is linear both with the number of observations M and in the number of particles N employed and displays an interaction between the two.

Once the precomputations have been performed, the computational complexity of the pSIS estimator is assessed by studying each step in Algorithm 2:

- 1) $N(c_p + 2)$ because the importance weights are computed by evaluating the initial distribution and performing two divisions per weight;
- 2) $N(2c_p + 3) + 2N - 1$ since the unnormalised weights are computed by evaluating two densities and performing three products or divisions, whereas the normalisation requires $N - 1$ sums and N divisions;
- 3) $2N - 1$ because the estimator is computed by N multiplications and $N - 1$ sums.

Step 2) is performed M times. Thus, the computational complexity of the pSIS is

$$C_{pSIS}(M, N) = N(c_p + 4) - 1 + M[N(2c_p + 5) - 1] = \mathcal{O}(N + M + MN), \quad (15)$$

which again is linear both with the number of observations M and in the number of particles N employed and showing the interaction between the two of them. In space tracking, the main saving stems from removing the term MNc_x , that is, sampling from the proposal distribution, which implies sparing MN new numerical integrations of the equations of motion every time the estimator is evaluated in the optimisation routine.

As derived in Appendix A, evaluating analytically the estimator derivatives requires an algorithmic complexity of $\mathcal{O}(MN^2 + MN\lambda c_{\partial p} + MNc_\pi)$. If the same quantities were to be computed by finite differences, the computational complexity would be $\mathcal{O}(MN\lambda c_\pi)$, that is evaluating the estimator λ times (or a small multiple of λ). Because of the quadratic term and because typically $c_{\partial p} > c_\pi$, the finite differences method has lower complexity in general. On the other hand, finite differences provide only an approximation of the actual derivative value and therefore are affected by truncation errors. Hence, the choice between an exact but slower derivative computation and a faster but less accurate finite difference approximation should be resolved for each specific test case. Factors influencing this trade-off are c_π , λ , $c_{\partial p}$ and the assessed accuracy of the finite difference approximation.

4. Polynomial propagator

Propagating the particles in time is the most computationally intensive step in space applications as it requires solving the initial value problem in Eq. (1) by numerical integration as many times as the number of particles employed. In this work, a polynomial surrogate of the dynamical flow is employed to speed up the uncertainty propagation step and enable a large number of particles to improve the estimate accuracy. A brief description of the polynomial propagation is provided in this section, while the full description of such procedure is reported in [27].

The propagation is split into two steps: first, a polynomial representation of all possible states at time t_{k+1} given the states at time t_k is created; then, the particles at time t_k are propagated to time t_{k+1} through the inexpensive polynomial evaluations.

Consider the set Ω_{ξ_k} of all possible uncertainty realisations $\xi_k = [\mathbf{x}_k, \mathbf{d}_k]^T$ at time t_k . The set \mathcal{F}_k^{k+1} of compatible states at time t_{k+1} can be defined as

$$\mathcal{F}_k^{k+1} = \{\mathbf{x}_{k+1}(\xi_k) \mid \xi_k \in \Omega_{\xi_k}\}, \quad (16)$$

where $\mathbf{x}_{k+1}(\xi_k)$ is defined as

$$\mathbf{x}_{k+1}(\xi_k) = \mathbf{x}_k + \int_{t_k}^{t_{k+1}} f(t, \mathbf{x}, \mathbf{d}_k) dt. \quad (17)$$

A truncated polynomial expansion of finite order q is constructed to approximate this mapping as

$$\mathcal{F}_k^{k+1}(\xi_k) \approx \tilde{\mathcal{F}}_k^{k+1}(\xi_k) = \sum_{i=0}^{N_q} \alpha_{k,i} \Psi_i(\xi_k), \quad (18)$$

where Ψ_i is the i -th multivariate polynomial basis, $\alpha_{k,i}$ its coefficient, and N_q the number of bases. In this work the coefficients are computed using stochastic collocation [28]: a number of samples \mathbf{x}_{k+1} is first evaluated over a structured grid of collocation points $\xi_k^{(j)}$; then a polynomial approximation $\tilde{\mathcal{F}}_k^{k+1}$ is built to fit the set of \mathbf{x}_{k+1} vectors. The Smolyak polynomial space variant is used to limit the growth of collocation points with increasing stochastic dimensions.

Intrusive polynomial algebra could be employed in place of non-intrusive alternatives when the dynamical model can be accessed and templated, as it provides good performance and scalability for high-dimensional problems [29] and it proved effective for uncertainty propagation in space applications [30].

C. Bound Estimator

The robust estimation method computes the bounds in Eq. (6) by taking advantage of the efficient estimator constructed above. In the following, only the routine to estimate the lower bound is presented, as the same approach is applicable to the computation of the upper bound.

The lower bound estimator and the corresponding epistemic parameter are denoted as

$$\hat{\underline{\theta}}(\mathcal{X}_{0:M}) = \min_{\lambda \in \Omega_\lambda} \hat{\theta}(\mathcal{X}_{0:M}, \lambda) \quad (19a)$$

$$\underline{\lambda}(\mathcal{X}_{0:M}) = \arg \min_{\lambda \in \Omega_\lambda} \hat{\theta}(\mathcal{X}_{0:M}, \lambda), \quad (19b)$$

that is, the minimum of the expectation estimator and the argument of the minimum given the set of samples $\mathcal{X}_{0:M}$. By employing the precomputation step, and Algorithm 2 to evaluate the estimator for a candidate λ , the estimator is written in the form of $\hat{\theta}(\mathcal{X}_{0:M}, \lambda) = \sum_{i=1}^N \hat{w}_M^{(i)}(\lambda) \phi^{(i)}$, with precomputed $\phi^{(i)}$, so that the optimisation process operates only on the importance weights $w_M^{(i)}(\lambda)$.

Eq. (19) entails two challenges: first, the optimisation problem is generally multi-modal since the objective function results from numerous operations between nonlinear functions as in Eq. (11); second, even if the global extremum is found, $\hat{\underline{\theta}}$ could still deviate from the true bound $\underline{\mathbb{E}}$. The former challenge is addressed with a convergent global optimisation scheme, presented in detail in Section II.D. The latter challenge is addressed with what follows.

The approximation of the bounds depends on two additional factors: the precision of the pSIS estimator (10) at approximating the expectation for a given epistemic parameter, and the accuracy of the bound estimator (19), at approximating the true bounds.

The former requires the estimator $\hat{\theta}(\mathcal{X}_{0:M}, \underline{\lambda})$ to suitably approximate the expectation $\mathbb{E}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}; \underline{\lambda}]$. A good approximation is ensured by an effective sample size.

The latter requires the estimator $\hat{\underline{\theta}}(\mathcal{X}_{0:M})$ to suitably approximate the bound $\underline{\mathbb{E}}[\phi(\mathbf{X}_{0:k}) | \mathbf{y}_{1:k}]$. Confidence bounds can be constructed for the bound estimator following the procedure in [23]. Let $\mathcal{X}_{0:M}^1, \dots, \mathcal{X}_{0:M}^{2l}$ be $2l$ sets of samples. Define the quantities

$$\begin{aligned} \hat{\underline{\theta}}^j &= \hat{\theta}(\mathcal{X}_{0:M}^j, \underline{\lambda}(\mathcal{X}_{0:M}^j)) \quad \text{for } j = 1, \dots, l \\ \hat{\underline{\theta}}^j &= \hat{\theta}(\mathcal{X}_{0:M}^j, \underline{\lambda}(\mathcal{X}_{0:M}^{j-l})) \quad \text{for } j = l+1, \dots, 2l, \end{aligned} \quad (20)$$

with $\underline{\lambda}(\mathcal{X}_{0:M})$ defined in Eq. (19b). For the first half, these are the classical bound estimators evaluated for different sets of samples. For the second half, these require evaluating $\hat{\theta}$ for a set of samples $\mathcal{X}_{0:M}^j$ with the epistemic parameter minimizing the estimator for another set of samples $\mathcal{X}_{0:M}^{j-l}$. Let $\mu_{\mathcal{X}}^{1H}$ and $\sigma_{\mathcal{X}}^{1H}$ be the mean and standard deviation of the first half of $\hat{\underline{\theta}}^j$, i.e. for $j = 1, \dots, l$, and $\mu_{\mathcal{X}}^{LH}$ and $\sigma_{\mathcal{X}}^{LH}$ the ones of the last half of $\hat{\underline{\theta}}^j$, that is for $j = l+1, \dots, 2l$. Then, under the assumption that enough particles have been used to keep the bias small and bounded for all λ , the confidence interval for the confidence level $1 - \alpha$ is

$$\left[\mu_{\mathcal{X}}^{1H} - t_{\alpha, l-1} \frac{\sigma_{\mathcal{X}}^{1H}}{\sqrt{N}}, \mu_{\mathcal{X}}^{LH} + t_{\alpha, l-1} \frac{\sigma_{\mathcal{X}}^{LH}}{\sqrt{N}} \right], \quad (21)$$

where $t_{\alpha, l-1}$ is the $1 - \alpha$ two-sided critical value of the t-distribution with $l - 1$ degrees of freedom.

The overall RPF algorithm to compute robust bounds under epistemic uncertainty is summarised in Algorithm 4. Line 1 is the precomputation step performed using a standard particle filter. The quantities $\mathcal{X}_{0:M}$, $\pi_{0:M}^{(i)}$, and $\phi^{(i)}$ were defined in Algorithm 1. Line 2 is the bound computation routine performed by a B&B optimisation of the expectation estimator $\hat{\theta}$.

Algorithm 4 Algorithmic scheme for the Robust Particle Filter.

Given:

- the estimation problem (4)
- the proposal distributions $\pi(\mathbf{x}_0)$ and $\pi(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{Y}_{1:k})$

1: Precomputation step by standard Particle Filter

Save collection of particles $\mathcal{X}_{0:M}$, densities $\pi_{0:M}^{(i)}$, and function evaluations $\phi^{(i)}$.

2: Bound computation by Branch & Bound

$$\begin{aligned}\underline{\hat{\theta}}(\mathcal{X}_{0:M}) &= \min_{\lambda \in \Omega_\lambda} \hat{\theta}(\mathcal{X}_{0:M}, \lambda) \\ \bar{\hat{\theta}}(\mathcal{X}_{0:M}) &= \max_{\lambda \in \Omega_\lambda} \hat{\theta}(\mathcal{X}_{0:M}, \lambda)\end{aligned}$$

D. Global search

The optimisation problem (19) is generally nonlinear, nonconvex and multi-modal because of the nonlinear dependencies of the distributions from the epistemic parameters and the nonlinear operations in the weights update step. Thus, a global optimisation routine is employed to explore the epistemic domain and find the extremum of $\hat{\theta}$.

In this work we developed a B&B approach using simplexes as subdomains and a Lipschitz-based lower bound estimation. The advantage of using B&B is that asymptotic convergence to the global optimum is granted and that an estimation of the distance from it is known at each iteration. The advantage of using simplexes is that the number of branched sub-simplexes is independent of the problem dimension rather than growing exponentially as with a simple lattice grid. An approximation of Lipschitz constant can be derived from the analytical derivatives of the estimator. Under the assumption that the derivatives exist and are bounded for $\lambda \in \Omega_\lambda$, the precomputed SIS estimator is continuous and differentiable with bounded derivatives in λ (from Eqs. (42) to (44)).

The B&B algorithm is presented in Appendix B. In this section, the bounding strategy is discussed in detail, along with a lemma on the bounding step needed for the proof of convergence. Then the branching strategy is presented together with a lemma showing that this approach eventually generates sub-domains of arbitrary small sizes and provides an upper bound on the number of branching steps. Finally, a theorem on the global convergence of the algorithm is stated.

1. Preliminary definitions

Let $\mathcal{S} \subset \mathbb{R}^{n_\lambda}$ be a n_λ -simplex with vertexes $[\lambda_0, \dots, \lambda_n]$. Let L , with $0 < L < \infty$, be the Lipschitz constant of $\hat{\theta}$ over Ω_λ such that

$$|\hat{\theta}(\chi_k, \lambda_1) - \hat{\theta}(\chi_k, \lambda_2)| \leq L \|\lambda_1 - \lambda_2\| \quad \forall \lambda_1, \lambda_2 \in \Omega_\lambda. \quad (22)$$

Let the diameter of a n_λ -simplex \mathcal{S} be the maximum distance between any two points belonging to the simplex, which is the maximum distance between two vertexes

$$\sigma(\mathcal{S}) = \max_{i,j \in [0:n_\lambda]} \|\lambda_i - \lambda_j\|,$$

and let λ_{i^*} and λ_{j^*} be the most distant vertexes such that $\sigma(\mathcal{S}) = \|\lambda_{i^*} - \lambda_{j^*}\|$. For each branching step $k = 1, \dots, K$, let \mathcal{L}_k be the list of disjoint simplexes covering the feasible domain, that is $\bigcup_j \mathcal{S}_k^{(j)} = \Omega$ for $\mathcal{S}_k^{(j)} \in \mathcal{L}_k$ and $\mathcal{S}_k^{(i)} \cap \mathcal{S}_k^{(j)} = \emptyset \quad \forall \mathcal{S}_k^{(i)}, \mathcal{S}_k^{(j)} \in \mathcal{L}_k$. Finally, let $lb : \mathcal{S} \rightarrow \mathbb{R}$ be a function bounding $\hat{\theta}$ from below in the simplex \mathcal{S}

$$lb(\lambda) \leq \hat{\theta}(\lambda) \quad \forall \lambda \in \mathcal{S}$$

and denote $lb(\mathcal{S})$ its minimum

$$lb(\mathcal{S}) = \min_{\lambda \in \mathcal{S}} lb(\lambda).$$

Thus, $lb(\mathcal{S})$ is the lower bound for the minimum of $\hat{\theta}$ in the simplex. Similarly, denote $ub(\mathcal{S})$ the upper bound for the minimum of $\hat{\theta}$ in the simplex.

2. Bounding

Because $\hat{\theta}$ is Lipschitz continuous, the lower bounding function $lb : \mathcal{S} \rightarrow \mathbb{R}$ can be defined as

$$lb(\lambda) = \max_j \hat{\theta}(\lambda_j) - L \|\lambda - \lambda_j\|, \quad (23)$$

where λ_j are the vertexes of the simplex. The lower bound value $lb(\mathcal{S})$ is therefore the minimum of $lb(\lambda)$ over the simplex \mathcal{S}

$$lb(\mathcal{S}) = \min_{\lambda \in \mathcal{S}} \max_j \hat{\theta}(\lambda_j) - L \|\lambda - \lambda_j\|. \quad (24)$$

For the upper bound of the minimum, the minimum of the fitness values at the simplex vertexes is chosen as trivial bound

$$ub(\mathcal{S}) = \min_j \hat{\theta}(\lambda_j). \quad (25)$$

Hence, we state a lemma to show that these values actually bound the minimum over the simplex. The proof is presented in Appendix C.

Lemma 1. *The bounds $lb(\mathcal{S})$ and $ub(\mathcal{S})$ in Eqs. (24) and (25) satisfy*

$$lb(\mathcal{S}) \leq \min_{\lambda \in \mathcal{S}} \hat{\theta}(\lambda) \leq ub(\mathcal{S}). \quad (26)$$

Another property required to prove convergence of Algorithm 5 is that as the simplex diameter goes to zero the bounds difference should uniformly converge to zero. This is proven in the following lemma for the B&B presented here. Again, the proof is discussed in Appendix C.

Lemma 2. *The bounds lb and ub satisfy*

$$\forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \text{ s.t. } \forall \mathcal{S} \subset \mathbb{R}^{n_\lambda}, \sigma(\mathcal{S}) \leq \delta \implies ub(\mathcal{S}) - lb(\mathcal{S}) \leq \varepsilon. \quad (27)$$

An analytical procedure has been developed to compute the lower bound $lb(\mathcal{S})$ for the Lipschitz bounding function in Eq. (24), and it is presented in Appendix D.

3. Branching

In order to prove B&B convergence, we need to show that the branching scheme eventually produces a simplex of arbitrary size δ . We adopt Longest Edge Bisection (LEB) as branching rule. The simplex \mathcal{S} is split into two simplexes by bisection along its longest edge, that is, the one connecting the vertices λ_{i^*} and λ_{j^*} , such that

$$\lambda_{ij}^* = \frac{\lambda_i^* + \lambda_j^*}{2} \quad (28)$$

is the new vertex. The two offspringing simplexes $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ share the same vertices of \mathcal{S} except that one has λ_{ij}^* in place of λ_i^* , whereas the other has λ_{ij}^* in place of λ_j^* . [31] showed that LEB reduces the diameter of a n_λ -simplex of at least $(\sqrt{3}/2)^{\lfloor k/n_\lambda \rfloor}$ after k splittings. On the other hand, as trivial lower bound, LEB reduces the diameter of a n_λ -simplex no more than half the size of the immediate parent after 1 split.

For each branching step $k = 1, \dots, K$, choose a simplex $\mathcal{S} \in \mathcal{L}_{k-1}$ and split it into 2 disjoint simplexes $\mathcal{S}^{(1)}$ and $\mathcal{S}^{(2)}$ by LEB. Define the next list of simplexes as $\mathcal{L}_k = (\mathcal{L}_{k-1} \setminus \mathcal{S}) \cup \mathcal{S}^{(1)} \cup \mathcal{S}^{(2)}$.

We now state that this branching rule eventually creates a simplex of arbitrary small diameter and define an upper bound on the number of iterations. The proof is reported in Appendix C.

Lemma 3. Let \mathcal{L}_0 be the initial list of N_0 n_λ -simplexes and

$$\sigma_0 = \max_{\mathcal{S}_0^{(i)} \in \mathcal{L}_0} \sigma \left(\mathcal{S}_0^{(i)} \right)$$

the largest diameter among them. For any $\delta \in \mathbb{R}^+$, there exist $K \in \mathbb{N}$ such that the LEB rule yields

$$\min_{\mathcal{S}_k^{(j)} \in \mathcal{L}_K} \sigma \left(\mathcal{S}_k^{(j)} \right) \leq \delta.$$

Specifically, the requested diameter is realised after at most K splitting of the initial simplexes with

$$K = N_0 n_\lambda \left\lceil \log_{\frac{2}{\sqrt{3}}} \frac{\sigma_0}{\delta} \right\rceil. \quad (29)$$

4. Convergence

The algorithm terminates when the difference between the upper and lower bounds on the most promising simplex \mathcal{S}^* , that is the one with the lowest bound for the minimum $lb(\mathcal{S}^*) \leq lb(\mathcal{S}_k^{(j)})$ for any j , is below a given threshold. Let

$$L_k = \min_{\mathcal{S}_k^{(j)} \in \mathcal{L}_k} lb(\mathcal{S}_k^{(j)})$$

$$U_k = \min_{\mathcal{S}_k^{(j)} \in \mathcal{L}_k} ub(\mathcal{S}_k^{(j)}),$$

then with the lemmas presented above, we can now state the theorem proving the convergence of Algorithm 5. The proof is detailed in Appendix C.

Theorem 1. Algorithm 5 converges in a finite number of steps to within a set threshold $\varepsilon \in \mathbb{R}^+$ of the global minimum $\hat{\theta}$. Specifically

$$\forall \varepsilon \in \mathbb{R}^+ \exists K_\varepsilon \text{ such that } U_{K_\varepsilon} - L_{K_\varepsilon} \leq \varepsilon, L_{K_\varepsilon} \leq \hat{\theta} \leq U_{K_\varepsilon}. \quad (30)$$

This ensures the global minimum to be within the interval $\hat{\theta} \in [lb(\mathcal{S}_{K_\varepsilon}^*), ub(\mathcal{S}_{K_\varepsilon}^*)]$ at convergence after K_ε steps.

5. Filter complexity with epistemic dimension

In this section, we quantify the computational complexity of the filter and its dependency on the number of dimensions of the epistemic space. This dependency gives a measure of the scalability of the filter with the number of epistemic parameters. We start by noting that the number of particles in the pSIS estimator depends on the dimension of the state space and on the shape of the distributions of the random variables \mathbf{X}_k , \mathbf{D} , and \mathbf{Y} but not directly on the dimension of the epistemic space. Similarly, the polynomial surrogate depends on the number of aleatory variables,

but not on the number of epistemic ones. Thus the only part of the filter whose complexity is directly affected by the number of epistemic parameters is the global optimisation algorithm.

Let n_λ be the number of epistemic variables. The initial number N_0 of simplexes in \mathcal{L}_0 covering the n_λ -dimensional epistemic space depends on the rule employed to partition Ω_λ . We currently use a standard algorithm for triangulation of a hyper-box into $N_0 = n_\lambda!$ simplexes [32]. Nonetheless, a more convenient decomposition that yields a lower number of initial simplexes could be employed. Among them, an initial circumscribing simplex with a fitness penalisation for points outside the domain yield $N_0 = 1$ regardless of n_λ [33, 34], a minimum-cardinality triangulation can be achieved by means of linear programming [35], or covering with intersecting simplexes could be employed in place of triangulation [36]. Therefore, in our complexity analysis we will retain $N_0 = N_0(n_\lambda)$ without making such dependence explicit.

In the first iteration, at most $N_0(n_\lambda + 1)$ evaluations of the pSIS estimator are required. For each LEB branching step, only one estimator evaluation is needed at the new common vertex λ_{ij}^* . By using the maximum number of branching steps K from Eq. (29), the upper bound on the filter complexity is

$$C_{\text{RPF}}(n_\lambda) = 2 [N_0(n_\lambda + 1) + K] C_{\text{pSIS}} = 2 \left[N_0 + N_0 n_\lambda \left(1 + \left\lceil \log_{\frac{2}{\sqrt{3}}} \frac{\sigma_0}{\delta} \right\rceil \right) \right] C_{\text{pSIS}}, \quad (31)$$

where C_{pSIS} was defined in Eq. (15). Therefore, the complexity upper bound is linear with the number of epistemic variables if N_0 is independent of n_λ , e.g. [33], otherwise it scales as $n_\lambda N_0(n_\lambda)$. By pruning the simplexes, as the B&B proceeds with the search, the actual filter complexity is significantly reduced from this conservative estimate.

6. Lipschitz constant estimation

The Lipschitz constant is estimated as the maximum of the estimator gradient evaluated on the vertexes of the simplexes. Specifically, at a generic iteration k , the domain is partitioned in disjoint simplexes $\mathcal{S}_k^{(i)}$ each with vertexes $[\lambda_0^{(i)}, \dots, \lambda_n^{(i)}]$. The Lipschitz constant is set to

$$L = \max_{i,j} \left\| \nabla_\lambda \hat{\theta}(\mathcal{X}_{0:k}, \lambda_j^{(i)}) \right\|. \quad (32)$$

This is an adaptive, although lower, approximation of the true Lipschitz constant. Therefore, it could lead to over-pruning potentially optimal simplexes. However, lower approximations are often employed in literature and have proven to be efficient on a large number of test cases [37–39].

III. Numerical Experiments

This section presents an assessment of the performance of the filter and of its ability to compute tight bounds on the probability of collision between two objects in space, in the presence of epistemic uncertainty. First, the

section introduces the working hypothesis and defines the conjunction scenario. Then, an experimental assessment of the pSIS wall time as a function of the number of particles and observations is presented and compared to the theoretical complexity derived in Sec.II.B.3. Finally, four instances of the conjunction scenario corresponding to different operational conditions are used to demonstrate the ability of the filter to bound all the realisations of the probability of collision.

A. Definition of the conjunction scenario

The conjunction scenario considered in this paper envisage a piece of debris on a potential collision course with a known operational satellite. SOCRATES (Satellite Orbital Conjunction Reports Assessing Threatening Encounters in Space) [40], an online service that provides twice-daily reports on the most likely collision events based on NORAD two-line elements (TLEs) and the SGP4 propagator, was employed to select two space objects with a low range at their close approach. The orbital elements of these two objects are reported in Table 1. The first object is an operational satellite, SENTINEL 2B (NORAD ID 42063), whose ephemerides are assumed to be known with a negligible degree of uncertainty as it carries a GPS receiver. The second object is a piece of debris (NORAD ID 30141) resulted from the anti-satellite test carried out on FENGYUN 1C DEB in 2007 [41]. Thus, the state of the debris is the latent variable \mathbf{x}_k that needs to be estimated. The reference Time of Closest Approach (rTCA) is the 13-Jan-2021 at 13:24:25 UTC.

Table 1 Spacecraft orbital elements at reference epoch from NORAD TLEs.

NORAD ID	EPOCH [UTC]	a [km]	e [-]	i [deg]	Ω [deg]	ω [deg]	M [deg]
42063	08-Jan-2021 01:17:15	7167.14	1.1e-4	98.57	85.33	81.09	279.04
30141	07-Jan-2021 16:24:07	7180.78	2.5e-3	99.08	183.01	252.25	107.59

Note that in general the uncertainty in both objects should be taken into account. However, the scope of this first example is to test the ability of the filter to generate robust bounds on the probability of collision rather than the assessment of the ability of the filter to handle high dimensional uncertainty spaces. For this reason we deliberately selected one of the two objects to be well-tracked, and assumed that the main source of epistemic uncertainty was the piece of debris. The application of our RPF to a scenario in which both objects are uncertain can be found in [42].

In this scenario, the motion of the body is described in Cartesian coordinates in an Earth-centered inertial reference frame. The dynamical model in Eq. (1) includes the following components [43]: Earth gravitational force derived from the EGM96 geopotential model up to degree and order 4; atmospheric drag, with atmospheric density computed with the Jacchia-Gill model; third-body disturbances due to the Moon and Sun gravitational attraction; solar radiation pressure with a conical shadow model for the Earth's eclipse.

The quantity to be bounded is the Probability of Collision (PoC). A collision is defined when the minimum distance

between two objects is smaller than a given threshold δ_{DCA} . This translates to the collision indicator function:

$$I_C(\mathbf{x}_F) = \begin{cases} 1 & \text{if } DCA(\mathbf{x}_F) \leq \delta_{DCA} \\ 0 & \text{if } DCA(\mathbf{x}_F) > \delta_{DCA}, \end{cases} \quad (33)$$

where DCA is the function extracting the Distance of Closest Approach (DCA), that is the minimum of the relative position norm between the debris state realisation and the known operational satellite. To detect the correct DCA for each sample, the relative distance at t_F is not used directly because different state realisations have different TCAs and, therefore, the distance at t_F is generally not the minimum one. On the contrary, the minimum DCA for a realisation \mathbf{x}_F is obtained by computing the pericenter distance of the relative hyperbolic trajectory between the two objects [44]. From here, the probability of collision is evaluated by computing the expectation of the indicator function

$$\text{PoC}(\boldsymbol{\lambda}) = \int I_C(\mathbf{x}_F) p(\mathbf{x}_F | \mathbf{y}_{1:M}; \boldsymbol{\lambda}) d\mathbf{x}_F. \quad (34)$$

Therefore, the goal is to compute robust bounds on the PoC as

$$\underline{\text{PoC}} = \min_{\boldsymbol{\lambda} \in \Omega_{\boldsymbol{\lambda}}} \text{PoC}(\boldsymbol{\lambda}) \quad (35a)$$

$$\overline{\text{PoC}} = \max_{\boldsymbol{\lambda} \in \Omega_{\boldsymbol{\lambda}}} \text{PoC}(\boldsymbol{\lambda}), \quad (35b)$$

with a specific interest for $\overline{\text{PoC}}$ which represents the epistemic worst-case scenario.

1. Initial state uncertainty

The uncertainty in the initial state of the piece of debris is constructed by adding a variance to its TLEs as reported in Table 1. The value of such uncertainty is defined by following the guidelines of the European Space Agency (ESA) [45]. For inclinations larger than 60 deg, the 1σ uncertainty in the radial, transversal and normal components of position and velocity are reported in Table 2. A Gaussian parametric family is employed to describe the initial state uncertainty.

Table 2 1σ position (\mathbf{r}) and velocity (\mathbf{v}) uncertainty of TLEs for orbits with $e < 0.1, i > 60$ deg, perigee altitude $\leq 800km$, in radial (U), transversal (V), and normal components (W).

$1\sigma_{rU}$ [m]	$1\sigma_{rV}$ [m]	$1\sigma_{rW}$ [m]	$1\sigma_{vU}$ [mm/s]	$1\sigma_{vV}$ [mm/s]	$1\sigma_{vW}$ [mm/s]
104	556	139	559	110	148

The covariance matrix $\Sigma_{\mathbf{x}_0}$ in inertial coordinates is computed using the Jacobian of the transformation from radial, transversal and normal to Cartesian coordinates. From here, the importance initial distribution is defined as a normal

distribution

$$\pi(\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_0; \boldsymbol{\mu}_{\mathbf{x}_0}, \boldsymbol{\Sigma}_{\mathbf{x}_0}\right) \quad (36)$$

where $\boldsymbol{\mu}_{\mathbf{x}_0}$ is the Cartesian state retrieved from the TLE. Epistemic uncertainty is introduced in the probability distribution associated with the initial conditions by using two epistemic parameters. The initial epistemic set is defined as:

$$\begin{aligned} \mathcal{P}_{\mathbf{x}_0} &= \{p(\mathbf{x}_0; \boldsymbol{\lambda}_0) : p(\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_0; \boldsymbol{\mu}_{\mathbf{x}_0}, \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}_0}; \boldsymbol{\lambda}_0\right), \\ \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}_0} &= \text{diag}(\lambda_{\mathbf{x}_{0-1}} \boldsymbol{\Sigma}_{\mathbf{x}_0}(1:3, 1:3), \lambda_{\mathbf{x}_{0-2}} \boldsymbol{\Sigma}_{\mathbf{x}_0}(4:6, 4:6)), \\ \lambda_{\mathbf{x}_{0-1}} &\in [1/5^2, 5^2], \lambda_{\mathbf{x}_{0-2}} \in [1/5^2, 5^2] \}, \end{aligned} \quad (37)$$

where $\boldsymbol{\lambda}_0 = [\lambda_{0-1}, \lambda_{0-2}]$ is the epistemic parameter on the initial distribution, $\boldsymbol{\Sigma}_{\mathbf{x}_0}(1:3, 1:3)$ and $\boldsymbol{\Sigma}_{\mathbf{x}_0}(4:6, 4:6)$ indicate respectively the position block and the velocity block of the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}_0}$, and the operator diag indicates a block-diagonal matrix. Therefore, the set $\mathcal{P}_{\mathbf{x}_0}$ is parameterised using two epistemic multipliers $\lambda_{\mathbf{x}_{0-1}}$ and $\lambda_{\mathbf{x}_{0-2}}$ which scale the covariance matrix, reducing the initial uncertainty for multipliers < 1 , or increasing the initial uncertainty for multipliers > 1 . The multiplier range $[1/5^2, 5^2]$ means that the standard deviations of the initial state may be reduced or increased by a factor 5 of their reference value computed from Table 2.

Note that the definition and parameterisation of the family of distributions are very much dependent on the nature of the epistemic uncertainty that one is considering. In this illustrative example, we maintain the parameterisation of a Gaussian type of family of distributions. More general parameterisations are also possible, see [46] for an example of the use of Bernstein polynomials to represent generic families on a bounded support. Although the inputs' distribution is parametric, the state distribution in time will have a generic form thanks to the particle-based method, which captures the dynamical and observational nonlinearities. The range and distribution of the uncertain parameters $\lambda_{\mathbf{x}_{0-1}}$ and $\lambda_{\mathbf{x}_{0-2}}$ can be derived from a statistical analysis of a time series of observations or Collision Data Messages. In this example, we take the more general case in which the parameters are defined within intervals, but their distribution is unspecified.

2. Observation model and errors

We assumed that a new observations are received in between the time at which the state of the piece of debris and operational satellite are first acquired and the rTCA. The new observations provide an indirect measurement of the state of the piece of debris. The measured quantities are the debris azimuth and elevation with respect to the equatorial plane [47], such that the ideal measurement model is expressed as

$$h(t_k, \mathbf{x}_k, \boldsymbol{\varepsilon}_k) = h(t_k, \mathbf{x}_k) + \boldsymbol{\varepsilon}_k = \begin{bmatrix} h_{az}(t_k, \mathbf{x}_k) \\ h_{el}(t_k, \mathbf{x}_k) \end{bmatrix} + \boldsymbol{\varepsilon}_k = \begin{bmatrix} \arctan \frac{\mathbf{x}_k(2)}{\mathbf{x}_k(1)} \\ \arcsin \frac{\mathbf{x}_k(3)}{\|\mathbf{x}_k(1:3)\|} \end{bmatrix} + \boldsymbol{\varepsilon}_k, \quad (38)$$

where $\mathbf{x}_k(i)$ indicates the i -th element of the vector \mathbf{x}_k , i.e. the Cartesian position elements for $i = 1, \dots, 3$, $h(t_k, \mathbf{x}_k)$ indicates the ideal model, and the noise term is assumed additive. The noise is modelled as zero-mean normal distribution with diagonal covariance $\Sigma_{\mathbf{y}_k}$ generated using the standard deviations in Table 3. The measurements $\bar{\mathbf{y}}_k$ are simulated

Table 3 1σ azimuth (az) and elevation (el) uncertainty for noisy measurements of debris.

$1\sigma_{az}$ [arcsec]	$1\sigma_{el}$ [arcsec]
10	10

starting from the debris reference trajectory, i.e. from the initial conditions $\boldsymbol{\mu}_{\mathbf{x}_0}$, and are affected by an error drawn from the distribution of ε_k . The observation covariance $\Sigma_{\mathbf{y}_k}$ is assumed to be partially unknown due to poor sensor characterisation. Therefore, the likelihood epistemic set $\mathcal{P}_{\mathbf{Y}_k|\mathbf{X}_k}$ is parameterised as

$$\begin{aligned} \mathcal{P}_{\mathbf{Y}_k|\mathbf{X}_k} &= \left\{ p(\bar{\mathbf{y}}_k|\mathbf{x}_k) : p(\bar{\mathbf{y}}_k|\mathbf{x}_k) = \mathcal{N}\left(h(t_k, \mathbf{x}_k); \bar{\mathbf{y}}_k, \tilde{\Sigma}_{\mathbf{y}_k}\right), \right. \\ &\quad \tilde{\Sigma}_{\mathbf{y}_k} = \text{diag}(\lambda_{\mathbf{y}_{\Sigma-1}} \Sigma_{\mathbf{y}_k}(1, 1), \lambda_{\mathbf{y}_{\Sigma-2}} \Sigma_{\mathbf{y}_k}(2, 2)), \\ &\quad \left. \lambda_{\mathbf{y}_{\Sigma-1}} \in [1/5^2, 5^2], \lambda_{\mathbf{y}_{\Sigma-2}} \in [1/5^2, 5^2] \right\}, \end{aligned} \quad (39)$$

where $\Sigma_{\mathbf{y}_k}(1, 1)$ and $\Sigma_{\mathbf{y}_k}(2, 2)$ indicate respectively the azimuth and elevation variance values of the reference covariance matrix $\Sigma_{\mathbf{y}_k}$ as resulting from the standard deviations in Table 3. The unknown epistemic parameters $\lambda_{\mathbf{y}_{\Sigma-1}}$ and $\lambda_{\mathbf{y}_{\Sigma-2}}$ are constant for all the acquired measurements $k = 1, \dots, M$ as they are intended to describe the lack of knowledge about the accuracy of the sensor and not the accuracy of a specific observation. The resulting ranges of standard deviation for the azimuth and elevation measurements encompass diverse figures found in literature [48–50].

B. Wall time analysis

In this section, the wall time T_{pSIS} of a pSIS call is used as a metric to numerically check the algorithm complexity in Eq. (15) for varying numbers of observations and particles. A two dimensional grid is constructed by using $M \in \{10 + 5i : i = 1, \dots, 8\}$ observations and $N = \{5000j : j = 1, \dots, 10\}$ particles. The upper value $N = 50000$ was selected experimentally as the number of particles after which the pSIS estimator value changed only marginally (less than 1% from the previous N). The filter is run for each pair and the average wall time for 100 pSIS calls is stored. The process is repeated 30 times, i.e. until their average varied only marginally for most of the pairs (less than 1% from the previous iteration), and the average time is stored. This simulation was performed on Matlab R2020b on a macOS Big Sur 3.5GHz Dual-Core i7. The same system will be employed for the successive simulations.

Hence, linear regressions on the average wall time are run with three different candidate models:

- a) *linear* $T_{pSIS} \sim \beta_0 + \beta_1 M + \beta_2 N$;
- b) *linear with interaction* $T_{pSIS} \sim \beta_0 + \beta_1 M + \beta_2 N + \beta_3 MN$;

c) *quadratic* $T_{pSIS} \sim \beta_0 + \beta_1 M + \beta_2 N + \beta_3 MN + \beta_4 M^2 + \beta_5 N^2$.

The results of the linear regression for each of this models are shown in Table 4. Each row reports: the Root Mean Squared Error (RMSE) on the residuals; R^2 , that is the ratio between the regression’s sum of squared residuals and the total sum of squares, thus it measures the variance proportion explained by the model; p -value, that corresponds to an F-test on the regression model, the change in R^2 from one model to another as well as the p -value of the F-test on these ΔR^2 .

Table 4 Linear regression of T_{pSIS} with different models.

Model	RMSE	R^2	Model p -value	ΔR^2	ΔR^2 p -value
<i>linear</i>	0.0578	0.928	<.001	0.928	<.001
<i>linear w/ int</i>	0.0360	0.972	<.001	0.045	<.001
<i>quadratic</i>	0.0357	0.973	<.001	0.001	.21

From the regression statistics, the linear model results significant given the observed wall times as well as the estimate of the coefficients. The R^2 value shows that the linear model explains 92.8% of the observed variance. The R^2 value increases to 97.2% with a significant change when the interaction term MN is included in the regression model, as expected from the theoretical complexity analysis. Also, the RMSE in this model decreased by approximately 37% with respect to the linear model. When compared to the linear p model with interaction, the R^2 increase is not significant and only marginal, and the RMSE decreased by less than 1%.

Thus, this analysis provides empirical evidence that the numerical complexity of a p SIS call is indeed linear with interaction on the number of observations and particles as derived in Eq. (15).

C. Results

Four main instances of the conjunction scenario are considered in this section. In the first, no observations are simulated and we compute the collision probability resulting from uncertainty propagation only. In the last three, we simulate observations up until 24 hours before the rTCA. In the first two instances, the true unknown trajectory of the FENGYUN 1C DEB debris eventually results in a collision with SENTINEL 2B. In the second, after the robust PoC analysis 24 hours before rTCA, the operator decides to take further measurements before deciding whether to manoeuvre or not. In the third one, the debris’ true unknown trajectory does not result in a collision although it passes close to SENTINEL. Similarly, extra measurements are employed to decide whether to implement a manoeuvre. The last instance is again a collision trajectory, but instead of taking new measurements, they are simulated to see how they would impact the PoC. Different subcases are tested for this last instance to test the filter scalability further. The filter settings and performance for each instance are reported in Table 5.

1. Uncertainty propagation

For instance A, an initial uncertainty propagation using $N = 50000$ particles, the maximum N used in the wall time analysis, and without any measurements, that is $M = 0$, is performed to compute the probability bounds resulting from uncertainty on the initial conditions (37). The PoC is computed using a Hard-Body Radius (HBR) corresponding to the threshold $\delta_{DCA} = 50\text{m}$, which is conservative for the combined satellites' dimensions. Fig. 3 represents 3σ ellipses in b-plane coordinates. The ellipses are reconstructed from the propagated samples and weights. The different ellipses correspond to different values of the epistemic parameters, which impact the weights. Besides, each sample is characterised by a different incoming asymptote, which correspond to different B-plane orientation [51]. Hence, Fig. 3 and successive ones represent all the ellipses in a unique plane, which is the average of all the realisations-specific b-planes for a given conjunction instance. The ellipses corresponding to the epistemic parameters $\underline{\lambda}$ and $\bar{\lambda}$, respectively yielding the lower $\underline{\text{PoC}}$ and upper $\bar{\text{PoC}}$, are plotted as well with the same line style.

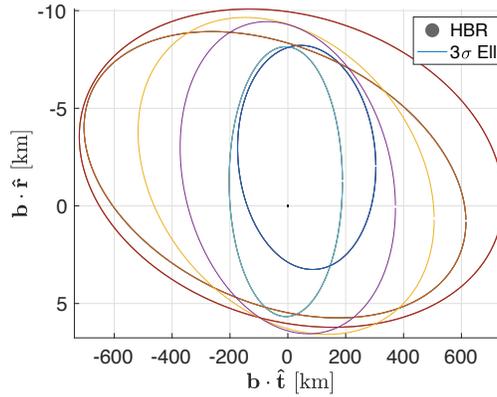


Fig. 3 Debris' HBR and b-plane 3σ ellipses for instance A.

One could also plot the Empirical Cumulative Distribution Function (ECDF) of the DCA from the propagated samples and weights as displayed in Fig. 4. The distribution corresponding to the lower probability, henceforth labelled as lower distribution, is displayed in green, whereas the one corresponding to the upper probability, labelled upper distribution, is represented in red. The dashed grey distributions are the ones resulting from different epistemic parameters within the imprecise sets. Fig. 4(a) displays the full distributions. Fig. 4(b) is a zoom of Fig. 4(a) for DCA up to 500m. For the threshold distance $\delta_{DCA} = 50\text{m}$, the collision probability is within the robust interval

$$\text{PoC}_{50\text{m}} \in [0, 4.97 \cdot 10^{-5}].$$

The filter performance for this instance A is reported in Table 5. The propagation step, including surrogate construction and particle propagations, requires approximately half of the total computational time. The time dedicated to construct the proposal is zero as no observations are employed in this instance.

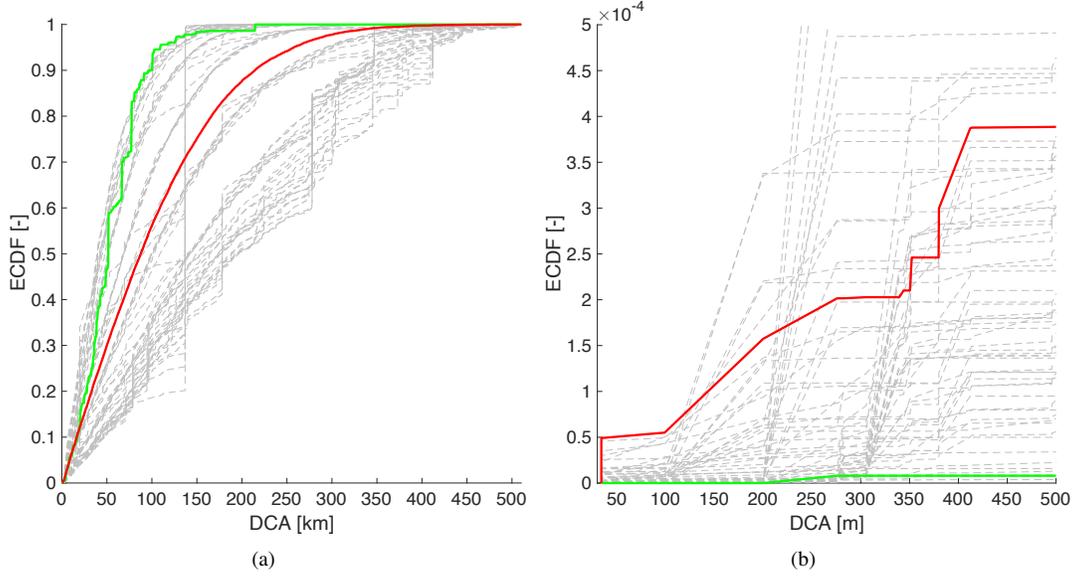


Fig. 4 DCA's ECDFs for instance A. Fig. 4(b) is a zoom of Fig. 4(a) for DCAs up to 500 meters.

It is important to underline that the lower and upper distributions do not necessarily result in lower and upper bounds for the collision probability at values of the threshold $\delta_{DCA} \neq 50\text{m}$. Specifically, the lower and upper distributions are not lower and upper envelopes for the ECDFs resulting from the imprecise set, but they are distributions belonging to the set itself bounding the expectation of a specific quantity of interest. From Fig. 4 we can see that the lower and upper distributions that bound the PoC for $\delta_{DCA} = 50\text{m}$ are not bounding the ECDFs for thresholds larger than $\approx 100\text{m}$. Even more, from Fig. 4(a) we can see that the green line is above the red one for DCA thresholds larger than $\approx 20\text{km}$.

The ellipses and ECDFs are reconstructed from the uncertainty propagation of the TLE-derived uncertainties. ESA's guidelines for pure Two-Line Elements (TLE) screening, set a threshold of $\text{PoC} > 10^{-4}$ for $\delta_{DCA} = 300\text{m}$ to trigger further analyses. For such distance, by visual inspection of Fig. 4(b), the upper probability is $\overline{\text{PoC}}_{300\text{m}} > 5 \cdot 10^{-4}$. Thus, the upper probability value flags up a potential collision that needs to be further investigated. On the other hand, the lower probability is $\underline{\text{PoC}}_{300\text{m}} < 1 \cdot 10^{-5}$, which would indicate a near-safe conjunction. The difference between these two bounds is the reason why it is critical to model and handle epistemic uncertainty in collision risk assessment. Specifically, the width of the probability interval indicates the degree of ignorance of the probability estimate. Thus, it should be used to support the decision-maker and decide whether to acquire new measurements or to manoeuvre. In the next sections, we will use the distance threshold of $\delta_{DCA} = 50\text{m}$ as it is tailored for the satellites under assessment.

2. Collision instance

For instance B, a tracking campaign with $M = 12$ noisy observations is considered between the satellites' epoch dates and 24 h before the rTCA. At this cut-off time, the RPF is run to estimate the PoC bounds. The ellipses resulting

from said simulation are displayed in Fig. 5(a). The collision probability now belongs to the robust interval

$$\text{PoC}_{50\text{m}} \in [0, 2.15 \cdot 10^{-4}] .$$

The effective sample size at the last propagation step is greater than 75%, and therefore it is deemed suitable for the estimator approximation given the high number of particles employed. The upper PoC increased by almost an order of magnitude when the new tracking observations are taken into account. Still, for some combination of the initial and likelihood epistemic parameters, the PoC could be zero. It is important to note how epistemic uncertainty on the initial distribution and observation likelihood result in B-plane ellipses, which vary in size, orientation, and central point. Thus, algorithms that tackle the dilution of collision probability by just tuning the covariance at TCA would fail to capture the actual large variety of distributions resulting from epistemic uncertainty.

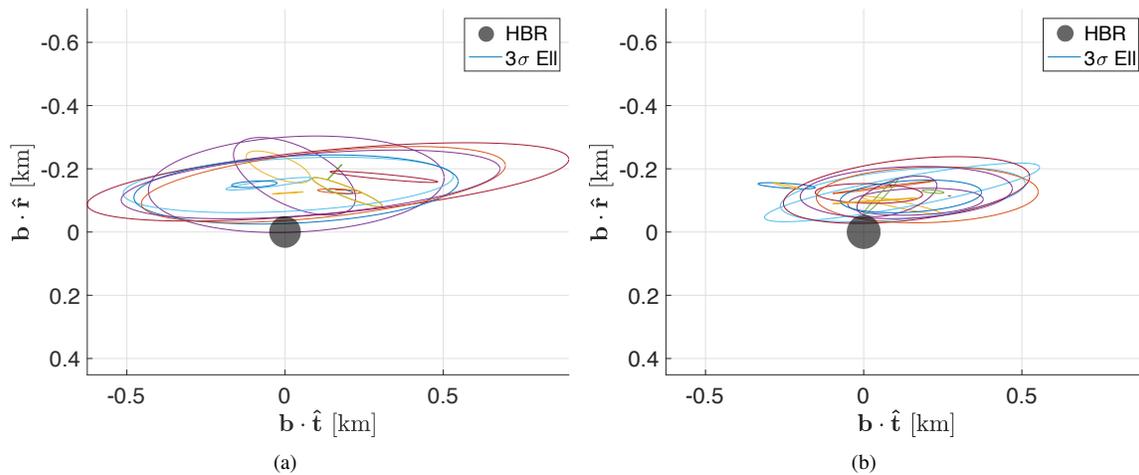


Fig. 5 Debris' HBR and b-plane 3σ ellipses for instance B with observations until (a) 24 h and (b) 9 h before rTCA.

We now investigate how these bounds change if new observations are taken into account. In particular, two new sets of azimuth and elevation measurements are employed until 9 h before the rTCA. This second cut-off time is chosen to leave enough time for further observations and the possible implementation of a manoeuvre. The new DCA distributions are shown in Fig. 5(b). The ellipses are more concentrated near the HBR of the operational satellite. Thus, the collision probability interval is now

$$\text{PoC}_{50\text{m}} \in [0, 9.69 \cdot 10^{-3}] .$$

The upper probability is almost 1%, which flags up a critical collision risk. Again, the lower bound is 0 although the unknown trajectory is actually a collision one, and the last observation is relatively close to the rTCA mainly because of probability dilution resulting from specific combinations of the initial condition and likelihood covariance matrices. This

difference further highlights the importance of properly quantifying epistemic uncertainty in collision risk assessment.

The filter performance for this instance B is reported in Table 5. The propagation requires the most computational time whereas the optimisation converges in less than 1000 iterations. When compared to instance A, the use of observations causes a higher computational time for: i) the surrogate construction as a new polynomial expansion needs to be constructed for each observation time; ii) the bound computation, despite the lower number of B&B iterations, as the pSIS estimator complexity increases with the number of measurements. The proposal requires a smaller share of time as it employs the surrogate to propagate the sigma points and it requires only few efficient matrix operations.

3. No-collision instance

For instance C, the true unknown initial velocity is slightly perturbed by ≈ 1 m/s to cause a reference miss distance of 1 km. Again a tracking campaign with $M = 12$ is considered up to 24 h before the rTCA. The resulting ellipses are shown in Fig. 6(a). The collision probability now belongs to

$$\text{PoC}_{50\text{m}} \in [0, 2.82 \cdot 10^{-4}] .$$

Although the true unknown trajectory does not result in a collision, we see how the upper PoC computed 24h before rTCA is similar to the one in the collision instance. This happens for two reasons: the measurements are affected by noise; the dynamics is highly nonlinear and evolves fast in time. The former causes larger uncertainties in the state estimates. The latter translates even small deviations at the cut-off time into significant differences at rTCA. Indeed, the satellites revolve more than 14 times in the 24h time span.

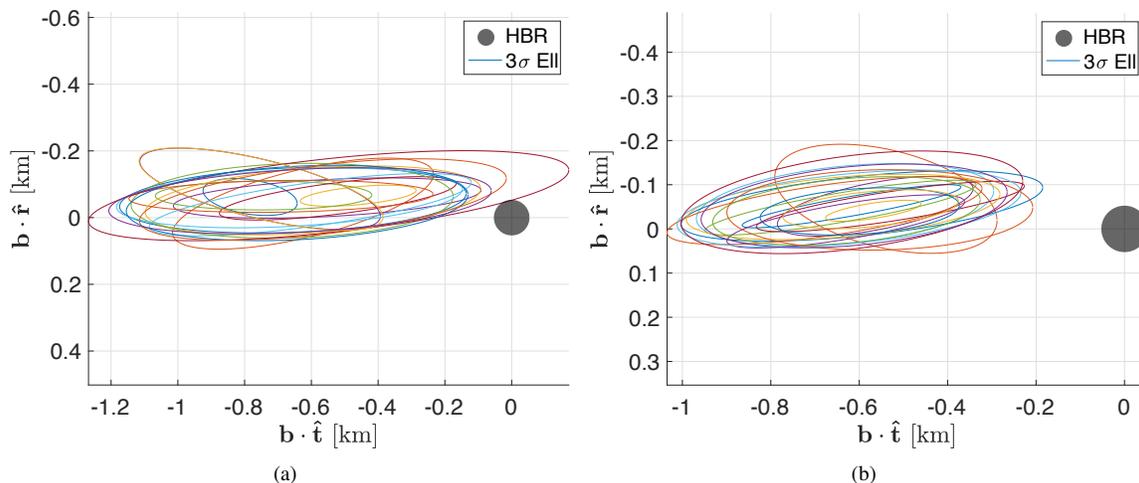


Fig. 6 Debris' HBR and b-plane 3σ ellipses for instance C with observations until (a) 24 h and (b) 9 h before rTCA.

Again, we investigate how the bounds change with the two new sets of measurements until 9 h before rTCA. The

new ellipses are plotted in Fig. 6(b), and the probability interval is now

$$\text{PoC}_{50\text{m}} \in [0, 0] .$$

The width of the probability interval is zero, meaning that the estimated probability is 0 no matter the value of the epistemic parameters. Hence, although the $\overline{\text{PoC}}$ at the 24 h cut-off was larger than the one in the collision instance, the PoC interval at the 9 h cut-off actually reflects the no-collision instance.

Again, the performance for this instance C is displayed in Table 5. In this case, the bound computation routine requires more iterations to estimate the robust bounds, whereas the propagation and proposal steps take approximately the same computational time.

4. Measurements simulation instance

The last instance D is again the collision one in Section III.C.2, but the measurements after the 24h cut-off are simulated rather than acquired. The simulations of future observations is useful to understand and quantify the impact of future measurements on the evolution of the PoC. If the PoC estimates do not change significantly with successive measurements, it would be optimal to manoeuvre, if necessary, as soon as possible without waiting for further measurements. If the probability interval changes significantly, and specifically becomes wider, successive observations can help to discriminate whether a manoeuvre is truly needed or not.

The likelihood mean in Eq. (39) was set to be the received measurement itself. Simulated measurements are characterised by further epistemic uncertainty on the likelihood mean because no observation has been received yet. Thus, for this instance, the likelihood epistemic set for the simulated future measurements $k = M + 1, M + 2$ is parameterised as

$$\begin{aligned} \mathcal{P}_{\mathbf{Y}_k|\mathbf{x}_k} &= \left\{ p(\mathbf{y}_k|\mathbf{x}_k) : p(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}\left(h(t_k, \mathbf{x}_k); \mathbf{y}_k, \tilde{\Sigma}_{\mathbf{y}_k}\right), \right. \\ \mathbf{y}_k &= \begin{bmatrix} h_{az}(t_k, \bar{\mathbf{x}}_k) + \lambda_{\mathbf{y}_{\mu_k-1}} \sqrt{\Sigma_{\mathbf{y}_k}(1, 1)} \\ h_{el}(t_k, \bar{\mathbf{x}}_k) + \lambda_{\mathbf{y}_{\mu_k-2}} \sqrt{\Sigma_{\mathbf{y}_k}(2, 2)} \end{bmatrix} \\ \tilde{\Sigma}_{\mathbf{y}_k} &= \text{diag}(\lambda_{\mathbf{y}_{\Sigma-3}} \Sigma_{\mathbf{y}_k}(1, 1), \lambda_{\mathbf{y}_{\Sigma-4}} \Sigma_{\mathbf{y}_k}(2, 2)), \\ \lambda_{\mathbf{y}_{\mu_k-1}} &\in [-5, 5], \lambda_{\mathbf{y}_{\mu_k-2}} \in [-5, 5] \\ \lambda_{\mathbf{y}_{\Sigma-3}} &\in [1/5^2, 5^2], \lambda_{\mathbf{y}_{\Sigma-4}} \in [1/5^2, 5^2] \left. \right\}, \end{aligned} \quad (40)$$

where $\lambda_{\mathbf{y}_{\mu_k-1}}$ and $\lambda_{\mathbf{y}_{\mu_k-2}}$ are two new epistemic parameters modelling the uncertainty on each observation value. These epistemic parameters regulate deviations from a reference measurement value $h(t_k, \bar{\mathbf{x}}_k)$ taken along the reference trajectory. The epistemic parameters $\lambda_{\mathbf{y}_{\Sigma-3}}$ and $\lambda_{\mathbf{y}_{\Sigma-4}}$ are the covariance multipliers which are constant for each new

simulated measurement as before.

For this instance, six subcases are run to investigate the filter scalability. The first three (D1, D2 and D3) employ a set of simulated measurements until 9 h before collision, while the last three (D4, D5 and D6) simulate an extra set of observations until 6 h before rTCA. Then, they differ for the B&B convergence thresholds ε as reported in Table 5.

The results of the RPF runs on instances D3 and D6 are shown in Fig. 7. Specifically, in Fig. 7(a) the measurements are simulated at the same time instance as in Section III.C.2. The probability interval for D3 is

$$\text{PoC}_{50\text{m}} \in [0, 2.94 \cdot 10^{-2}] .$$

By comparing this interval with the corresponding one in Fig. 5(b), we can see that the value of the upper probability is larger, up to almost 3%. Likely, this is due to the noise on the measurement in the first instance, whereas here the observation corresponding to the collision trajectory is included in the epistemic set.

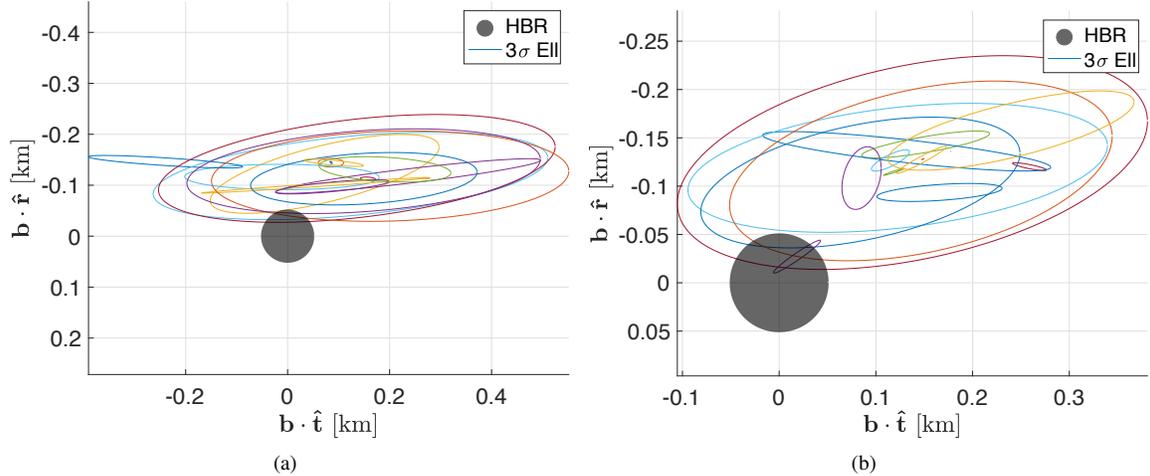


Fig. 7 Debris' HBR and b-plane 3σ ellipses for instance (a) D3 and (b) D6.

Fig. 7(b) shows the ellipses for D6 when the measurements are simulated until 6 h before rTCA, that is less than 4 revolutions before the possible impact. The probability interval becomes

$$\text{PoC}_{50\text{m}} \in [1.20 \cdot 10^{-8}, 9.94 \cdot 10^{-1}] .$$

The upper probability shows that for specific measurement values within the imprecise set, the PoC estimate would be very close to 100%, therefore predicting the collision. On the other hand, the lower probability bound indicates that future measurements can still imply an extremely low collision risk. The width of such interval indicates a great degree of ignorance. In such a case, new better measurements should be acquired to narrow the PoC interval and to allow a better informed decision.

The filter performances for these instances are shown in Table 5 as well. The precomputation times are in line with instances B and C. The time to compute the bounds sharply increases primarily with the number of epistemic dimensions, mainly due to the initial triangulation complexity, and then with a more demanding convergence threshold. Nonetheless, the time increase is not factorial as the filter upper bound complexity thanks to early pruning.

Table 5 Filter performance for different conjunction instances.

Instance	n_λ	ε	L	Time				pSIS calls
				Surrogate [s]	Proposal [s]	Optimisation [s]	Total [min]	
A	2	10^{-6}	$6.03 \cdot 10^{-7}$	76.38	.00	80.02	2.60	1023
B	4	10^{-6}	$2.63 \cdot 10^{-6}$	150.62	26.97	117.91	4.93	903
C	4	10^{-6}	$1.67 \cdot 10^{-6}$	150.04	32.42	280.39	7.71	2774
D1		10^{-4}				784.82	16.20	4140
D2	8	10^{-5}	$7.83 \cdot 10^{-5}$	154.27	33.17	1265.25	23.76	6227
D3		10^{-6}				1466.72	27.57	7376
D4		10^{-4}				1555.66	29.16	9752
D5	10	10^{-5}	$2.57 \cdot 10^{-4}$	159.58	34.44	1736.89	32.18	11636
D6		10^{-6}				1922.37	35.27	13043

IV. Conclusions

The main contribution of this work is a Robust Particle Filter that can handle both aleatory and epistemic uncertainty. It was demonstrated how this filter can be effectively used to compute robust bounds on the probability of collision between satellites and debris.

The filtering problem is formulated in terms of the expectation on the realisations of a quantity of interest, given a mix of aleatory and epistemic uncertainty in prior, likelihood, and model parameters' distributions. The output is an interval bounding the expectation from above and below.

By exploiting a precomputation of the set of possible realisations, performed with a standard particle filtering approach, the RPF proposed in this paper was shown to efficiently compute an estimation of the bounds by an inexpensive tuning of the importance weights. Since the computation of the upper and lower bounds requires the solution of a global optimisation problem, we proposed a simplex-based B&B optimiser exploiting the estimator's Lipschitz continuity. We theoretically derived the computational complexity of the estimator and the filter and demonstrated the convergence of the B&B optimiser with an estimation of the error on the exact value of the bounds.

When applied to the robust estimation of the PoC between SENTINEL 2B and FENGYUN 1C debris, the RPF was shown to efficiently compute robust probability bounds under both aleatory and epistemic uncertainty in the prior and likelihood distributions. The combined use of a UKF-based proposal and the polynomial propagator ensured a large

and satisfactory effective sample size for the estimators' computation. The last instance, in which we assumed that the measurements at later times were simulated rather than acquired, demonstrated how the RPF could be used to make robust decisions on whether to implement a manoeuvre or acquire new measurements in highly uncertain situations. With the numerical experiments we confirmed the correctness of the theoretical prediction of the performance of the filter and demonstrated how the filter can compute tight bounds enclosing all realisations of the PoC for a given threshold. We also demonstrated that in some cases epistemic uncertainty can lead to fairly wide bounds, which suggests that a poor consideration for epistemic uncertainty would give operators a false confidence in the value of the PoC.

The conjunction instances analysed in this paper illustrated how the filter could handle epistemic uncertainty. However, more realistic and general scenarios would need to include epistemic uncertainty on both objects and on the prior mean. The applicability of the filter to these more general and complex cases was recently demonstrated and will appear in future publications.

Given the results in this paper, different aspects would deserve further investigations: the use of a resampling strategy to enhance the the representation accuracy of the estimator if the effective sample size degrades; an analysis of alternative methods to construct the proposal; a more scalable initial domain decomposition strategy to reduce the computational complexity with the number of epistemic dimensions; the use of alternative optimisation approaches in place of the pure branch and bound method used in this paper. A furthermore element of future investigation will be to consider additional information on the uncertainty associated to the epistemic parameters. In fact, in this paper we considered the case in which no information on the distribution of the epistemic parameters was available. This corresponds to a single pair of boundaries for the probability of collisions with confidence equal to 1. If more information is available, the domain of the epistemic parameters can be partitioned in subdomains with associated probability mass and the bounds on the probability of collision can be computed with the same machinery presented in this paper but with an associated confidence that depends on the distribution of the epistemic parameters. Last but not least future analyses will consider more complex collision models, e.g., accounting for the objects' shape, orientation, and associated uncertainties, to provide more accurate risk metrics.

A. Estimator Derivatives

The derivatives of the pSIS estimator can be computed analytically. The derivative knowledge is extremely valuable in the bound computation (6) as it enables the use of efficient gradient descent methods to improve the exploitation stage in the optimisation. Furthermore, because the precomputed SIS works on fixed samples $\mathcal{X}_{0:k}$, there would be no noise due to sampling in the derivative information.

Let us assume that we can compute the derivatives of the density functions in Equation (4), that is, we can evaluate $\nabla_{\lambda_0} p(\mathbf{x}_0; \lambda_0)$, $\nabla_{\lambda_x} p(\mathbf{x}_k | \mathbf{x}_{k-1}; \lambda_x)$, and $\nabla_{\lambda_y} p(\mathbf{y}_k | \mathbf{x}_k; \lambda_y)$. The quantity to compute is the gradient of the estimator

with respect to the epistemic parameters

$$\nabla_{\lambda} \hat{\theta} = \left[\nabla_{\lambda_0} \hat{\theta}, \nabla_{\lambda_x} \hat{\theta}, \nabla_{\lambda_y} \hat{\theta} \right]^T. \quad (41)$$

By the linearity of the derivative operator and the chain rule, the gradient can be computed as

$$\nabla_{\lambda} \hat{\theta} = \sum_{i=1}^N \phi^{(i)} \nabla_{\lambda} \hat{w}_M^{(i)} \quad (42)$$

since the precomputed function evaluations are independent from λ .

Hence, the estimator gradient is obtained by computing the derivative of the weights. By Eq. (11) and chain rule again, the weight derivative $\nabla_{\lambda} \hat{w}_k^{(i)}$ can be computed by the previous weight derivative $\nabla_{\lambda} \hat{w}_{k-1}^{(i)}$ as

$$\nabla_{\lambda} w_k^{(i)} = \begin{bmatrix} 0 \\ p(\mathbf{y}_k | \mathbf{x}_k^{(i)}; \lambda_y) \nabla_{\lambda_x} p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}; \lambda_x) \\ \nabla_{\lambda_y} p(\mathbf{y}_k | \mathbf{x}_k^{(i)}; \lambda_y) p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}; \lambda_x) \end{bmatrix} \frac{\hat{w}_{k-1}^{(i)}}{\pi_k^{(i)}} + \frac{p(\mathbf{y}_k | \mathbf{x}_k^{(i)}; \lambda_y) p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}; \lambda_x)}{\pi_k^{(i)}} \nabla_{\lambda} \hat{w}_{k-1}^{(i)} \quad (43a)$$

$$\nabla_{\lambda} \hat{w}_k^{(i)} = \sum_{j=1}^N \frac{\partial w_k^{(i)}}{\partial w_k^{(j)}} \nabla_{\lambda} w_k^{(j)}, \quad (43b)$$

with

$$\frac{\partial w_k^{(i)}}{\partial w_k^{(j)}} = -\frac{w_k^{(i)}}{\left(\sum_{k=1}^N w_k^{(k)}\right)^2} + \frac{\delta_{ij}}{\sum_{k=1}^N w_k^{(k)}} \quad (44)$$

where $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ otherwise. The initial conditions for the weights derivative computation are $\nabla_{\lambda} w_0^{(i)} = \left[\nabla_{\lambda_0} p(\mathbf{x}_0^{(i)}; \lambda_0) / \pi_0^{(i)}, 0, 0 \right]^T$ because the initial distribution does not depend on the transition and likelihood epistemic parameters.

Let $c_{\partial p}$ be the computational complexity of evaluating one of the derivative (assumed equal complexity), and $\lambda = \lambda_0 + \lambda_x + \lambda_y$ the total number of epistemic parameters. The computational complexity of evaluating the derivative of the precomputed SIS estimator is

$$C_{\partial p \text{ sis}}(M, N) = \mathcal{O}(MN\lambda c_{\partial p} + MN^2) = \mathcal{O}(MN + MN^2), \quad (45)$$

where the quadratic term comes from Eq. (43b), and the quantities already computed in Algorithm 2 have not been recomputed, e.g. evaluating the densities and the computation of the weights. The terms $c_{\partial p}$ and λ have been retained in the first big O notation because their values influence the algorithm complexity significantly.

B. B&B Algorithm

The algorithmic flow of the global optimisation is presented in Algorithm 5. The algorithm is a classical B&B

Algorithm 5 Algorithmic scheme for B&B over simplexes.

Let:

- $\varepsilon \in \mathbb{R}^+$ a termination threshold
- \mathcal{L}_0 initial list of simplexes covering Ω
- $L_0 = \min_{\mathcal{S}_0^{(i)} \in \mathcal{L}_0} lb(\mathcal{S}_0^{(i)})$ initial lower bound of $\hat{\theta}$
- $U_0 = \min_{\mathcal{S}_0^{(i)} \in \mathcal{L}_0} ub(\mathcal{S}_0^{(i)})$ initial upper bound of $\hat{\theta}$
- $k = 0$ the iterator

while $U_k - L_k > \varepsilon$

- 1: Branch most promising simplex $lb(\mathcal{S}_k^*) = L_k$
 $\mathcal{S}_k^* \rightarrow \mathcal{S}_{k+1}^{*1}, \mathcal{S}_{k+1}^{*2}$
- 2: Construct new list of simplexes
 $\mathcal{L}_{k+1} = (\mathcal{L}_k \setminus \mathcal{S}_k^*) \cup \mathcal{S}_{k+1}^{*1} \cup \mathcal{S}_{k+1}^{*2}$
- 3: Compute new lower and upper bounds of $\hat{\theta}$
 $L_{k+1} = \min_{\mathcal{S}_{k+1}^{(j)} \in \mathcal{L}_{k+1}} lb(\mathcal{S}_{k+1}^{(j)})$
 $U_{k+1} = \min_{\mathcal{S}_{k+1}^{(j)} \in \mathcal{L}_{k+1}} ub(\mathcal{S}_{k+1}^{(j)})$
- 4: Discard non-optimal simplexes
 $\mathcal{L}_{k+1} = \mathcal{L}_{k+1} \setminus \mathcal{S}_{k+1}^{(j)}$ if $lb(\mathcal{S}_{k+1}^{(j)}) > U_{k+1}$
- 5: Update iterator
 $k = k + 1$

end while

starting with an initial list of simplexes \mathcal{L}_0 covering the search space Ω . On each simplex $\mathcal{S}_0^{(i)}$, the function minimum is bounded from below $lb(\mathcal{S}_0^{(i)})$ and from above $ub(\mathcal{S}_0^{(i)})$. The minimum of both is kept in the variables L_0 and U_0 respectively, such that $\hat{\theta}$ is bounded between them $L_0 \leq \hat{\theta} \leq U_0$. Then, starting from $k = 0$, the B&B iterates until the upper and lower bound are closer than the set threshold ε . The most promising simplex is branched into two sub-simplexes in each iteration according to the branching rule employed. The two new simplexes are added to the list \mathcal{L}_{k+1} and the old one removed. The bounding procedure is repeated on the new simplexes, and the values L_{k+1} and U_{k+1} updated. Because $L_{k+1} \leq \hat{\theta} \leq U_{k+1}$, the simplexes whose lower bound $lb(\mathcal{S}_{k+1}^{(j)})$ is larger than U_{k+1} can be automatically deemed as non-optimal and therefore discarded. Finally, the iterator is updated and the loop repeated until convergence.

C. B&B proofs

The proof of Lemma 1 is as follows.

Proof. As for the lower bound, from Eq. (22) we have $n + 1$ inequalities $\hat{\theta}(\boldsymbol{\lambda}) \geq \hat{\theta}(\boldsymbol{\lambda}_j) - L\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_j\|$, for every $\boldsymbol{\lambda}_j$. Therefore, $\hat{\theta}(\boldsymbol{\lambda})$ must be larger than the largest r.h.s.

$$\hat{\theta}(\boldsymbol{\lambda}) \geq \max_j \hat{\theta}(\boldsymbol{\lambda}_j) - L\|\boldsymbol{\lambda} - \boldsymbol{\lambda}_j\|. \quad (46)$$

This r.h.s. is lb so Eq. (46) holds for every point in the simplex. In the argument of the minimum of $\hat{\theta}$ over the simplex, call it $\underline{\lambda}_{\mathcal{S}}$ (which may be non-unique), we have that $\hat{\theta}_{\mathcal{S}} = \hat{\theta}(\underline{\lambda}_{\mathcal{S}}) \geq lb(\underline{\lambda}_{\mathcal{S}})$ and by Eq. (24) we have that $lb(\underline{\lambda}_{\mathcal{S}}) \geq lb_{\mathcal{S}}^*$ resulting in the lower bound inequality $\hat{\theta}_{\mathcal{S}} \geq lb_{\mathcal{S}}^*$.

For the upper bound of the minimum, we have the trivial relationship $\hat{\theta}(\lambda_0) = ub(\mathcal{S}) \geq \hat{\theta}_{\mathcal{S}}$ because λ_0 is in the simplex, proving the upper bound inequality. □

The proof of Lemma 2 is presented below.

Proof. Let $\lambda_{lb}^* = \arg \min_{\lambda \in \mathcal{S}} lb(\lambda)$ be the argument of the minimum of lb over the simplex. Equation $\sigma(\mathcal{S}) \leq \delta$ implies that $\|\lambda_{lb}^* - \lambda\| \leq \delta$. Hence, let us consider the bound difference

$$ub(\mathcal{S}) - lb(\mathcal{S}) = |\hat{\theta}(\lambda_0) - lb(\lambda_{lb}^*)| = \hat{\theta}(\lambda_0) - \max_j [\hat{\theta}(\lambda_j) - L\|\lambda_{lb}^* - \lambda_j\|].$$

For each λ_j , we have

$$|\hat{\theta}(\lambda_0) - \hat{\theta}(\lambda_j) + L\|\lambda_{lb}^* - \lambda_j\|| \leq |\hat{\theta}(\lambda_0) - \hat{\theta}(\lambda_j)| + L\|\lambda_{lb}^* - \lambda_j\| \leq L\|\lambda_0 - \lambda_j\| + L\|\lambda_{lb}^* - \lambda_j\| \leq L\delta + L\delta,$$

where the first inequality comes from the triangle inequality, the second one from the Lipschitz condition and the last one from $\sigma(\mathcal{S}) \leq \delta$. Therefore, also for the λ_j argument of the maximum in the lb expression, we have that $ub(\mathcal{S}) - lb(\mathcal{S}) = |\hat{\theta}(\lambda_0) - lb(\lambda_{lb}^*)| \leq 2L\delta$. Hence, there exists δ such that $ub(\mathcal{S}) - lb(\mathcal{S}) \leq 2L\delta \leq \varepsilon$, and it can be simply chosen as $0 < \delta \leq \varepsilon/2L$. □

The proof of Lemma 3 is constructed as follows.

Proof. The diameter of a n_λ -simplex $\mathcal{S}_0^{(i)} \in \mathcal{L}_0$ is reduced at least by a factor $\sqrt{3}/2$ every n_λ splits [31] such that

$$\min_{\mathcal{S}_k^{(j)} \in \mathcal{L}_k} \sigma(\mathcal{S}_k^{(j)}) \leq \left(\frac{\sqrt{3}}{2}\right)^{\lfloor k/n_\lambda \rfloor} \sigma(\mathcal{S}_0^{(i)})$$

holds true. For increasing k the right hand side goes to zero because $(\sqrt{3}/2)^{\lfloor k/n_\lambda \rfloor}$ is a monotonically decreasing sequence. Therefore the minimum diameter of the simplexes generated by $\mathcal{S}_0^{(i)}$ converges to within the tolerance δ in a finite number of steps. By repeating this argument for each $\mathcal{S}_0^{(i)}$ of the N_0 simplexes in \mathcal{L}_0 we get the first part of the lemma.

By focusing again on a single $\mathcal{S}_0^{(i)} \in \mathcal{L}_0$, we can find the number of splittings $K^{(i)}$ to achieve a minimum diameter smaller or equal to a given threshold δ . Using again the bound in [31], we can write

$$\min_{\mathcal{S}_k^{(j)} \in \mathcal{L}_k} \sigma(\mathcal{S}_k^{(j)}) \leq \left(\frac{\sqrt{3}}{2}\right)^{\lfloor k/n_\lambda \rfloor} \sigma(\mathcal{S}_0^{(i)}) \leq \delta.$$

By rearranging the terms and taking the logarithm, one obtains $\lfloor k/n_\lambda \rfloor \geq \log_{\frac{2}{\sqrt{3}}} \sigma(\mathcal{S}_0^{(i)})/\delta$, which is equivalent to $k \geq n_\lambda \left\lceil \log_{\frac{2}{\sqrt{3}}} \sigma(\mathcal{S}_0^{(i)})/\delta \right\rceil$. For being conservative, one can employ σ_0 in place of $\sigma(\mathcal{S}_0^{(i)})$, that is using the largest initial diameter for every initial simplex $\mathcal{S}_0^{(i)}$ and set an upper bound on the number of simplex splits as $K^{(i)} = n_\lambda \left\lceil \log_{\frac{2}{\sqrt{3}}} \sigma_0/\delta \right\rceil$. By repeating this argument for each of the N_0 initial simplexes, we obtain a conservative upper bound on the number of LEB splits to obtain the required diameter threshold δ as

$$K = N_0 n_\lambda \left\lceil \log_{\frac{2}{\sqrt{3}}} \frac{\sigma_0}{\delta} \right\rceil,$$

thus concluding the proof. □

The proof of the convergence Theorem 1 is outlined below.

Proof. Let $\mathcal{S}_K^* = \arg \min_{\mathcal{S}_K \in \mathcal{L}_K} \sigma(\mathcal{S}_K)$ be the (possibly non-unique) simplex with smallest diameter at the k -th iteration of the B&B. Let $\mathcal{S}_{K_\varepsilon}^* \in \mathcal{L}_{K_\varepsilon}$ be the parent simplex which was split at the K_ε -th iteration, with $K_\varepsilon < k$, to obtain \mathcal{S}_k^* . From Lemma 2, there exists $\delta \in \mathbb{R}^+$ such that

$$\forall \mathcal{S} \subset \mathbb{R}^n, \sigma(\mathcal{S}) \leq 2\delta \implies ub(\mathcal{S}) - lb(\mathcal{S}) \leq \varepsilon.$$

Hence, select $K \in \mathbb{N}$ sufficiently large such that $\sigma(\mathcal{S}_K^*) \leq \delta$, which is granted to happen in a finite number of steps by Lemma 3. Then, because of the selected LEB branching rule, we have a lower bound on the diameter change as

$$\sigma(\mathcal{S}_{K_\varepsilon}^*) \leq 2\sigma(\mathcal{S}_K^*) \leq 2\delta.$$

As shown before, this implies that

$$ub(\mathcal{S}_{K_\varepsilon}^*) - lb(\mathcal{S}_{K_\varepsilon}^*) \leq \varepsilon.$$

Since $\mathcal{S}_{K_\varepsilon}^*$ was split, from passage 1 of Algorithm 5, we have that $L_{K_\varepsilon} = lb(\mathcal{S}_{K_\varepsilon}^*)$, and by definition $U_{K_\varepsilon} \leq ub(\mathcal{S}_{K_\varepsilon}^*)$.

Thus, we conclude that

$$U_{K_\varepsilon} - L_{K_\varepsilon} \leq ub(\mathcal{S}_{K_\varepsilon}^*) - lb(\mathcal{S}_{K_\varepsilon}^*) \leq \varepsilon$$

as stated in the first part of the theorem.

Finally, we need to show that $L_{K_\varepsilon} \leq \hat{\theta} \leq U_{K_\varepsilon}$. First, we need to show that the simplex containing the optimum has not been discarded in any iteration up to K_ε . Let the argument of the minimum $\lambda^* = \arg \min_{\lambda \in \Omega} \hat{\theta}(\lambda)$ be attained in the simplex $\mathcal{S}_k^* \in \mathcal{L}_k$ at a generic iteration $k \in \mathbb{N}$, that is the optimal simplex has not been discarded. By Lemma 1, in this simplex we have

$$lb(\mathcal{S}_k^*) \leq \hat{\theta} \leq ub(\mathcal{S}_k^*).$$

For all the simplexes in the list $\mathcal{S}_k \in \mathcal{L}_k$, we have by the definition of ub and U_k

$$U_k \leq ub(\mathcal{S}_k).$$

Hence, we have that the simplex containing λ^* cannot be discarded because

$$lb(\mathcal{S}_k^*) \not\leq U_k.$$

Finally, because $\lambda^* \in \mathcal{S}_k^*$, $\hat{\theta} \geq L_{K_\varepsilon}$ and $\hat{\theta} \leq U_{K_\varepsilon}$ it follows that

$$U_{K_\varepsilon} - \hat{\theta} \leq U_{K_\varepsilon} - L_{K_\varepsilon} \leq \varepsilon$$

$$\hat{\theta} - L_{K_\varepsilon} \leq U_{K_\varepsilon} - L_{K_\varepsilon} \leq \varepsilon.$$

These two inequalities directly imply

$$L_{K_\varepsilon} \leq \hat{\theta} \leq U_{K_\varepsilon},$$

thus concluding the proof. □

D. Lower bound computation

The function $\hat{\theta}(\lambda_j) - L\|\lambda - \lambda_j\|$ can be seen as the boundary of a hyper-cone C_j with apex in $[\lambda_j, \hat{\theta}(\lambda_j)]$ and half-aperture $\tan^{-1}(1/L)$. The lower bounding function $lb(\lambda)$ in a simplex \mathcal{S} (see Eq. (23)) results from the maximum of the cones generated from all the vertexes $[\lambda_0, \dots, \lambda_{n_\lambda}]$. By construction, the minimum $lb(\mathcal{S})$ is either attained at the common intersection of these $n + 1$ hyper-cones, call it λ_\cap , if this intersection is inside the simplex, or at the boundary of the simplex, call it λ_∂ , if the intersection is outside. In the latter case, the common intersection has a lower function value than the point on the boundary, that is $lb(\lambda_\cap) < lb(\lambda_\partial)$, because $lb(\lambda_\cap)$ is the minimum of the convex hull $\mathbb{C}\mathbb{H} = \text{conv}\{\lambda_0, \dots, \lambda_{n_\lambda}, \lambda_\cap\}$ and $\mathcal{S} \subset \mathbb{C}\mathbb{H}$ [52].

The intersection point λ_{\cap} can be found by reformulating Eq. (24) by introducing a slack variable $s \in \mathbb{R}$

$$\min_s s \quad (47a)$$

$$\text{s.t.} \quad \|\lambda - \lambda_j\| \geq (\hat{\theta}(\lambda_j) - s)/L, \quad (47b)$$

where Eq. (47b) holds for $j = 0, \dots, n_{\lambda}$. Problem (47) tells us that we need to find for the minimum s such that there exists a feasible point outside the n_{λ} -dimensional hyper-spheres centred in λ_j and with radius $(\hat{\theta}(\lambda_j) - s)/L$. This value is achieved at the common intersection of the boundary S_j of such hyper-spheres for the minimum feasible s .

For a given s , the intersection $\lambda_{\cap_{ij}}$ between two hyper-spheres S_i and S_j lays on an hyperplane perpendicular to the line connecting λ_i and λ_j . Hence, the intersection belongs to the plane parameterised as

$$\lambda_{\cap_{ij}} \in \frac{(\lambda_j - \lambda_i)^T}{\|\lambda_j - \lambda_i\|} \lambda - \lambda_{\cap_{0j}} = 0$$

where $\lambda_{\cap_{0j}} \in \mathbb{R}$. This holds between each pair of vertexes of the simplex. To discard redundant pairs, we can write everything with respect to $\lambda_i = \lambda_0$ for $j = 1, \dots, n_{\lambda}$. To find the value of $\lambda_{\cap_{0j}}$, let us suppose for simplicity that λ_0 is in the origin and that $\lambda_j = [d_{0j}, 0, \dots, 0]$, with $d_{0j} = \|\lambda_j - \lambda_0\|$. The intersection $\lambda_{\cap_{0j}}$ between S_0 and S_j comes from the solution of the system

$$\begin{aligned} \lambda_0^2 + \lambda_1^2 + \dots + \lambda_{n_{\lambda}-1}^2 &= (\hat{\theta}(\lambda_0) - s)^2/L^2 \\ (\lambda_0 - d_{0j})^2 + \lambda_1^2 + \dots + \lambda_{n_{\lambda}-1}^2 &= (\hat{\theta}(\lambda_1) - s)^2/L^2, \end{aligned}$$

where λ_0, λ_1 and so on are the components of the λ vector. By subtracting the equations and re-arranging the terms, we obtain

$$\lambda_{\cap_{0j}} := \lambda_0 = m_{0j} s + c_{0j} = \frac{\hat{\theta}(\lambda_j) - \hat{\theta}(\lambda_0)}{L^2 d_{0j}} s + \left(\frac{\hat{\theta}(\lambda_0)^2 - \hat{\theta}(\lambda_j)^2}{2L^2 d_{0j}} + \frac{d_{0j}}{2} \right),$$

that is, the intersection lies on a hyperplane with coordinate $\lambda_0 = \lambda_{\cap_{0j}}$. This holds true for generic λ_0 and λ_j because we can interpret $\lambda_{\cap_{0j}}$ as the distance of the hyperplane with normal $\lambda_j - \lambda_0$ from the vertex λ_0 . Hence, we can write the system of linear equations $\lambda_{\cap_{ij}}$ needs to satisfy as

$$\left[\frac{(\lambda_1 - \lambda_i)^T}{\|\lambda_1 - \lambda_i\|}, \dots, \frac{(\lambda_{n_{\lambda}} - \lambda_i)^T}{\|\lambda_{n_{\lambda}} - \lambda_i\|} \right]^T \lambda = [m_{01}, \dots, m_{0n}]^T s + [c_{01}, \dots, c_{0n}]^T \quad (48)$$

which is a classical linear equation in the form $\mathbf{dV}\lambda = \mathbf{m}s + \mathbf{c}$. The hyperplane intersection is linear in s . Indeed, it is worth underlining that the hyperplanes' intersection generally differs from the spheres' intersection because Eq. (48) is a set of necessary but not sufficient conditions for $\lambda_{\cap_{ij}}$, and they coincide only for a specific value of s . Hence, to

find a unique value, we need to add an additional constraint in s . This can be solved by requiring that the hyperplanes' intersection lays on the hyper-sphere centred in λ_0 as

$$\begin{cases} \mathbf{dV}\lambda = \mathbf{m}s + \mathbf{c} \\ \|\lambda - \lambda_0\|^2 = (\hat{\theta}(\lambda_0) - s)^2/L^2. \end{cases}$$

This system has a closed-form solution. Again, without loss of generality, let us suppose that λ_0 coincides with the origin. Then, let $\lambda = \tilde{\mathbf{m}}s + \tilde{\mathbf{c}}$ with $\tilde{\mathbf{m}} = \mathbf{dV}^{-1}\mathbf{m}$ and $\tilde{\mathbf{c}} = \mathbf{dV}^{-1}\mathbf{c}$. Plug λ into the quadratic equation to get

$$\sum_{j=1}^n (\tilde{m}_{0j}s + \tilde{c}_{0j})^2 = (\hat{\theta}(\lambda_0) - s)^2/L^2,$$

that is a simple second-order equation in s that can be solved analytically. The smallest root s_\cap out of the two should be selected because the largest one corresponds to the upper half of the double cone. The spheres' intersection can then be computed as $\lambda_\cap = \tilde{\mathbf{m}}s_\cap + \tilde{\mathbf{c}}$. By construction, the sought lower bound is

$$lb(\mathcal{S}) = lb(\lambda_\cap) = s_\cap.$$

Funding Sources

This work was funded by the European Commission's H2020 programme, through the H2020-MSCA-ITN-2016 UTOPIAE Marie Curie Innovative Training Network, grant agreement 722734.

References

- [1] Simon, D., *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*, John Wiley & Sons, 2006, Chap. 11, pp. 333–372.
- [2] Weiss, R., "An approach to Bayesian sensitivity analysis," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, No. 4, 1996, pp. 739–750. <https://doi.org/10.1111/j.2517-6161.1996.tb02112.x>.
- [3] Sarkka, S., *Bayesian Filtering and Smoothing*, 1st ed., Cambridge University Press, New York, 2013, Chap. 7, pp. 116–133. <https://doi.org/10.1017/CBO9781139344203>.
- [4] Gordon, N. J., Salmond, D. J., and Smith, A. F., "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE proceedings F (radar and signal processing)*, Vol. 140 (2), 1993, pp. 107–113. <https://doi.org/10.1049/ip-f-2.1993.0015>.
- [5] Wills, A. G., Hendriks, J., Renton, C., and Ninness, B., "A Bayesian filtering algorithm for Gaussian mixture models," *arXiv preprint arXiv:1705.05495*, 2017.

- [6] The Consultative Committee for Space Data Systems, “Recommendation for space data system standards: Conjunction Data Message,” Tech. rep., CCSDS 133.0-B-1. Blue Book, 2013.
- [7] Poore, A. B., Aristoff, J. M., Horwood, J. T., Armellin, R., Cerven, W. T., Cheng, Y., Cox, C. M., Erwin, R. S., Frisbee, J. H., Hejduk, M. D., et al., “Covariance and uncertainty realism in space surveillance and tracking,” Tech. rep., Numerica Corporation Fort Collins United States, 2016.
- [8] Carpenter, J. R., “Covariance Realism Is Not Enough,” *AAS/AIAA Astrodynamics Specialist Conference*, 2019. <https://ntrs.nasa.gov/citations/20190032049>.
- [9] Balch, M. S., Martin, R., and Ferson, S., “Satellite conjunction analysis and the false confidence theorem,” *Proceedings of the Royal Society A*, Vol. 475, No. 2227, 2019. <https://doi.org/10.1098/rspa.2018.0565>.
- [10] Li, W., and Jia, Y., “H-infinity filtering for a class of nonlinear discrete-time systems based on unscented transform,” *Signal Processing*, Vol. 90, No. 12, 2010, pp. 3301–3307. <https://doi.org/10.1016/j.sigpro.2010.05.023>.
- [11] Delande, E., Houssineau, J., and Jah, M., “Physics and human-based information fusion for improved resident space object tracking,” *Advances in Space Research*, Vol. 62, No. 7, 2018, pp. 1800–1812. <https://doi.org/10.1016/j.asr.2018.06.033>.
- [12] Delande, E., Jah, M., and Jones, B., “A New Representation of Uncertainty for Collision Assessment,” *AAS/AIAA Spaceflight Mechanics 2019*, 2019. <http://www.univelt.com/book=7406>.
- [13] Cai, H., Hussein, I., and Jah, M., “Possibilistic admissible region using outer probability measure theory,” *Acta Astronautica*, Vol. 177, 2020, pp. 246–257. <https://doi.org/10.1016/j.actaastro.2020.07.041>.
- [14] Wang, W.-S., and Orshansky, M., “Robust estimation of parametric yield under limited descriptions of uncertainty,” *Proceedings of the 2006 IEEE/ACM international conference on Computer-aided design*, 2006, pp. 884–890. <https://doi.org/10.1109/ICCAD.2006.320093>.
- [15] Masreliez, C., and Martin, R., “Robust Bayesian estimation for the linear model and robustifying the Kalman filter,” *IEEE transactions on Automatic Control*, Vol. 22, No. 3, 1977, pp. 361–371. <https://doi.org/10.1109/TAC.1977.1101538>.
- [16] Morrell, D. R., and Stirling, W. C., “Set-values filtering and smoothing,” *IEEE transactions on systems, man, and cybernetics*, Vol. 21, No. 1, 1991, pp. 184–193. <https://doi.org/10.1109/21.101148>.
- [17] Smets, P., and Ristic, B., “Kalman filter and joint tracking and classification in the TBM framework,” *Proceedings of the Seventh International Conference on Information Fusion*, Vol. 1, Citeseer, 2004, pp. 46–53. <https://doi.org/10.1016/j.inffus.2005.06.004>.
- [18] Noack, B., Klumpp, V., Brunn, D., and Hanebeck, U. D., “Nonlinear Bayesian estimation with convex sets of probability densities,” *2008 11th International Conference on Information Fusion*, IEEE, 2008, pp. 1–8.
- [19] Benavoli, A., Zaffalon, M., and Miranda, E., “Robust filtering through coherent lower previsions,” *IEEE Transactions on Automatic Control*, Vol. 56, No. 7, 2010, pp. 1567–1581. <https://doi.org/10.1109/TAC.2010.2090707>.

- [20] Augustin, T., Coolen, F. P. A., De Cooman, G., and Troffaes, M., *Introduction to Imprecise Probabilities*, John Wiley & Sons, 2014, Chap. 2, pp. 28–55. <https://doi.org/10.1002/9781118763117>.
- [21] Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, 1st ed., Bellman R. Academic Press, New York, 1970, Mathematics in Science and Engineering, Vol. 64, Chap. 6, pp. 162–194.
- [22] Doucet, A., de Freitas, N., and Gordon, N., “An Introduction to Sequential Monte Carlo Methods,” *Sequential Monte Carlo Methods in Practice*, edited by A. Doucet, N. de Freitas, and N. Gordon, Springer New York, New York, NY, 2001, Chap. 1, pp. 3–14. https://doi.org/10.1007/978-1-4757-3437-9_1.
- [23] Troffaes, M. C., “Imprecise Monte Carlo simulation and iterative importance sampling for the estimation of lower previsions,” *International Journal of Approximate Reasoning*, Vol. 101, 2018, pp. 31–48. <https://doi.org/10.1016/j.ijar.2018.06.009>.
- [24] Liu, J. S., and Chen, R., “Blind deconvolution via sequential imputations,” *Journal of the american statistical association*, Vol. 90, No. 430, 1995, pp. 567–576. <https://doi.org/10.2307/2291068>.
- [25] Van Der Merwe, R., Doucet, A., De Freitas, N., and Wan, E. A., “The unscented particle filter,” *Advances in neural information processing systems*, 2001, pp. 584–590.
- [26] Julier, S., Uhlmann, J., and Durrant-Whyte, H., “A new method for the nonlinear transformation of means and covariances in filters and estimators,” *IEEE Transactions on Automatic Control*, Vol. 45, No. 3, 2000, pp. 477–482. <https://doi.org/10.1109/9.847726>.
- [27] Greco, C., Campagnola, S., and Vasile, M., “Robust Space Trajectory Design using Belief Optimal Control,” *Journal of Guidance, Control, and Dynamics*, Manuscript submitted for publication.
- [28] Bäck, J., Nobile, F., Tamellini, L., and Tempone, R., “Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison,” *Spectral and High Order Methods for Partial Differential Equations*, Lecture Notes in Computational Science and Engineering, Vol. 76, edited by J. Hesthaven and E. Ronquist, Springer, 2011, pp. 43–62. https://doi.org/10.1007/978-3-642-15337-2_3.
- [29] Vasile, M., Ortega, C. A., and Riccardi, A., “Set propagation in dynamical systems with generalised polynomial algebra and its computational complexity,” *Communications in Nonlinear Science and Numerical Simulation*, Vol. 75, 2019, pp. 22–49. <https://doi.org/10.1016/j.cnsns.2019.03.019>.
- [30] Greco, C., Di Carlo, M., Vasile, M., and Epenoy, R., “Direct multiple shooting transcription with polynomial algebra for optimal control problems under uncertainty,” *Acta Astronautica*, Vol. 170, 2020, pp. 224–234. <https://doi.org/10.1016/j.actaastro.2019.12.010>.
- [31] Kearfott, B., “A proof of convergence and an error bound for the method of bisection in \mathbb{R}^n ,” *Mathematics of Computation*, Vol. 32, No. 144, 1978, pp. 1147–1153. <https://doi.org/10.2307/2006341>.
- [32] Žilinskas, J., “Branch and bound with simplicial partitions for global optimization,” *Mathematical Modelling and Analysis*, Vol. 13, No. 1, 2008, pp. 145–159.

- [33] Clausen, J., and Žilinskas, A., “Subdivision, sampling, and initialization strategies for simplicial branch and bound in global optimization,” *Computers & Mathematics with Applications*, Vol. 44, No. 7, 2002, pp. 943–955. [https://doi.org/10.1016/S0898-1221\(02\)00205-5](https://doi.org/10.1016/S0898-1221(02)00205-5).
- [34] Witczyński, K., “n-Cubes inscribed in simplices,” *Journal of Geometry*, Vol. 81, No. 1-2, 2004, pp. 192–198. <https://doi.org/10.1007/s00022-004-1714-6>.
- [35] Hughes, R. B., and Anderson, M. R., “Simplexity of the cube,” *Discrete Mathematics*, Vol. 158, No. 1, 1996, pp. 99–150. [https://doi.org/10.1016/0012-365X\(95\)00075-8](https://doi.org/10.1016/0012-365X(95)00075-8).
- [36] Bliss, A., and Su, F. E., “Lower bounds for simplicial covers and triangulations of cubes,” *Discrete & Computational Geometry*, Vol. 33, No. 4, 2005, pp. 669–686. <https://doi.org/10.1007/s00454-004-1128-0>.
- [37] Malherbe, C., and Vayatis, N., “Global optimization of lipschitz functions,” *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, JMLR. org, 2017, pp. 2314–2323.
- [38] Gimbutas, A., and Žilinskas, A., “On global optimization using an estimate of Lipschitz constant and simplicial partition,” *AIP Conference Proceedings*, Vol. 1776 (1), AIP Publishing LLC, 2016. <https://doi.org/10.1063/1.4965346>.
- [39] Sergeyev, Y. D., Strongin, R. G., and Lera, D., *Introduction to global optimization exploiting space-filling curves*, Springer Science & Business Media, 2013, Chap. 3, pp. 47–90. <https://doi.org/10.1007/978-1-4614-8042-6>.
- [40] “SOCRATES - Satellite Orbital Conjunction Reports Assessing Threatening Encounters in Space,” , 2021. <http://celestrak.com/SOCRATES/>, accessed: 08-Jan-2021.
- [41] Johnson, N. L., Stansbery, E., Liou, J.-C., Horstman, M., Stokely, C., and Whitlock, D., “The characteristics and consequences of the break-up of the Fengyun-1C spacecraft,” *Acta Astronautica*, Vol. 63, No. 1, 2008, pp. 128–135. <https://doi.org/10.1016/j.actaastro.2007.12.044>.
- [42] Greco, C., Sánchez, L., and Vasile, M., “A Robust Bayesian Agent for Optimal Collision Avoidance Manoeuvre Planning,” *IEEE Transactions on Aerospace and Electronic Systems*, Manuscript in preparation.
- [43] Montenbruck, O., and Gill, E., *Satellite orbits: models, methods and applications*, 1st ed., Springer Science & Business Media, 2000, Chap. 3, pp. 53–116. <https://doi.org/10.1007/978-3-642-58351-3>.
- [44] Wakker, K. F., *Fundamentals of Astrodynamics*, TU Delft Library, 2015, Chap. 8, pp. 191–202.
- [45] Klinkrad, H., Alarcon, J., and Sanchez, N., “Collision avoidance for operational ESA satellites,” *4th European Conference on Space Debris*, Vol. 587, 2005. <https://conference.sdo.esoc.esa.int/proceedings/sdc4/paper/49>.
- [46] Vasile, M., and Tardioli, C., “On the Use of Positive Polynomials for the Estimation of Upper and Lower Expectations in Orbital Dynamics,” *Stardust Final Conference*, edited by M. Vasile, E. Minisci, L. Summerer, and P. McGinty, Springer International Publishing, Cham, 2018, pp. 99–107. https://doi.org/10.1007/978-3-319-69956-1_6.

- [47] Tapley, B. D., Schutz, B. E., and Born, G. H., *Statistical orbit determination*, Elsevier, 2004, Chap. 3, pp. 93–158. <https://doi.org/10.1016/B978-0-12-683630-1.X5019-X>.
- [48] Sang, J., and Smith, C., “An analysis of observations from EOS space debris tracking system,” *11th Australian Space Science Conference, Canberra, Australia*, Citeseer, 2011, pp. 179–189.
- [49] Bennett, J. C., Sang, J., Smith, C., and Zhang, K., “An analysis of very short-arc orbit determination for low-Earth objects using sparse optical and laser tracking data,” *Advances in Space Research*, Vol. 55, No. 2, 2015, pp. 617–629. <https://doi.org/10.1016/j.asr.2014.10.020>.
- [50] Li, B., Sang, J., and Zhang, Z., “A real-time orbit determination method for smooth transition from optical tracking to laser ranging of debris,” *Sensors*, Vol. 16, No. 7, 2016, pp. 1–14. <https://doi.org/10.3390/s16070962>.
- [51] Farnocchia, D., Eggl, S., Chodas, P. W., Giorgini, J. D., and Chesley, S. R., “Planetary encounter analysis on the B-plane: a comprehensive formulation,” *Celestial Mechanics and Dynamical Astronomy*, Vol. 131, No. 8, 2019, p. 36. <https://doi.org/10.1007/s10569-019-9914-4>.
- [52] Casado, L. G., García, I., Tóth, B., and Hendrix, E. M., “On determining the cover of a simplex by spheres centered at its vertices,” *Journal of Global Optimization*, Vol. 50, No. 4, 2011, pp. 645–655. <https://doi.org/10.1007/s10898-010-9524-x>.