

AN ADAPTIVE GAUSSIAN MIXTURE APPROXIMATION FOR RECURSIVE BAYESIAN SMOOTHING OF NONLINEAR SYSTEMS

Aditya H. Desai^{*†}, B. Schneiderheinze^{*‡}, A. De Vittori[§], and Keith A. LeGrand[¶]

Cislunar systems—often characterized by nonlinear dynamics and observation models—pose challenging state estimation problems. These orbital regimes can give rise to significantly non-Gaussian distributions, particularly in regions of high nonlinearity or over periods of measurement sparsity. Adaptive Gaussian mixture filters dynamically adjust their mixture resolution to systematically approximate non-Gaussian uncertainty, offering a balance between computational efficiency and estimation accuracy. Nonlinear smoothing algorithms, which have received considerably less attention than their filter counterparts, improve estimates by refining past filter outputs retrodictively. Smoothing algorithms have demonstrated significant estimation improvements in many different systems, with particularly compelling applications to space domain awareness and navigation. While closed-form smoothing solutions are available for linear systems, recursive Bayesian smoothing in nonlinear, non-Gaussian settings poses additional theoretical and computational challenges. This work presents a new Bayesian recursive smoothing algorithm that refines Gaussian mixture posteriors produced by a forward adaptive Gaussian mixture filter. This approach uses a projection-based dimension reduction of a nonlinear corrector function alongside higher-order Taylor series expansions. A Mahalanobis-distance-based mixture of backward correctors is also proposed that circumvents the mixand lineage bookkeeping requirements of existing approaches. The proposed smoother’s estimation capabilities are demonstrated on a cislunar orbit determination problem and shown to greatly improve filtered estimates and uncertainty representations and match the performance of existing smoothing methods at a significantly reduced computational burden.

INTRODUCTION

System state estimation is a fundamental problem with applications spanning financial forecasting,^{1,2} target tracking,³ mechanical systems,⁴ and, critically, space situational awareness.⁵ Space object tracking specifically poses significant estimation challenges due to nonlinear dynamics and sparse measurements, and with the rapidly growing population of satellites, rocket bodies, and debris in Earth orbit, accurate tracking and state characterization is essential. These challenges become even more pronounced in the cislunar domain, where spacecraft are subject to multi-body dynamics, prolonged propagation intervals, and low-thrust maneuvers, often resulting in highly non-Gaussian

*These authors contributed equally to this work.

[†]M.S. Student, School of Aeronautics and Astronautics, Purdue University, 701 W. Stadium Ave, West Lafayette, IN, 47907

[‡]Ph.D. Student, School of Aeronautics and Astronautics, Purdue University, 701 W. Stadium Ave, West Lafayette, IN, 47907

[§]Postdoctoral Researcher, School of Aeronautics and Astronautics, Purdue University, 701 W. Stadium Ave, West Lafayette, IN, 47907.

[¶]Assistant Professor, School of Aeronautics and Astronautics, Purdue University, 701 W. Stadium Ave, West Lafayette, IN, 47907.

and multi-modal uncertainty. Consequently, maintaining reliable estimates in this regime necessitates a shift toward advanced estimation frameworks. Such systems must robustly handle irregular observation cadences,⁶ anticipate maneuvering intent,⁷ and utilize adaptive probabilistic methods⁸ to propagate state uncertainty through the inherent nonlinearities of the cislunar environment.⁹

In this context, state estimation typically relies on recursive filtering techniques, which fuse system dynamics with noisy and partial observations to produce real-time state estimates. While filtering enables sequential inference, early estimates within an observation window are often poorly informed due to limited data. In applications where immediacy is not required, smoothing provides a retrodictive alternative by estimating system states using the complete set of observations, both past and future, yielding more uniformly informed and accurate estimates across the time horizon.

Modern filtering methods largely originated in the 1960s with the development of the Kalman filter (KF),^{10,11} which provides the optimal linear state estimator. Extensions such as the extended Kalman filter (EKF),¹² unscented Kalman filter (UKF),¹³ and cubature Kalman filter (CKF)¹⁴ address moderate nonlinearities through local approximations. However, these approaches may degrade when system dynamics or uncertainties deviate significantly from Gaussian assumptions. Particle filter (PF) methods^{15,16} offer improved flexibility by representing the posterior through weighted samples, but often suffer from degeneracy and poor scalability in high-dimensional settings.^{17,18} Gaussian mixture filters (GMFs)^{19,20} provide a computationally tractable alternative by approximating non-Gaussian posteriors as weighted sums of Gaussian components. Adaptive variants dynamically adjust the mixture complexity through splitting and reduction operations, forming the basis of adaptive Gaussian mixture (AGM) filtering,²¹ with numerous strategies proposed in the literature.

Extending these filtering approaches, smoothing for AGM estimates is most naturally addressed through forward-backward formulations. In particular, two-filter smoothers rely on information-form representations that are incompatible with Gaussian mixtures,²² whereas forward-backward smoothers exploit the forward-pass mixture structure and introduce corrections via a backward corrector (BC), originally developed for linear systems²³ and later extended to nonlinear models.²⁴

Building on this framework, this work develops a Bayesian forward-backward AGM smoother for nonlinear systems. The proposed approach accommodates time-varying mixture cardinality without enforcing one-to-one mixand correspondence across time. Its key contributions include a local projection-based compression of the BC to reduce computational complexity, a predicted-density-transport mechanism to mitigate linearization errors between smoothing steps, and a Mahalanobis-distance-based association weighting scheme that quantifies inter-temporal mixand influence in a principled manner.

The paper is organized as follows. It first introduces the problem formulation with underlying assumptions, then reviews relevant background material. The proposed smoothing methodology is presented next, followed by an evaluation of its estimation performance within the circular restricted 3-body problem (CR3BP). Finally, the main findings are summarized and directions for future research are outlined.

PROBLEM FORMULATION

This work considers the estimation of a continuous-time dynamical system. The system is described by a state vector $\mathbf{x} \in \mathbb{R}^{n_x}$ evolving over some timespan from initial time t_0 until some final

time t_T according to the system of stochastic ordinary differential equations (ODEs) given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t) + \mathbf{\Gamma}(t)\mathbf{w}(t), \quad (1)$$

where $\mathbf{f} : \mathbb{R}^{n_x}, \mathbb{R} \rightarrow \mathbb{R}^{n_x}$ is the dynamics model, $\mathbf{\Gamma}(t)$ is the noise input matrix, and $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_c)$ is white, Gaussian continuous process noise representing unknown disturbances in the system with power spectral density \mathbf{Q}_c . The corresponding continuous-time process noise covariance is expressed as

$$\mathbf{Q}_c \delta(t - \tau) = \mathbb{E}[\mathbf{w}(t)\mathbf{w}^\top(\tau)]. \quad (2)$$

The solution flow of the deterministic portion of the system is defined as

$$\varphi_{\Delta t}(\mathbf{x}(t_0)) = \mathbf{x}(t_0 + \Delta t) = \mathbf{x}(t_0) + \int_{t_0}^{t_0 + \Delta t} \mathbf{f}(\mathbf{x}(\tau), \tau) d\tau. \quad (3)$$

For nonlinear stochastic ODEs, the moment equations give rise to an infinite dimensional system of fully-coupled ODEs and thus are generally unavailable.²⁵ Thus, an approximate discrete stochastic process is used instead, expressed as

$$\mathbf{x}_k = \varphi_{t_k - t_{k-1}}(\mathbf{x}_{k-1}) + \mathbf{\Gamma}_{k-1}\mathbf{w}_{k-1}, \quad (4)$$

where

$$\mathbb{E}[\mathbf{w}_k \mathbf{w}_{k'}^\top] = \mathbf{Q}_k \delta_k^{k'}. \quad (5)$$

The stochastic portion of the system requires some approximation both to convert it from a continuous-time form to a discrete-time form and propagate it for a given Δt . For analysis performed in this work, the discrete process noise integral is approximated for nonlinear systems as

$$\mathbf{Q}_k \approx \int_{t_k}^{t_{k+1}} \mathbf{\Phi}(t_{k+1}, \tau) \mathbf{\Gamma}_k \mathbf{Q}_c \mathbf{\Gamma}_k^\top \mathbf{\Phi}(t_{k+1}, \tau)^\top d\tau, \quad (6)$$

where $\mathbf{\Phi}$ is the state transition matrix (STM), which describes the first-order sensitivity of deviations in the output state to perturbations to the input state and is defined as

$$\mathbf{\Phi}(t_{k+1}, t_k) = \frac{\partial \mathbf{x}_{k+1}}{\partial \mathbf{x}_k}. \quad (7)$$

This system is also characterized by measurements $\mathbf{z} \in \mathbb{R}^{n_z}$, defined by the nonlinear function

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\nu}_k, \quad (8)$$

where $\mathbf{h}_k : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_z}$ is the measurement model and $\boldsymbol{\nu}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ is additive, white, Gaussian measurement noise. The corresponding measurement noise covariance is defined as

$$\mathbf{R}_k = \mathbb{E}[\boldsymbol{\nu}_k \boldsymbol{\nu}_k^\top]. \quad (9)$$

BACKGROUND

Gaussian Mixtures

In nonlinear dynamical systems, the assumption that the underlying probability density function (pdf) remains approximately Gaussian is typically valid only over short time horizons. As the system evolves and nonlinear effects accumulate, this approximation rapidly deteriorates, leading to degraded estimation accuracy. Consequently, representations capable of capturing strongly non-Gaussian and multimodal pdfs become necessary. One approach is to approximate the nonlinear filtering distribution with a weighted sum of Gaussian components, forming a Gaussian mixture (GM).^{19,20} This representation preserves the analytical tractability of Gaussian densities while enabling the mixture to capture multimodal structures that arise in nonlinear systems. In effect, each Gaussian component performs the role of many particles, providing a more computationally efficient yet expressive approximation of the true distribution. Mathematically, the GM approximation yields an estimated pdf of the form

$$p(\mathbf{x}) \approx \sum_{n=1}^N w^{(n)} \mathcal{N}(\mathbf{x}; \mathbf{m}^{(n)}, \mathbf{P}^{(n)}), \quad (10)$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is the system state, $\mathbf{m}^{(n)} \in \mathbb{R}^{n_x}$ and $\mathbf{P}^{(n)} \in \mathbb{S}_{++}^{n_x}$. The weights $w^{(n)} \in [0, 1]$ satisfy the normalization condition

$$\sum_{n=1}^N w^{(n)} = 1. \quad (11)$$

Note that even from this mixture form, the overall distribution mean and covariance of a GM can be recovered easily by the expressions

$$\mathbf{m} = \mathbb{E}[\mathbf{x}] = \sum_{n=1}^N w^{(n)} \mathbf{m}^{(n)}, \quad (12)$$

$$\mathbf{P} = \text{Cov}[\mathbf{x}] = \sum_{n=1}^N w^{(n)} \left[\mathbf{P}^{(n)} + \left(\mathbf{m}^{(n)} - \mathbf{m} \right) \left(\mathbf{m}^{(n)} - \mathbf{m} \right)^\top \right]. \quad (13)$$

Bayesian Filtering

In Bayesian forward-backward AGM smoothers, a prior filtering step is required, during which the system's state estimates are updated as new measurements become available. Let $\mathbf{x}_k \in \mathbb{R}^{n_x}$ denote the system state at discrete time t_k , and let $\mathbf{z}_k \in \mathbb{R}^{n_z}$ be the measurement obtained at the same time. The complete set of measurements up to time t_k is denoted by

$$\mathbf{z}_{1:k} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k\}. \quad (14)$$

The estimate of the state at time t_k , conditioned on all measurements up to that time, is described by the posterior probability density

$$p_{k|k}(\mathbf{x}) = p(\mathbf{x}_k | \mathbf{z}_{1:k}). \quad (15)$$

Although the Bayesian framework does not impose any specific form on the probability distribution, it is common to represent the estimate either as a single Gaussian or a GM. In highly nonlinear

systems, a GM representation is preferred due to its ability to approximate general non-Gaussian densities to arbitrary precision. This GM takes the form²⁶

$$p(\mathbf{x}_k | \mathbf{z}_{1:k}) = \sum_{n=1}^{N_k} w_k^{(n)} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_k^{(n)}, \mathbf{P}_k^{(n)}), \quad (16)$$

where $w_k^{(n)}$, $\mathbf{m}_k^{(n)}$, and $\mathbf{P}_k^{(n)}$ are the weight, mean, and covariance of the n -th component ($n = 1, \dots, N_k$) of the GM.

The system is affected by several sources of uncertainty, including the process noise w_k , representing uncertainties in the system dynamics, the measurement noise v_k , characterizing uncertainties in the observations, and the initial state uncertainty \mathbf{x}_0 , expressing prior knowledge of the system state. The noises are assumed to be conditionally independent, which allows the system to be described as a Markov process.²⁶ Under this assumption, the state evolution satisfies the Markov property, meaning that the future state \mathbf{x}_{k+1} depends only on the current state \mathbf{x}_k and the process model, not on the full history of past states. Consequently, the state estimate at time t_{k+1} can be recursively updated using only the previous estimate and the current measurement.

Whenever a new measurement \mathbf{z}_{k+1} becomes available, the posterior distribution $p_{k+1|k+1}(\mathbf{x})$ is obtained through a *prediction step*, based on the Chapman–Kolmogorov equation, followed by an *update step*, based on Bayes’ rule:^{27,28}

$$p(\mathbf{x}_{k+1} | \mathbf{z}_{1:k}) = \int p(\mathbf{x}_k | \mathbf{z}_{1:k}) f_{k+1|k}(\mathbf{x}_{k+1} | \mathbf{x}_k) d\mathbf{x}_k, \quad (17)$$

$$p(\mathbf{x}_{k+1} | \mathbf{z}_{1:k+1}) = p(\mathbf{x}_{k+1} | \mathbf{z}_{1:k}) \frac{g(\mathbf{z}_{k+1} | \mathbf{x}_{k+1})}{\int g(\mathbf{z}_{k+1} | \mathbf{x}_{k+1}) p(\mathbf{x}_{k+1} | \mathbf{z}_{1:k}) d\mathbf{x}_{k+1}}, \quad (18)$$

where $f_{k+1|k}(\cdot)$ denotes the transition kernel propagating the state from \mathbf{x}_k to \mathbf{x}_{k+1} , and $g(\cdot)$ denotes the measurement likelihood for a measurement \mathbf{z}_{k+1} at time t_{k+1} given state \mathbf{x}_{k+1} . Together, these equations constitute the standard Bayesian filtering recursion.

Bayesian Smoothing

While Bayesian filtering provides an estimate of the system state based solely on measurements available up to the current time t_k , i.e., $\mathbf{z}_{1:k}$, it does not exploit information contained in future measurements. In applications where the complete measurement sequence is available *a posteriori*, the filtering framework can be extended to a Bayesian smoothing formulation, in which the state at time t_k is estimated by conditioning on the entire set of measurements $\mathbf{z}_{1:T}$, with $T \geq k$.²⁹ In contrast to filtering, which proceeds forward in time from t_0 to t_T , Bayesian smoothing is performed backwards, starting from t_T and recursively propagating information toward the initial time t_0 . The backward recursion is initialized at the final time with a smoothed posterior equal to the filtered posterior since there is no future measurement information to incorporate. For any other time index $k < T$, the smoothed distribution $p(\mathbf{x}_k | \mathbf{z}_{1:T})$ is

$$p(\mathbf{x}_k | \mathbf{z}_{1:T}) = p(\mathbf{x}_k | \mathbf{z}_{1:k}) \int \frac{p(\mathbf{x}_{k+1} | \mathbf{z}_{1:T})}{p(\mathbf{x}_{k+1} | \mathbf{z}_{1:k})} p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{z}_{1:k}) d\mathbf{x}_{k+1}. \quad (19)$$

Unfortunately, common representations for estimated state distributions are incompatible with this probabilistic form of the smoothing equation due to the quotient of densities. If a GM form of

the state distribution is chosen, the resulting quotient of GMs is not generally a GM, which prohibits the direct application of the smoothing equation. Instead, the problematic terms are grouped into a function called the BC, which takes the following form:²³

$$B_{k|T}(\mathbf{x}_k) = \int \frac{p(\mathbf{x}_{k+1}|\mathbf{z}_{1:T})}{p(\mathbf{x}_{k+1}|\mathbf{z}_{1:k})} p(\mathbf{x}_{k+1}|\mathbf{x}_k, \mathbf{z}_{1:k}) d\mathbf{x}_{k+1}. \quad (20)$$

In this form, the quotient of densities problem is circumvented because the BC at time t_{k-1} can be computed using the BC at a future time t_k and $L_k(\cdot)$, the ‘‘pseudo-likelihood’’:

$$L_k(\mathbf{z}_k; \cdot) = \frac{g(\mathbf{z}_k|\mathbf{x}_k)}{\int g(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) d\mathbf{x}_k}, \quad (21)$$

$$B_{k-1|T}(\mathbf{x}_{k-1}) = \int B_{k|T}(\mathbf{x}_k)L_k(\mathbf{z}_k; \mathbf{x}_k)f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) d\mathbf{x}_k. \quad (22)$$

In this recursive form, there is no quotient of densities, since the information carried in the quotient from (19) is preserved in the BC from future times. The backward recursion is initialized at the final time step by setting $B_{T|T}(\mathbf{x}_T) = 1$, which reflects the fact that no future measurements are available beyond time t_T . The BC at earlier time instants is then obtained by propagating information backward in time.

Nonlinear Gaussian Mixture Smoother

While the linear²³ GM smoother admits a closed-form solution, no exact closed form solution is available in the nonlinear setting, requiring approximations.²⁴ The BC, generalized for nonlinear systems, takes the following form as a function of the state:

$$B_{k|T}^{(n)}(\mathbf{x}_k) = \frac{1}{r_k} \mathcal{N}(\boldsymbol{\eta}_k; \boldsymbol{\zeta}_k(\mathbf{x}_k), \mathbf{D}_k(\mathbf{x}_k)), \quad (23)$$

where r_k is a normalization constant, $\boldsymbol{\eta}_k$ is a stacked vector of the future measurements, $\boldsymbol{\zeta}_k(\mathbf{x}_k)$ is the corresponding stacked vector of predicted future measurements, and $\mathbf{D}_k(\mathbf{x}_k)$ is the covariance matrix associated with $\boldsymbol{\zeta}_k(\mathbf{x}_k)$. The predicted future measurement vector is constructed from the current state \mathbf{x}_k , the nonlinear solution flow $\boldsymbol{\varphi}(\cdot)$, and the measurement functions $\mathbf{h}_k(\cdot)$. These future measurement vectors are written explicitly as

$$\boldsymbol{\eta}_k = \begin{bmatrix} \mathbf{z}_T \\ \mathbf{z}_{T-1} \\ \vdots \\ \mathbf{z}_{k+1} \end{bmatrix}, \quad \boldsymbol{\zeta}_k(\mathbf{x}_k) = \begin{bmatrix} \mathbf{h}_T(\boldsymbol{\varphi}_k^T(\mathbf{x}_k)) \\ \mathbf{h}_{T-1}(\boldsymbol{\varphi}_k^{T-1}(\mathbf{x}_k)) \\ \vdots \\ \mathbf{h}_{k+1}(\boldsymbol{\varphi}_k^{k+1}(\mathbf{x}_k)) \end{bmatrix}, \quad (24)$$

where $\boldsymbol{\varphi}_k^i(\cdot)$ denotes the solution flow from t_k to t_i . By these definitions, the BC at time t_{k-1} is given by²⁴

$$B_{k-1|T}(\mathbf{x}_{k-1}) = \frac{1}{r_k \nu_k} \mathcal{N}(\boldsymbol{\eta}_{k-1}; \boldsymbol{\zeta}_{k-1}(\mathbf{x}_{k-1}), \mathbf{D}_{k-1}(\mathbf{x}_{k-1})). \quad (25)$$

The normalization factor ν_k is given by

$$\nu_k = \sum_{n=1}^{N_k} w_{k|k-1}^{(n)} \mathcal{N}\left(\mathbf{z}_k; \mathbf{H}_k|_{\mathbf{m}_{k|k-1}} \mathbf{m}_{k|k-1}^{(n)}, \mathbf{H}_k|_{\mathbf{m}_{k|k-1}} \mathbf{P}_{k|k-1}^{(n)} \mathbf{H}_k^\top|_{\mathbf{m}_{k|k-1}} + \mathbf{R}_k\right), \quad (26)$$

where $(\cdot)_{k|k-1}$ denotes quantities of the prior filter distribution $p(\mathbf{x}_k | \mathbf{z}_{1:k-1})$ and \mathbf{H}_k is the Jacobian of the measurement function $\mathbf{h}_k(\mathbf{x}_k)$. The predicted future measurement covariance $\mathbf{D}_{k-1}(\mathbf{x}_{k-1})$ is defined recursively as

$$\begin{aligned} \mathbf{D}_{i-1}(\mathbf{x}_{k-1}) &= \begin{bmatrix} \mathbf{D}_i(\mathbf{x}_{k-1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_i \end{bmatrix} \\ &+ \begin{bmatrix} \mathbf{C}_i(\boldsymbol{\varphi}_{k-1}^i(\mathbf{x}_{k-1})) \\ \mathbf{H}_i|_{\boldsymbol{\varphi}_{k-1}^i(\mathbf{x}_{k-1})} \end{bmatrix} \mathbf{Q}_{i-1}|_{\boldsymbol{\varphi}_{k-1}^{i-1}(\mathbf{x}_{k-1})} \begin{bmatrix} \mathbf{C}_i(\boldsymbol{\varphi}_{k-1}^i(\mathbf{x}_{k-1}))^\top & \mathbf{H}_i^\top|_{\boldsymbol{\varphi}_{k-1}^i(\mathbf{x}_{k-1})} \end{bmatrix}, \end{aligned} \quad (27a)$$

$$\mathbf{D}_{T-1}(\mathbf{x}_{k-1}) = \mathbf{R}_T + \mathbf{H}_T|_{\boldsymbol{\varphi}_{k-1}^T(\mathbf{x}_{k-1})} \mathbf{Q}_{T-1}|_{\boldsymbol{\varphi}_{k-1}^{T-1}(\mathbf{x}_{k-1})} \mathbf{H}_T^\top|_{\boldsymbol{\varphi}_{k-1}^T(\mathbf{x}_{k-1})}, \quad (27b)$$

with $i \in \{T-1, \dots, k\}$. The matrix $\mathbf{C}_i(\boldsymbol{\varphi}_{k-1}^i(\mathbf{x}_{k-1}))$ corresponds to the Jacobian of the predicted future measurement vector, evaluated at the state $\boldsymbol{\varphi}_{k-1}^i(\mathbf{x}_{k-1})$. The general expression at k is obtained by evaluating the Jacobian at \mathbf{x}_k and $i = k$:

$$\mathbf{C}_k(\mathbf{x}_k) = \frac{\partial \boldsymbol{\zeta}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k} = \begin{bmatrix} \mathbf{H}_T|_{\boldsymbol{\varphi}_k^T(\mathbf{x}_k)} \boldsymbol{\Phi}_k^T|_{\mathbf{x}_k} \\ \mathbf{H}_{T-1}|_{\boldsymbol{\varphi}_k^{T-1}(\mathbf{x}_k)} \boldsymbol{\Phi}_k^{T-1}|_{\mathbf{x}_k} \\ \vdots \\ \mathbf{H}_{k+1}|_{\boldsymbol{\varphi}_k^{k+1}(\mathbf{x}_k)} \boldsymbol{\Phi}_k^{k+1}|_{\mathbf{x}_k} \end{bmatrix}, \quad (28)$$

where $\boldsymbol{\Phi}_k^i|_{\mathbf{x}_k}$ denotes the STM from t_k to t_i , using \mathbf{x}_k as initial condition on the state. Consequently, $\mathbf{C}_k(\mathbf{x}_k)$ stacks the linearized measurement sensitivities from t_{k+1} to t_T , combining the effects of state propagation and measurement linearization:

$$\mathbf{C}_{k-1}(\mathbf{x}_{k-1}) = \begin{bmatrix} \mathbf{C}_k(\boldsymbol{\varphi}_{k-1}^k(\mathbf{x}_{k-1})) \\ \mathbf{H}_k|_{\boldsymbol{\varphi}_{k-1}^k(\mathbf{x}_{k-1})} \end{bmatrix} \boldsymbol{\Phi}_{k-1}^k|_{\mathbf{x}_{k-1}}. \quad (29)$$

With the BC available at each timestep, the smoothed GM estimate is

$$p(\mathbf{x}_k | \mathbf{z}_{1:T}) = \sum_{n=1}^{N_k} \frac{w_k^{(n)} q_k^{(n)}}{r_k} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_k^{*(n)}, \mathbf{P}_k^{*(n)}), \quad (30a)$$

$$q_k^{(n)} = \mathcal{N}(\boldsymbol{\eta}_k; \boldsymbol{\zeta}_k(\mathbf{m}_k^{(n)}), \mathbf{C}_k(\mathbf{m}_k^{(n)}) \mathbf{P}_k^{(n)} \mathbf{C}_k(\mathbf{m}_k^{(n)})^\top + \mathbf{D}_k(\mathbf{m}_k^{(n)})), \quad (30b)$$

$$\mathbf{K}_k^{(n)} = \mathbf{P}_k^{(n)} \mathbf{C}_k(\mathbf{m}_k^{(n)})^\top (\mathbf{C}_k(\mathbf{m}_k^{(n)}) \mathbf{P}_k^{(n)} \mathbf{C}_k(\mathbf{m}_k^{(n)})^\top + \mathbf{D}_k(\mathbf{m}_k^{(n)}))^{-1}, \quad (30c)$$

$$\mathbf{m}_k^{*(n)} = \mathbf{m}_k^{(n)} + \mathbf{K}_k^{(n)} (\boldsymbol{\eta}_k - \boldsymbol{\zeta}_k(\mathbf{m}_k^{(n)})), \quad (30d)$$

$$\mathbf{P}_k^{*(n)} = (\mathbf{I} - \mathbf{K}_k^{(n)} \mathbf{C}_k(\mathbf{m}_k^{(n)})) \mathbf{P}_k^{(n)} (\mathbf{I} - \mathbf{K}_k^{(n)} \mathbf{C}_k(\mathbf{m}_k^{(n)}))^\top + \mathbf{K}_k^{(n)} \mathbf{D}_k(\mathbf{m}_k^{(n)}) (\mathbf{K}_k^{(n)})^\top. \quad (30e)$$

Since the matrices $\mathbf{C}_k(\mathbf{x}_k)$ and $\mathbf{D}_k(\mathbf{x}_k)$ are nonlinear functions of the state \mathbf{x}_k , they explicitly depend on the chosen mixand mean and therefore cannot be directly reused from the smoothing at time t_{k+1} . As the smoothing recursion proceeds backward in time, the BC incorporates information from an increasing number of future time instants, leading to a growth in both dimensionality and computational burden.

One approach to mitigate the growth of the BC is to perform a local least squares (LS) approximation of the BC around each mixand of the filtering distribution.²⁴ This approximation introduces a reduced-dimension quantity, referred to as a *pseudo-measurement*, denoted by $\hat{\boldsymbol{\eta}}_k$. The pseudo-measurement is designed to encapsulate the information contained in the full stack of true measurements $\boldsymbol{\eta}_k$ in the exact BC, while drastically reducing its dimensionality:

$$B_{k|T}(\mathbf{x}_k) \approx \hat{B}_{k|T}^{(n)}\left(\hat{\boldsymbol{\eta}}_k; \hat{\boldsymbol{\zeta}}_k\left(\mathbf{m}_k^{(n)}\right), \hat{\mathbf{D}}_k^{(n)}\left(\mathbf{m}_k^{(n)}\right)\right), \quad (31)$$

where $\hat{\boldsymbol{\zeta}}_k\left(\mathbf{m}_k^{(n)}\right)$ is the predicted pseudo-measurement and $\hat{\mathbf{D}}_k^{(n)}$ is the associated covariance matrix. In principle, this pseudo-measurement can be defined through an arbitrary function, but the expressions for $\hat{\boldsymbol{\eta}}_k$, $\hat{\boldsymbol{\zeta}}_k^{(n)}$, $\hat{\mathbf{C}}_k^{(n)}$, and $\hat{\mathbf{D}}_k^{(n)}$ using the LS approximation are detailed in Lee and Campbell²⁴ and the approach is summarized below.

Equation (31) provides a local approximation of the function $B_{k|T}(\mathbf{x}_k)$ around the mixand means. Since the pseudo-measurement is common to all mixands, it cannot be inferred from any single mixand in isolation. Instead, it is estimated by combining the information provided by all mixands through a LS procedure. In practice, each mixand induces a local linearized relationship between the unknown pseudo-measurement and the corresponding predicted measurement. For a given mixand, this relationship can be expressed in terms of the residual

$$\Delta \mathbf{m}_k^{(n)} = \mathbf{m}_k^{*(n)} - \mathbf{m}_k^{(n)}, \quad (32)$$

defined as the difference between the smoothed mixand mean and its prior. The LS problem is constructed by stacking all mixand-wise residual equations into a single linear system, whose solution provides a compact pseudo-measurement that, when incorporated, approximates the true smoothing update resulting from the original, larger-dimensional stacked measurement vector.

A generic step of this smoothing algorithm proceeds as follows. First, the BCs from time t_k are each appended a measurement and propagated back to t_{k-1} according to (25). Then, these BCs are used to smooth the distribution at time t_k according to (30), resulting in $p(\mathbf{x}_k | \mathbf{z}_{1:T})$. From there, local LS approximations of each BC about their mixand means $\mathbf{m}_k^{(i)}$ are computed using a common pseudo-measurement $\hat{\boldsymbol{\eta}}_k$. Finally, these approximate BCs are then appended a measurement and propagated back to t_{k-2} and the recursion proceeds from there.

METHODOLOGY

The proposed generalized AGM smoother builds upon the BC framework developed for linear and nonlinear, non-adaptive GM smoothers.^{23,24} The main departure from these approaches is the introduction of a dimension-reduction step that replaces the existing least-squares formulation with a direct, projection-based compression. This strategy enables a scalable and numerically stable implementation for nonlinear systems, reducing computational cost while preserving the positive definiteness of the BC covariance matrices. To maintain consistency of the BC, characterized by an arbitrary number of mixands, across consecutive smoothing steps, a first-order weighted partial derivative tensor (PDT) is employed to shift the relevant parameters between expansion points. By construction, this formulation avoids the memory growth associated with the fully nonlinear approach in (23)–(30), thereby mitigating practical hardware limitations.

Crucially, both the projection-based compression and the association of BCs across time steps require the BC to admit a locally linear parametrization with respect to the state. For this reason,

each mixand-specific BC is approximated by a first-order expansion of the true nonlinear BC about the corresponding mixand mean. This linear representation makes the dependence of the BC on the state explicit and allows its parameters to be compressed, shifted, and matched across mixands without re-evaluating the full backward recursion. Accordingly, the BC as a function of \mathbf{x}_k , local to the i -th mixand mean $\mathbf{m}_k^{(i)}$, is written as

$$B_{k|T}(\mathbf{x}_k; \mathbf{m}_k^{(i)}) = \frac{1}{r_k^{(i)}(\mathbf{m}_k^{(i)})} \mathcal{N}\left(\boldsymbol{\eta}_k; \boldsymbol{\zeta}_k^{(i)}(\mathbf{m}_k^{(i)}) + \mathbf{C}_k^{(i)}(\mathbf{m}_k^{(i)})(\mathbf{x}_k - \mathbf{m}_k^{(i)}), \mathbf{D}_k^{(i)}(\mathbf{m}_k^{(i)})\right). \quad (33)$$

This local BC is the approximation of the true nonlinear BC used to arrive at the nonlinear smoothing equations in (30), and so preserving this local BC is sufficient to preserve the smoothing performance of the true nonlinear BC for its specific mixand. The ensuing sections outline a method to approximate $\{B_{k-1|T}(\mathbf{x}_{k-1}; \mathbf{m}_{k-1}^{(n)})\}_{n=1}^{N_{k-1}}$ using the priorly computed $\{B_{k|T}(\mathbf{x}_k; \mathbf{m}_k^{(i)})\}_{i=1}^{N_k}$ at a fixed computation and dimension, allowing smoothing to be performed at time t_{k-1} according to (30).

Projection-Based Backward Corrector Compression

Even in the most general formulation given in (22), the information captured by the BC grows unbounded as additional future measurements are incorporated. The LS approximation of the BC in (31) reduces the effective dimensionality by reformulating the future information in a lower-dimensional space. However, this approach requires solving an additional optimization problem at each iteration k , which adds computational overhead. A projection-based compression of the BC exploits the Gaussian structure of the BC parameters. In particular, the future information can be equivalently represented in a lower-dimensional space through a deterministic projection, without incurring any loss of information. By applying the proposed projection, the local BC in (33) can be expressed in compressed form as

$$\hat{B}_{k|T}^{(i)}(\mathbf{x}_k; \mathbf{m}_k^{(i)}) \approx \frac{1}{\hat{r}_k^{(i)}(\mathbf{m}_k^{(i)})} \mathcal{N}\left(\hat{\boldsymbol{\eta}}_k; \hat{\boldsymbol{\zeta}}_k^{(i)}(\mathbf{m}_k^{(i)}) + \hat{\mathbf{C}}_k^{(i)}(\mathbf{m}_k^{(i)})(\mathbf{x}_k - \mathbf{m}_k^{(i)}), \hat{\mathbf{D}}_k^{(i)}(\mathbf{m}_k^{(i)})\right), \quad (34)$$

where $\hat{(\cdot)}$ denotes quantities defined in the compressed space. Before deriving explicit expressions for the compressed parameters, we first introduce the mathematical framework underlying a lossless compression of Gaussian information. The compression strategy adopted relies on the equality stated in Proposition 1, proved in the Appendix. This result shows that a Gaussian density defined in the backward information space can be equivalently expressed in a lower-dimensional projected space through a deterministic linear transformation. As a consequence, local BCs can be represented using a reduced-dimensional set of parameters without any loss of information relevant to the state.

Proposition 1. *If $\mathbf{C} \in \mathbb{R}^{n_\eta \times n_x}$ has full column rank and $\mathbf{D} \in \mathbb{R}^{n_\eta \times n_\eta}$ is positive definite, then*

$$\mathcal{N}(\boldsymbol{\eta}; \boldsymbol{\zeta} + \mathbf{C}(\mathbf{x} - \mathbf{m}), \mathbf{D}) = \alpha \mathcal{N}\left(\mathbf{0}; \mathbf{T}(\boldsymbol{\zeta} - \boldsymbol{\eta}) + \mathbf{T}\mathbf{C}(\mathbf{x} - \mathbf{m}), \mathbf{T}\mathbf{D}\mathbf{T}^\top\right), \quad (35)$$

where

$$\mathbf{T} = \mathbf{C}^\top \mathbf{D}^{-1}, \quad (36)$$

$$\alpha = (2\pi)^{-(n_\eta - n_x)} \frac{|\mathbf{C}^\top \mathbf{D}^{-1} \mathbf{C}|^{1/2}}{|\mathbf{D}|^{1/2}} \exp\left[\frac{1}{2}(\boldsymbol{\zeta} - \boldsymbol{\eta})^\top \mathbf{D}^{-1}(\mathbf{M} - \mathbf{I})(\boldsymbol{\zeta} - \boldsymbol{\eta})\right], \quad (37)$$

$$\mathbf{M} = \mathbf{C} \left(\mathbf{C}^\top \mathbf{D}^{-1} \mathbf{C}\right)^{-1} \mathbf{C}^\top \mathbf{D}^{-1}. \quad (38)$$

Thus, by Proposition 1, the compressed parameters appearing in (34) are obtained analytically according to the following equations, where for notational brevity the functional dependence of these quantities on the mixand mean $\mathbf{m}_k^{(i)}$ is omitted throughout:

$$\mathbf{T}_k^{(i)} = \left(\mathbf{C}_k^{(i)}\right)^\top \left(\mathbf{D}_k^{(i)}\right)^{-1}, \mathbf{F}_k^{(i)} = \left(\mathbf{C}_k^{(i)}\right)^\top \left(\mathbf{D}_k^{(i)}\right)^{-1} \mathbf{C}_k^{(i)}, \mathbf{M}_k^{(i)} = \mathbf{C}_k^{(i)} \left(\mathbf{F}_k^{(i)}\right)^{-1} \mathbf{T}_k^{(i)}, \quad (39a)$$

$$\alpha_k^{(i)} = \frac{\left|\mathbf{F}_k^{(i)}\right|^{1/2}}{2\pi^{(n_\eta - n_x)} \left|\mathbf{D}_k^{(i)}\right|^{1/2}} \exp\left[\frac{1}{2} \left(\boldsymbol{\zeta}_k^{(i)} - \boldsymbol{\eta}_k\right)^\top \left(\mathbf{D}_k^{(i)}\right)^{-1} \left(\mathbf{M}_k^{(i)} - \mathbf{I}\right) \left(\boldsymbol{\zeta}_k^{(i)} - \boldsymbol{\eta}_k\right)\right], \quad (39b)$$

$$\hat{\boldsymbol{\eta}}_k = \mathbf{0}, \quad (39c)$$

$$\hat{\boldsymbol{\zeta}}_k^{(i)} = \mathbf{T}_k^{(i)} \left(\boldsymbol{\zeta}_k^{(i)} - \boldsymbol{\eta}_k\right), \quad (39d)$$

$$\hat{\mathbf{C}}_k^{(i)} = \mathbf{T}_k^{(i)} \mathbf{C}_k^{(i)} = \mathbf{F}_k^{(i)}, \quad (39e)$$

$$\hat{\mathbf{D}}_k^{(i)} = \mathbf{T}_k^{(i)} \mathbf{D}_k^{(i)} \left(\mathbf{T}_k^{(i)}\right)^\top = \mathbf{F}_k^{(i)}, \quad (39f)$$

$$\hat{r}_k^{(i)} = \frac{r_k}{\alpha_k^{(i)}}. \quad (39g)$$

Under the proposed compression, the transformed future measurement vector $\hat{\boldsymbol{\eta}}_k = \mathbf{0}$ for all mixands by construction. As a result, it remains common to all mixands and independent of the BC expansion point $\mathbf{m}_k^{(i)}$. However, because the normalization constant $\alpha_k^{(i)}$ has an explicit dependence on the expansion point through $\mathbf{C}_k^{(i)}$ and $\mathbf{D}_k^{(i)}$, the scaling factor $\hat{r}_k^{(i)}$ gains a dependence on the expansion point through this compression. Importantly, under this compression, a local BC still remains mixand-specific only through its expansion point. This property does not hold for the least-squares compression in (31). As a result, matching existing BCs to new mixands across timesteps is significantly simplified, since transferring a BC to a new mixand requires only an adjustment of its expansion point rather than a full recomputation of its parameters.

Parameter Expansion Point Adjustment in Backward Corrector

A key challenge in GM smoothing is the association of mixands across timesteps, as adaptive GM filters may split or merge components, precluding a direct one-to-one correspondence. Because there is no direct correspondence between mixands, no direct correspondence exists between the mixand-specific BCs at time t_k and the mixands at time t_{k-1} to which the BCs must be matched. These mixand relationships across timesteps must instead be inferred algorithmically. Since a local BC is mixand-specific only through its expansion point, transferring a BC between mixands across timesteps can be achieved simply by updating this point. In particular, transferring a BC from mixand i at time t_k to mixand n at time t_{k-1} involves adjusting the expansion point from $\mathbf{m}_k^{(i)}$ to $\boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})$. This expansion point adjustment can be achieved through the use of a first-order Taylor approximation of the BC parameters about the current mixand mean $\mathbf{m}_k^{(i)}$. This method circumvents the recursion of (27) and (29) that would be required to compute the BC at the new expansion point directly, and is much less computationally intensive as a result. Using this method,

the BC for mixand n at time t_{k-1} can be obtained from the BC for mixand i at time t_k as

$$B_{k|T}^{(n)}(\mathbf{x}_k; \boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})) \approx \check{B}_{k|T}^{(i,n)}(\mathbf{x}_k; \boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})), \quad (40)$$

$$= \frac{1}{\check{r}_k^{(i,n)}} \mathcal{N}\left(\check{\boldsymbol{\eta}}_k; \check{\boldsymbol{\zeta}}_k^{(i,n)} + \check{\mathbf{C}}_k^{(i,n)}(\mathbf{x}_k - \boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})), \check{\mathbf{D}}_k^{(i,n)}\right), \quad (41)$$

where $\check{(\cdot)}$ denotes shifted parameters and the functional dependence of the shifted BC parameters on the new expansion point $\boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})$ is omitted for brevity. Their expressions are reported below, with $\Delta \mathbf{m}_k^{(i,n)} = \left(\boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)}) - \mathbf{m}_k^{(i)}\right)$:

$$\begin{aligned} \check{\zeta}_k^{(i,n),[j]} \Big|_{\boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})} &= \zeta_k^{(i),[j]} \Big|_{m_k^{(i)}} + C_{k,[l]}^{(i),[j]} \Big|_{m_k^{(i)}} \Delta m_k^{(i,n),[l]} \\ &\quad + \frac{1}{2} \frac{\partial C_{k,[l]}^{(i),[j]}}{\partial x^{[o]}} \Big|_{m_k^{(i)}} \Delta m_k^{(i,n),[l]} \Delta m_k^{(i,n),[o]} \end{aligned} \quad (42a)$$

$$\check{C}_{k,[l]}^{(i,n),[j]} \Big|_{\boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})} = C_{k,[l]}^{(i),[j]} \Big|_{m_k^{(i)}} + \frac{\partial C_{k,[l]}^{(i),[j]}}{\partial x^{[o]}} \Big|_{m_k^{(i)}} \Delta m_k^{(i,n),[o]}, \quad (42b)$$

$$\check{D}_k^{(i,n),[jq]} \Big|_{\boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})} = D_k^{(i),[jq]} \Big|_{m_k^{(i),[p]}} + \frac{\partial D_k^{(i),[jq]}}{\partial x^{[l]}} \Big|_{m_k^{(i)}} \Delta m_k^{(i,n),[l]}, \quad (42c)$$

$$\check{r}_k^{(i,n)} \Big|_{\boldsymbol{\varphi}_{k-1}^k(\mathbf{m}_{k-1}^{(n)})} = r_k^{(i)} \Big|_{m_k^{(i)}}, \quad (42d)$$

where quantities indexed by square brackets $[\cdot]$ denote tensor elements, and repeated bracketed indices imply summation according to the Einstein convention. The quantities x and $\Delta m_k^{(i,n)}$ are first-order contravariant tensors of type $(1, 0)$, the covariance quantities $\check{D}_k^{(i,n),[\cdot]}$ and $D_k^{(i,n),[\cdot]}$ are second-order contravariant tensors of type $(2, 0)$, and the sensitivities $\check{C}_{k,[\cdot]}^{(i,n),[\cdot]}$ and $C_{k,[\cdot]}^{(i,n),[\cdot]}$ are mixed tensors of type $(1, 1)$. The remainder of this section concerns the computation of these PDTs for \mathbf{C}_k and \mathbf{D}_k .

Within the context of the algorithm, the parameter expansion point adjustment occurs following the compression step of the previous iteration. Under the projection-based BC compression employed in this work, the quantities $\hat{\mathbf{C}}_k$, $\hat{\mathbf{D}}_k$, and \mathbf{F}_k in (34) are element-wise identical, despite being expressed using different tensor representations. The quantity \mathbf{F}_k can be related recursively across timesteps using the recursive relations for \mathbf{C}_k and \mathbf{D}_k in (29) and (27a), resulting in the following simplified relationship:

$$\mathbf{F}_k(\mathbf{x}_k) = (\Phi_k^{k+1})^\top \Big|_{\mathbf{x}_k} \mathbf{J}_k(\mathbf{x}_k) \Phi_k^{k+1} \Big|_{\mathbf{x}_k}, \quad (43a)$$

$$\mathbf{J}_k(\mathbf{x}_k) = \mathbf{W}_k(\mathbf{x}_k) \mathbf{B}_k(\mathbf{x}_k)^{-1}, \quad (43b)$$

$$\mathbf{B}_k(\mathbf{x}_k) = \mathbf{I} + \mathbf{Q}_k \Big|_{\mathbf{x}_k} \mathbf{W}_k(\mathbf{x}_k), \quad (43c)$$

$$\mathbf{W}_k(\mathbf{x}_k) = \mathbf{F}_{k+1}(\mathbf{x}_k) + \mathbf{H}_{k+1}^\top \Big|_{\boldsymbol{\varphi}_k^{k+1}(\mathbf{x}_k)} \mathbf{R}_{k+1}^{-1} \mathbf{H}_{k+1} \Big|_{\boldsymbol{\varphi}_k^{k+1}(\mathbf{x}_k)}. \quad (43d)$$

After taking the derivatives of each term with respect to the state \mathbf{x}_k , the resulting PDTs are found by the following equations, where each quantity is evaluated at x_k unless otherwise notated and Ψ

denotes the state transition tensor (STT):

$$\begin{aligned} \frac{\partial F_{k,[jq]}}{\partial x^{[o]}} &= \Psi_{k,[jo]}^{k+1,[l]} J_{k,[lm]} \Phi_{k,[q]}^{k+1,[m]} + \Phi_{k,[j]}^{k+1,[l]} \frac{\partial J_{k,[lm]}}{\partial x_k^{[o]}} \Phi_{k,[q]}^{k+1,[m]} \\ &+ \Phi_{k,[j]}^{k+1,[l]} J_{k,[lm]} \Psi_{k,[qo]}^{k+1,[m]}, \end{aligned} \quad (44a)$$

$$\frac{\partial J_{k,[jq]}}{\partial x^{[o]}} = \frac{\partial W_{k,[jl]}}{\partial x^{[o]}} (B_k^{-1})_{[q]}^{[l]} + W_{k,[jl]} (B_k^{-1})_{[m]}^{[l]} \frac{\partial B_{k,[n]}^{[m]}}{\partial x^{[o]}} (B_k^{-1})_{[q]}^{[n]}, \quad (44b)$$

$$\frac{\partial B_{k,[q]}^{[j]}}{\partial x^{[o]}} = \frac{\partial Q_k^{[jl]}}{\partial x^{[o]}} W_{k,[lq]} + Q_k^{[jl]} \frac{\partial W_{k,[lq]}}{\partial x^{[o]}}, \quad (44c)$$

$$\begin{aligned} \frac{\partial W_{k,[jq]}}{\partial x^{[o]}} &= \frac{\partial F_{k+1,[jq]}}{\partial x^{[o]}} + \Phi_{k,[j]}^{k+1,[l]} \frac{\partial H_{k+1,[l]}^{[m]}}{\partial x^{[o]}} \Bigg|_{\varphi_k^{k+1}(x_k)} (R^{-1})_{[mn]} H_{k+1,[q]}^{[n]} \Bigg|_{\varphi_k^{k+1}(x_k)} \\ &+ H_{k+1,[j]}^{[l]} \Bigg|_{\varphi_k^{k+1}(x_k)} (R^{-1})_{[lm]} \frac{\partial H_{k+1,[n]}^{[m]}}{\partial x^{[o]}} \Bigg|_{\varphi_k^{k+1}(x_k)} \Phi_{k,[q]}^{k+1,[n]}. \end{aligned} \quad (44d)$$

The exact form of this PDT shifting requires \mathbf{F}_{k+1} and its PDT to be computed recursively until the last measurement timestep is reached. To avoid this recursion, they can instead be approximated by a weighted average of their evaluated values as given by

$$F_{k+1,[jq]}^{(i)} \Bigg|_{\varphi_k^{k+1}(m_k^{(i)})} \approx \sum_{l=1}^{N_{k+1}} \varepsilon_{k+1}^{(l,i)} \left[F_{k+1,[jq]}^{(l)} \Bigg|_{m_{k+1}^{(l)}} + \frac{\partial F_{k+1,[jq]}^{(l)}}{\partial x^{[o]}} \Bigg|_{m_{k+1}^{(l)}} \Delta m_{k+1}^{(l,i),[o]} \right], \quad (45a)$$

$$\frac{\partial F_{k+1,[jq]}^{(i)}}{\partial x^{[o]}} \Bigg|_{\varphi_k^{k+1}(m_k^{(i)})} \approx \sum_{l=1}^{N_{k+1}} \varepsilon_{k+1}^{(l,i)} \frac{\partial F_{k+1,[jq]}^{(l)}}{\partial x^{[o]}} \Bigg|_{m_{k+1}^{(l)}}, \quad (45b)$$

where $\varepsilon_k^{(i,n)}$ is a weight that associates mixand l at time t_{k+1} to mixand i at time t_k . This weight scores how close a propagated component mean of the distribution at t_{k+1} is to components of the distribution at t_k based on the Malanobis distance (MD), where closer components are given a larger weight. This weight is given by

$$\Delta \mathbf{m}_{k+1}^{(l,i)} = \varphi_k^{k+1}(\mathbf{m}_k^{(i)}) - \mathbf{m}_{k+1}^{(l)}, \quad (46a)$$

$$\mathbf{P}_{k+1|k}^{(i)} = \left(\Phi_k^{k+1} \Big|_{\mathbf{m}_k^{(i)}} \right) \mathbf{P}_k^{(i)} \left(\Phi_k^{k+1} \Big|_{\mathbf{m}_k^{(i)}} \right)^\top, \quad (46b)$$

$$\varepsilon_{k+1}^{(l,i)} = \frac{\left(\left(\Delta \mathbf{m}_{k+1}^{(l,i)} \right)^\top \left(\mathbf{P}_{k+1|k}^{(i)} \right)^{-1} \Delta \mathbf{m}_{k+1}^{(l,i)} \right)^{-1}}{\sum_{j=1}^{N_{k+1}} \left(\left(\Delta \mathbf{m}_{k+1}^{(j,i)} \right)^\top \left(\mathbf{P}_{k+1|k}^{(i)} \right)^{-1} \Delta \mathbf{m}_{k+1}^{(j,i)} \right)^{-1}}. \quad (46c)$$

With this, the PDTs for \mathbf{C}_k and \mathbf{D}_k can be computed for mixand i at time t_k and the BC can be shifted to mixand n at time t_{k-1} according to (42).

Backward Corrector Component Matching

Because there is no direct one-to-one correspondence between the mixands in the filtered distribution at times t_k and t_{k-1} , there are no mixand relationships to inform to which pairings of mixands between times t_k and t_{k-1} the expansion point adjustment should be performed. As such, the expansion point adjustment is performed between all pairs of mixands between times t_k and t_{k-1} , resulting in N_k BCs corresponding to each mixand n of the filtered distribution at time t_{k-1} . Because each of these BCs are expanded about the same mixand mean $\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})$ as a result of the expansion point adjustment, then they are equivalent up to the errors of this adjustment. To counteract these potential errors, a new BC is composed using a weighted average of their parameters. This is written explicitly as

$$B_{k|T}^{(n)}(\mathbf{x}_k; \varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})) \approx \overline{B}_{k|T}^{(n)}(\mathbf{x}_k; \varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})) \quad (47)$$

$$= \frac{1}{\overline{r}_k^{(n)}} \mathcal{N}\left(\boldsymbol{\eta}_k; \overline{\boldsymbol{\zeta}}_k^{(n)} + \overline{\mathbf{C}}_k^{(n)}(\mathbf{x}_k - \varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})), \overline{\mathbf{D}}_k^{(n)}\right), \quad (48)$$

where $\overline{(\cdot)}$ denotes averaged parameters and the functional dependence of the averaged BC parameters on $\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})$ is omitted for brevity. The averaged BC parameters are obtained using the component-matching weights from (46) as

$$\overline{\boldsymbol{\zeta}}_k^{(n)}(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})) = \sum_{i=1}^{N_k} \varepsilon_k^{(i,n)} \check{\boldsymbol{\zeta}}_k^{(i,n)}(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})), \quad (49a)$$

$$\overline{\mathbf{C}}_k^{(n)}(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})) = \sum_{i=1}^{N_{k+1}} \varepsilon_k^{(i,n)} \check{\mathbf{C}}_k^{(i,n)}(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})), \quad (49b)$$

$$\overline{\mathbf{D}}_k^{(n)}(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})) = \sum_{i=1}^{N_k} \varepsilon_k^{(i,n)} \check{\mathbf{D}}_k^{(i,n)}(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})), \quad (49c)$$

$$\overline{r}_k^{(n)}(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})) = \sum_{i=1}^{N_k} \varepsilon_k^{(i,n)} \check{r}_k^{(i,n)}(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})), \quad (49d)$$

This averaged BC is a local approximation of the true nonlinear BC at time t_k for a mixand n at time t_{k-1} . Equations (24)-(29) can then be applied to this averaged BC at time t_k to produce the associated BC at time t_{k-1} , written explicitly as

$$B_{k-1|T}^{(n)}(\mathbf{x}_{k-1}; \mathbf{m}_{k-1}^{(n)}) \approx \overline{B}_{k-1|T}^{(n)}(\mathbf{x}_{k-1}; \mathbf{m}_{k-1}^{(n)}) \quad (50)$$

$$= \frac{1}{\overline{r}_k^{(n)} \nu_k} \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\eta}_k \\ \mathbf{z}_k \end{bmatrix}; \begin{bmatrix} \overline{\boldsymbol{\zeta}}_k^{(n)} \\ \mathbf{h}_k(\varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)})) \end{bmatrix} + \overline{\mathbf{C}}_{k-1}^{(n)}(\mathbf{x}_{k-1} - \mathbf{m}_{k-1}^{(n)}), \overline{\mathbf{D}}_{k-1}^{(n)}\right), \quad (51)$$

where the functional dependence of the BC parameters on $\mathbf{m}_{k-1}^{(n)}$ is omitted for brevity. Once this BC at time t_{k-1} is computed for each mixand i at time t_{k-1} , then smoothing can be performed on the filtered distribution at time t_{k-1} according to (30). Finally, each of the BCs are compressed and the algorithm repeats to t_{k-2} .

Nonlinear Adaptive Gaussian Mixture Smoother

The generalized nonlinear AGM smoother is presented in Algorithm 1. Each cycle of this algorithm requires the filter distributions, the set of compressed approximations of the BC from the previous smoothing step, and the approximate PDT shifting recursion values from the previous smoothing step. It returns the smoothed distribution, the next set of compressed approximations of the BC for the smoothing that just occurred, and the approximate PDT shifting recursion values for the next smoothing step.

Algorithm 1 Generalized AGM Smoother Step, $k < T$

Require: Filter estimates $p(\mathbf{x}_k|\mathbf{z}_{1:k}), p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}),$ BCs $\{\hat{B}_{k|T}^{(i)}(\mathbf{x}_k; \mathbf{m}_k^{(i)})\}_{i=1}^{N_k}, \{\mathbf{F}_{k+1}^{(i)}(\mathbf{m}_k^{(i)})\}_{i=1}^{N_k},$
and $\{\nabla_{\mathbf{x}}\mathbf{F}_{k+1}^{(i)}(\mathbf{m}_k^{(i)})\}_{i=1}^{N_k}$
for $n = 1, \dots, N_{k-1}$ **do**
 $\{w_{k-1}^{(n)}, \mathbf{m}_{k-1}^{(n)}, \mathbf{P}_{k-1}^{(n)}\} \leftarrow p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$
 for $i = 1, \dots, N_k$ **do**
 Obtain $\nabla_{\mathbf{x}}\mathbf{F}_k^{(i)}(\mathbf{m}_k^{(i)}) \equiv \nabla_{\mathbf{x}}\hat{\mathbf{C}}_k^{(i)}(\mathbf{m}_k^{(i)}) \equiv \nabla_{\mathbf{x}}\hat{\mathbf{D}}_k^{(i)}(\mathbf{m}_k^{(i)})$ via (44)
 Obtain shifted $\check{B}_{k|T}^{(i,n)}(\mathbf{x}_k; \varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)}))$ via (41) and (42)
 Obtain $\varepsilon_k^{(i,n)}$ via (46)
 end for
 Obtain $\mathbf{F}_k^{(n)}(\mathbf{m}_{k-1}^{(n)})$ and $(\nabla_{\mathbf{x}}\mathbf{F}_k^{(n)})(\mathbf{m}_{k-1}^{(n)})$ via (45)
 Obtain $\bar{B}_{k|T}^{(n)}(\mathbf{x}_k; \varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)}))$ using $\{\varepsilon_k^{(i,n)}, \check{B}_{k|T}^{(i,n)}(\mathbf{x}_k; \varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)}))\}_{i=1}^{N_k}$ via (48) and (49)
 Obtain $\bar{B}_{k-1|T}^{(n)}(\mathbf{x}_{k-1}; \mathbf{m}_{k-1}^{(n)})$ from $\bar{B}_{k|T}^{(n)}(\mathbf{x}_k; \varphi_{k-1}^k(\mathbf{m}_{k-1}^{(n)}))$ via (51)
 Obtain $\{q_{k-1}^{(n)}, \mathbf{m}_{k-1}^{*(n)}, \mathbf{P}_{k-1}^{*(n)}\}_{i=1}^{N_k}$ via (30b)-(30e)
 end for
 Assemble $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:T})$ using $\{q_k^{(n)}, \mathbf{m}_{k-1}^{*(n)}, \mathbf{P}_{k-1}^{*(n)}\}_{n=1}^{N_{k-1}}$ via (30a) ▷ Smoothing complete
 for $n = 1, \dots, N_{k-1}$ **do**
 Obtain $\hat{B}_{k-1|T}^{(n)}(\mathbf{x}_{k-1}; \mathbf{m}_{k-1}^{(n)})$ via (34) and (39)
 end for ▷ BCs for next smoothing step computed
return $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:T}), \{\hat{B}_{k-1|T}^{(n)}(\mathbf{x}_{k-1}; \mathbf{m}_{k-1}^{(n)})\}_{n=1}^{N_{k-1}}, \{\mathbf{F}_k^{(n)}(\mathbf{m}_{k-1}^{(n)})\}_{n=1}^{N_{k-1}}, \{\nabla_{\mathbf{x}}\mathbf{F}_k^{(n)}(\mathbf{m}_{k-1}^{(n)})\}_{n=1}^{N_{k-1}}$

RESULTS

The proposed AGM smoother is assessed on a space object tracking problem involving a target in a southern Earth–Moon L_2 halo orbit. Its estimation performance is compared with that of the forward filter and with the uncompressed nonlinear AGM smoother²⁴ that the proposed approach approximates while maintaining constant computational complexity.

The state is represented by Cartesian position and velocity in the synodic frame as

$$\mathbf{x}_k = [\mathbf{r}_k^\top \quad \mathbf{v}_k^\top]^\top = [x_k \quad y_k \quad z_k \quad \dot{x}_k \quad \dot{y}_k \quad \dot{z}_k]^\top. \quad (52)$$

For modeling cislunar dynamics, the assumptions of the CR3BP are adopted. To ensure that position and velocity are of similar magnitudes, nondimensional length and time units are employed

according to the following conversions for the Earth-Moon system:

$$1[\text{LU}] = 379,632.95[\text{km}], \quad (53)$$

$$1[\text{TU}] = 368,232.68[\text{s}]. \quad (54)$$

The ratio of the mass of the Moon to the sum of the masses of the Earth and the Moon is assumed to be $\mu = 0.0122$. The dynamics of the target with respect to the synodic frame are then described by the following system of continuous-time differential equations:

$$\ddot{x} - 2\dot{y} = \frac{\partial U}{\partial x} \Big|_{\mathbf{r}}, \quad \ddot{y} + 2\dot{x} = \frac{\partial U}{\partial y} \Big|_{\mathbf{r}}, \quad \ddot{z} = \frac{\partial U}{\partial z} \Big|_{\mathbf{r}}, \quad (55)$$

$$U(\mathbf{r}) = \frac{1-\mu}{\mathbf{r}/\text{Earth}} + \frac{\mu}{\mathbf{r}/\text{Moon}} + \frac{x^2 + y^2}{2}, \quad (56)$$

where U is the pseudo-potential and \mathbf{r}/Earth and \mathbf{r}/Moon are the relative positions of the target with respect to the Earth and Moon respectively. The positions of the Earth and Moon are constant in the synodic frame and given by

$$\mathbf{r}_{\text{Earth}} = [-\mu \ 0 \ 0]^\top, \quad (57)$$

$$\mathbf{r}_{\text{Moon}} = [1-\mu \ 0 \ 0]^\top. \quad (58)$$

The discrete-time dynamics function $\varphi(\cdot)$ is given by the solution flow of the CR3BP equations of motion. The process noise power spectral density and process noise mapping matrix are given by

$$\mathbf{Q}_c = \begin{bmatrix} q_{w_x}^2 & 0 & 0 \\ 0 & q_{w_y}^2 & 0 \\ 0 & 0 & q_{w_z}^2 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \mathbf{0}_{3 \times 3} \\ \mathbf{I}_{3 \times 3} \end{bmatrix}, \quad (59)$$

where it is assumed that $q_{w_x} = q_{w_y} = q_{w_z} = 10^{-5}[\text{LU}/\text{TU}^{5/2}]$.

The measurements consist of azimuth and elevation angles taken by a space-based observer in a distant retrograde orbit (DRO). The observer has a body frame defined by

$$\mathcal{B} = (\mathbf{r}_{\text{obs}}, \hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \hat{\mathbf{b}}_3), \quad (60)$$

where \mathbf{r}_{obs} is the position of the observer in the synodic frame and $\hat{\mathbf{b}}_1$, $\hat{\mathbf{b}}_2$, and $\hat{\mathbf{b}}_3$ form an orthonormal basis. The observer's camera is assumed to be aligned with $\hat{\mathbf{b}}_1$. With this, the measurement function is given by

$$\mathbf{h}_k(\mathbf{x}) = \begin{bmatrix} \arctan\left(\frac{\hat{\mathbf{u}} \cdot \hat{\mathbf{b}}_1}{\hat{\mathbf{u}} \cdot \hat{\mathbf{b}}_2}\right) \\ \arcsin\left(\hat{\mathbf{u}} \cdot \hat{\mathbf{b}}_3\right) \end{bmatrix}, \quad \hat{\mathbf{u}} = \frac{\mathbf{r} - \mathbf{r}_{\text{obs}}}{\|\mathbf{r} - \mathbf{r}_{\text{obs}}\|}, \quad (61)$$

and the measurements are assumed to be corrupted by white Gaussian noise with a standard deviation of 10 [arcsec].

The initial positions and velocities of the target and observer are given in Table 1. The target's state begins with an initial uncertainty of 1[km] in each position state and 0.01[m/s] in each velocity

Table 1: Initial states for the target and observer.

State	Value
\mathbf{r}_0	$[1.0193276 \ 0 \ -0.1801721]^\top$ [LU]
\mathbf{v}_0	$[0 \ -0.09731243 \ 0]^\top$ [LU/TU]
$\mathbf{r}_{\text{obs},0}$	$[1.0193276 \ 0 \ -0.1801721]^\top$ [LU]
$\mathbf{v}_{\text{obs},0}$	$[0 \ -0.09731243 \ 0]^\top$ [LU/TU]

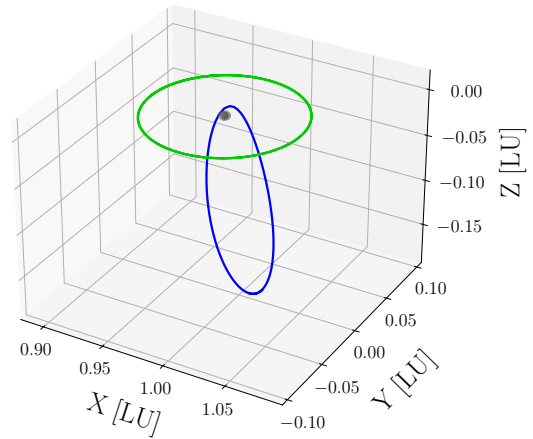


Figure 1: Orbits for cislunar scenario, where the Moon is gray, the target’s southern L_2 halo orbit is blue, and the observer’s DRO is green.

state. The target is tracked for 14 days, during which the observer takes a single measurement every six hours. This constant measurement cadence is a simplifying assumption, as the true visibility conditions for cislunar space would restrict measurement opportunities.^{30,31} The orbits of the target and observer are shown in Figure 1, where the target is in blue and the observer is in green.

Prior to smoothing, the filtered state estimate is obtained using an adaptive Gaussian mixture-based extended Kalman filter.^{30,32} Nonlinearity and uncertainty-informed moment-matching Gaussian splitting is applied before both the prediction and update steps, using the uncertainty-scaled first-order stretching (US-FOS) and whitened uncertainty-scaled second-order linearization change (W-US-SOLC) splitting metrics, respectively.^{33,34} Following the Bayes’ update, the posterior densities are reduced using Runnall’s algorithm.³⁵ The reduced mixture size is chosen randomly and arbitrarily as three, four, or five to demonstrate the AGM smoother’s ability to handle smoothing between mixtures of differing sizes.

Figure 2 shows the evolution of the y - z marginal distributions over a 14-day interval for the filtered, uncompressed smoothed, and compressed smoothed estimates, identified by white, yellow, and blue text boxes, respectively. The true state is denoted by the red and white diamond. Given that the target’s orbital period is 6.5 days and that the initial condition corresponds to apolune, the epochs $t - t_0 = 3.5$ [day] and $t - t_0 = 10.5$ [day] correspond to times shortly after perilune, while $t - t_0 = 7$ [day] corresponds to a time shortly after apolune. The compressed smoothed estimate exhibits a substantially reduced uncertainty relative to the filtered solution, and shows strong agreement with the uncompressed smoothed estimate over the entire time horizon considered. The marginals at $t - t_0 = 3.5$ [day] and $t - t_0 = 10.5$ [day] show some variation between the compressed and uncompressed smoothed estimates, which is likely due to the first-order expansion point adjustment being less accurate near perilune.

In order to more clearly see this uncertainty improvement compared to the filtered estimate, Fig. 3 shows the ratios of the marginal standard deviations of the uncertainty between the filtered and smoothed estimates for each state variable. These ratios for the filtered and uncompressed smoothed estimate are in yellow, while the ratios for the filtered and compressed smoothed estimates are in

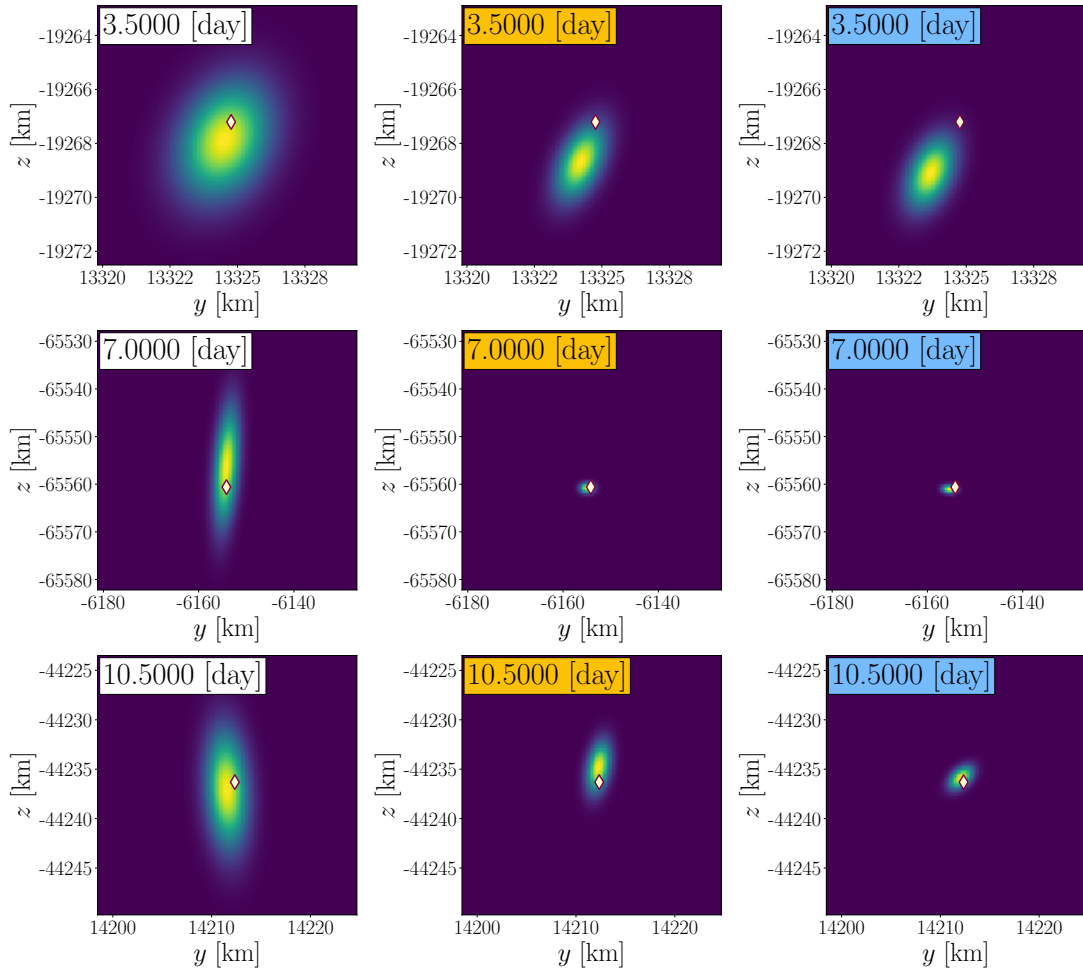


Figure 2: Time evolution of y - z marginal densities for filtered, uncompressed smoothed, and compressed smoothed estimates, denoted by the white, yellow, and blue shading of the text boxes respectively. The truth is shown by the red and white diamonds.

blue. There is a very strong agreement in the uncertainty reduction between the two smoothers, which indicates that the compressed smoother is approximating the uncompressed smoother well. The z and \dot{z} states show both the highest uncertainty reduction and the weakest agreement between the smoothers. This is due to more future measurements with more accumulated expansion point adjustment error retaining more of their value for these states due to the vertical nature of the target's orbit, which amplifies the approximation error of the compressed smoother. However, the compressed smoothed estimate still remains consistent under these conditions, indicating that the increased uncertainty reduction is not entirely unfounded.

CONCLUSION

This paper presents a novel adaptive Gaussian mixture (AGM) smoothing algorithm for nonlinear systems. The algorithm utilizes a novel projection-based approximation of a nonlinear corrector function alongside expansion point shifting via higher-order Taylor series expansions to approximate current state-of-the-art smoothing algorithms with escalating computational complexities at a

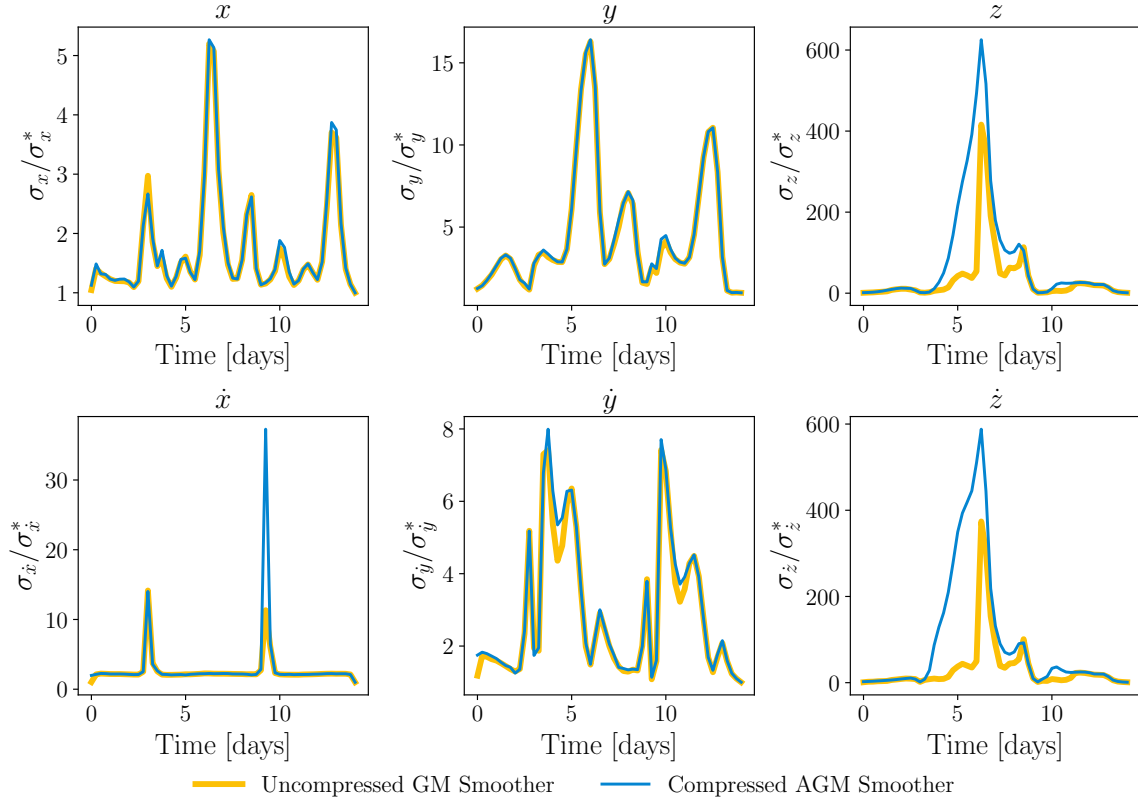


Figure 3: Uncertainty improvement over time of smoothed estimates compared to filtered estimate, where the uncompressed smoothed estimate is in yellow and the compressed smoothed estimate is in blue.

constant computational complexity. When applied to tracking a cislunar space object in an southern L_2 halo orbit, the AGM smoother exhibited massively improved estimation accuracy and uncertainty representation compared to the filtered distribution, and performed nearly identically to the computationally intensive smoothing algorithms being approximated.

ACKNOWLEDGMENT

Part of this research was sponsored by the United States Air Force Research Laboratory and the United States AFRL Regional Hub and was accomplished under Cooperative Agreement Number FA8750-22-2-0501. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

APPENDIX: PROOF OF PROPOSITION 1

Proof. Let $\xi = (\zeta - \eta) + \mathbf{C}(\mathbf{x} - \mathbf{m})$ such that $\mathbf{T}(\zeta - \eta) + \mathbf{TC}(\mathbf{x} - \mathbf{m}) = \mathbf{T}\xi$. Then:

$$(\mathbf{T}\xi)^\top (\mathbf{TDT}^\top) (\mathbf{T}\xi) = \xi^\top \mathbf{D}^{-1} \mathbf{M} \xi = \xi^\top \mathbf{D}^{-1} \xi + \xi^\top \mathbf{D}^{-1} (\mathbf{M} - \mathbf{I}) \xi$$

Since \mathbf{C} is full column rank and \mathbf{D} is invertible, then $\mathbf{C}^\top \mathbf{D}^{-1} \mathbf{C}$ is invertible. Thus, $(\mathbf{M} - \mathbf{I})\mathbf{C}(\mathbf{x} - \mathbf{m}) = \mathbf{0}$, and:

$$\boldsymbol{\xi}^\top \mathbf{D}^{-1} (\mathbf{M} - \mathbf{I}) \boldsymbol{\xi} = (\boldsymbol{\zeta} - \boldsymbol{\eta})^\top \mathbf{D}^{-1} (\mathbf{M} - \mathbf{I}) (\boldsymbol{\zeta} - \boldsymbol{\eta})$$

With these results, we get:

$$\begin{aligned} \mathcal{N}(\mathbf{0}; \mathbf{T}\boldsymbol{\xi}, \mathbf{T}\mathbf{D}\mathbf{T}^\top) &= \frac{1}{(2\pi)^{n_\eta/2} |\mathbf{T}\mathbf{D}\mathbf{T}^\top|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{T}\boldsymbol{\xi})^\top (\mathbf{T}\mathbf{D}^{-1}\mathbf{T}^\top) (\mathbf{T}\boldsymbol{\xi})\right) \\ &= \frac{1}{(2\pi)^{n_\eta/2} |\mathbf{C}^\top \mathbf{D}^{-1} \mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\xi}^\top \mathbf{D}^{-1} \boldsymbol{\xi} - \frac{1}{2} (\boldsymbol{\zeta} - \boldsymbol{\eta})^\top \mathbf{D}^{-1} (\mathbf{M} - \mathbf{I}) (\boldsymbol{\zeta} - \boldsymbol{\eta})\right) \\ &= \frac{(2\pi)^{n_x/2} |\mathbf{D}|^{1/2}}{(2\pi)^{n_\eta/2} |\mathbf{C}^\top \mathbf{D}^{-1} \mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{\zeta} - \boldsymbol{\eta})^\top \mathbf{D}^{-1} (\mathbf{M} - \mathbf{I}) (\boldsymbol{\zeta} - \boldsymbol{\eta})\right) \\ &\quad \times \frac{1}{(2\pi)^{n_x/2} |\mathbf{D}|^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\xi}^\top \mathbf{D}^{-1} \boldsymbol{\xi}\right) \\ &= \frac{1}{\alpha} \mathcal{N}(\boldsymbol{\eta}; \boldsymbol{\zeta} + \mathbf{C}(\mathbf{x} - \mathbf{m}), \mathbf{D}) \end{aligned}$$

which is equivalent to the desired equality. \square

REFERENCES

- [1] G. Pasricha, "Kalman Filter and its Economic Applications," *Munich Personal RePEc Archive, Ludwig Maximilian University of Munich*, 01 2006.
- [2] E. Ghysels and M. Marcellino, *Applied Economic Forecasting Using Time Series Methods*. Oxford University Press, 2018.
- [3] P. Zarchan and H. Musoff, *Fundamentals of Kalman Filtering: A Practical Approach*. Progress in Astronautics and Aeronautics, American Institute of Aeronautics and Astronautics, 2000.
- [4] M. A. Azzam, U. Batool, and H. Fauzi, "Design of an Helical Spring using Single-solution Simulated Kalman Filter Optimizer," *Mekatronika*, Vol. 1, July 2019, pp. 93–97, 10.15282/mekatronika.v1i2.4990.
- [5] S. Kazemi, N. L. Azad, K. A. Scott, H. B. Oqab, and G. B. Dietrich, "Orbit determination for space situational awareness: A survey," *Acta Astronautica*, Vol. 222, 2024, pp. 272–295, <https://doi.org/10.1016/j.actaastro.2024.06.015>.
- [6] S. Fedeler, M. Holzinger, and W. Whitacre, "Sensor tasking in the cislunar regime using Monte Carlo Tree Search," *Advances in Space Research*, Vol. 70, No. 3, 2022, pp. 792–811, <https://doi.org/10.1016/j.asr.2022.05.003>.
- [7] D. Lin and K. Oguri, "Cislunar Maneuvering Low-thrust Spacecraft Tracking with Adaptive Optimal Control Estimation," 08 2025.
- [8] F. Giraldo-Grueso, A. A. Popov, U. D. Hanebeck, and R. Zanetti, "Optimal Sampling for Point Mass Filtering with Applications to Cislunar Orbit Determination," *Journal of the Astronautical Sciences*, Vol. 72, 2025, p. 44, 10.1007/s40295-025-00521-7.
- [9] G. Siciliano, K. LeGrand, and J. Kulik, "Deferred Higher-Order Splitting for Adaptive Gaussian Mixture Orbit Uncertainty Propagation," 07 2025, 10.23919/FUSION65864.2025.11123926.
- [10] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Vol. 82, Mar. 1960, pp. 35–45, 10.1115/1.3662552.
- [11] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering*, Vol. 83, Mar. 1961, pp. 95–108, 10.1115/1.3658902.
- [12] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," *Signal Processing, Sensor Fusion, and Target Recognition VI* (I. Kadar, ed.), Vol. 3068, International Society for Optics and Photonics, SPIE, 1997, pp. 182 – 193, 10.1117/12.280797.
- [13] E. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, 10 2000, pp. 153–158, 10.1109/ASSPCC.2000.882463.
- [14] I. Arasaratnam and S. Haykin, "Cubature Kalman Filters," *IEEE Transactions on Automatic Control*, Vol. 54, 6 2009, pp. 1254–1269, 10.1109/TAC.2009.2019800.

- [15] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, Vol. 140, 4 1993, pp. 107–113, 10.1049/ip-f-2.1993.0015.
- [16] A. Doucet and A. Johansen, "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," *Handbook of Nonlinear Filtering*, Vol. 12, 01 2008.
- [17] F. Daum and J. Huang, "Particle degeneracy: root cause and solution," *Signal Processing, Sensor Fusion, and Target Recognition XX*, Vol. 8050, SPIE, 5 2011, pp. 367–377, 10.1117/12.877167.
- [18] F. Daum and J. Huang, "Curse of dimensionality and particle filters," *2003 IEEE Aerospace Conference Proceedings (Cat. No.03TH8652)*, Vol. 4, 3 2003, pp. 4_1979–4_1993, 10.1109/AERO.2003.1235126.
- [19] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, Vol. 17, 8 1972, pp. 439–448, 10.1109/TAC.1972.1100034.
- [20] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica (Oxf.)*, Vol. 7, July 1971, pp. 465–479.
- [21] U. D. Hanebeck, K. Briechle, and A. Rauh, "Progressive Bayes: a new framework for nonlinear state estimation," *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2003* (B. V. Dasarathy, ed.), SPIE, Apr. 2003.
- [22] G. Kitagawa, "The Two-Filter Formula for Smoothing and an Implementation of the Gaussian-Sum Smoother," *Annals of the Institute of Statistical Mathematics*, 1994.
- [23] B.-N. Vo, B.-T. Vo, and R. P. S. Mahler, "Closed Form Solutions to Forward-Backward Smoothing," *IEEE Transactions on Signal Processing*, Vol. 60, No. 1, 2012, pp. 2–17.
- [24] D. Lee and M. Campbell, "Smoothing Algorithm for Nonlinear Systems Using Gaussian Mixture Models," *Journal of Guidance, Control, and Dynamics*, Vol. 38, Aug. 2015, pp. 1438–1451, 10.2514/1.G000603.
- [25] C. Kuehn, *Moment Closure—A Brief Review*, pp. 253–271. Cham: Springer International Publishing, 2016, 10.1007/978-3-319-28028-8.
- [26] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*. New York: Chapman and Hall/CRC, 2 ed., Oct. 2011, 10.1201/b11154.
- [27] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Analytical Mechanics Associates, Inc., 1970.
- [28] B. Anderson and J. Moore, *Optimal Filtering*. Dover Books on Electrical Engineering, Dover Publications, 2012.
- [29] S. Särkkä and L. Svensson, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2nd ed., 2023.
- [30] J. Iannamorelli and K. LeGrand, "Adaptive Gaussian Mixture Filtering for Multi-sensor Maneuvering Cislunar Space Object Tracking," *The Journal of the Astronautical Sciences*, Vol. 72, 01 2025, 10.1007/s40295-024-00478-z.
- [31] M. Klonowski, N. Owens-Fahrner, C. Heidrich, and M. Holzinger, "Cislunar Space Domain Awareness Architecture Design and Analysis for Cooperative Agents," *The Journal of the Astronautical Sciences*, 2024.
- [32] K. LeGrand, A. Khilnani, and J. Iannamorelli, "Bayesian Angles-Only Cislunar Space Object Tracking," *Proceedings of the 33rd AAS/AIAA Space Flight Mechanics Meeting*, Jan. 2023.
- [33] J. Kulik and K. LeGrand, "Nonlinearity and Uncertainty Informed Moment-Matching Gaussian Mixture Splitting," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. PP, 01 2025, pp. 1–21, 10.1109/TAES.2025.3632242.
- [34] G. A. Siciliano, K. A. LeGrand, and J. Kulik, "Higher-Order Tensor-Based Deferral of Gaussian Splitting for Orbit Uncertainty Propagation," July 2025, 10.48550/arXiv.2507.01771.
- [35] A. Runnalls, "Kullback-Leibler Approach to Gaussian Mixture Reduction," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 43, 7 2007, pp. 989–999.