AAS 15-575

GAUSSIAN MIXTURE APPROXIMATION OF THE BEARINGS-ONLY INITIAL ORBIT DETERMINATION LIKELIHOOD FUNCTION

Mark L. Psiaki*, Ryan M. Weisman†, and Moriba K. Jah‡

A method is developed to approximate the bearings-only orbit determination likelihood function using a Gaussian mixture to incorporate information about an admissible region. The resulting probability density function can provide the *a priori* information for a Gaussian mixture orbit determination filter. The new technique starts with a nonlinear batch least-squares solution. The solution enforces soft constraints on an admissible region defined in terms of minimum periapsis and maximum apoapsis. This admissible region information can compensate for poor observability from a short arc of bearings-only data. Although this soft-constrained solution lies in or near the admissible region, it does not characterize that region well. It provides a starting point to develop a Gaussian mixture approximation of the batch least-squares likelihood function as modified through multiplication by a finite-support function that is zero outside the admissible region and equal to one in that region. This Gaussian mixture is optimized to fit the resulting probability density in the 2-dimensional subspace of position/velocity space that has the most uncertainty. This optimal fitting allows the Gaussian mixture to use a low number of mixands while fitting the finite-support probability density function well. By approximating the product of a finite-support function and the original likelihood function, the new method gains the capability to transition smoothly between regimes where the admissibility constraints dominate, i.e., high-altitude/short-measurement-arc cases, and those where they are irrelevant, i.e., low-altitude/long-measurement-arc cases.

INTRODUCTION

Orbit Determination (OD) based on bearings measurements has been studied since the time of Brahe and Kepler, and several of the greats of science have contributed methods.[1,2] If only a few data points are available from a single observing station over a brief time window, then the resulting orbital uncertainties can be very large,[3] sometime so large that hyperbolic orbits lie within the realm of possibility. The largest uncertainties tend to be in the range and range rate.

A method of dealing with this problem is to add information about the orbit's admissible regions.[4,5] References [2] and [3] use this technique with an admissible region that is defined by upper bounds on the orbital eccentricity and on the semi-major axis, or equivalently, on the orbital specific energy. Reference [2] uses this admissible region in order to constrain its orbital hypotheses. Reference [3] develops a Gaussian mixture that approximates a flat distribution which covers the projection of the admissible region onto the range/range-rate subspace.

*Professor, Sibley School of Mechanical & Aerospace Engineering, Cornell University, 224 Upson Hall, Ithaca, NY 14853-7501.

†Research Aerospace Engineer, Guidance, Navigation, and Controls Group, Air Force Research Lab Space Vehicles Directorate, 3550 Aberdeen Ave. SE, Albuquerque, NM 87117.

‡Technical Advisor, Guidance, Navigation, and Controls Group, Air Force Research Lab Space Vehicles Directorate, 3550 Aberdeen Ave. SE, Albuquerque, NM 87117.

Preprint, 2015 AAS/AIAA Astrodynamics Specialist Conf., 9-13 August 2015, Vail, CO.

A useful capability is that of determining a reasonable Gaussian mixture Initial Orbit Determination (IOD) state probability density function. Such a function can be used to initialize a Gaussian mixture OD filter. A Gaussian mixture filter can provide a powerful means to closely approximate the true Bayesian solution to a nonlinear filtering problem if the individual mixands are narrow enough and if there are enough of them.[6,7,8] Sufficiently narrow mixands allow the Bayesian calculations to be implemented using Taylor series approximations that reduce to simple EKF or UKF calculations on a per-mixand basis. These calculations are followed by a re-weighting among the mixands after the measurement update.[6,8] Recent advances have been made on the key lagging technology for Gaussian mixture filtering, a mixture re-sampling algorithm that re-approximates a distribution accurately while limiting its mixand covariances.[6,7] These advances provide the motivation for seeking OD filter initialization strategies that are consistent with Gaussian mixture filtering.

This paper develops an advanced version of the admissible-region-based Gaussian mixture IOD method of Ref. [3]. It makes five new contributions. First, it uses different criteria to define an admissible region. The new constraints are a minimum orbital periapsis and a maximum orbital apoapsis. These constraints are developed using Keplerian dynamics, and they remain non-singular at zero orbital eccentricity. The second contribution is a nonlinear batch least squares solver for an initial rough approximation of the position/velocity state. This solver includes admissible region "soft" constraints that are implemented by adding penalty functions to the least-squares cost function. This paper's third contribution is to generalize the definition of the subspace in which the IOD uncertainty is characterized primarily by the admissible region. Instead of using range and range rate, this subspace is defined as the 2-dimensional sub-space of position/velocity state space that has the most uncertainty after processing the bearings-only data in a batch least-squares filter. This subspace will lie primarily in the range/range-rate space for the short-data-arc/single-observing-station case, despite the lack of an explicit constraint that it correspond exactly to range and range rate. The fourth contribution is to characterize the IOD probability density in the high-uncertainty 2-D subspace by a combination of admissible region constraints and a Gaussian approximation of the bearings-only likelihood function. This feature enables the method to operate sensibly in high-orbit/short-arc cases and in low-orbit/long-arc cases. The former cases rely heavily on the admissible region constraints, while the latter cases achieve good IOD primarily from the bearings measurements. The new method is agnostic to this issue and transitions smoothly between these extremes. The fifth contribution is a new way of forming the Gaussian mixture approximation in the high-uncertainty 2-D subspace. The method of Ref. [3] uses a product of 1-D Gaussian mixtures that have been partially optimized to fit a flat distribution – only their standard deviations have been optimized. This paper's method directly fits a 2-D mixture to the underlying admissible-region/likelihood-function distribution, and it optimizes the weights, mean values, and covariances of all its 2-D mixands. This approach can reduce the number of mixands needed to obtain a given approximation accuracy.

This paper develops and explains its new IOD methods in six main sections followed by a section that summarizes and gives conclusions. Counting from this introductory first section, the second section presents the periapsis and apoapsis constraints for the IOD admissible region. The third section develops the nonlinear batch least-squares problem with penalty function constraints, and it outlines the solution algorithm. The fourth section defines the new 2D subspace that generalizes the range/range-rate subspace in which the admissibility constraints provide information. The fifth section defines the theoretical IOD probability density function as a product of two terms. One is a Gaussian approximation of the bearings-only likelihood function projected onto the 4-

dimensional compliment of the most uncertain subspace. The other term is itself the product of two terms: the same likelihood approximation projected onto the 2-D high uncertainty subspace and a finite-support function on the admissible region's projection onto that subspace. The sixth section develops a Gaussian mixture approximation of the IOD probability density function in the 2D maximal uncertainty subspace. It also defines an optimization problem to fit of this Gaussian mixture to the theoretical probability density function, and it presents an algorithm for solving this problem. The seventh section assembles the results of the preceding sections to form the Gaussian mixture approximation of the IOD probability density function in the full 6-dimensional position/velocity OD state space. The final section gives a summary of this paper's contributions along with its conclusions.

ADMISSIBLE REGION BASED ON ORBITAL PERIAPSIS AND APOAPSIS

The IOD admissible region is defined using a minimum periapsis body-centric radius and a maximum apoapsis body-centric radius. This definition represents a departure from the maximum eccentricity and maximum semi-major axis constraints of Refs. [2] and [3]. The justification for this departure is that maximum constraints on eccentricity and semi-major axis can allow orbits that impact the attracting body. If such orbits are allowed by the resulting IOD calculations, then an extra *ad hoc* method must be employed to rule out the offending orbits. No extra *ad hoc* discrimination is required if a minimum periapsis radius is enforced as part of the admissible region definition, provided that it is sufficiently larger than the radius of the attracting central body. Thus, the new admissible regions constraints seem more natural.

The new constraints pose a challenge. A square root operation is required in order to calculate the Keplerian orbital eccentricity as a function of inertial position and velocity. Therefore, it's first derivatives with respect to position and velocity have a singularity at zero eccentricity. The enforcement of a maximum eccentricity constraint can skirt this difficulty by enforcing a maximum constraint on the square of the eccentricity. The periapsis and apoapsis formulas, on the other hand, cannot directly use the square of the orbital eccentricity. Thus, simple constraints on periapsis and apoapsis could lead to numerical difficulties when using them in an algorithm such as a batch least-squares solver.

It is possible to transform a minimum constraint on the periapsis and a maximum constraint on the apoapsis into nonsingular forms. The following two constraints do exactly that:

$$0 \geq \left(\frac{\mathcal{E}}{\mu}\right)\rho_{min}^2 + \rho_{min} - \frac{\|\mathbf{h}\|^2}{2\mu} \tag{1}$$

$$0 \geq \left(\frac{\mathcal{E}}{\mu}\right)\rho_{max}^2 + \rho_{max} - \frac{\|\mathbf{h}\|^2}{2\mu} \tag{2}$$

where ρ_{min} is the minimum body-centric radius of the periapsis, ρ_{max} is the maximum body-centric radius of the apoapsis, μ is the attracting body's central gravitational constant, $\mathcal{E}$ is the orbital energy per unit mass, and $\|\mathbf{h}\|^2$ is the square of the magnitude of the orbital angular momentum per unit mass. The first of these equations of the peripasis constraint, and the second is the apoapsis constraint. The specific energy and the square of the specific angular momentum are related to the

Cartesian position vector $\mathbf{r}$ and velocity vector $\mathbf{v}$ as follows:

$$\begin{aligned} \mathcal{E} &= \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} - \frac{\mu}{\sqrt{\mathbf{r}^{\mathrm{T}}\mathbf{r}}} = -\frac{\mu}{2a} \\ \|\mathbf{h}\|^2 &= \|\mathbf{r}\times\mathbf{v}\|^2 \\ &= (\mathbf{v}^{\mathrm{T}}\mathbf{v})(\mathbf{r}^{\mathrm{T}}\mathbf{r}) - (\mathbf{v}^{\mathrm{T}}\mathbf{r})^2 \end{aligned} \tag{3}$$

where a is the orbital semi-major axis. The only possible singularity in these expressions is $\sqrt{\mathbf{r}^{\mathrm{T}}\mathbf{r}}$ at $\mathbf{r} = 0$, a situation unlikely to be encountered in practice.

The constraints in Eqs. (1) and (2) are ambiguous. It is possible to satisfy them both if the entire orbit lies above the radius ρ_{max} or below ρ_{min}. Note that removal of the "min" and "max" subscripts from ρ in these two equations results in identical terms on their right-hand sides, terms that are quadratic in ρ. The set of possible orbital radii are those for which this quadratic function is non-negative. The ρ values that cause this function to equal zero are the periapsis and the apoapsis. The desired condition is that the non-negative peak value of this quadratic function and all non-negative values lie in the range $\rho_{min} \leq \rho \leq \rho_{max}$. The anomalous situations have all the non-negative values lying either in the range $\rho_{max} \leq \rho$ or in the range $\rho \leq \rho_{min}$.

The ambiguity in Eqs. (1) and (2) is removed by adding two more constraints: that the orbital semi-major axis a also lie between ρ_{min} and ρ_{max}. These constraints can be written as constraints on the orbital specific energy:

$$0 \geq -\frac{\mu}{2\rho_{min}} - \mathcal{E} \tag{4}$$

$$0 \geq \frac{\mu}{2\rho_{max}} + \mathcal{E} \tag{5}$$

These latter constraints are normally satisfied as strict inequalities. They are added in order to preclude this paper's batch least squares algorithm from allowing an orbit that violates one of the constraints in Eqs. (1) and (2) over its entire period.

It is helpful to re-write the 4 inequality constraints in Eqs. (1), (2), (4), and (5) in the following normalized vector inequality form:

$$0 \geq \mathbf{c}(\begin{bmatrix}\mathbf{r}\\ \mathbf{v}\end{bmatrix}) = \mathbf{c}(\mathbf{x}) = \begin{bmatrix} c_1(\mathbf{x}) \\ c_2(\mathbf{x}) \\ c_3(\mathbf{x}) \\ c_4(\mathbf{x}) \end{bmatrix} \tag{6}$$

where $\mathbf{x} = [\mathbf{r}^{\mathrm{T}}, \mathbf{v}^{\mathrm{T}}]^{\mathrm{T}}$ is the 6-dimensional orbital position/velocity state vector and where the 4

constraint function components are

$$
\begin{aligned}
c_1(\begin{bmatrix}\mathbf{r}\\ \mathbf{v}\end{bmatrix}) &= \frac{\left[\frac{\mathcal{E}(\mathbf{r},\mathbf{v})}{\mu}\right]\rho_{min}^2 + \rho_{min} - \frac{\|\mathbf{h}(\mathbf{r},\mathbf{v})\|^2}{2\mu}}{\frac{1}{2}\left[1 - \left(\frac{\rho_{min}}{\rho_{max}}\right)\right]\sigma_\rho} \\
c_2(\begin{bmatrix}\mathbf{r}\\ \mathbf{v}\end{bmatrix}) &= \frac{\left[\frac{\mathcal{E}(\mathbf{r},\mathbf{v})}{\mu}\right]\rho_{max}^2 + \rho_{max} - \frac{\|\mathbf{h}(\mathbf{r},\mathbf{v})\|^2}{2\mu}}{\frac{1}{2}\left[\left(\frac{\rho_{max}}{\rho_{min}}\right) - 1\right]\sigma_\rho} \\
c_3(\begin{bmatrix}\mathbf{r}\\ \mathbf{v}\end{bmatrix}) &= \frac{-\rho_{min} - \left(\frac{2\rho_{min}^2}{\mu}\right)\mathcal{E}(\mathbf{r},\mathbf{v})}{\sigma_\rho} \\
c_4(\begin{bmatrix}\mathbf{r}\\ \mathbf{v}\end{bmatrix}) &= \frac{\rho_{max} + \left(\frac{2\rho_{max}^2}{\mu}\right)\mathcal{E}(\mathbf{r},\mathbf{v})}{\sigma_\rho}
\end{aligned}
\tag{7}
$$

where the parameter σ_ρ serves as an inverse penalty weight given in distance units. Each of these four inequality constraints is non-dimensional. They reproduce the four right-hand sides of Eqs. (1), (2), (4), and (5), but with re-scaling factors that non-dimensionalize them and that severely penalize constraint violations if their range-equivalent values exceed zero by magnitudes on the order of σ_ρ or larger.

NONLINEAR BATCH LEAST-SQUARES IOD PROBLEM AND SOLVER BASED ON BEARINGS DATA AND ADMISSIBLE REGION CONSTRAINTS

Batch Measurement Model

The initial operation of this paper's bearings-only IOD algorithm is to solve a batch least-squares problem that involves the bearings measurements and the periapsis and apoapsis admissible region bounds in Eq. (6). The right-ascension and declination bearing measurements at a sequence of times can be modeled in the following nonlinear batch measurement model form:

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \boldsymbol{\nu} \tag{8}$$

where $\mathbf{y}$ is a $2N$-dimensional vector that contains the stacked right-ascension and declination measurements at N samples times. The function $\mathbf{h}(\mathbf{x})$ is a nonlinear measurement model function. It includes an orbit propagator from the epoch time of the position/velocity state $\mathbf{x}$ to each individual bearing measurement time. This model includes the inverse trigonometric calculation of each right-ascension/declination pair. Its computations use the body-fixed coordinates of each ground station from which data are available along with the attracting body's known orientation time history relative to inertial space. The $2N$-dimensional vector $\boldsymbol{\nu}$ models the bearing measurement errors. It is assumed to be a zero-mean, Gaussian random vector. Its square-root information matrix is R_ν, which implies that its measurement error covariance is $P_\nu = R_\nu^{-1}R_\nu^{-\mathrm{T}}$. Typically R_ν is diagonal with its diagonal elements equal to the reciprocals of the corresponding right-ascension and declination measurement error standard deviations.

The orbit propagator used to implement $\mathbf{h}(\mathbf{x})$ can be a simple Keplerian calculation or something more complicated. The present study uses the Kepler 2-body orbital motion model. It performs propagation using non-singular equinoctial elements.[9]

Batch Least-Squares Problem

The initial rough orbit estimate is determined by solving the following nonlinear weighted batch least-squares problem with penalty terms:

$$\begin{aligned} &\text{find: } \mathbf{x} \\ &\text{to minimize: } J(\mathbf{x}) = \frac{1}{2}[\mathbf{y} - \mathbf{h}(\mathbf{x})]^{\mathrm{T}} R_{\nu}^{\mathrm{T}} R_{\nu}[\mathbf{y} - \mathbf{h}(\mathbf{x})] \\ &\qquad\qquad + \frac{1}{2}\sum_{i=1}^{4}\{\max[0, c_i(\mathbf{x})]\}^2 \end{aligned} \tag{9}$$

The problem's cost function $J(\mathbf{x})$ contains two terms. The first term is the usual weighted sum of the squared errors between the actual bearing measurements and their modeled values based on the epoch state $\mathbf{x}$. The second term penalizes violations of the 4 inequality constraints from Eq. (6). The maximization operation $\max[0, c_i(\mathbf{x})]$ ensures that violations are penalized only when the corresponding $c_i(\mathbf{x})$ inequality constraint component is positive.

Note that the first term in the cost function in Eq. (9) is the negative log likelihood cost of the bearings-only measurement problem. Typical batch estimators would include only this term. In the case of short data arcs from a single ground station and high orbital altitude, the minimum of this term is not sufficiently well defined to provide the sole basis for orbit determination, hence the extra penalty terms.

Batch Least-Squares Solution Algorithm

An algorithm with special features has been developed to solve the problem in Eq. (9) for the soft-constrained bearings-only IOD problem. It is a Levenberg-Marquardt method[10] that includes features which ensure convergence to the global minimum and features which treat the penalty terms in Eq. (9). One feature is a search step adjustment within the Levenberg-Marquardt algorithm that ensures a decrease of $J(\mathbf{x})$ with every change that seeks to improve upon the current guess of the $\mathbf{x}$ solution. The use of such adjustments ensures that the method converges at least to a local minimum of $J(\mathbf{x})$.

Another feature of the algorithm is an active set method for dealing with the inequality penalty terms in Eq. (9). If the i^{th} constraint is satisfied, then it is ignored in the standard Levenberg-Marquardt computation of the search direction in $\mathbf{x}$ space. If the constraint is violated, then the Levenberg-Marquardt computation appends the equation $0 = c_i(\mathbf{x})$ to the over-determined system of equations for which it seeks a least-squares solution. During the iterations that are involved in a Levenberg-Marquardt solution, a given inequality constraint could switch from inactive to active and back multiple times.

Global minimization of the cost function $J(\mathbf{x})$ is ensured by re-seeding of the Levenberg-Marquardt method with multiple randomly chosen first guesses. The solution that produces the lowest cost is taken as the global solution. The first guesses are chosen by using one of the bearing measurement pairs and the corresponding observation station location in order to determine two of the six unknowns in $\mathbf{x}$. The remaining 4 unknowns are the range to the object along the measured direction and the 3 velocity components. The range to the object, its velocity magnitude, and the angle between its velocity and position vectors are seeded randomly by using a random number generator in order to select the periapsis and apoapsis radial distances along with the orbital mean

anomaly. The former two quantities are sampled from a uniform distribution in the range between ρ_{min} and ρ_{max}. The mean anomaly is sampled from a uniform distribution in the range 0 to 2π. The resulting state guess takes the form

$$\mathbf{x}_g = \begin{bmatrix} \mathbf{r}_g \\ \left(v_{g\parallel}\hat{\mathbf{r}}_g + v_{g\perp}\cos\theta_g\,\hat{\mathbf{r}}_{a\perp} + v_{g\perp}\sin\theta_g\,\hat{\mathbf{r}}_{b\perp}\right) \end{bmatrix} \tag{10}$$

where the guessed epoch-state position is

$$\mathbf{r}_g = \mathbf{r}_{station} + \rho_g \begin{bmatrix} \cos\alpha\,\cos\delta \\ \sin\alpha\,\cos\delta \\ \sin\delta \end{bmatrix} \tag{11}$$

with ρ_g being the guessed range from the observing station location $r_{station}$, α being the measured right ascension, and δ being the measured declination. The unit direction vector parallel to the guessed position vector is $\hat{\mathbf{r}}_g = \mathbf{r}_g/\sqrt{\mathbf{r}_g^T\mathbf{r}_g}$. The other two unit direction vectors, $\hat{\mathbf{r}}_{a\perp}$ and $\hat{\mathbf{r}}_{b\perp}$, are chosen to be perpendicular to $\hat{\mathbf{r}}_g$ and to each other. The component of the guessed velocity parallel to $\hat{\mathbf{r}}_g$ is $v_{g\parallel}$, and the component perpendicular to $\hat{\mathbf{r}}_g$ is $v_{g\perp}$. The guessed values ρ_g, $v_{g\parallel}$, and $v_{g\perp}$ are derived based on Kepler's laws and using the randomly sampled periapsis, apoapsis, and mean anomaly.

The guessed rotation of the $v_{g\perp}$ velocity component about $\hat{\mathbf{r}}_g$ is θ_g. The guess of this angle is determined via brute-force optimization of $J[\mathbf{x}_g(\theta_g)]$. This optimization holds ρ_g, $v_{g\parallel}$, and $v_{g\perp}$ constant, and it evaluates $J[\mathbf{x}_g(\theta_g)]$ on a grid of θ_g values that span the range from 0 to 2π. The θ_g grid value that gives the lowest $J[\mathbf{x}_g(\theta_g)]$ is chosen as the guessed value. This method of choosing θ_g aligns the initial guess of the orbital plane about $\hat{\mathbf{r}}_g$ in the way that best fits the bearing measurement time history contained in the $\mathbf{y}$ vector.

The batch least-squares algorithm tends to converge from a first guess $\mathbf{x}_g$ to a local minimum in 10's to 100's of Levenberg-Marquardt iterations when using 3-40 pairs of (α, δ) bearing measurements. Fewer iterations tend to be required if the penalty terms are not active. It typically requires about 15 or fewer independent $\mathbf{x}_g$ guesses in order to locate the global minimum. These calculations can be carried out in execution times that range from a few seconds up to 4 minutes when working in MATLAB on a 3 GHz laptop computer. The robustness of this batch algorithm, when coupled with its lack of simplifying assumptions, makes it a powerful alternative to existing methods that address this problem, including those developed by Gauss, Laplace, and others.[1,2]

Batch Least-Squares Solution Analysis

Let $\hat{\mathbf{x}}_{cmle}$ be the estimate that solves the penalty-function-constrained maximum-likelihood estimation problem in Eq. (9). It is useful to develop a square-root information equation[11] that is a projection of the linearized measurement model in Eq. (8) onto the range space of the pre-whitened Jacobian matrix of the measurement model. The linearization is computed at the batch solution $\hat{\mathbf{x}}_{cmle}$. The desired square-root information equation takes the form:

$$R_{cmle}(\mathbf{x} - \hat{\mathbf{x}}_{cmle}) = \mathbf{z}_{cmle} - \boldsymbol{\nu}_{cmle} \tag{12}$$

The 6-by-6 square-root information matrix R_{cmle} is computed using orthonormal/upper-triangular (QR) factorization[10] of the pre-whitened Jacobian matrix of the measurements:

$$Q_{cmle}R_{cmle} = R_\nu \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}_{cmle}} \tag{13}$$

with Q_{cmle} being a $2N$-by-6 matrix that has orthonormal columns. It is also an output of the QR factorization. This latter matrix and the measurement residuals are used to compute the 6-by-1 non-homogeneous term of Eq. (12):

$$\mathbf{z}_{cmle} = Q_{cmle}^{\mathrm{T}} R_{\nu} [\mathbf{y} - \mathbf{h}(\hat{\mathbf{x}}_{\mathbf{cmle}})] \tag{14}$$

This term will be zero if none of the inequality constraint penalties in Eq. (9) are active at $\hat{\mathbf{x}}_{cmle}$. It will be non-zero when one or more of the constraint terms is active. In situations with non-zero $\mathbf{z}_{cmle}$, an approximation of the unconstrained maximum-likelihood IOD state estimate is

$$\hat{\mathbf{x}}_{mle} = \hat{\mathbf{x}}_{cmle} + R_{cmle}^{-1} \mathbf{z}_{cmle} \tag{15}$$

The vector $\boldsymbol{\nu}_{cmle}$ in Eq. (12) is modeled as being a zero-mean, identity-covariance Gaussian random vector. Given this assumption, the modeled error covariance of the maximum-likelihood estimate in Eq. (15) is

$$\begin{aligned} E\{(\mathbf{x} - \hat{\mathbf{x}}_{mle})(\mathbf{x} - \hat{\mathbf{x}}_{mle})^{\mathrm{T}}\} &= P_{mle} \\ &= \left[\left(R_{\nu} \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_{cmle}} \right)^{\mathrm{T}} \left(R_{\nu} \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_{cmle}} \right) \right]^{-1} \\ &= R_{cmle}^{-1} R_{cmle}^{-\mathrm{T}} \end{aligned} \tag{16}$$

For the problems under consideration, this covariance often has two very large components. They are so large that the approximate maximum likelihood estimate $\hat{\mathbf{x}}_{mle}$ is useless. The remainder of this paper seeks to add information to the bearing information of Eq. (8) in order to develop a more useful IOD estimate.

SUBSPACE OF MAXIMUM BEARINGS-ONLY IOD UNCERTAINTY

For a challenging IOD problem with high altitude and a short data arc, the two most uncertain components of the state vector can have very large error standard deviations of their maximum-likelihood estimates. In a typical example, their standard deviations are larger by factors of 2,000 and 20,000, respectively, than the standard deviation of the next most uncertain component. These two components define the subspace that requires IOD aiding based on admissible region analysis.

It is necessary to make the units of the IOD state vector $\mathbf{x}$ uniform in order to determine the most uncertain subspace of $\mathbf{x}$. Otherwise, one is comparing "apples" and "oranges" because the first 3 elements of $\mathbf{x}$ are given in distance units and the last 3 are given in velocity units. One can use the orbit's Keplerian mean motion in order to transform into consistent units. Using the estimated state $\hat{\mathbf{x}}_{cmle} = [\hat{\mathbf{r}}_{cmle}^{\mathrm{T}}, \hat{\mathbf{v}}_{cmle}^{\mathrm{T}}]^{\mathrm{T}}$ and Keplerian dynamics, the mean motion is

$$\begin{aligned} n_{0cmle} &= \sqrt{\frac{\mu}{[a(\hat{\mathbf{r}}_{cmle}, \hat{\mathbf{v}}_{cmle})]^3}} \\ &= \sqrt{\mu \left(\frac{2}{\sqrt{\hat{\mathbf{r}}_{cmle}^{\mathrm{T}} \hat{\mathbf{r}}_{cmle}}} - \frac{\hat{\mathbf{v}}_{cmle}^{\mathrm{T}} \hat{\mathbf{v}}_{cmle}}{\mu} \right)^3} \end{aligned} \tag{17}$$

The mean motion can be used to define a transformed square-root information matrix that applies for **r** given in some sensible distance units, e.g., meters or km, and for **v** given in those same distance units per radian of mean motion. This transformed square-root information matrix is

$$\tilde{R}_{cmle} = R_{cmle} \begin{bmatrix} I_{3\times 3} & 0_{3\times 3} \\ 0_{3\times 3} & n_{0cmle} I_{3\times 3} \end{bmatrix} = R_{cmle} T_{cmle} \tag{18}$$

where the 6-by-6 diagonal matrix T_{cmle} in the right-most expression is defined to equal the block diagonal partial non-dimensionalization matrix in the middle expression.

The most uncertain 2-dimensional subspace can be determined by computing the singular value decomposition of $\tilde{R}_{cmle}$:

$$\tilde{U}\tilde{\Sigma}^{-1}\tilde{V}^{\mathrm{T}} = \tilde{R}_{cmle} \tag{19}$$

where $\tilde{U}$ and $\tilde{V}$ are 6-by-6 orthonormal matrices and $\tilde{\Sigma}^{-1}$ is a diagonal matrix with positive singular values on its diagonal in descending order. Suppose that the diagonal elements of $\tilde{\Sigma}^{-1}$ are $1/\tilde{\sigma_j}$ for j = 1,...,6 so that $\tilde{\sigma_1} \leq \tilde{\sigma_2} \leq ... \leq \tilde{\sigma_6}$ because their reciprocals are the descending diagonal elements of $\tilde{\Sigma}^{-1}$. Let the corresponding columns of $\tilde{V}$ be designated as $\tilde{v}_j$ for j = 1,...,6.

The $\tilde{\sigma_j}$ singular value reciprocals are maximum-likelihood estimation error standard deviations for the corresponding state-space directions in $\tilde{V}$. Large values correspond to large uncertainties. Therefore, the two must uncertain IOD directions are those associated with the last two rows and columns of $\tilde{\Sigma}^{-1}$. In position/velocity state space with the original units, these two directions are

$$\Delta\mathbf{x}_5 = T_{cmle}\tilde{v}_5 \quad \text{and} \quad \Delta\mathbf{x}_6 = T_{cmle}\tilde{v}_6 \tag{20}$$

THEORETICAL IOD PROBABILITY DENSITY FUNCTION BASED ON BEARINGS-ONLY LIKELIHOOD AND ADMISSIBLE REGION

A reasonable way to incorporate the admissible region information is to postulate that the initial orbit is equally likely to lie anywhere in this region. Suppose that the maximum-likelihood linearized square-root information equation in Eq. (12) contains all of the IOD information from the bearing measurements. Then the Bayesian fusion of these two sources of information yields the following IOD probability density function

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}, \rho_{min}, \rho_{max}) = C \exp\{-\frac{1}{2}[R_{cmle}(\mathbf{x}-\hat{\mathbf{x}}_{cmle}) - \mathbf{z}_{cmle}]^{\mathrm{T}}[R_{cmle}(\mathbf{x}-\hat{\mathbf{x}}_{cmle}) - \mathbf{z}_{cmle}]\} \\ \times \Pi_{np}[c_1(\mathbf{x})] \times \Pi_{np}[c_2(\mathbf{x})] \end{aligned} \tag{21}$$

where $\Pi_{np}(\eta)$ is the scalar finite support function over non-positive values of its scalar input argument:

$$\Pi_{np}(\eta) = \begin{cases} 1 & \text{if} \quad \eta \leq 0 \\ 0 & \text{if} \quad 0 < \eta \end{cases} \tag{22}$$

and where C is a normalization constant defined by:

$$\begin{aligned} \frac{1}{C} = \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}[R_{cmle}(\mathbf{x}-\hat{\mathbf{x}}_{cmle}) - \mathbf{z}_{cmle}]^{\mathrm{T}}[R_{cmle}(\mathbf{x}-\hat{\mathbf{x}}_{cmle}) - \mathbf{z}_{cmle}]\} \\ \times \Pi_{np}[c_1(\mathbf{x})] \times \Pi_{np}[c_2(\mathbf{x})] d\mathbf{x} \end{aligned} \tag{23}$$

The effect of the two $\Pi_{np}(*)$ factors in Eq. (21) is to null out the probability density of all states that lie outside the admissible region. If this region is very large and if the Gaussian likelihood

probability density function lies mostly in this region, then the admissible region terms will not have much impact. The IOD probability density function will be nearly equal to the exponential likelihood term in Eq. (21), after multiplication by an appropriate normalization constant. This will be the case for LEO orbits with a sufficiently long bearings-only data arc. If the admissible region is small relative to the extent of the Gaussian likelihood term, then the two finite-support functions will have significant effects on the final form of the conditional distribution.

Recall from Eqs. (1)-(7) that the constraint functions $c_1(\mathbf{x})$ and $c_2(\mathbf{x})$ are used to model, respectively, the minimum periapsis constraint and the maximum apoapsis constraint. Recall, also, that these two constraints alone can be ambiguous, as per the discussion associated with Eqs. (4) and (5). This potential ambiguity will be ignored for the remainder of the paper under the reasonable assumption that its resolution will be obvious in the vicinity of solution $\hat{\mathbf{x}}_{cmle}$ of the batch estimation problem of Eq. (9).

The remainder of this paper seeks to develop an accurate and efficient Gaussian mixture approximation of the conditional probability density function given in Eq. (21). Accuracy means that the approximation lies close to this theoretical probability density function. Efficiency means that it uses a low number of Gaussian components – also known as mixands – in order to achieve an accurate approximation.

GAUSSIAN MIXTURE APPROXIMATION OF THE IOD PROBABILITY DENSITY FUNCTION IN THE 2D UNCERTAIN SUBSPACE

Transformation and Projection onto the Maximally Uncertain 2D Subspace

In theory, The IOD conditional probability density in Eq. (21) could be directly approximated by a Gaussian mixture. In practice, it would be very difficult to develop and implement an algorithm that could optimize such a Gaussian mixture to fit the theoretical probability density function well in the 6-dimensional $\mathbf{x}$ space. Therefore, a variable transformation and a 2-dimensional projection are implemented in order to simplify the problem of optimizing the fit of the Gaussian mixture. As will be demonstrated, forming an accurate and efficient Gaussian mixture in just 2 dimensions is difficult enough.

The change of coordinates employs the singular decomposition of Eq. (19) and the soft-constrained rough state estimate $\hat{\mathbf{x}}_{cmle}$. Suppose that the new coordinates are α_j for j = 1,...,6. Then the original position/velocity state vector can be written as follows in terms of these new coordinates:

$$\mathbf{x} = \hat{\mathbf{x}}_{cmle} + T_{cmle}\tilde{V} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix} = \hat{\mathbf{x}}_{cmle} + \Delta X \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix} \tag{24}$$

where the 6-by-6 matrix $\Delta X = T_{cmle}\tilde{V}$ and where its j^{th} column is $\Delta\mathbf{x}_j$ for j = 1,...,6, consistent with Eq. (20).

The inverse of this coordinate transformation takes the form

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix} = (\Delta X)^{-1}(\mathbf{x} - \hat{\mathbf{x}}_{cmle}) \tag{25}$$

The required inverse of ΔX is easy to compute because this matrix is the product of a diagonal matrix and an orthonormal matrix. This latter transformation demonstrates that the new α_j coordinates keep track of state perturbations from the soft-constrained batch least-squares estimate $\hat{\mathbf{x}}_{cmle}$. These new coordinates measure perturbations along the different columns of the orthonormal matrix $\tilde{V}$ in the partially non-dimensionalized position/velocity state space. Each of them has identical units, length units.

In order to derive the transformed probability density, it is necessary to define the transformed functions that implement the minimum periapsis constraint and the maximum apoapsis constraint. These transformed functions take the forms

$$\tilde{c}_i(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = c_i(\hat{\mathbf{x}}_{cmle} + \Delta X \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix}) \quad \text{for } i = 1, 2 \tag{26}$$

The transformed unconstrained maximum-likelihood estimates of the α_j values are also useful for deriving the transformed probability density. Starting from Eq. (15), one can apply various matrix relationships in order to show that these estimates are

$$\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \hat{\alpha}_5 \\ \hat{\alpha}_6 \end{bmatrix} = \tilde{\Sigma}\, \tilde{U}^{\mathrm{T}} \hat{\mathbf{z}}_{cmle} \tag{27}$$

Given the foregoing definitions and transformations, the transformed IOD conditional probability density takes the form

$$\begin{aligned} p(\alpha_1, \ldots, \alpha_6 | \mathbf{y}, \rho_{min}, \rho_{max}) = n_{0cmle}^3 C \exp\{-\frac{1}{2} \sum_{i=1}^{6} \left(\frac{\alpha_i - \hat{\alpha}_i}{\tilde{\sigma}_i} \right)^2 \} \\ \times \Pi_{np}[\tilde{c}_1(\alpha_1, \ldots, \alpha_6)] \times \Pi_{np}[\tilde{c}_2(\alpha_1, \ldots, \alpha_6)] \end{aligned} \tag{28}$$

Next, the assumption is made that the α_i for i = 1,2,3,4 do not vary enough from their unconstrained maximum likelihood estimates to have much effect on the finite-support constraint terms in Eq. (28). This is a reasonable assumption because $\tilde{\sigma}_i$ for i = 1,2,3,4 tend to be much smaller than $\tilde{\sigma}_5$

and $\tilde{\sigma}_6$ for cases in which the admissible set terms play any important role. Using this assumption, the following approximation is made of the transformed probability density function:

$$\begin{aligned} p(\alpha_1, \ldots, \alpha_6 | \mathbf{y}, \rho_{min}, \rho_{max}) &\approx n_{0cmle}^3 C \exp\{-\frac{1}{2}\sum_{i=1}^{6}\left(\frac{\alpha_i - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\} \\ &\quad \times \Pi_{np}[\tilde{c}_1(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)] \\ &\quad \times \Pi_{np}[\tilde{c}_2(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)] \\ &\approx p(\alpha_1, \alpha_2, \alpha_3, \alpha_4 | \mathbf{y}) \times p(\alpha_5, \alpha_6 | \mathbf{y}, \rho_{min}, \rho_{max}) \end{aligned} \tag{29}$$

where

$$p(\alpha_1, \alpha_2, \alpha_3, \alpha_4 | \mathbf{y}) = \frac{1}{(2\pi)^2 \prod_{i=1}^{4} \tilde{\sigma}_i} \exp\{-\frac{1}{2}\sum_{i=1}^{4}\left(\frac{\alpha_i - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\} \tag{30}$$

and

$$\begin{aligned} p(\alpha_5, \alpha_6 | \mathbf{y}, \rho_{min}, \rho_{max}) &= \tilde{C} \exp\{-\frac{1}{2}\sum_{i=5}^{6}\left(\frac{\alpha_i - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\} \\ &\quad \times \Pi_{np}[\tilde{c}_1(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)] \\ &\quad \times \Pi_{np}[\tilde{c}_2(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)] \end{aligned} \tag{31}$$

and where $\tilde{C}$ is another normalizing constant that is computed as follows:

$$\begin{aligned} \frac{1}{\tilde{C}} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{-\frac{1}{2}\sum_{i=5}^{6}\left(\frac{\alpha_i - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\} \\ &\quad \times \Pi_{np}[\tilde{c}_1(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)] \\ &\quad \times \Pi_{np}[\tilde{c}_2(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)] d\alpha_5 d\alpha_6 \\ &= \int_S \exp\{-\frac{1}{2}\sum_{i=5}^{6}\left(\frac{\alpha_i - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\} d\alpha_5 d\alpha_6 \end{aligned} \tag{32}$$

The set S is the region of (α_5,α_6) space where the probability density has support. This is the region where $\tilde{c}_1(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)$ and $\tilde{c}_2(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)$ are both non-positive.

As a reminder, the subspace (α_1,α_2,α_3,α_4), in which a sample Gaussian applies, maps roughly to the right-ascension and declination bearing and bearing rate for typical high-altitude, short-data-arc, single-ground-station cases. The highly uncertain subspace (α_5,α_6) maps roughly to range and range rate in this situation. Thus, this decomposition approach has many similarities to the approach used in Ref. [3].

Visualization of 2-Dimensional Finite-Support Region and Probability Density Function

The projection of the admissible region onto (α_5,α_6) is depicted in Fig. 1 for a simulated example problem. This example is the same as the first test case in Ref. [3]. Its orbital properties are semi-major axis = 43,000 km, eccentricity = 0.03, inclination = 3°, right-ascension of the ascending node = 0°, argument of perigee = 0°, and mean anomaly at epoch = 0°. The data arc is 19.5 seconds long. The admissible region is defined to have a minimum periapsis radius of 30,000 km and a maximum apoapsis radius of 47,300. These values encompass the true peripasis and apoapsis, respectively

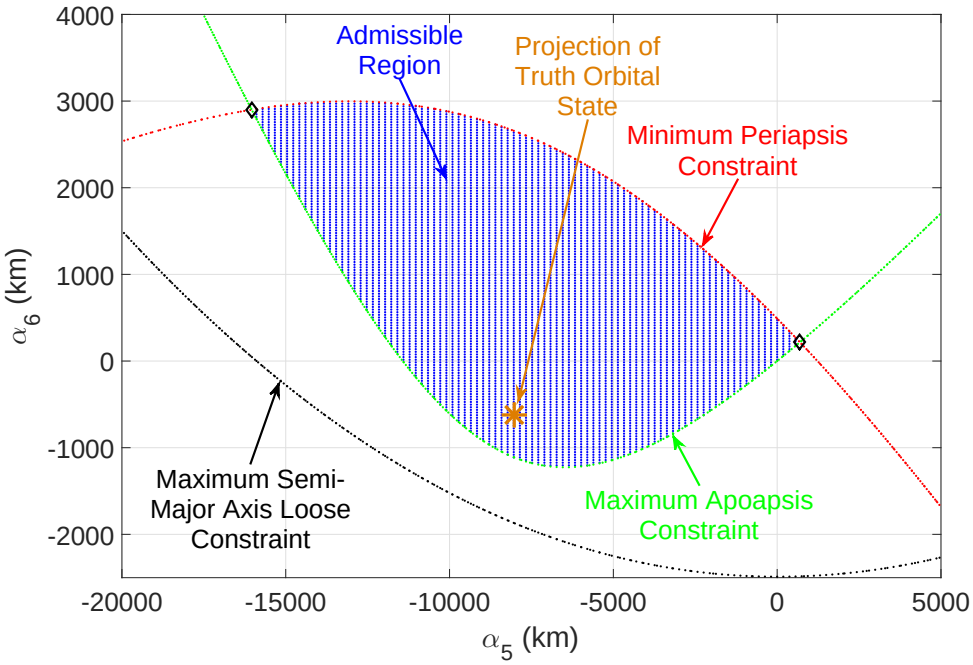

Figure 1. Example of admissible region projection into (α_5,α_6) space.

41,710 km and 44,290 km. The constraint weight parameter σ_ρ used in Eqs. (7) has been tuned to the value 500 km.

Figure 1 shows the projection of the admissible region onto (α_5,α_6) space. The figure's horizontal axis is the α_5 axis, and its vertical axis is the α_6 axis. The blue region is the admissible set S that has been defined in connection with the integral in the last line of Eq. (32). The point (α_5,α_6) = (0,0) lies on the boundary of this region. This is the point solved for in the rough IOD batch least-squares problem in Eq. (9). The fact that (0,0) lies on the maximum apoapsis boundary indicates that this constraint was active at the constrained batch least-squares solution. The "truth" state's projection onto (α_5,α_6) space is indicated by the brown asterisk in Fig. 1. It lies within the admissible region, as expected. It lies much closer the the green maximum apoapsis constraint than it does to the red minimum periapsis constraint, consistent with the relationship of the "truth" periapsis and apoapsis values to the constraint values. The black maximum semi-major axis constraint in this figure does not border the admissible region, consistent with expectations. Recall that this constraint has been included only to remove ambiguity about the active constraints during the batch least-squares solution of the problem in Eq. (9).

Figure 2 shows a 3-dimensional plot of the projected probability density function $p(\alpha_5, \alpha_6|\mathbf{y}, \rho_{min}, \rho_{max})$ vs. α_5 and α_6. Consistent with its design in Eq. (31), it is zero outside the (α_5,α_6) admissible region. Inside the region its variations are dictated by the exponential factor in Eq. (31). This variation is based on the fit to the bearing measurements. It distinguishes this approach from that of Ref. [3]. In that reference, the equivalent projected probability density function is exactly constant over the admissible region. The ability to allow probability density variations over the admissible region enables this paper's method to transition smoothly from cases where the admissible region constraints provide significant information to those where they do not.

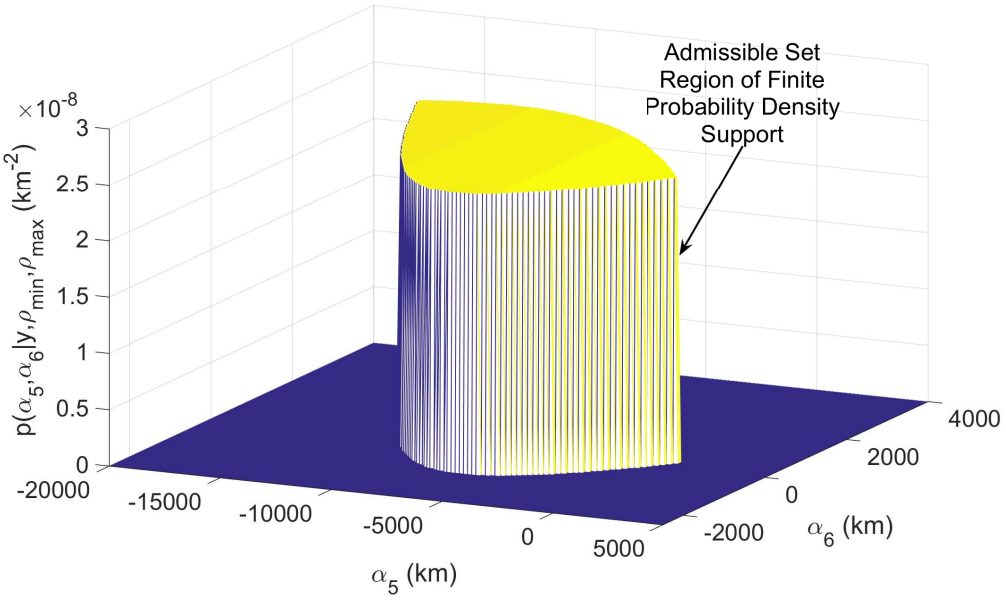

Figure 2. Example theoretical probability density function in (α_5,α_6) space – distribution that will be approximated by a Gaussian mixture.

The goal of the remainder of this section is to develop methods to approximate probability density functions such as the one in Fig. 2. These approximations will take the form of Gaussian mixtures.

Gaussian Mixture Definition

The Gaussian mixture approximation of the $p(\alpha_5, \alpha_6|\mathbf{y}, \rho_{min}, \rho_{max})$ probability density function takes the form:

$$p_{gm}(\alpha_5, \alpha_6) = \sum_{k=1}^{K} w_k \mathcal{N}_{sr2}(\alpha_5, \alpha_6; \boldsymbol{\mu}_{\alpha k}, R_{\alpha k}) \tag{33}$$

The $\mathcal{N}_{sr2}$ function is 2-dimensional Gaussian distribution given in square-root-information form:

$$\mathcal{N}_{sr2}(\alpha_5, \alpha_6; \boldsymbol{\mu}_\alpha, R_\alpha) = \frac{|\det(R_\alpha)|}{2\pi} \exp\{-\frac{1}{2}(\begin{bmatrix}\alpha_5 \\ \alpha_6\end{bmatrix} - \boldsymbol{\mu}_\alpha)^{\mathrm{T}} R_\alpha^{\mathrm{T}} R_\alpha (\begin{bmatrix}\alpha_5 \\ \alpha_6\end{bmatrix} - \boldsymbol{\mu}_\alpha)\} \tag{34}$$

with the 2-by-1 mean value vector $\boldsymbol{\mu}_\alpha$ and the 2-by-2 square-root information matrix R_α. The mixand weights w_k of the K mixands in Eq. (33) are non-negative, and they sum to 1:

$$\begin{aligned} 0 &\leq w_k && \text{for } k = 1, ..., K \\ 1 &= \sum_{k=1}^{K} w_k \end{aligned} \tag{35}$$

The 2-by-2 square-root information matrix $R_{\alpha k}$ has 4 elements, but only 3 affect the corresponding Gaussian mixand independently. This is true because $R_{\alpha k}$ can be altered through left-multiplication by any orthonormal 2-by-2 matrix without changing the corresponding mixand. There-

fore, one can arbitrarily choose $R_{\alpha k}$ to have a zero-valued (2,1) element without loss of generality:

$$R_{\alpha k} = \begin{bmatrix} R_{11\alpha k} & R_{12\alpha k} \\ 0 & R_{22\alpha k} \end{bmatrix} \tag{36}$$

Thus, each of the K mixands is characterized by a weight and 5 other independent parameters. It is convenient to lump the other parameters into the following 5-by-1 vector

$$\mathbf{p}_k = \begin{bmatrix} \boldsymbol{\mu}_{\alpha k} \\ R_{11\alpha k} \\ R_{12\alpha k} \\ R_{22\alpha k} \end{bmatrix} \tag{37}$$

It is also convenient to lump all of the individual mixand weights into a K-by-1 vector of weights and all of the 5-element parameter vectors into a $5K$-by-1 vector of parameters. The resulting two vectors are

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_2 \\ \vdots \\ w_K \end{bmatrix} \qquad \text{and} \qquad \mathbf{p} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \\ \vdots \\ \mathbf{p}_K \end{bmatrix} \tag{38}$$

Given these two definitions, one can formally denote the dependence of the Gaussian mixture in Eq. (33) on these vectors by writing it in the form $p_{gm}(\alpha_5, \alpha_6; \mathbf{w}, \mathbf{p})$. The goal of the Gaussian mixture fitting operation is to choose the $\mathbf{w}$ weight vector and the $\mathbf{p}$ parameter vector so that $p_{gm}(\alpha_5, \alpha_6; \mathbf{w}, \mathbf{p})$ optimally approximates $p(\alpha_5, \alpha_6 | \mathbf{y}, \rho_{min}, \rho_{max})$ from Eq. (31).

Optimal Approximation of Projected Probability Density by a Gaussian Mixture

The Gaussian mixture optimal fit problem takes the form

$$\begin{aligned} &\text{find: } \mathbf{w} \text{ and } \mathbf{p} \\ &\text{to minimize: } J_{gm}(\mathbf{w}, \mathbf{p}) = \frac{1}{2} \int_{-\infty}^{\infty} d\alpha_5 \int_{-\infty}^{\infty} d\alpha_6 [p(\alpha_5, \alpha_6 | \mathbf{y}, \rho_{min}, \rho_{max}) \\ &\qquad\qquad - p_{gm}(\alpha_5, \alpha_6; \mathbf{w}, \mathbf{p})]^2 \\ &\text{subject to: } 0 \leq \mathbf{w} \\ &\qquad 1 = \mathbf{e}^{\mathrm{T}} \mathbf{w} \end{aligned} \tag{39}$$

where $\mathbf{e} = [1,...,1]^{\mathrm{T}}$ is a K-by-1 vector of all 1's.

The cost function in this problem statement is the integral-squared difference (ISD) between the two probability distributions. It equals half the square of the function-space 2-norm of the difference between these two distributions. It has been used as a fit metric in other Gaussian mixture studies.[3,7,12] It can be re-written in the form

$$J_{gm}(\mathbf{w}, \mathbf{p}) = J_{gm0} + [\mathbf{g}_w(\mathbf{p})]^{\mathrm{T}} \mathbf{w} + \frac{1}{2} \mathbf{w}^{\mathrm{T}} H_{ww}(\mathbf{p}) \mathbf{w} \tag{40}$$

The constant term in this latter fit error cost expression is

$$J_{gm0} = \frac{1}{2}\int_{-\infty}^{\infty} d\alpha_5 \int_{-\infty}^{\infty} d\alpha_6 \, [p(\alpha_5, \alpha_6|\mathbf{y}, \rho_{min}, \rho_{max})]^2$$
$$= \frac{\tilde{C}^2}{2}\int_S d\alpha_5 d\alpha_6 \exp\{-\frac{1}{2}\sum_{i=5}^{6}\left(\frac{\alpha_i - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\}^2 \tag{41}$$

The k^{th} element of the K-by-1 gradient vector $\mathbf{g}_w(\mathbf{p})$ is

$$[\mathbf{g}_w(\mathbf{p})]_k = -\int_{-\infty}^{\infty} d\alpha_5 \int_{-\infty}^{\infty} d\alpha_6 \, p(\alpha_5, \alpha_6|\mathbf{y}, \rho_{min}, \rho_{max}) \mathcal{N}_{sr2}(\alpha_5, \alpha_6; \boldsymbol{\mu}_{\alpha k}, R_{\alpha k})$$
$$= -\tilde{C}\int_S d\alpha_5 d\alpha_6 \exp\{-\frac{1}{2}\sum_{i=5}^{6}\left(\frac{\alpha_i - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\} \mathcal{N}_{sr2}(\alpha_5, \alpha_6; \boldsymbol{\mu}_{\alpha k}, R_{\alpha k}) \tag{42}$$

The $(jk)^{\text{th}}$ element of the K-by-K Hessian matrix is[7]

$$[H_{ww}(\mathbf{p})]_{jk} = \int_{-\infty}^{\infty} d\alpha_5 \int_{-\infty}^{\infty} d\alpha_6 \, \mathcal{N}_{sr2}(\alpha_5, \alpha_6; \boldsymbol{\mu}_{\alpha j}, R_{\alpha j}) \mathcal{N}_{sr2}(\alpha_5, \alpha_6; \boldsymbol{\mu}_{\alpha k}, R_{\alpha k})$$
$$= \frac{|det(R_{\alpha j}) det(R_{\alpha k})|}{2\pi\sqrt{det(R_{\alpha j}^{\mathrm{T}} R_{\alpha j} + R_{\alpha k}^{\mathrm{T}} R_{\alpha k})}}$$
$$\times \exp\{-\frac{1}{2}(\boldsymbol{\mu}_{\alpha j} - \boldsymbol{\mu}_{\alpha k})^{\mathrm{T}} (R_{\alpha j}^{\mathrm{T}} R_{\alpha j} + R_{\alpha k}^{\mathrm{T}} R_{\alpha k})(\boldsymbol{\mu}_{\alpha j} - \boldsymbol{\mu}_{\alpha k})]\} \tag{43}$$

Note that the $[\mathbf{g}_w(\mathbf{p})]_k$ gradient element only depends on the 5 $\mathbf{p}_k$ elements of the Gaussian mixture parameter vector $\mathbf{p}$. Similarly, the $[H_{ww}(\mathbf{p})]_{jk}$ Hessian element depends only on the 5 $\mathbf{p}_j$ elements and the 5 $\mathbf{p}_k$ elements of $\mathbf{p}$.

A challenge in evaluating this cost function arises because the integrals over the admissible region S cannot be performed analytically. These needed integrals occur in Eqs. (32) and (42). Equation (41) also contains such an integral, but it does not need to be evaluated in order to solve the problem in Eq. (39) because that cost term is independent of the optimization variables $\mathbf{w}$ and $\mathbf{p}$. The next subsection describes a strategy for approximating the needed integrals.

A variant of Newton's method[10] is used to solve the optimization problem in Eq. (39). It works with guesses of the solution vectors $\mathbf{w}$ and $\mathbf{p}$. It computes improved guesses by optimizing a quadratic Taylor series expansion of $J_{gm}(\mathbf{w}, \mathbf{p})$ about the current guess. The gradient and Hessian of $J_{gm}(\mathbf{w}, \mathbf{p})$ with respect to the full $6N$-by-1 unknown vector [$\mathbf{w}$;$\mathbf{p}$] are both computed analytically – to the extent that the integrals over the admissible set S can be approximated analytically. The particular implementation of Newton's method used here ensures cost function descent on each iteration by adding a common non-negative scalar to each diagonal element of the Hessian matrix before computing the approximate optimal increment to the [$\mathbf{w}$;$\mathbf{p}$] solution vector. This non-negative scalar acts like a Levenberg-Marquardt parameter in a nonlinear least-squares solver.[10]

The first guess of the $\mathbf{w}$ weights vector and all solution increments to $\mathbf{w}$ are explicitly constrained to satisfy the unit-normalization condition 1 = $\mathbf{e}^{\mathrm{T}}\mathbf{w}$. The non-negativity constraints on the weights are enforced by starting with an all-positive guess of $\mathbf{w}$ and by enforcing an *ad hoc* restriction on the size of the search step. If needed, the restriction is enforced through an increase of the Levenberg-Marquardt-like parameter that gets added to the diagonal Hessian elements before computing the search step.

This implementation of Newton's method can make slow progress when far from the optimal [**w**;**p**] vector. It tends to converge rapidly in the vicinity of the optimum. Its *ad hoc* method of ensuring non-negativity of the Gaussian mixture weights tends to work well. It is conjectured that the optimization would rather adjust the mean and square-root information matrix of a given mixand in order to achieve a better probability density fit instead of deleting the mixand by driving its weight to zero. Having more mixands with non-zero weights tends to enable a better fit.

It is important to use reasonable scalings during the Newton's method optimization calculations. Reasonable scalings can be achieved through non-dimensionalization of the two elements of each mixand mean $\boldsymbol{\mu}_{\alpha k}$ and the 3 elements of each mixand square-root information matrix $R_{11\alpha k}$, $R_{12\alpha k}$, and $R_{22\alpha k}$. The mixand means can be non-dimensionalized through division by the respective maximum α_5 and α_6 extents of the admissible region S, as in Fig. 1. Non-dimensionalization of $R_{11\alpha k}$ can be achieved through multiplication by the same maximum α_5 extent of S. Non-dimensionalization of $R_{12\alpha k}$ and $R_{22\alpha k}$ can be achieved through multiplication by the maximum α_6 extent of S.

Practical Approximation of Integrals over the Admissible Set S

Calculation of the integrals over the set S in Eqs. (32) and (42) is carried out using various approximations. The first approximation is to the actual set S. The boundary of this set is defined implicitly by the equations 0 = $\tilde{c}_1(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)$ and 0 = $\tilde{c}_2(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)$. These equations are nonlinear in α_5 and α_6, and they do not admit closed-form solution. Instead of exact equations, quadratic approximations are used for these equations. These approximations are based on quadratic Taylor series expansions of the original constraint functions $c_1(\mathbf{x})$ and $c_2(\mathbf{x})$ that are defined in position/velocity state space. The two Taylor series are calculated about the rough batch-least squares solution point $\hat{\mathbf{x}}_{cmle}$. These quadratic approximations are then used in place of $c_1(\mathbf{x})$ and $c_2(\mathbf{x})$ in Eq. (26) in order to develop quadratic approximations of $\tilde{c}_1(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)$ and $\tilde{c}_2(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4, \alpha_5, \alpha_6)$. Using these approximations, the corresponding approximate boundaries of S are conic sections. In fact, these approximations have been used to generate the red and green hyperbolic approximate boundary curves in Fig. 1. The exact nonlinear constraints have been used to generate the blue shaded region. The correspondence of the approximate boundaries with the shaded region demonstrate the accuracy of the quadratic constraint approximation in this representative case.

Next, a coordinate transformation is introduced. New integration variables β_1 and β_2 replace α_5 and α_6 in the integrals of Eqs. (32) and (42). Figure 3 shows the β_1 and β_2 axes super-imposed on an example plot of the S region given in the original (α_5,α_6) coordinates. This figure shows the same admissible S region as in Fig. 1. The origin of the new (β_1,β_2) lies at the intersection point of the approximate $\tilde{c}_1$ and $\tilde{c}_2$ constraints that is nearest to the origin of the (α_5,α_6) coordinate system. The positive β_1 axis points from this intersection point to the next nearest intersection point of the two approximate admissibility constraints. These two intersection points are indicated by black diamonds in Fig. 3. Their (α_5,α_6) coordinates can be determined by solving a 4$^{\text{th}}$-order polynomial that is derivable from the intersection conditions for the two constraints' conic sections. The β_2 axis is defined to be perpendicular to the β_1 and to obey the right-hand rule. In Fig. 3 the β_2 axis does not appear to be perpendicular because of distortion in the figure: The figure's α_5 and α_6 scales are not equal. The mathematical transformation between the two coordinate systems takes the form

$$\begin{bmatrix} \alpha_5 \\ \alpha_6 \end{bmatrix} = \begin{bmatrix} \alpha_{50} \\ \alpha_{60} \end{bmatrix} + \tilde{Q} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \alpha_{50} \\ \alpha_{60} \end{bmatrix} + \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \tag{44}$$

where the point $[\alpha_{50}, \alpha_{60}]^{\mathrm{T}}$ is the right-most constraint intersection point shown in Fig. 3 and $\tilde{Q}$ is a 2-by-2 orthonormal rotation matrix with determinant equal to 1 and with elements $\tilde{Q}_{ij}$ for i,j = 1,2.

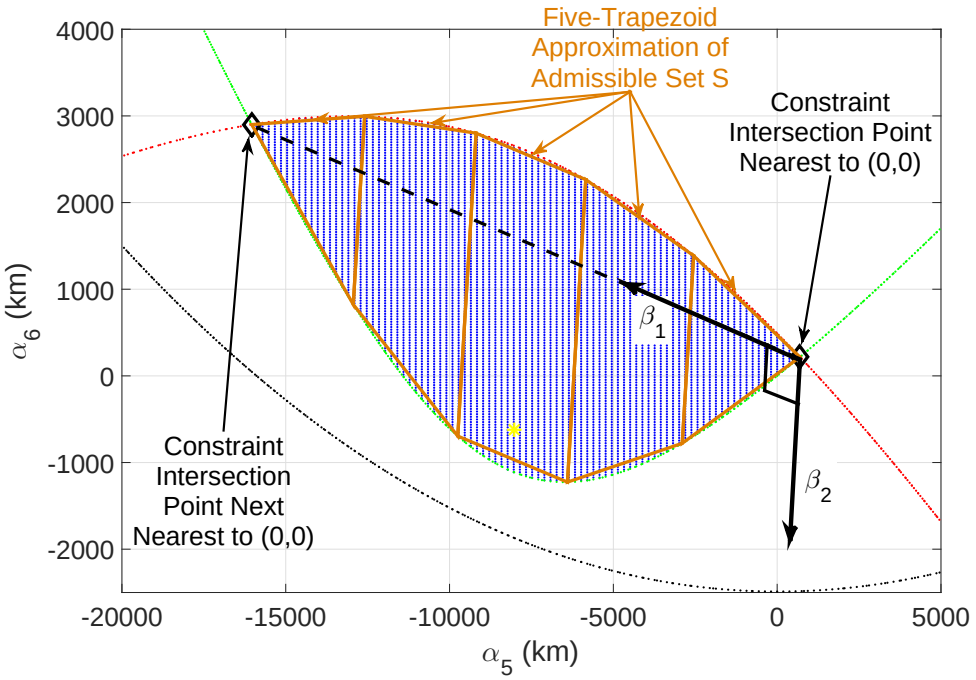

Figure 3. Coordinate transformation from (α_5,α_6) space to (β_1,β_2) space with admissible region approximation by 5 trapezoids that are distributed along the β_1 axis and aligned parallel to the β_2 axis.

The coordinate transformation in Eq. (44) can be applied directly to evaluation of the integrals in Eqs. (32) and (42). They become

$$\frac{1}{\tilde{C}} = \int_{\beta_{1min}}^{\beta_{1max}} d\beta_1 \int_{\beta_{2min}(\beta_1)}^{\beta_{2max}(\beta_1)} d\beta_2 \exp\{-\frac{1}{2}\sum_{i=5}^{6}\left(\frac{\alpha_{i0} + \tilde{Q}_{(i-4)1}\beta_1 + \tilde{Q}_{(i-4)2}\beta_2 - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\} \tag{45}$$

$$[\mathbf{g}_w(\mathbf{p})]_k = -\tilde{C}\int_{\beta_{1min}}^{\beta_{1max}} d\beta_1 \int_{\beta_{2min}(\beta_1)}^{\beta_{2max}(\beta_1)} d\beta_2 \exp\{-\frac{1}{2}\sum_{i=5}^{6}\left(\frac{\alpha_{i0} + \tilde{Q}_{(i-4)1}\beta_1 + \tilde{Q}_{(i-4)2}\beta_2 - \hat{\alpha}_i}{\tilde{\sigma}_i}\right)^2\} \times \mathcal{N}_{sr2}(\beta_1, \beta_2; \boldsymbol{\mu}_{\beta k}, R_{\beta k}) \tag{46}$$

The limits of the outer β_1 integrals, β_{1min} and β_{1max}, are determined from the two constraint intersection point solutions that are used to define the β_1 axis, as in Fig. 3. The β_1-dependent limits of the inner β_2 integrals, $\beta_{2min}(\beta_1)$ and $\beta_{2max}(\beta_1)$, can be determined by solving the quadratic equation that is the projection of each quadratic constraint approximation along the β_2 axis at the given β_1 value. The Gaussian mixand means and square-root information matrices can be transformed

from (α_5,α_6) space into (β_1,β_2) space in order to yield

$$\boldsymbol{\mu}_{\beta k} = \tilde{Q}^{\mathrm{T}} \left(\boldsymbol{\mu}_{\alpha k} - \begin{bmatrix} \alpha_{50} \\ \alpha_{60} \end{bmatrix} \right)$$
$$R_{\beta k} = R_{\alpha k} \tilde{Q} \tag{47}$$

for all k = 1,...,K.

The integral formulas in Eqs. (45) and (46) are not yet in a state amenable to closed-form evaluation. The inner integrals with respect to β_2 can be evaluated in closed form by using the standard error function from statistics, *erf*(*). The outer β_1 integral cannot be evaluated analytically because the two *erf*(*) terms that result from the inner integral are evaluated at arguments which depend on β_1 through the integration limit functions $\beta_{2min}(\beta_1)$ and $\beta_{2max}(\beta_1)$.

Two approximations are made in order enable evaluation of the outer β_1 integrals in Eqs. (45) and (46). The first approximates the admissible region S as a union of trapezoids. An example of such a union is depicted in Fig. 3. It consists of 5 trapezoids with their vertices on the quadratic constraint approximations and with their parallel axes aligned along the β_2 axis. (Actually there are 3 trapezoids plus 2 triangles at the ends – a triangle is a trapezoid with one parallel edge that has zero length.) This union of trapezoids model amounts to approximating the $\beta_{2min}(\beta_1)$ and $\beta_{2max}(\beta_1)$ functions as being piecewise linear in β_1 with a common set of β_1 nodes defining their respective linear domains. The 5 trapezoids shown in Fig. 3 give a somewhat coarse approximation of S. The use of more trapezoids with smaller β_1 dimensions can improve the multi-trapezoid approximation of S to any desired accuracy.

The second approximation is to model each of the *erf*(*) terms from the inner β_2 integral using a Taylor series in β_1. A unique approximation is used for each trapezoid. The Taylor series expansion is performed about the β_1 midpoint of the corresponding trapezoid. This last approximation, when coupled with the multi-trapezoid approximation of S, enables analytic evaluation of the outer β_1 integrals. The number of terms in the β_1 Taylor series for each *erf*(*) is user selectable. For the present paper, all terms in the Taylor series out to 5$^{\text{th}}$ order have been retained. If a very fine β_1 grid is used to break S up into many trapezoids, then it should be possible to use a lower order Taylor series approximation.

The resulting analytic formals for $1/\tilde{C}$ and $[\mathbf{g}_w(\mathbf{p})]_k$ in Eqs. (45) and (46) are very complicated. The Newton's method solution algorithm for the Gaussian mixture fit optimization problem in Eq. (39) also requires first and second partial derivatives of each $[\mathbf{g}_w(\mathbf{p})]_k$ with respect to the relevant elements of $\mathbf{p}$. Although complicated, it is possible to derive suitable partial derivative formulas using the chain rule, formulas that are practical to evaluate within a numerical optimization routine's software.

A 2-Dimensional Gaussian Mixture Fit Example

The Gaussian mixture fit optimization problem in Eq. (39) has been solved to fit the example probability density function shown in Fig. 2. One of the fits used K = 1 mixands, and the other fit used K = 13 mixands. The 1-mixand fit is shown in Fig. 4, and the 13-component fit is shown in Fig. 5. Both figures show the original distribution in their left-hand panels for comparison purposes, the one from Fig. 2. Their corresponding Gaussian mixture approximations are shown in their right-hand panels. Each fit involves 6 parameters per Gaussian mixand and a total of 1 equality constraint

on the sum of the mixand weights. Therefore, the fit in Fig. 4 involves 6 - 1 = 5 optimized free parameters, and the fit in Fig. 5 involves 13x6 - 1 = 77 optimized free parameters.

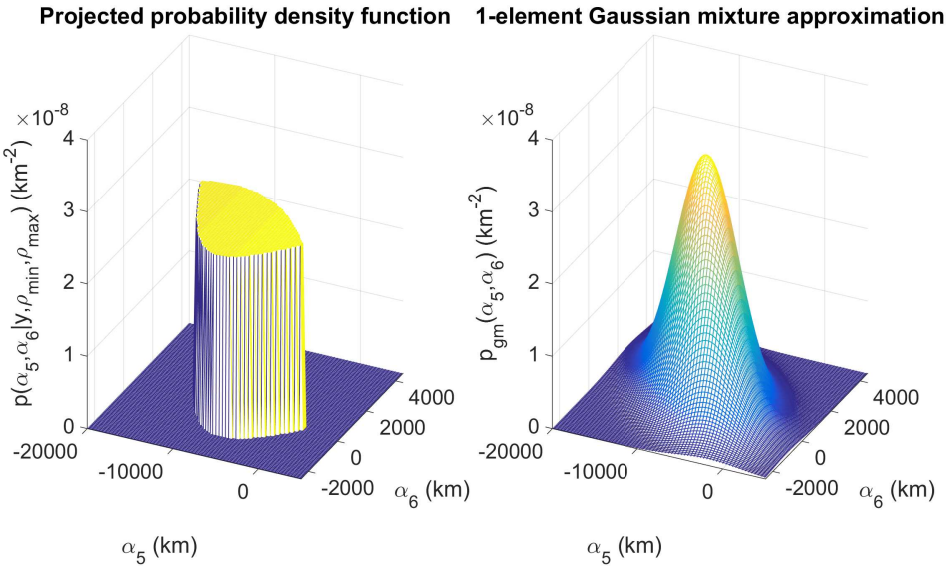

Figure 4. Optimal approximation of a 2-dimensional distribution by a 1-component Gaussian mixture.

A comparison of these two figures shows the superiority of the 13-mixand fit to the 1-mixand fit. The flat top and vertical sides of the true distribution are much better approximated by the 13-mixand fit in Fig. 5 than by the 1-mixand fit in Fig. 4. The cost function fit error from Eq. (39) tells a similar tale. The 13-mixand optimized cost equals 19% the optimized 1-mixand cost. The 13-mixand cost is only 4% of the cost value that results from setting $\mathbf{w} = 0$ in the $J_{gm}(\mathbf{w}, \mathbf{p})$ formula of Eq. (39). Thus, the functional 2-norm of the error between $p_{gm}(\alpha_5, \alpha_6)$ and $p(\alpha_5, \alpha_6|\mathbf{y}, \rho_{min}, \rho_{max})$ is only a factor of $\sqrt{0.04} = 0.2$ as large as the functional 2-norm of the original distribution $p(\alpha_5, \alpha_6|\mathbf{y}, \rho_{min}, \rho_{max})$.

The fit of the 13-component Gaussian mixand is further illustrated by Fig. 6. This figure shows a plan view of the admissible region in (α_5,α_6) space. The mixands are indicated by their 1-σ contours (the dotted green ellipses). The boundaries of the admissible region are the solid red and blue curves. There are no "holes" between the mixands because the mixands do not immediately vanish to zero outside their 1-σ contours. This figure shows the efficient arrangement of the 13 mixands. The mixands towards the middle have large variances in both dimensions. The mixands near the boundaries of the admissible region are narrow in their dimensions that are perpendicular to the boundary, which is necessary in order to develop a good approximation of the steep probability density fall-off at the boundary. Different from the mixtures of Ref. [3], there is no way to decompose this mixture along grid lines so that it can be represented as a product of one-dimensional mixtures. This new arrangement enables the current method to achieve accurate approximation of the original distribution with many fewer mixands, 10's of mixands rather than 100's.

This paper has made no attempt to analyze the efficacy of the 13 mixands of Figs. 5 and 6 as the initial probability density function for a Gaussian mixture OD Kalman filter. It is possible, perhaps likely, that further splitting of its mixands would be needed in order to approximate the dynamic propagation of these mixands accurately in the face of orbital dynamics nonlinearities. That question is left for a future study, one that anticipates using the techniques of Ref. [7] in order

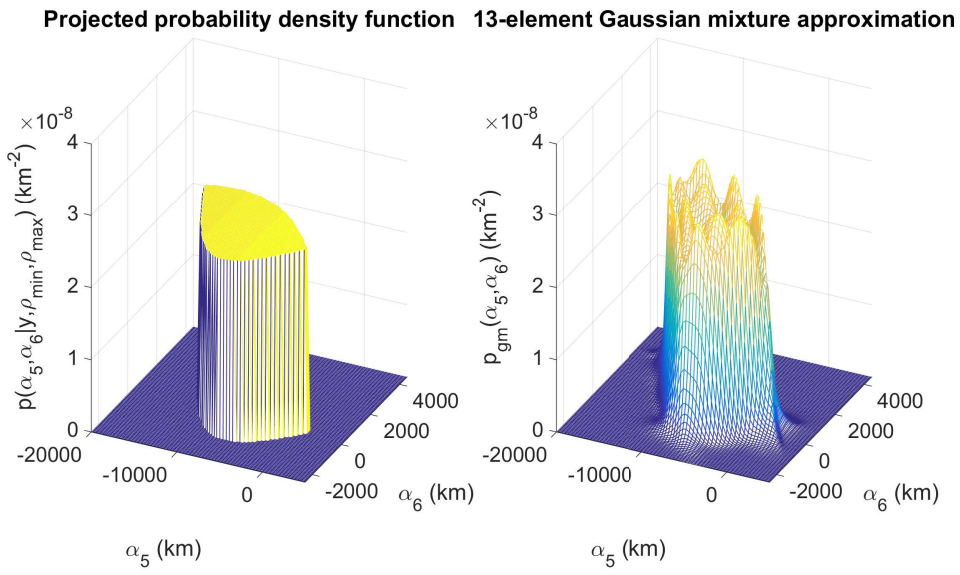

Figure 5. A 13-component optimal Gaussian mixture approximation of a 2-dimensional distribution.

to efficiently plan the needed further splits of the original 13 mixands.

Solution of the optimization problem in Eq. (39) requires a first guess to seed Newton's method. In the case of the 1-mixand problem, seeding of the first guess is easy. The guessed mixand is placed near the center of the admissible region, and its square-root information matrix is sized to give a 1-σ contour that nearly covers the admissible region. Development of a sensible first guess for the 13-mixand case is more difficult. It starts with the solution to the 1-mixand case. The optimal result for that case is used to develop a first guess for a K = 7 mixand case by distributing 6 smaller initial guess mixands around the initial 1-mixand optimal solution. The optimal 7-mixand result is then used to develop a first guess for the K = 13 mixand case. The 6 additional mixands are designed "by-hand" by placing them in "holes" in a figure like Fig. 6 that applies for the optimal 7-mixand solution. Future work on this algorithm will seek to automate this process in a way that produces better first guesses for higher numbers of mixands.

IOD GAUSSIAN MIXTURE IN POSITION/VELOCITY STATE SPACE

A Gaussian mixture in the original $\mathbf{x}$ position/velocity state space can be constructed by starting from the 2-dimensional transformed and projected Gaussian mixture in Eq. (33). First, one calculates the mixture's optimal-fit weights, mean values, and square-root information matrices by solving the problem in Eq. (39). Next, one forms an approximation of $p(\alpha_1, \ldots, \alpha_6|\mathbf{y}, \rho_{min}, \rho_{max})$ from Eq. (28) by multiplying the optimized (α_5,α_6) mixture by the Gaussian $p(\alpha_1, \alpha_2, \alpha_3, \alpha_4|\mathbf{y})$ from Eq. (30). Finally, one performs a probability density coordinate transformation from [α_1, α_2, α_3, α_4, α_5, α_6]$^{\mathrm{T}}$ coordinates to $\mathbf{x}$ coordinates using the transformation in Eq. (25).

The resulting Gaussian mixture IOD conditional probability density function takes the form:

$$p(\mathbf{x}|\mathbf{y}, \rho_{min}, \rho_{max}) = \sum_{k=1}^{K} w_k \mathcal{N}_{sr6}(\mathbf{x}; \boldsymbol{\mu}_{xk}, R_{xk}) \tag{48}$$

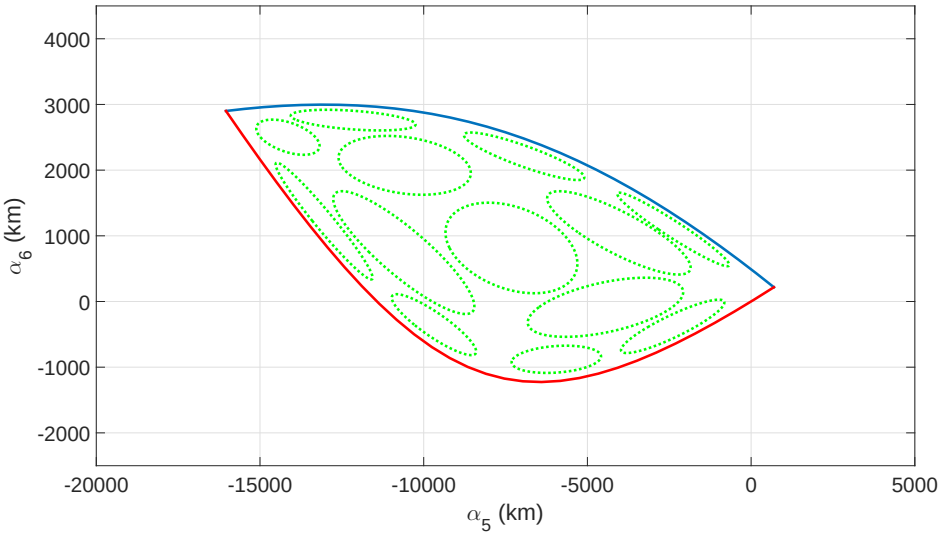

Figure 6. Arrangement of 13 components of an optimal Gaussian mixture approximation in (α_5,α_6) plane.

where the 6-dimensional Gaussian distributions in square-root-information form are:

$$\mathcal{N}_{sr6}(\mathbf{x};\boldsymbol{\mu}_{xk},R_{xk}) = \frac{|\det(R_{xk})|}{(2\pi)^3}\exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_{xk})^T R_{xk}^T R_{xk}(\mathbf{x}-\boldsymbol{\mu}_{xk})\} \tag{49}$$

The mixand means are

$$\boldsymbol{\mu}_{xk} = \hat{\mathbf{x}}_{cmle} + \Delta X \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \hat{\alpha}_4 \\ \boldsymbol{\mu}_{\alpha k} \end{bmatrix} \quad \text{for } k = 1,\ldots,K \tag{50}$$

and the mixand square-root information matrices are

$$R_{xk} = \begin{bmatrix} (1/\tilde{\sigma}_1) & 0 & 0 & 0 & 0 \\ 0 & (1/\tilde{\sigma}_2) & 0 & 0 & 0 \\ 0 & 0 & (1/\tilde{\sigma}_3) & 0 & 0 \\ 0 & 0 & 0 & (1/\tilde{\sigma}_4) & 0 \\ 0 & 0 & 0 & 0 & R_{\alpha k} \end{bmatrix} (\Delta X)^{-1} \quad \text{for } k = 1,\ldots,K \tag{51}$$

This position/velocity IOD Gaussian mixture is in a form suitable for initialization of a Gaussian mixture OD filter, perhaps an OD version of the general Gaussian mixture algorithm described in Ref. [6].

SUMMARY AND CONCLUSIONS

This paper has developed a new means of approximating an initial orbit determination conditional state probability density function based solely on bearing measurements and a known admissible region. It defines its admissible region as the set of position/velocity states that respect a lower

bound on the Keplerian orbital periapsis and an upper bound on the corresponding apoapsis. The method starts by computing a rough orbit estimate using a batch least-squares fit to the bearing measurements that includes soft constraint terms which penalize violations of the admissible region constraints. Special nonlinear optimization techniques are employed to ensure that this initial calculation determines the true global minimum of its optimal estimation problem. This rough initial estimate provides a starting point for developing a refined probabilistic characterization of the state. This refined characterization is based on a projection of the admissible region onto the maximally uncertain 2-dimensional subspace of position/velocity space. The relative uncertainties of subspaces are assessed based only on the information in the bearing measurements. For a 1-station observation and a very short data arc, this uncertain subspace translates roughly into range and range-rate. The bearings-only likelihood function in this subspace is modified by re-setting its probability to zero outside the admissible region projection. Next, a 2-dimensional Gaussian mixture is optimally fit to the modified likelihood function. Finally, a Gaussian mixture in the full position/velocity space is developed by forming the product of this 2-dimensional Gaussian mixture with a projection of the bearings-only Gaussian likelihood function onto the complimentary least-uncertain 4-dimensional subspace.

This paper's Gaussian mixture initial orbit determination algorithm has been encoded and tested on an example problem. In one test case, a 13-mixand probability distribution forms a good approximation of the initial admissible-region-based distribution. The Gaussian mixture comes close to the mesa-like structure of the true admissible-region-constrained distribution: nearly vertical sides and a gently sloping top.

Future efforts on this algorithm are planned. One effort will refine the Gaussian mixture approximation algorithm's method for generating first guess mixtures. Another effort will evaluate the efficacy of the generated mixtures for seeding a Gaussian mixture orbit determination filter.

ACKNOWLEDGMENT

Mark Psiaki's work on this project has been supported by the National Research Council through a Senior Research Associate appointment at the Air Force Research Lab Space Vehicles Directorate, Kirtland AFB, Albuquerque, NM.

REFERENCES

[1] A. Schaeperkoetter, "A Comprehensive Comparison between Angles-Only Initial Orbit Determination Techniques," m.s. thesis, Aerospace Engineering, Texas A & M, Dec. 2011.

[2] R. Weisman and M. Jah, "Uncertainty Quantification for Angles-Only Initial Orbit Determination," *Proceedings of the AIAA/AAS Astrodynamics Specialist Conf.*, No. AAS 14-434, San Diego, CA, AIAA, Aug. 2014.

[3] K. DeMars and M. Jah, "Probabilistic Initial Orbit Determination Using Gaussian Mixture Models," *Journal of Guidance, Control, and Dynamics*, Vol. 36, Sept.-Oct. 2013, pp. 1324–1335.

[4] A. Milani, G. Gronchi, M. Vitturi, and Z. Knežević, "Orbit Determination with Very Short Arcs: I. Admissible Regions," *Celestial Mechanics and Dynamical Astronomy*, Vol. 90, No. 1, 2004, pp. 57–85.

[5] G. Tommei, A. Milani, and A. Rossi, "Orbit Determination of Space Debris: Admissible Regions," *Celestial Mechanics and Dynamical Astronomy*, Vol. 97, No. 4, 2007, pp. 289–304.

[6] M. Psiaki, "The 'Blob' Filter: Gaussian Mixture Nonlinear Filtering with Re-Sampling for Mixand Narrowing," *Proceedings of the IEEE/ION PLANS 2014 Conf.*, Monterey, CA, ION, May 2014, pp. 393–406.

[7] M. Psiaki, J. Schoenberg, and I. Miller, "Gaussian Sum Reapproximation for Use in a Nonlinear Filter," *Journal of Guidance, Control, and Dynamics*, Vol. 38, Feb. 2015, pp. 292–303.

[8] K. DeMars, R. Bishop, and M. Jah, "Entropy-Based Approach for Uncertainty Propagation of Nonlinear Dynamical Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 36, July-Aug. 2013, pp. 1047–1057.

[9] R. Broucke and P. Cefola, "On the Equinoctial Orbit Elements," *Celestial Mechanics*, Vol. 5, 1972, pp. 303–310.

[10] P. Gill, W. Murray, and M. Wright, *Practical Optimization*. New York: Academic Press, 1981.

[11] G. Bierman, *Factorization Methods for Discrete Sequential Estimation*. New York: Academic Press, 1977.

[12] J. Williams and P. Maybeck, "Cost-Function-Based Gaussian Mixture Reduction for Target Tracking," *Proceedings of the 6th International Conference on Information Fusion*, IEEE, 2003, pp. 1047–1054.