Finite Horizon Maximum Likelihood Estimation for Integrated Navigation with RF Beacon Measurements*

Sharad Shankar Kenan Ezal João P. Hespanha

Abstract

We develop a Finite Horizon Maximum Likelihood Estimator (FHMLE) that fuses Inertial Measurement Unit (IMU) and radio frequency (RF) measurements over a sliding window of finite length for three-dimensional navigation. Available RF data includes pseudo–ranges, angles of transmission (AoT), and Doppler shift measurements. The navigation estimates are obtained by solving a finite-dimensional nonlinear optimization using a primal-dual interior point algorithm (PDIP). The benefits of the proposed estimation method are highlighted using simulations results comparing the FHMLE approach with an Unscented Kalman Filter (UKF), in a scenario where an aircraft approaches a carrier, with RF measurements from beacons aboard the carrier, and low-cost IMU measurements aboard the aircraft. When the Geometric Dilution of Precision is large, we found that the FHMLE is able to achieve smaller estimation errors than the UKF, which tends to carry a bias throughout the trajectory.

Keywords: Navigation, Maximum likelihood estimation, Moving horizon estimation, Kalman filtering, GNSS.

1 Introduction

Integrated navigation here refers to the estimation of the position and orientation of a moving rigid body, based on a fusion of measurements collected from radio frequency (RF) beacons and an Inertial Measurement Unit (IMU). The RF beacons provide timing information from which we extract pseudo-range measurements (i.e., ranges up to an unknown constant due a clock offset), angle of transmission (AoT) measurements constructed by processing the signals transmitted/received by an antenna array, and velocity measurements obtained from the Doppler shift on the RF carrier frequency.

The fusion of pseudo-range with IMU measurements is a well-studied problem for Global Navigation Satellite Systems (GNSS) [4], in which the satellites play the role of the RF beacons. However, our setup presents a few key differences: First, we are interested in scenarios where the RF beacons are localized to a relatively small region of space, which would lead to very poor Geometric Dilution of Precision (GDoP) [4] in a GNSS setting. Second, we are also interested in scenarios where the number of beacons is small, in particular, fewer than the minimum of four beacons required for a system operating just with pseudo-range measurements and asynchronous clocks. Third, in addition to the usual pseudo-range measurements, we also have angle of transmission measurements,

*This material is based upon work partially supported by the National Science Foundation under Grant No. ECCS-1608880.

which as we will show, can compensate for the lack of beacons and poor GDoP. The use of angle of transmission or angle of arrival measurements in integrated navigation is discussed in [13], while localization using Doppler measurements is discussed in [2]. This work is especially relevant to GPS-denied scenarios, where only a small number of ground-based beacons are available, for example when navigating using Signals of Opportunity [18].

In this paper, we construct a maximum-likelihood estimator (MLE) for the position and orientation of a moving rigid body that fuses all available measurements (RF and IMU) over a finite window of time with constant length T, and ending at the current time t_0. Considering a finite window of time (and consequently a finite number of measurements) and finitely parameterizing the position/orientation over the time window $(t_0 - T, t_0]$, enables us to compute the maximum likelihood estimate numerically using a nonlinear programming solver. In this work, we use a primal-dual interior-point (PDIP) method that has been optimized for real-time computations [11].

A UKF approach for navigation in the GNSS setting appears in [26] and an attitude observer and EKF combination is considered in [6]. Our approach is in contrast to this type of Kalman filtering based approaches that recursively compute estimates and require approximations to nonlinear dynamics and non-Gaussian noise. Extended Kalman Filters (EKFs) locally linearize nonlinear dynamics and measurement equations, while the Unscented Kalman Filter (UKF) assumes that the prior and posterior distributions are adequately approximated by a finite set of "sigma points" [14]. A key benefit of the approach proposed here is that we need not assume that the posterior distribution of the state estimate is Gaussian, which is an important assumption in the derivations of both the EKF and UKF. Nonlinear measurements or dynamics often result in non-Gaussian posterior distributions, which can cause Kalman filter-based methods to diverge or converge to sub-optimal solutions, depending on initialization and measurement history [7, 23]. However, in a FHMLE approach, we are able to re-initialize independently at each step, as well as compute solutions for multiple initializations. This prevents the filter from getting "stuck" on a suboptimal trajectory. Furthermore, EKFs and UKFs are designed for additive Gaussian noises, however such an assumption on the AoT measurements is problematic because an error in direction should be constrained to the unit-sphere. There are several well-studied unit-sphere distributions from the field of Directional Statistics, including the von Mises-Fisher and Kent distributions.

Finite horizon MLE techniques also differ from Kalman filtering-based approaches in that the former discards old measurements. While this appears to be a disadvantage at first, it also means that the effect of outliers or poor initializations is quickly attenuated. In the context of our problem, we will see that the use of a finite horizon helps the system rapidly recover from the effect of measurements taken over unfavorable geometries (i.e. all beacons localized to a small region of space). Our finite horizon MLE is heavily inspired by the literature on Moving Horizon Estimation (MHE) [24], which also considers a finite window of measurements, but typically also includes a penalty term to account to the "missing" measurements. A sliding window approach to GNSS/INS integration appeared in [9]. The effect of changing the length of the sliding window is considered in [28].

One final benefit of an approach that reduces state estimation to a numerical optimization lies in the ability to incorporate constraints into the design of the estimator. These constraints can be used, e.g., to restrict estimates to known bounds. For example, we can add constraints on bias terms for IMU sensors or on the position/orientation/velocity of the rigid body. Such constraints add additional information that can greatly improve the estimation accuracy. We note that equality constrained versions of both EKFs and UKFs also exist [24], but they rely on projections, and suffer from many of the challenges mentioned above.

To illustrate the potential of our approach we present simulation results obtained from a scenario in which an aerial vehicle lands on an aircraft carrier and estimates its positions using an affordable IMU unit and measurements from RF beacon on board of the aircraft. The vehicle starts a few kilometers away from the aircraft, which presents a very unfavorable geometry [4] since all RF beacons are localized very close to each other (in comparison to the distance between aircraft and beacons). Our results show that the FHMLE is especially attractive to estimate the parameters mostly affected by this challenging geometry, consistently leading to smaller estimation error in regions with poor geometry.

This paper is organized as follows: In Section 2, we introduce the measurement models and likelihood functions to formulate the FHMLE optimization. The conversion of the maximum likelihood estimation to a numerical optimization is discussed in Section 3, together with a brief description of the nonlinear programming solver that we use. In Section 4, we review the construction of an Unscented Kalman Filter that could also be used to solve the problem considered here. Finally, Section 5 presents simulation results for an integrated aerial navigation system based on RF beacons aboard an aircraft carrier. We use these simulations to compare the performance of the FHMLE and the UKF.

2 Finite Horizon MLE for Navigation

Consider a coordinate frame B attached to a rigid body that moves with respect to an inertial frame C and denote by $(p, R) \in \mathrm{SE}(3)$ the instantaneous position and orientation of the frame B with respect to C. Our goal is to estimate $p(t) \in \mathbb{R}^3$ and $R(t) \in \mathrm{SO}(3)$ over an interval $t \in (t_0 - T, t_0]$ based on a set of measurements $Y_{(t_0-T,t_0]}$ taken by sensors attached to B during the interval $(t_0 - T, t_0]$. Here, $SE(3)$ denotes the special Euclidean group and $SO(3)$ the special orthogonal group in three dimensions [19].

We denote by $\ell\big(Y_{(t_0-T,t_0]} \mid p, R\big)$ the likelihood of the set of measurements $Y_{(t_0-T,t_0]}$ given the time evolutions p and R of the position and orientation of B with respect to C, respectively. This notation enables us to express the maximum-likelihood estimator (MLE) as

$$(\hat{p}, \hat{R}) \coloneqq \arg\max_{p \in \mathcal{P}, R \in \mathcal{R}} \ell\big(Y_{(t_0-T,t_0]} \mid p, R\big),$$

where $\mathcal{P}$ and $\mathcal{R}$ denote "admissible" sets for the functions $p(t)$ and $R(t)$, $t \in (t_0 - T, t_0]$ that define the position and orientation of B with respect to C, respectively. For p and R, to be "admissible", they must be compatible with the dynamics of the rigid body, with the understanding that in order to solve (8) numerically, the sets $\mathcal{P}, \mathcal{R}$ must be parameterized using a finite number of parameters. We defer further discussion on the parameterization of these sets to Section 3.1 and focus the remainder of this section on the construction of the likelihood function that appears in (8).

2.1 Pseudorange Measurements

The pseudorange measurement $\rho_k(t)$ associated with the reception of an RF message from beacon k at time t is typically defined as

$$\rho_k(t) \coloneqq c(t - t^{\mathrm{sent}}),$$

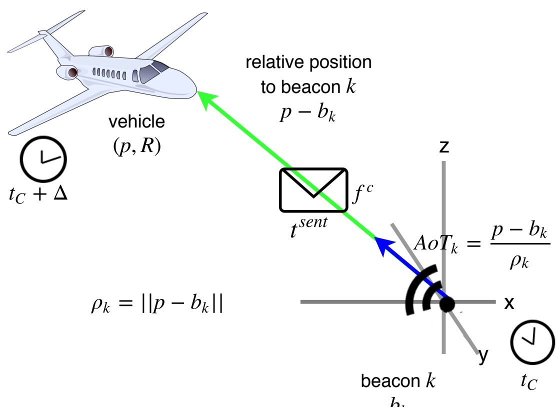


Figure 1: RF measurements: pseudorange, angle of transmission, Doppler shift.

where t^{sent} denotes the time at which the message was sent and c is the RF propagation speed. Let b_k be the position in the coordinate frame C of the beacon that transmitted the message. Assuming that the receive antenna is at the origin of the coordinate frame B, under ideal conditions we would have

$$\rho_k(t) = \|p(t) - b_k\|,$$

where $\|\cdot\|$ denote the Euclidean norm of a vector (see Figure 1). However, the send time t^{sent} and the receive time t are measured with respect to two clocks that are not perfectly synchronized. In addition, the RF detection circuit introduces a stochastic error in the measurement of the receive time. Denoting by $\Delta(t)$ the offset between the transmitter and receiver clocks and assuming a zero-mean Normal distribution with variance σ_ρ^2 for the stochastic component of the error, the pseudorange measurement $\rho_k(t)$ is a Normal random variable with mean $\|p - b(t))\| + c\Delta(t)$ and variance σ_ρ^2, leading to the following likelihood function:

$$\ell\big(\rho_k(t)|p(t), \Delta(t)\big) = \frac{\exp\left(-\frac{\left(\rho_k(t) - \|p(t) - b_k\| + c\Delta(t)\right)^2}{2\sigma_\rho^2}\right)}{\sqrt{2\pi}\sigma_\rho}. \tag{1}$$

2.2 Angle of Transmission Measurements

The angle of transmission measurement $AoT_k(t)$ associated with the reception of an RF message from beacon k at time t is defined by the unit vector pointing from the beacon to the receive antenna. Under ideal conditions we would have

$$AoT_k(t) = \frac{p(t) - b_k}{\|p(t) - b_k\|} \tag{2}$$

(see Figure 1), but RF detection errors will also introduce a stochastic error in this measurement. We assume that the measurement follows a von Mises-Fisher unimodal isotropic distribution on the unit sphere in $\mathbb{R}^3$, with mean direction given by (2) and concentration parameter $\kappa > 0$ [1], which leads to the following likelihood function:

$$\ell\big(AoT_k(t)|p(t)\big) = \frac{\kappa \exp\left(\kappa \frac{(p(t) - b_k)' AoT_k(t)}{\|p(t) - b_k\|}\right)}{2\pi(e^\kappa - e^{-\kappa})}. \tag{3}$$

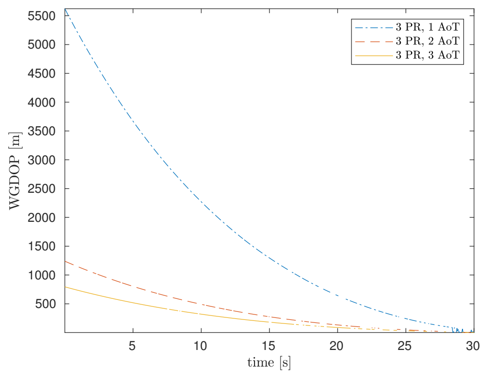


Figure 2: Weighted GDOP for various measurements

Note that a high value for the concentration parameter $\kappa > 0$ of a von Mises-Fisher distribution corresponds to a distribution tightly clustered around the mean, and consequently less noise on AoT_k, whereas the limiting case $\kappa \to 0^+$ corresponds to the uniform distribution. For non-isotropic distributions of the AoT error, one could base the likelihood on the Kent distribution instead [15].

Figure 2 shows the Weighted Geometric Dilution of Precision (WGDOP) for various sets of measurement types over the sample trajectory and beacon positions used in the simulations of Section 5. WGDOP is an approximation of the position and clock offset estimation error that is obtained from the inverse of the Fisher Information matrix of the weighted least squares problem with linearized measurements [22]. Note that GDOP is infinite with only three pseudorange measurements and no AoT. Adding AoT measurements significantly improves WGDOP, with diminishing returns between two and three angle measurements.

2.3 Doppler Shift Measurements

The relative motion between transmitter and receiver antennas creates a Doppler shift

$$\Delta f_k(t) \coloneqq f_k^{rec}(t) - f_k^c$$

between the carrier frequency $f_k^{rec}(t)$, measured by the RF detector at time t, and the original carrier frequency f_k^c transmitted by beacon k. Under ideal conditions we would have

$$\Delta f_k(t) = -\frac{f_k^c}{c}\left(\frac{d}{dt}\|p(t) - b_k\|\right) = -\frac{f_k^c}{c}\frac{(p(t) - b_k)'\dot{p}(t)}{\|p(t) - b_k\|},$$

but due to measurement errors we assume that the perceived Doppler shift is a zero-mean Normal distribution with variance σ_f^2, leading to the following likelihood function:

$$\ell\big(\Delta f_k(t)|\dot{p}(t)\big) = \frac{\exp\left(-\frac{\left(\Delta f_k(t) + \frac{f_k^c}{c}\frac{(p(t)-b_k)'\dot{p}(t)}{\|p(t)-b_k\|}\right)^2}{2\sigma_f^2}\right)}{\sqrt{2\pi}\sigma_f}. \tag{4}$$

2.4 IMU Measurements

An accelerometer attached to the frame B produces a measurement $a_{\mathrm{m}}(t)$ of the proper acceleration expressed in B, which is given by

$$a_{\mathrm{m}}(t) \coloneqq R(t)(\ddot{p}(t) - g) + a_{\mathrm{b}}(t),$$

where g denotes the local gravity vector expressed in the inertial frame C, for example $g = \begin{bmatrix} 0 & 0 & 9.8065 \end{bmatrix}$ in the North-West-Up (NWU) inertial navigation frame. Additionally, $a_{\mathrm{b}}(t)$ is an additive bias introduced by the sensor. For simplicity, we assume that the accelerometer is calibrated to remove the effect of other deterministic errors such as scale factor and non-orthogonality errors [25], and that this measurement is further corrupted by additive zero-mean Gaussian noise with covariance matrix Σ_a, so that we obtain the following likelihood function

$$\ell\big(a_{\mathrm{m}}(t)|\ddot{p}(t), R(t)\big) = \frac{\exp\Big(-\frac{1}{2}\left\|a_{\mathrm{m}}(t) - R(t)(\ddot{p}(t) + g) - a_{\mathrm{b}}(t)\right\|^2_{\Sigma_a^{-1}}\Big)}{\sqrt{2\pi^3 \det \Sigma_a}} \tag{5}$$

where $\|x\|^2_\Lambda$ denotes the quadratic form $x'\Lambda x$.

A rate-gyro attached to B produces a measurement $\omega_{\mathrm{m}}(t)$ of the angular velocity expressed in B, given by

$$\omega_{\mathrm{m}}(t) \coloneqq R(t)\omega(t) + \omega_{\mathrm{b}}(t)$$

where the entries of the angular velocity vector $\omega = \left[\begin{smallmatrix} \omega_x & \omega_y & \omega_z \end{smallmatrix}\right]$ are can obtained from

$$\dot{R}(t)R'(t) = \left[\begin{smallmatrix} 0 & -\omega_z(t) & \omega_y(t) \\ \omega_z(t) & 0 & -\omega_x(t) \\ -\omega_y(t) & \omega_x(t) & 0 \end{smallmatrix}\right]$$

and $\omega_{\mathrm{b}}(t)$ denotes an additive bias introduced by the sensor. We similarly assume that the gyroscope is calibrated to remove other deterministic errors, and is corrupted by additive zero-mean Gaussian noise with covariance matrix Σ_ω, and we obtain the following likelihood function:

$$\ell\big(\omega_{\mathrm{m}}(t)|R(t), \dot{R}(t)\big) = \frac{\exp\Big(-\frac{1}{2}\left\|\omega_{\mathrm{m}}(t) - R(t)\omega(t) - \omega_{\mathrm{b}}(t)\right\|^2_{\Sigma_\omega^{-1}}\Big)}{\sqrt{2\pi^3 \det \Sigma_\omega}}. \tag{6}$$

A magnetometer attached to B produces a measurement $m_{\mathrm{m}}(t)$ of the local magnetic field expressed in B, given by

$$m_{\mathrm{m}}(t) \coloneqq R(t)m$$

where m denotes the direction/intensity of the local magnetic north expressed in the inertial frame C. Assuming that this measurement is corrupted by additive zero-mean Gaussian noise with covariance matrix Σ_m, we obtain the following likelihood function:

$$\ell\big(m_{\mathrm{m}}(t)|R(t)\big) = \frac{\exp\Big(-\frac{1}{2}\left\|m_{\mathrm{m}}(t) - R(t)m\right\|^2_{\Sigma_m^{-1}}\Big)}{\sqrt{2\pi^3 \det \Sigma_m}}. \tag{7}$$

For simplicity, we assume here that the magnetometer has been calibrated to remove the effect of any soft and hard iron distortions.

2.5 Total Likelihood Function

While our main goal is to estimate the position $p(t)$ and orientation $R(t)$ of the rigid body attached to the coordinate frame B, the likelihood of the sensor measurements that we have introduced above depend on three other quantities that need to be estimated in addition to $p(t)$ and $R(t)$: the offset $\Delta(t) \in \mathbb{R}$ between the clock of the beacons and the clock of the RF receiver, the accelerometer bias $a_{\mathrm{b}}(t) \in \mathbb{R}^3$, and the rate-gyro bias $\omega_{\mathrm{b}}(t) \in \mathbb{R}^3$. In light of this, our maximum-likelihood estimator (MLE) is actually of the form

$$(\hat{p}, \hat{R}, \hat{\Delta}, \hat{a}_{\mathrm{b}}, \hat{\omega}_{\mathrm{b}}) \coloneqq \underset{\substack{p\in\mathcal{P}, R\in\mathcal{R}, \Delta\in\mathcal{D}, \\ a_{\mathrm{b}}\in\mathcal{A}_{\mathrm{b}}, \\ \omega_{\mathrm{b}}\in\Omega_{\mathrm{b}}}}{\arg\max} \ \ell\big(Y_{(t_0-T,t_0]} \mid p, R, \Delta, a_{\mathrm{b}}, \omega_{\mathrm{b}}\big), \tag{8}$$

where $\mathcal{D}$, $\mathcal{A}_{\mathrm{b}}$, and Ω_{b} denote admissible sets for the functions $\Delta(t)$, $a_{\mathrm{b}}(t)$, and $\omega_{\mathrm{b}}(t)$.

Assuming that the random variables associated with the likelihood functions given by (1), (3), (4), (5), (6), and (7) are all conditionally independent, and given the trajectories $p \in \mathcal{P}, R \in \mathcal{R}, \Delta \in \mathcal{D}, a_{\mathrm{b}} \in \mathcal{A}_{\mathrm{b}}$, and $\omega_{\mathrm{b}} \in \Omega_{\mathrm{b}}$, the likelihood function $\ell\big(Y_{(t_0-T,t_0]} \mid p, R, \Delta, a_{\mathrm{b}}, \omega_{\mathrm{b}}\big)$ in (8) is obtained as the product of the pdfs in (1), (3), (4), (5), (6), (7) for each of the measurements available in $Y_{(t_0-T,t_0]}$ taken by sensors in the interval $(t_0 - T, t_0]$.

Since each of the pdfs in (1), (3), (4), (5), (6), (7) is from the exponential family, rather than performing the maximization in (8), we instead solve the following optimization

$$(\hat{p}, \hat{R}, \hat{\Delta}, \hat{a}_{\mathrm{b}}, \hat{\omega}_{\mathrm{b}}) \coloneqq \underset{\substack{p\in\mathcal{P}, R\in\mathcal{R}, \Delta\in\mathcal{D}, \\ a_{\mathrm{b}}\in\mathcal{A}_{\mathrm{b}}, \\ \omega_{\mathrm{b}}\in\Omega_{\mathrm{b}}}}{\arg\min} \ -\log \ell\big(Y_{(t_0-T,t_0]} \mid p, R, \Delta, a_{\mathrm{b}}, \omega_{\mathrm{b}}\big), \tag{9}$$

which has the same set of minima $\{p, R, \Delta, a_{\mathrm{b}}, \omega_{\mathrm{b}}\}$.

3 Numerical Optimization

This section addresses two key aspects of involved in solving (9): first, the parameterization of the sets of admissible functions, and second, the numerical method used to solve the resulting optimization.

3.1 Trajectory Parameterization

To solve (9) using a numerical solver, we need to use finite parameterizations for the sets of admissible functions $\mathcal{P}, \mathcal{R}, \mathcal{D}, \mathcal{A}_{\mathrm{b}}, \Omega_{\mathrm{b}}$.

For temperature compensated IMU sensors, the accelerometer and rate-gyro biases drift very slowly so we can typically neglect changes in the biases over time intervals of length T on the order of tens of seconds [20]. This corresponds to sets $\mathcal{A}_{\mathrm{b}}, \Omega_{\mathrm{b}}$ consisting of signals in $\mathbb{R}^3$ that are constant over the whole interval $(t_0 - T, t_0]$ and can be parameterized, for example by

$$\mathcal{A}_{\mathrm{b}} \coloneqq \big\{a \in \mathbb{R}^3 : \|a\| \leqslant L_a\big\}, \qquad \Omega_{\mathrm{b}} \coloneqq \big\{\omega \in \mathbb{R}^3 : \|\omega\| \leqslant L_\omega\big\},$$

where L_a and L_ω denote maximum norms for the accelerometer and biases, typically obtained from the manufacturers datasheets.

Temperature-compensated clock circuits typically exhibit a very small drift over time intervals of tens of seconds [17]. However, pseudoranges are extremely sensitive to such errors (recall that, at light speed, 1 nano second corresponds to about 0.3 meters), so we consider a richer class of admissible functions for the clock offset $\mathcal{D}$. Specifically, we allow the set $\mathcal{D}$ of admissible clock offsets to contain not only constant offsets but also constant clock drifts, denoted by β, over the time interval of interest. This corresponds to a set of linear functions of the form:

$$\mathcal{D} \coloneqq \big\{\beta(t-t_0)+\phi : |\beta - 1| \leqslant L_\Delta, \phi \in \mathbb{R}\big\},$$

where L_Δ denotes the maximum mismatch between the frequencies of the two clocks.

For a model-free parameterization of the sets $\mathcal{P}$ and $\mathcal{R}$, we draw inspiration from direct collocation methods for trajectory optimization [3], and parameterize admissible trajectories $p(t)$, $t \in (t_0 - T, t_0]$ with the values of the function $p(t)$ at a fixed set of N times

$$\tau_1 \coloneqq t_0 - T < \tau_2 < \cdots < \tau_{N-1} < \tau_N \coloneqq t_0, \tag{10}$$

which means that $\mathcal{P}$ is parameterized by

$$\big(p(\tau_1), p(\tau_2), \ldots, p(\tau_N)\big) \in \mathbb{R}^{3N}. \tag{11}$$

Following the collocation methods terminology, we refer to the τ_i as *knots*. While splines provide a very general and attractive option to interpolate the value of $p(t)$ and its derivatives between the knots, for numerical efficiency, the results presented here use a different and more computationally efficient approach: We take the knots, τ_i, to be the times at which we have measurements, meaning that we do not need to interpolate $p(t)$ between the knots, but we do need to compute the first and second derivatives of $p(t)$ for the likelihood functions of the Doppler and accelerometer measurements in (4) and (5), respectively. We evaluate these derivatives using the Lyness and Moler algorithm [16], which for second-order polynomials and equally spaced knots $\tau_i - \tau_{i-1} \eqqcolon h$ leads to

$$\dot{p}(\tau_i) = \frac{p(\tau_{i+1}) - p(\tau_{i-1})}{2h}, \qquad \ddot{p}(\tau_i) = \frac{p(\tau_{i+1}) - 2p(\tau_i) + p(\tau_{i-1})}{h^2}, \qquad \forall k \in \{2, \ldots, N-1\}. \tag{12}$$

Formally, this means that $\mathcal{P}$ is the linear subspace of functions from $(t_0 - T, t_0]$ to $\mathbb{R}^3$ whose first and second derivatives at each knot τ_i are equal to those of the third-order Lagrange polynomial that interpolates the function at the three knots closest to τ_i. In view of (12), this particular set $\mathcal{P}$ has the desirable feature that the values of $p(\tau_i)$ and its derivatives $\dot{p}(\tau_i)$ and $\ddot{p}(\tau_i)$ are all linear combinations of at most three of the optimization parameters in (11).

To parameterize the admissible set of orientations $\mathcal{R}$, we follow a similar approach, except that we represent rotation matrices using unit quaternions. In particular, we parameterize an admissible trajectory $R(t)$, $t \in (t_0 - T, t_0]$ by a sequence of unit quaternions,

$$\big(q(\tau_1), q(\tau_2), \ldots, q(\tau_N)\big) \in SU(2)^N$$

corresponding to the orientations at the knots in (10). We then take $\mathcal{R}$ to be the linear subspace of functions from $(t_0 - T, t_0]$ to $SU(2)^N$ whose first and second derivatives at each knot τ_i are equal to those of the 3rd order Lagrange polynomial that interpolates the function at the 3 knots closest to τ_i, which means that we can again obtain simple formulas for $\dot{q}(\tau_i)$ and $\dot{R}(\tau_i)$ using the Lyness and Moler algorithm. These derivatives are needed for the likelihood function of the rate-gyro measurements in (6).

Remark 1 (Known dynamics). The above approach to constructing the sets $\mathcal{P}$ and $\mathcal{R}$ has the advantage that it does not require knowledge of the specific rigid body dynamics. However, when such a model is known incorporating these dynamics may potentially decrease the estimation errors. To understand how this could be accomplished within the framework of this paper, suppose that the motion of the rigid body attached to the coordinate frame B follows dynamics of the following general form:

$$\dot{x} = f(x,t), \qquad p(t) = G\big(x(t)\big), \qquad R(t) = H\big(x(t)\big), \qquad x \in \mathbb{R}^n. \tag{13}$$

The sets $\mathcal{P}$ and $\mathcal{R}$ are naturally parameterized by the initial conditions to (13) as follows:

$$\mathcal{P} \coloneqq \{p_{x_0} : x_0 \in \mathbb{R}^n\}, \qquad \mathcal{R} \coloneqq \{R_{x_0} : x_0 \in \mathbb{R}^n\},$$

where

$$p_{x_0}(t) \coloneqq G(\varphi(t; x_0, t_0 - T)), \qquad R_{x_0}(t) \coloneqq H(\varphi(t; x_0, t_0 - T)), \tag{14}$$

and $\varphi(t; x_0, t_0 - T)$ denotes the solution to (13) with initial condition $x(t_0 - T) = x_0$. To follow this approach, we would not need to solve (13) explicitly as (14) seems to imply. Instead, and inspired again by direct collocation methods, we would parameterize the sets by the values of $x(\tau_i)$ at the knot points and regard (13) as optimization constraints at the knot points. $\square$

3.2 Primal-Dual Interior-Point Solver

For the parameterizations of the admissible sets discussed in Section 3.1, the optimization in (9) is of the form

$$f(\hat{x}) = \min \big\{ f(x) : F(x) \geqslant 0, G(x) = 0, x \in \mathbb{R}^{n_x} \big\}, \tag{15}$$

where $\hat{x}$ denotes the optimizer of f. We solve this optimization numerically using a second-order Primal-Dual Interior Point (PDIP) method. While this class of methods only guarantees convergence when the function $f(x)$ is convex and the constraints $F(x) \geqslant 0, G(x) = 0$ define a convex set [5], in practice PDIP methods they perform well on a much wider class of problems. The particular solver that we use is generated by the MATLAB® toolbox `TensCalc` [10, 11]. This toolbox generates specialized C-code solvers for optimization problems like (15) and is especially fast for optimizations up to a few thousands of optimization variables and constraints. Short solve times are achieved by a combination of features that include automatically detecting and exploring the sparsity structure of the Hessian matrix of $f(x)$ and the Jacobian matrices of $F(x)$ and $G(x)$ to speed up each iteration of the PDIP iteration, reusing intermediate computations within and across iterations, and the use of code that improves the efficiency of micro-processor pipelining and caching [11].

3.2.1 Solver Complexity for Sparse Problems

The key challenge with FHMLE is that, at each time step, we need to solve the optimization in (9). However, it turns that the specialized C-code solvers produced by the MATLAB® toolbox `TensCalc` [10, 11] are very efficient for the type of optimizations needed by FHMLE. Due to the independence of measurements across time and the numerical method used to handle dynamics in (12), we note that the Hessian matrix associated with the optimization in (15) is sparse because

$$\frac{\partial^2 f}{\partial(p,R)(\tau_i)\partial(p,R)(\tau_j)} = 0,$$

for all indices i and j such that $|i-j|>1$. This means that the number of nonzero entries of the Hessian increases linearly with the number of knots. Solve times for state-of-the-art methods of inverting sparse matrices scale essentially linear with the number of nonzero elements [8]. The computational complexity of a second-order PDIP method is determined by the complexity of inverting the Hessian matrix $\frac{\partial^2 f}{\partial x x'}$, therefore PDIP solve times increase linearly with the size of time window considered [5].

3.2.2 Implementation for FHMLE

One limitation of second-order PDIP methods is that the optimization criterion $f(x)$ and the constraint function $F(x)$, $G(x)$ must be twice-differentiable. This is not the case in our problem, because the likelihood functions associated with the angle of transmission in (3) and the Doppler shift in (4) contain a division by $\|p(t)-b_k\|$ that is not differentiable at the point $p(t)=b_k$. However, one can confidently exclude such points from the set of admissible $p(t)$, because this would imply that the rigid body was collocated with the beacon. Nevertheless, we have found that the convergence of the solver was improved by introducing the additional slack optimization variables

$$r_k(t) \coloneqq \frac{(p(t)-b_k)' AoT_k(t)}{\|p(t)-b_k\|}, \tag{16}$$

which enabled us to rewrite (3) as

$$f\big(AoT_k(t)|p(t)\big) = \frac{\kappa \exp(\kappa r_k(t))}{2\pi(e^{\kappa}-e^{-\kappa})} \tag{17}$$

subject to the following constraint that enforces (16):

$$\|p(t)-b_k\|\, r_k(t) = (p(t)-b_k)' AoT_k(t). \tag{18}$$

Note that both (17) and (18) are smooth everywhere, as opposed to the original likelihood in (3). A similar procedure can be applied to smooth the likelihood of the Doppler measurements in (4). While the use of slack variables increases the size of the optimization, the added numerical stability generally decreases the number of iterations for the solver, and overall results in smaller solve times.

Remark 2 (Regularization and a-priori information). Solver convergence is sometimes improved by introducing L_2-regularization terms on some variables, e.g., in the IMU biases a_b and ω_b. From a Bayesian perspective, this is equivalent to placing a Gaussian prior on those variables and simply amounts to including additional quadratic additive terms in (9). Similarly, box inequality constraints placed on optimization variables, which may also be viewed as priors, can improve solver convergence. □

4 Unscented Kalman Filter Estimator

To test the effectiveness of our approach against the prior state of the art, we used an Unscented Kalman Filter (UKF) to solve the same estimation problem. In essence, the UKF uses the unscented transform to extend Kalman Filtering to scenarios with nonlinear dynamics or measurements. The unscented transform uses a set of "sigma points" to represent the first two moments of a given distribution. In a UKF, the unscented transform takes the estimated mean $\hat{x}(t)$ and variance $\hat{P}_x(t)$ at time t, and computes sigma points $s^1(t), s^2(t), ..., s^{n+1}(t)$ each in $\mathbb{R}^n$, for n the dimension of the state x. It can be shown that a correct choice of the $n + 1$ sigma points fully describes the first two moments of a distribution [14]. These sigma points are simulated through the noisy process and measurements to obtain sigma points of the predicted state and measurement at time at a subsequent time. The UKF then performs the same update as the usual Kalman filter through the inverse unscented transform. The key benefit of an UKF is that the it is generally less sensitive to filter divergence [7] than Kalman filters extensions that rely on local linearization like the EKF [27]. The UKF in this simulation was designed with the same modeling assumptions as the FHMLE, except that the AoT measurements were modeled as independent and identical Gaussian noises in the azimuth and elevation angles of the origin of B with respect to C. The resulting simulated distribution is similar to the von Mises-Fisher distribution, especially for large values of κ, and preserves the notion of angle as belonging to the unit sphere.

5 Simulations

The FHMLE and a tightly integrated UKF were compared in a simulated scenario where an aerial vehicle approaches an aircraft carrier from the south in a North-West-Up (NWU) inertial frame. Three RF beacons were placed on board the aircraft carrier: two at the start of the landing strip (one on each side separated by 40 meters) and one on the control tower (44 meters north of the western forward beacon and 35 meters above the deck). The noise parameters for the sensors are summarized in Table 1. The values corresponding to the RF measurements (pseudorange, angle of transmission, and Doppler shift) were obtained through detailed Monte Carlo simulations of the RF signal processing that leads to these measurements and the IMU sensor parameters were obtained from InvenSense's MPU-6050 MEMS IMU datasheets [12].

A FHMLE horizon length of one second with a measurement frequency of 20 Hz leads to an optimization over 20 knots. The accelerometer bias, gyroscope bias, and clock drift are assumed constant over each window, since each changes relatively slowly to the scale of a second. This leads to an optimization in 268 variables with 140 equality constraints, corresponding to the simulation displayed in Figure 3. Despite the Hessian being 408×408, at most 9000 elements are nonzero, or around 5.4% of the matrix entries. Since the number of nonzero matrix entries scales linearly, while the size of the matrix scales quadratically, we expect the percentage of nonzero elements to decrease with horizon length. No regularization was used for the FHMLE, however box constraints were placed on the position and bias terms to improve convergence.

5.1 Filter Initialization

The FHMLE and the UKF estimators both depend on how they are initialized: At every time, the FHMLE nonlinear solver needs to be initialized with a state estimate that is the starting point for

Parameter	Symbol	Value
RF propagation speed	c	299792458 m/s
RF carrier frequency	f_k^c	5.4 GHz
pseudorange noise variance	σ_ρ	1 m
angle of transmission noise concentration	κ	$2 \cdot 10^4$
Doppler shift noise variance	σ_f	.1 Hz
gravity field	g	$\begin{bmatrix} 0 \\ 0 \\ 9.8065 \end{bmatrix}$ m/s^2
accelarometer noise covariance matrix	Σ_a	$.15\, I_{3\times3}$ [m/s^2]
rate-gyro noise covariance matrix	Σ_ω	$.05\, I_{3\times3}$ [rad/s]
local magnetic field	m	$\begin{bmatrix} 40000 \\ 0 \\ 0 \end{bmatrix}$ [nT]
magnetometer noise covariance matrix	Σ_m	$400\, I_{3\times3}$ [nT]
measurements sampling frequency	h	20Hz
FHMLE horizon length	T	1 sec

Table 1: Simulation parameters.

the PDIP iterations. The state of the UKF filter needs to be initialized at the start of an experiment (and whenever the filter diverges).

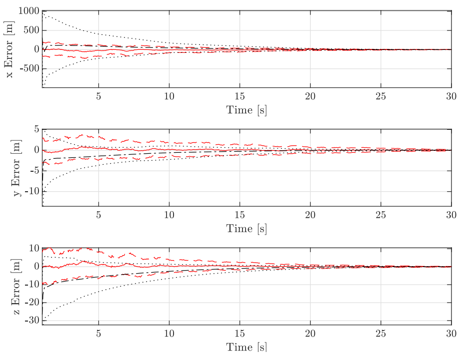
Each pair (AoT_i, AoT_j) of angle of transmission measurements, corresponding to beacons i and j, provides a position estimate $\bar{p}_{ij}$ at the point p that is closest to the corresponding AoT rays:

$$\bar{p}_{ij} \coloneqq \min_{p,\alpha_i,\alpha_j} \|p - b_i + \alpha_i AoT_i\|^2 + \|p - b_j + \alpha_j AoT_j\|^2.$$

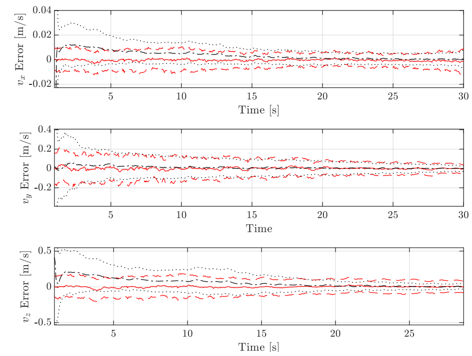
We initialized the FHMLE estimator using the straight line trajectory that better fits these $\bar{p}_{ij}$ and that is compatible with the Doppler velocity measurements (which is a simple least squares problem). While this initial estimate may not be particularly good at fitting the pseudo-range measurements and the IMU measurements, the PDIP was able to very reliably converge from this initialization to the maximum likelihood estimate in a small number of iterations. We note that with three angle of transmission measurements, we can get three estimates $\bar{p}_{ij}$ by selecting (i, j) equal to $(1, 2)$, $(2, 3)$, or $(2, 3)$. Initializing with these 3 options, we selected the one which converged to the largest value for the likelihood function. The same algorithm is then used to initialize the UKF at the start of each trial.

5.2 Results and Discussion

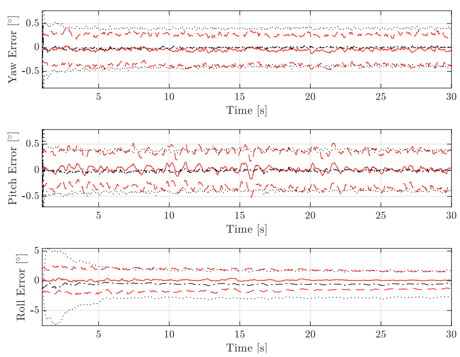
Figure 3 compares the estimation errors and 1-σ bounds for the UKF and FHMLE over 1000 Monte Carlo trials of measurements generated as in Section 2. These plots show that the UKF has a significant initial position bias and is noisier towards the beginning of the trajectory. Figure 4 shows the empirical distributions of the norms of UKF and FHMLE position errors over two segments of trajectory: the the first 10 seconds of the trajectory where geometry is poor, as well as the final 20 seconds where geometry is improved. This figure confirms that the UKF is heavier tailed in the first 10 seconds, while the performance of the two methods is comparable in the final 20. This is likely due to a combination of poor geometry and the fact that the UKF uses an approximation of the von Mises-Fisher distribution, while the FHMLE is able to handle the exact equations. Figure 5 demonstrates that by increasing κ to $2.5 \cdot 10^4$ and using a higher cost IMU (In this example a Honeywell HG-1700 IMU [20] with $\Sigma_a = 0.03 I_{3\times3}\ m/s^2$), we enter a regime where the UKF is



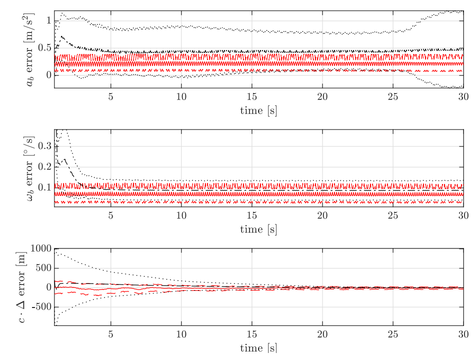
(a) position estimation errors



(b) velocity estimation errors



(c) attitude estimation errors



(d) bias estimation errors

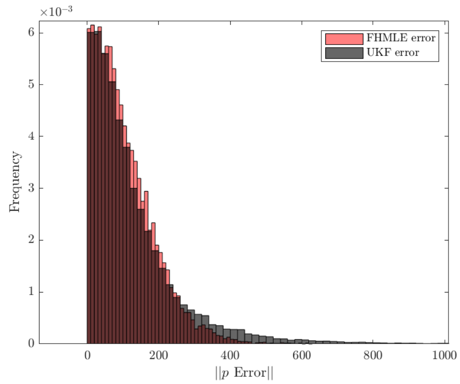
Figure 3: Monte Carlo positions, velocities, orientation, and bias average errors with sigma bounds (solid red: FHMLE average error, dashed red: FHMLE 1-σ error, dot-dashed black: UKF average error, dotted black: UKF 1-σ error).

actually more effective than the FHMLE at estimating position, especially in the final 20 seconds. Figure 3 also shows that FHMLE estimates of attitude, velocity, and bias terms are all either comparable or clearly better than those of the UKF.

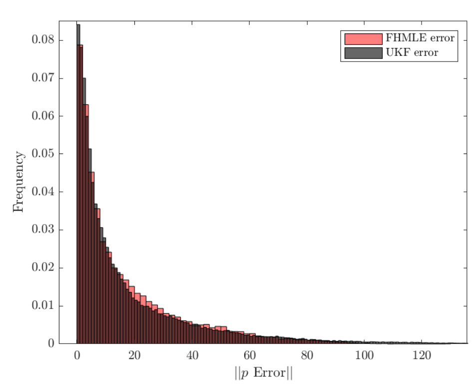
Figure 6 shows a sample MC trial for position and attitude errors, demonstrating that the UKF position estimates are smoother than the FHMLE estimates. However this comes with the drawback that the effects of noisy measurements taken in segments with poor geometry on estimates persist for longer than in the FHMLE, which can recover more quickly from unexpectedly poor quality measurements. This is the primary tradeoff between the two methods. Both FHMLE and UKF attitude estimates are nonsmooth, but errors are quite small.

Figure 7 shows the errors in position and velocity of the FHMLE without angle measurements. Because the initialization described above is not available without angle measurements, the filter is initialized with truth at the start of the trials and then warm starts are used throughout the rest of the trajectory. Regularization as in Remark 2 was observed to be necessary for convergence of the FHMLE without AoT measurements. Comparing Figures 3 and 7 demonstrates that the FHMLE performs significantly better in position and velocity estimates with AoT measurements than without. This matches the intuition given by Weighted GDOP in Figure 2.

To validate the usefulness of the proposed FHMLE algorithm for more general maneuvers, we

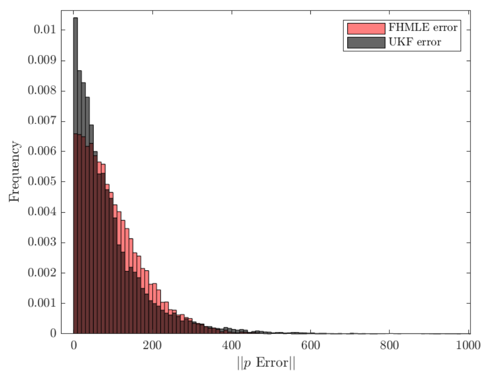


(a) first 10 seconds

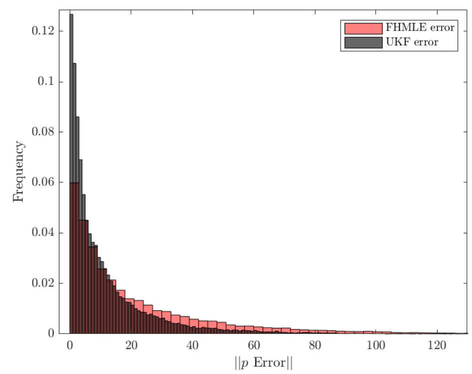


(b) final 20 seconds

Figure 4: Empirical distribution of $\|\hat{p} - p\|$ in the first 10 and final 20 seconds respectively. Red represents the FHMLE error, grey the UKF error, and dark red sections are where the two distributions overlap.



(a) first 10 seconds

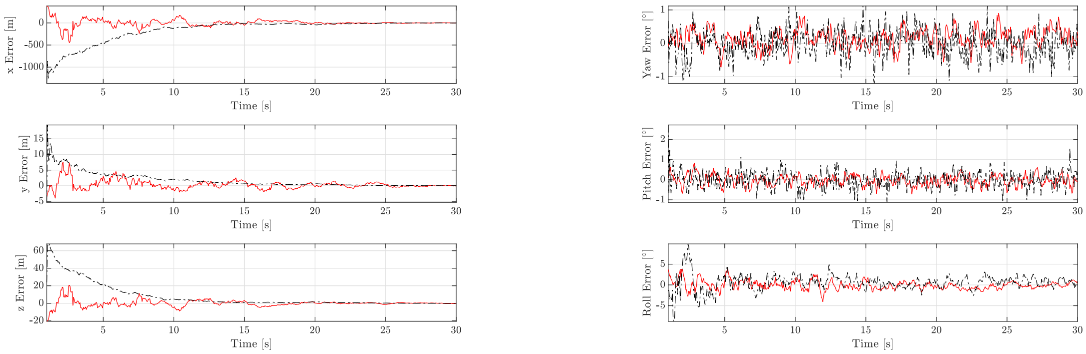


(b) final 20 seconds

Figure 5: Empirical distribution of $\|\hat{p} - p\|$ in the first 10 and final 20 seconds respectively, with $\kappa = 2.5 \cdot 10^4$ and $\Sigma_a = 0.03\, I_{3\times 3}\ m/s^2$. Red represents the FHMLE error, grey the UKF error, and dark red sections are where the two distributions overlap.

consider the case where the aircraft completes a banked circle of diameter 600 meters, centered 1500 meters from the deck of the carrier. Such a maneuver is common in the lead up to the carrier approach. The Monte Carlo results over 100 Monte Carlo trials are shown in Figure 8. We see the effect of bias and multimodal behavior in the UKF estimates. The UKF estimates appear to have lower variance in most cases than the FHMLE estimates, while the FHMLE estimates appear to be unbiased.

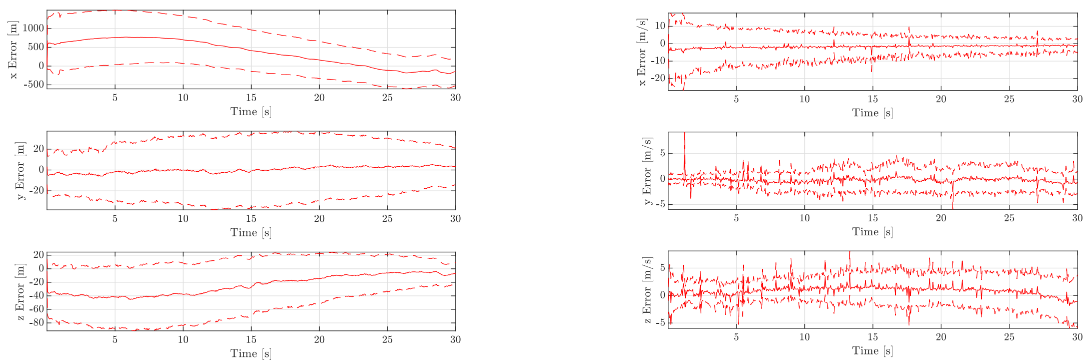
Figure 9 shows the effect of horizon length on runtime and root mean-squared error (RMSE). The FHMLE runtime for the one second horizon used for Figures 3-7 is around 11ms, while the UKF runtime in Matlab is around 5ms, and both are well below our sampling time for RF measurements (50ms). This plot shows that we can use a window up to around 2 seconds in length without average runtimes exceeding 30 ms. The computation times also grow approximately linearly with the window length, as suggested by the discussion in Section 3.2.1 The RMSE also decreases with horizon length, however these gains are offset by the latency introduced in estimation, especially for



(a) position estimation errors

(b) attitude estimation errors

Figure 6: Position and attitude estimation errors over a singe MC trial (solid red: FHMLE error, dot-dashed black: UKF error).



(a) position error without AoT

(b) velocity error without AoT

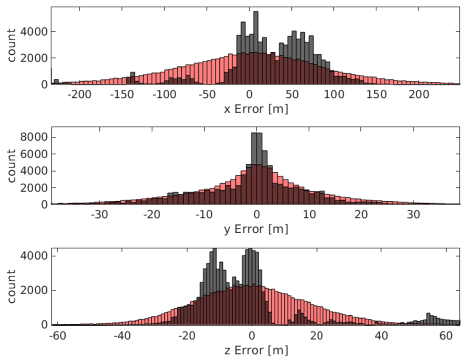
Figure 7: Position and velocity Monte Carlo estimation errors without angle of transmission measurements (solid red: FHMLE average error, dashed red: FHMLE 1-σ error).

a rapidly moving vehicle. The determination of optimal window lengths is explored further in [28].

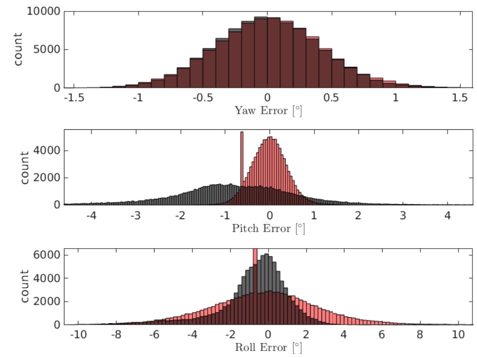
6 Conclusions

Navigation using RF beacons is useful in GPS-denied environments or when relying on Signals of Opportunity. We derived a Finite Horizon Maximum Likelihood Estimator for integrated navigation using realistic RF measurements. Benefits of the finite horizon approach include avoiding linearization, naturally accommodating constraints, and allowing for a richer class of noise models. This paper shows that, by using appropriate numerical methods, FHMLE in integrated navigation is effective and compares favorably to a UKF in when using low-cost IMUs and in situations with poor geometry, especially when making use of AoT measurements.

Our simulation results show that the FHMLE tends to produce estimates that, while having a smaller estimation error can be less smooth than those obtained from a UKF. In moving horizon estimation (MHE), this is often resolved by adding to the optimization cost terms that penalize the

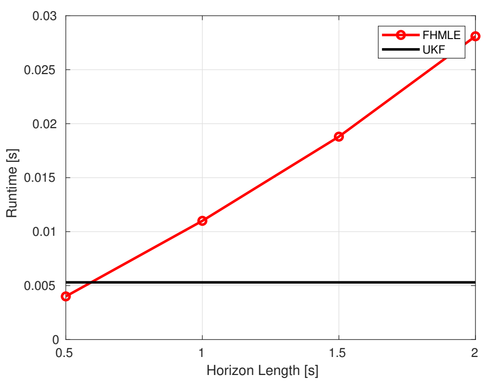


(a) position estimation errors

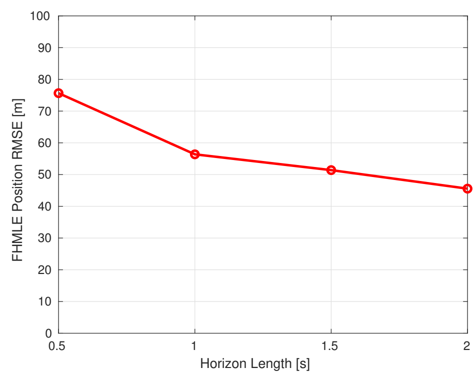


(b) attitude estimation errors

Figure 8: Position and attitude estimation errors over 100 MC trials of a 600m diameter banked circle. Red represents the FHMLE error, grey the UKF error, and dark red sections are where the two distributions overlap..



(a) Average runtimes for various horizon lengths



(b) RMSE for various horizon lengths

Figure 9: Average Runtimes and RMSE for various horizon lengths.

distance between the current estimate and the one obtained with the previous window of measurements [9, 21]. Experimenting with this option and understanding its implications in the context of maximum likelihood estimation is an important topic for future research. Other topics for future research include the development of FHMLE approaches to the "loose integration" of RF and IMU measurements (as in [4]) and the simultaneous estimation of position/orientation and the parameters of the noise distributions (covariance matrices for Gaussian distributions and concentration parameters for the von Mises-Fisher distribution).

References

[1] A. Abdi, J. A. Barger, and M. Kaveh. A parametric model for the distribution of the angle of arrival and the associated correlation function and power spectrum at the mobile station. *IEEE Transactions on Vehicular Technology*, 51(3):425–434, May 2002. (cited in p. 4)

[2] A. Amar and A. J. Weiss. Localization of narrowband radio emitters based on Doppler frequency shifts. *IEEE Transactions on Signal Processing*, 56(11):5500–5508, Nov 2008. (cited in p. 2)

[3] U. M. Ascher and L. R. Petzold. *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1st edition, 1998. (cited in p. 8)

[4] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Jan. 2001. (cited in p. 1, 3, 16)

[5] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. (cited in p. 9, 10)

[6] T. H. Bryne, J. M. Hansen, R. H. Rogne, N. Sokolova, T. I. Fossen, and T. A. Johansen. Nonlinear observers for integrated INS/GNSS navigation: Implementation aspects. *IEEE Control Systems Magazine*, 37(3):59–86, June 2017. (cited in p. 2)

[7] R. Fitzgerald. Divergence of the Kalman filter. *IEEE Transactions on Automatic Control*, 16 (6):736–747, Dec 1971. (cited in p. 2, 11)

[8] J. Gilbert, C. Moler, and R. Schreiber. Sparse matrices in matlab: Design and implementation. *SIAM Journal on Matrix Analysis and Applications*, 13(1):333–356, 1992. doi: 10.1137/0613024. URL `https://doi.org/10.1137/0613024`. (cited in p. 10)

[9] H. U. Gul, Y. D. Kai, and M. Khan. Multi-sensor integrated filtering for highly dynamic system using recursive moving horizon estimation technique. In *2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pages 351–356, Jan 2017. (cited in p. 2, 16)

[10] J. P. Hespanha. `tenscalc` — a Matlab toolbox for nonlinear optimization using symbolic tensor calculus. `https://github.com/hespanha/tenscalc`, June 2017. (cited in p. 9)

[11] J. P. Hespanha. TensCalc — A toolbox to generate fast code to solve nonlinear constrained minimizations and compute Nash equilibria. Technical report, University of California, Santa Barbara, Santa Barbara, June 2017. Available at `http://www.ece.ucsb.edu/~hespanha/techrep.html`. (cited in p. 2, 9)

[12] InvenSense. Mpu-6050 mems imu. `https://www.invensense.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf`, 2013. (cited in p. 11)

[13] J. T. Isaacs, K. O. Ezal, and J. P. Hespanha. Local carrier-based precision approach and landing system. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 6284–6290, Dec 2016. (cited in p. 2)

[14] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, Mar 2004. (cited in p. 2, 11)

[15] J. T. Kent. The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(1):71–80, 1982. (cited in p. 5)

[16] J. N. Lyness and C. B. Moler. Van der Monde systems and numerical differentiation. *Numerische Mathematik*, 8(5):458–464, Aug. 1966. (cited in p. 8)

[17] R. J. Matthys. *Crystal oscillator circuits*. Wiley-Interscience, New York, 1983. (cited in p. 8)

[18] J. Morales, P. Roysdon, and Z. Kassas. Signals of opportunity aided inertial navigation. In *Proceedings of ION GNSS Conference*, pages 1492–1501, 2016. (cited in p. 2)

[19] R. Murray, Z. Li, and S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. Taylor & Francis, 1994. URL https://books.google.com/books?id=D_PqGKRo7oIC. (cited in p. 3)

[20] Novatel. Hg1700 span. https://www.novatel.com/assets/Documents/Papers/HG1700_SPAN62.pdf, 2008. (cited in p. 7, 12)

[21] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, 48(2):246–258, Feb 2003. (cited in p. 16)

[22] H. Sairo, D. Akopian, and J. Takala. Weighted dilution of precision as quality measure in satellite positioning. *IEE Proceedings - Radar, Sonar and Navigation*, 150(6):430–436, Dec 2003. ISSN 1350-2395. doi: 10.1049/ip-rsn:20031008. (cited in p. 5)

[23] F. H. Schlee, C. J. Standish, and N. F. Toda. Divergence in the Kalman filter. *AIAA Journal*, 5:1114–1120, June 1967. (cited in p. 2)

[24] D. Simon. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory Applications*, 4(8):1303–1318, August 2010. (cited in p. 2, 3)

[25] D. Unsal and K. Demirbas. Estimation of deterministic and stochastic imu error parameters. In *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, pages 862–868, April 2012. doi: 10.1109/PLANS.2012.6236828. (cited in p. 6)

[26] R. Van Der Merwe, E. Wan, and S. I Julier. Sigma-point Kalman filters for nonlinear estimation and sensor-fusion-applications to integrated navigation. In *Proceedings of the AIAA Guidance, Navigation & Control Conference*, volume 3, 08 2004. (cited in p. 2)

[27] E. A. Wan and R. V. D. Merwe. The unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pages 153–158, 2000. (cited in p. 11)

[28] Z. Zhou and B. Li. GNSS windowing navigation with adaptively constructed dynamic model. *GPS Solut.*, 19(1):37–48, Jan. 2015. ISSN 1080-5370. doi: 10.1007/s10291-014-0363-y. URL http://dx.doi.org/10.1007/s10291-014-0363-y. (cited in p. 2, 15)