Maximum Likelihood Identification of an Ornstein-Uhlenbeck Model and Its CRLB

Shida Ye
Dept. of ECE
University of Connecticut
Storrs, CT
shida.ye@uconn.edu

Yaakov Bar-Shalom
Dept. of ECE
University of Connecticut
Storrs, CT
yaakov.bar-shalom@uconn.edu

Peter Willett
Dept. of ECE
University of Connecticut
Storrs, CT
peter.willett@uconn.edu

Ahmed Zaki
Naval Undersea Warfare Center
Newport, RI
ahmed.s.zaki.civ@us.navy.mil

***Abstract*—This paper applies Maximum Likelihood Estimation (MLE) to the identification of a stochastic error model of a gyroscope. The error model used for illustration features an Ornstein-Uhlenbeck process with an unknown time constant driven by a process noise with unknown variance, and a white measurement noise also with unknown variance. As the setup of MLE, the likelihood function (LF) is derived in the steady-state Kalman filter framework and is defined in reference to the parameters of the Kalman filter gain and innovation variance. The resulting log-likelihood function (LLF) is a quadratic function of the measurements, facilitating the evaluation of the Cramér-Rao Lower Bound (CRLB) and makes it possible to confirm the statistical efficiency, i.e., optimality, of the ML estimator presented in this paper.**

***Index Terms*—inertial sensors, noise, system identification, stochastic systems, maximum likelihood estimation, Cramer-Rao bound**

I. Introduction

The gyroscope is an inertial sensor (Inertial Measurement Unit—IMU) which measures rotation rate used in conjunction with an accelerometer for navigation by dead reckoning. Because the rate gyroscope output is susceptible to a random error (also known as random drift), the error must be modeled and its statistics obtained to characterize measurement uncertainty as the basis of accurate navigation. Distinct from deterministic sensor errors, including bias, misalignment and scale factor uncertainty, the random drift is typically modeled as a stochastic process. In particular, a stochastic model of two noise terms can be formulated as a state-space model with one noise term representing the state and the other noise term the measurement noise. Thus, inertial sensor noise modeling is essentially a noise identification problem, a topic that should interest both navigation and estimation researchers [8]. This is especially the case if the model involves a Gauss-Markov process, which can both represent IMU random drift and diverse dynamics in the estimation literature. The first-order Gauss-Markov (GM-1) process (characterized by a time constant) is also called the Ornstein-Uhlenbeck (OU) process. The OU process has been used to model target acceleration in the Singer model [9] as well as in long term target dynamics [10], and its bounded uncertainty property makes it useful for IMU random drift [2]. In fact, both the standalone OU process and the sum of multiple OU are used to model common noise types in IMU [4][17].

The classic approach to IMU noise identification is the Allan variance method [5]. A time-domain analysis technique, Allan variance computes differences in time averages between different time intervals and is suitable for identifying the noise statistics in a stochastic model driven by prespecified noise types. Estimators of the noise parameters can then be obtained graphically from the log-log plot of Allan variance versus averaging time, or more rigorously by fitting the Allan variance to a function of the parameters [19]. While this method has modest computational cost, it is suboptimal and is not applicable to OU processes [16] which have an unknown time constant. Maximum likelihood estimation, using the expectation-maximization algorithm, is an optimal approach, but the algorithm has a convergence problem with complex likelihood functions [16]. On the other hand, direct maximization of the likelihood function is generally difficult but still feasible for simpler problems, and [11] is a recent example with state of the art performance. The present work adopts a similar approach but relies on the standard angular rate measurement model, dispensing with integrated measurements as the basis of the LF.

In this paper, the likelihood function (LF) is derived using the state space model in the steady-state Kalman filter (KF) theory framework. The LF is given by the joint pdf of the measurements conditioned on the filter parameters, which in this case are the system transition function (determined by the time constant), filter gain and innovation variance [1]. The whiteness of the innovations allows the likelihood function to be written as a product of the pdfs of the innovations and ultimately as a quadratic function of the measurements. The innovation-based LF was pioneered by [15] with an application in gyroscope error modeling, and is preferable to the non-KF likelihood function in [3] which requires the inversion of a large matrix. Because the LF depends only implicitly on the noise parameters, relations derived from steady-state filter equations are used to map the filter parameter estimates (gain and innovation variance) to noise parameters. This two-

S. Ye, Y. Bar-Shalom, P. Willett, *et al.*, "Maximum likelihood identification of an Ornstein-Uhlenbeck model and its CRLB," in *Proc. 2024 27th International Conference on Information Fusion (FUSION)*, Jul. 2024, forthcoming. Work supported by NIUVT/ONR COMP 49.

step approach finds a precedent in [6]. The implementation is offline using an off-the-shelf optimization algorithm.

The Cramér-Rao Lower Bound (CRLB) is the gold standard for assessing the optimality of unbiased estimators of parameters related to inertial sensor calibration [13]. If the mean square errors of the estimates attain the CRLB, then the estimators are said to be statistically efficient. So far there have been no CRLB results for the joint estimation of system parameter and noise variances. Since the log-likelihood function (LLF) is a quadratic function of the measurements, the explicit form CRLB can be evaluated analytically, enabling a significant novelty of the paper. Statistical tests are used to verify the statistical efficiency of the parameters' estimates. The importance of the CRLB is that it quantifies the best possible estimation accuracy for a given system and data length; and, if it is achieved by an algorithm, this algorithm is optimal and unsurpassable.

II. Mathematical Model

The continuous-time differential equation characterizing the scalar OU process model is

$$\dot{x}(t) = -ax(t) + \tilde{v}(t) \tag{1}$$

where x is the state, a is the inverse time constant and $\tilde{v}$ is a zero-mean white process noise with autocorrelation

$$E[\tilde{v}(t)\tilde{v}(\tau)] = \tilde{Q}\delta(t-\tau) \tag{2}$$

with power spectral density (PSD) $\tilde{Q}$. The OU process $x(t)$ models an exponentially correlated random drift and is observed via the measurement model

$$z(t) = x(t) + \tilde{w}(t) \tag{3}$$

where the zero-mean white measurement noise $\tilde{w}(t)$, also termed the angle random walk (ARW), has the autocorrelation

$$E[\tilde{w}(t)\tilde{w}(\tau)] = \tilde{R}\delta(t-\tau) \tag{4}$$

with the PSD $\tilde{R}$. The angle random walk coefficient is defined as $\sqrt{\tilde{R}}$ with units deg/hr$^{1/2}$ [5].

Under the uniform sampling interval

$$T = t_{i+1} - t_i, \quad i = 1, 2, \ldots \tag{5}$$

the discretized OU process corresponding to (1) is

$$x(t_{i+1}) = e^{-aT}x(t_i) + v(t_i) \tag{6}$$

where e^{-aT} is the transition function and the discrete-time process noise is obtained by discretization as [1, Ch. 4]

$$v(t_i) = \int_{t_{i-1}}^{t_i} e^{-a(t_i-t)}\tilde{v}(t)dt \approx \int_0^T \tilde{v}(t)dt \tag{7}$$

for $T \ll \frac{1}{a}$, with the variance

$$\begin{aligned} E(v(t_i)^2) &= \int_0^T \int_0^T E[\tilde{v}(t)\tilde{v}(\tau)]d\tau dt \\ &= \int_0^T \int_0^T \tilde{Q}\delta(\tau-t)d\tau dt \\ &= \int_0^T \tilde{Q}dt = \tilde{Q}T \triangleq Q \end{aligned} \tag{8}$$

The continuous-time measurement $z(t)$ is averaged over the sampling interval to obtain the discretized measurement,[1] i.e.,

$$\begin{aligned} z(t_i) &= \frac{1}{T}\int_{t_{i-1}}^{t_i} z(t)dt = \frac{1}{T}\int_{t_{i-1}}^{t_i} x(t)dt + \frac{1}{T}\int_{t_{i-1}}^{t_i} \tilde{w}(t)dt \\ &\approx x(t_i) + w(t_i) \end{aligned} \tag{9}$$

where the discrete-time measurement noise is a white Gaussian noise sequence defined as

$$w(t_i) \triangleq \frac{1}{T}\int_{t_{i-1}}^{t_i} \tilde{w}(t)dt \tag{10}$$

Hence its variance is

$$\begin{aligned} E(w(t_i)^2) &= \frac{1}{T^2}\int_0^T \int_0^T E[\tilde{w}(t)\tilde{w}(\tau)]d\tau dt \\ &= \frac{1}{T^2}\int_0^T \int_0^T \tilde{R}\delta(\tau-t)d\tau dt \\ &= \frac{1}{T^2}\int_0^T \tilde{R}dt = \frac{\tilde{R}}{T} \triangleq R \end{aligned} \tag{11}$$

The measurement model assumes independence between the measurement noise and the process noise. The variances R and Q have units deg^2/hr^2. Together with the inverse time constant a whose unit is hr^{-1}, they constitute the set of parameters to be estimated from the sensor output with a known sampling interval T and total duration NT.

III. Maximum Likelihood Estimation of the Model Parameters

The sequence of innovations is employed to construct the likelihood function for maximum likelihood estimation of the target parameters. This approach is grounded in the understanding that the filter parameters' likelihood function—as detailed in Chapter 5 of [1]—is the joint pdf of the measurements conditioned on the parameters, and this, in turn, is equivalent to the joint pdf of the innovations. The steady-state filter is suitable for processes that are stationary [15][2]. To start deriving the likelihood function, the state estimate $\hat{x}_i$ of the OU process in (6) is expressed using the standard equations of the Kalman filter, with simplified notations and time indices

$$\hat{x}_i = \bar{x}_i + W(z_i - \bar{z}_i) \tag{12}$$

[1]This is a major difference from the commonly used measurement model in tracking where remote sensors (radar, sonar) provide discrete time measurements from matched filters, with noise variances that do not depend on the sampling (revisit) interval T. As shown in the sequel, the measurement noise variance here depends on T.

[2]One cannot estimate noise variances during the transient of the KF. However, the KF reaches steady state (under certain conditions) typically, after a small number of samples.

where W is the s. s. filter gain. The predicted state $\bar{x}_i$ and predicted measurement $\bar{z}_i$ can be substituted with their expressions, resulting in

$$\begin{aligned}\hat{x}_i &= e^{-aT}\hat{x}_{i-1} + W(z_i - e^{-aT}\hat{x}_{i-1}) \\ &= e^{-aT}(1-W)\hat{x}_{i-1} - Wz_i \\ &= W\sum_{j=1}^{i}(e^{-aT}(1-W))^{i-j}z_j \end{aligned} \tag{13}$$

The innovation is then

$$\nu_i = z_i - e^{-aT}\hat{x}_{i-1} = z_i - e^{-aT}W\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{i-j-1}z_j \tag{14}$$

When the gain of the steady-state filter is optimal, the sequence of innovations is independent and identically distributed under the white noise assumption, according to which the discrete-time measurement is driven by Gaussian noise. Consequently, the joint LF of a, the inverse time constant, W, the optimal steady-state gain, and S, the steady-state innovation variance, based on N measurements, is expressed as follows:

$$\begin{aligned}\Lambda(a,W,S) &= p(z_1, z_2, \ldots, z_N|a,W,S) \\ &= p(\nu_1,\nu_2,\ldots,\nu_N|a,W,S) \\ &= (2\pi)^{-N/2}S^{-N/2}\exp\left\{-\frac{1}{2S}\sum_{i=1}^{N}\nu_i^2\right\}\end{aligned} \tag{15}$$

Hence the log-likelihood function (LLF) is

$$\begin{aligned}&\ln\Lambda(a,W,S) = \frac{N}{2}\ln S - \frac{1}{2S}\sum_{i=1}^{N}\nu_i^2 + \text{const.} \\ &= -\frac{N}{2}\ln S - \frac{1}{2S}\sum_{i=1}^{N}\Bigg(z_i^2 \\ &- 2e^{-aT}Wz_i\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{i-j-1}z_j \\ &+ e^{-2aT}W^2\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}[e^{-aT}(1-W)]^{2i-j-k-2}z_jz_k\Bigg) + \text{const.}\end{aligned} \tag{16}$$

It is important to highlight that the LLF (16) is exact and analytically dependent solely upon the transition function, W and S. This aspect stands as the principal novelty of the present work: the estimation of parameters is conducted through a, W and S, and the derivation of the CRLB serves a dual purpose: it not only demonstrates the statistical efficiency of the estimates but also quantifies the best achievable estimation accuracy for parameters for a given system and data length.

With the LLF as the objective function, the constrained optimization problem can be formulated as

$$\min_{a>0, W\in(0,1), S>0}[-\ln\Lambda(a,W,S)] \tag{17}$$

The search algorithm used is an interior point method for nonlinear programming that uses the primal-dual approach and a barrier function to negotiate infeasibility [12]. Given the optimal solution presented in (17), denoted as (a^*, W^*, S^*), the steady-state filter equations facilitate a one-to-one mapping to the optimal system parameter triplet (a^*, Q^*, R^*). This process mirrors the derivation of the alpha-beta filter equations, as outlined in [1], and has the following results:

$$Q^* = (W^* - e^{-2a^*T}W^* + e^{-2a^*T}(W^*)^2)S^* \tag{18}$$

$$R^* = (1-W^*)S^* \tag{19}$$

The details of derivation of (18)–(19) can be found in [22]. The LLF for the Wiener process model, presented in [22] is a particular case of (16) with $a = 0$.

IV. The CRLB for a, Q and R

Upon defining the estimators for a, Q and R, we advance to obtain their Cramér-Rao Lower Bounds (CRLB) to demonstrate their statistical efficacy. Although the CRLB can typically be computed numerically through the average of the Hessian matrix associated with the LLF, the high complexity of the present scenario involving as many as tens of thousands of measurements necessitates an analytical solution. Similar to the process in the last section, results are first obtained for the parameters a, W, S and then transformed into results for a, Q, R.

A. Procedure for Derivation of the FIM of a, W and S

Denote the LLF as

$$\lambda \triangleq \ln\Lambda(a,W,S) \tag{20}$$

Let $\theta = (a,W,S)$, then the Fisher information matrix (FIM) for θ is defined in terms of the second partial derivatives of the LLF [1]

$$\mathbf{J}_\theta = -\begin{bmatrix} E\left[\frac{\partial^2\lambda}{\partial a^2}\right] & E\left[\frac{\partial^2\lambda}{\partial a\partial W}\right] & E\left[\frac{\partial^2\lambda}{\partial a\partial S}\right] \\ E\left[\frac{\partial^2\lambda}{\partial a\partial W}\right] & E\left[\frac{\partial^2\lambda}{\partial W^2}\right] & E\left[\frac{\partial^2\lambda}{\partial W\partial S}\right] \\ E\left[\frac{\partial^2\lambda}{\partial a\partial S}\right] & E\left[\frac{\partial^2\lambda}{\partial W\partial S}\right] & E\left[\frac{\partial^2\lambda}{\partial S^2}\right]\end{bmatrix} \tag{21}$$

In this section, the derivation of $E\left[\frac{\partial^2\lambda}{\partial a^2}\right]$ is explained in details and the results of this and other terms in (21) are stated in Appendix A. By starting from the expressions of the innovation (14), one can obtain the partial derivatives of the innovations w.r.t. a. The first and second derivatives are

$$\begin{aligned}\frac{\partial\nu_i}{\partial a} &= Te^{-aT}W\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{i-j-1}z_j \\ &+ Te^{-2aT}W(1-W)\sum_{j=1}^{i-1}(i-j-1)[e^{-aT}(1-W)]^{i-j-2}z_j\end{aligned} \tag{22}$$

$$
\frac{\partial^2 \nu_i}{\partial a^2} = -T^2 e^{-aT} W \sum_{j=1}^{i-1} [e^{-aT}(1-W)]^{i-j-1} z_j
$$
$$
- 3T^2 e^{-2aT} W(1-W) \sum_{j=1}^{i-1} (i-j-1)[e^{-aT}(1-W)]^{i-j-2} z_j
$$
$$
- T^2 e^{-3aT} W(1-W)^2 \sum_{j=1}^{i-1} (i-j-1)(i-j-2)
$$
$$
\cdot [e^{-aT}(1-W)]^{i-j-3} z_j \tag{23}
$$

In view of these two expressions, taking the second derivative of the LLF (16) results in

$$
\frac{\partial^2 \ln \Lambda}{\partial a^2} = -\frac{1}{S} \sum_{i=1}^{N} \left(\frac{\partial \nu_i}{\partial a} \right)^2 + \nu_i \frac{\partial^2 \nu_i}{\partial a^2}
$$
$$
= -\frac{1}{S} \sum_{i=1}^{N} \Bigg[2T^2 e^{-2aT} W^2
$$
$$
\cdot \sum_{j=1}^{i-1} \sum_{j=k}^{i-1} [e^{-aT}(1-W)]^{2i-j-k-2} z_j z_k
$$
$$
+ 5T^2 e^{-3aT} W^2 (1-W)
$$
$$
\cdot \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} (i-j-1)[e^{-aT}(1-W)]^{2i-j-k-3} z_j z_k
$$
$$
+ T^2 e^{-4aT} W^2 (1-W)^2
$$
$$
\cdot \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} (i-j-1)(i-k-1)[e^{-aT}(1-W)]^{2i-j-k-4} z_j z_k
$$
$$
- T^2 e^{-aT} W \sum_{j=1}^{i-1} [e^{-aT}(1-W)]^{i-j-1} z_i z_j
$$
$$
- 3T^2 e^{-2aT} W(1-W)
$$
$$
\cdot \sum_{j=1}^{i-1} (i-j-1)[e^{-aT}(1-W)]^{i-j-2} z_i z_j
$$
$$
- T^2 e^{-3aT} W(1-W)^2
$$
$$
\cdot \sum_{j=1}^{i-1} (i-j-1)(i-j-2)[e^{-aT}(1-W)]^{i-j-3} z_i z_j
$$
$$
+ T^2 e^{-4aT} W^2 (1-W)^2
$$
$$
\cdot \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} (i-j-1)(i-j-2)[e^{-aT}(1-W)]^{2i-j-k-4} z_j z_k \Bigg] \tag{24}
$$

To apply expectation on the expression above, the relevant random variables are the measurements. The expected value of the product of two measurements is

$$
E(z_i z_j) = E(x_i x_j) + \delta_{ij} R \tag{25}
$$

where δ_{ij} is the Kronecker delta function and

$$
E(x_i x_j) = E\left(\sum_{k=1}^{i} F^{i-k} \nu_k \sum_{h=1}^{j} F^{j-h} \nu_h \right)
$$
$$
= \sum_{k=1}^{\min(i,j)} F^{i+j-2k} Q \tag{26}
$$

We next substitute $F = e^{-aT}$ above, hence the geometric series evaluates to

$$
E(x_i x_j) = \frac{e^{-(i+j-2)aT} - e^{-(i+j-2-2n)aT}}{1 - e^{2aT}} Q \triangleq G(i,j) Q \tag{27}
$$

where $n = \min(i,j)$. Out of convenience, the function $G(i,j)$ denotes the coefficient of Q for fixed a and T. In particular, we have

$$
G(i,j) = \frac{e^{-(i+j-2)aT} - e^{-(i-j-2)aT}}{1 - e^{2aT}} \quad \text{if } j < i \tag{28}
$$
$$
G(i,i) = \frac{e^{-(2i-2)aT} - e^{2aT}}{1 - e^{2aT}} \tag{29}
$$

Finally, we can apply the expression of $E(z_i z_j)$ to obtain $E\left[\frac{\partial^2 \ln \Lambda}{\partial a^2}\right]$, shown in Appendix A. The corresponding expressions for other terms in the FIM can be found there.

B. CRLB Transformation from One Set of Parameters to Another

Given the FIM for the parameter vector $\theta = (a, W, S)$ from the previous subsection $\mathbf{J}_\theta$, we perform the matrix transformation required to obtain the CRLB for $\phi = (a, Q, R)$ according to [7] as

$$
\mathbf{B} = J_{\phi,\theta} \mathbf{J}_\theta^{-1} J'_{\phi,\theta} \tag{30}
$$

where the Jacobian matrix of ϕ with respect to θ is

$$
J_{\phi,\theta} = \begin{bmatrix} \frac{\partial a}{\partial a} & \frac{\partial a}{\partial W} & \frac{\partial a}{\partial S} \\ \frac{\partial Q}{\partial a} & \frac{\partial Q}{\partial W} & \frac{\partial Q}{\partial S} \\ \frac{\partial R}{\partial a} & \frac{\partial R}{\partial W} & \frac{\partial R}{\partial S} \end{bmatrix}
= \begin{bmatrix} 1 & 0 & 0 \\ \frac{\partial Q}{\partial a} & \frac{\partial Q}{\partial W} & \frac{\partial Q}{\partial S} \\ 0 & -S & 1-W \end{bmatrix} \tag{31}
$$

whose component terms are

$$
\frac{\partial Q}{\partial a} = 2TS e^{-2aT} (W - W^2) \tag{32}
$$
$$
\frac{\partial Q}{\partial W} = (1 - e^{-2aT} + 2e^{-2aT} W) S \tag{33}
$$
$$
\frac{\partial Q}{\partial S} = W - e^{-2aT} + e^{-2aT} W^2 \tag{34}
$$

Finally, the CRLBs for a, Q, R are the diagonal terms of the matrix $\mathbf{B}$ and σ_{CRLB} is used to denote the square roots of these diagonal terms.

V. Simulation Results

The performance of the ML estimators is tested with simulated data based on the parameter values from real sensors as reported in [20]. The simulated gyroscope signals have the true parameter values (PSD) of $\tilde{Q} = 5.8$ deg^2/hr^3 and $\tilde{R} = 2.5\text{e}{-3}$ deg^2/hr. Discrete time signals are randomly generated at two different sampling intervals, $T = 10^{-2}$ hour and $T = 10^{-3}$ hour, covering a total duration of 48 hours, according to (6) and (9). Consequently, the PSD values are mapped to two sets of variances according to the value of T. In MATLAB, the fmincon function is applied to optimize the log likelihood function, subject to predefiend constraints. A total of $N_{MC} = 1000$ Monte Carlo runs have been conducted and the root mean square (RMS) error of the estimates has been computed. The results are summarized in Tables I-II.

TABLE I
Maximum likelihood estimation results for $D = 48$ hr from 1000 MC trials for $T = 10^{-3}$ hr

Param.	true	avg. est.	$\frac{\text{RMSE}}{\text{true}}\%$	σ_{CRLB}
Q	5.8e$-$3	5.80e$-$3	6.77	3.95e$-$4
R	2.5	2.50	0.665	0.0166
a	3.6	3.64	12.6	0.451

The units of a are hr^{-1} and the units of R and Q are deg^2/hr^2.

TABLE II
Maximum likelihood estimation results for $D = 48$hr from 1000 MC trials for $T = 10^{-2}$ hr

Param.	true	avg. est.	$\frac{\text{RMSE}}{\text{true}}\%$	σ_{CRLB}
Q	5.8e$-$2	5.84e$-$2	7.12	4.15e$-$3
R	0.25	0.250	2.81	7.08e$-$3
a	3.6	3.65	13.1	0.460

The units of a are hr^{-1} and the units of R and Q are deg^2/hr^2.

The average estimates of noise variances are practically the same as the true values, confirming the unbiasedness of the MLE and the applicability of the CRLB. The RMSE of the variances are relatively small compared to the true values of the parameters: at $T = 10^{-3}$ hr, the relative error is 6.77% for Q and 0.665% for R; at $T = 10^{-2}$ hr, the relative error is 7.12% for Q and 2.81% for R. For the inverse time constant a, the RMSE of 12.6% at $T = 10^{-3}$ hr and 13.1% at $T = 10^{-2}$ hr are relatively large though the biases are very small. The difficulty of estimating a is a well known issue in the literature [14][18]. This is because the "available information" is low, yielding a relatively large CRLB (which is unbeatable).

Whether the estimators are efficient depends on whether the RMSE values are as low as the minimum given by the CRLB. While ML estimators are asymptotically efficient, hypothesis testing can be used verify that the RMSE computed from a finite number of trials is within the acceptance interval around σ_{CRLB}. By the definition of RMSE, if the estimator is efficient, the normalized estimation error squared (NEES) is chi-square distributed[3] as follows [1, Sec. 3.7.6]:

$$\frac{\text{RMSE}^2}{\sigma^2_{CRLB}} \sim \frac{1}{N_{MC}}\chi^2_{N_{MC}} \tag{35}$$

where N_{MC} is the number of MC trials. It follows that the 95% probability interval of the NEES is $[0.914, 1.090]$, thus the ratio RMSE$/\sigma_{\text{CRLB}}$ should be in the interval given by the corresponding square roots, i.e., $[0.956, 1.044]$. The numbers in Table III all fall inside this range ($\leq 4.4\%$). Therefore the hypothesis test accepts that the CRLB is reached for all parameters.

TABLE III
RMSE and the probability regions around CRLB

T (hr)	Param.	RMSE	$0.96\sigma_{\text{CRLB}}$	$1.04\sigma_{\text{CRLB}}$
10^{-3}	Q	3.93e$-$4	3.78e$-$4	4.13e$-$4
	R	0.0166	0.0158	0.0173
	a	0.453	0.432	0.471
10^{-2}	Q	4.13e$-$3	3.97e$-$3	4.33e$-$3
	R	7.03e$-$3	6.77e$-$3	7.39e$-$3
	a	0.47	0.439	0.480

The units of a are hr^{-1} and the units of R and Q are deg^2/hr^2.

VI. Conclusion

In the inertial sensor literature, optimal identification of sensor errors is reputed to be expensive and suboptimal approaches are commonly used. This paper shows that optimal estimation of unknown noise variances and model parameters is achievable with reasonable computational costs, and the case in point is gyroscope random drift modeled by an Ornstein-Uhlenbeck process and white measurement noise. This paper presents a maximum likelihood estimator derived from the exact likelihood function under steady-state Kalman filter assumption. This well substantiated foundation facilitates the derivation of the CRLB for all parameters under consideration. By comparing the sample RMSE of the estimates with the CRLB of the estimators, the statistical test confirms that the true RMSE can be accepted via a statistical test as equal to the CRLB, indicating the optimality of the ML estimators. The approach introduced here is not confined to simple low-order bias drift models; plans for future research include extending this methodology to more complex models incorporating multiple OU processes as well as more general dynamic systems.

Appendix A
FIM of a, W and S

Applying the expectation over (24) leads to

[3]Assuming the errors are Gaussian (their actual distribution is not known; however this assumption is verified in practice [1]).

$$E\left[\frac{\partial^2 \ln\Lambda}{\partial a^2}\right] = -\frac{1}{S}\sum_{i=1}^{N}\left(\frac{\partial \nu_i}{\partial a}\right)^2 + \nu_i\frac{\partial^2 \nu_i}{\partial a^2}$$

$$= -\frac{1}{S}\sum_{i=1}^{N}\Bigg(2T^2e^{-2aT}W^2\sum_{j=1}^{i-1}\sum_{j=k}^{i-1}[e^{-aT}(1-W)]^{2i-j-k-2}$$
$$\cdot G(j,k)Q$$
$$+ 2T^2e^{-2aT}W^2R\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{2i-2j-2}$$
$$+ 5T^2e^{-3aT}W^2(1-W)\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}(i-j-1)$$
$$\cdot [e^{-aT}(1-W)]^{2i-j-k-3}G(j,k)Q$$
$$+ 5T^2e^{-3aT}W^2(1-W)R\sum_{j=1}^{i-1}(i-j-1)$$
$$\cdot [e^{-aT}(1-W)]^{2i-2j-3}$$
$$+ T^2e^{-4aT}W^2(1-W)^2\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}(i-j-1)(i-k-1)$$
$$\cdot [e^{-aT}(1-W)]^{2i-j-k-4}G(j,k)Q$$
$$+ T^2e^{-4aT}W^2(1-W)^2R\sum_{j=1}^{i-1}(i-j-1)^2$$
$$\cdot [e^{-aT}(1-W)]^{2i-2j-4}$$
$$- T^2e^{-aT}W\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{i-j-1}G(i,j)Q$$
$$- 3T^2e^{-2aT}W(1-W)\sum_{j=1}^{i-1}(i-j-1)$$
$$\cdot [e^{-aT}(1-W)]^{i-j-2}G(i,j)Q$$
$$- T^2e^{-3aT}W(1-W)^2\sum_{j=1}^{i-1}(i-j-1)(i-j-2)$$
$$\cdot [e^{-aT}(1-W)]^{i-j-3}G(i,j)Q$$
$$+ T^2e^{-4aT}W^2(1-W)^2\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}(i-j-1)(i-j-2)$$
$$[e^{-aT}(1-W)]^{2i-j-k-4}G(j,k)Q$$
$$+ T^2e^{-4aT}W^2(1-W)^2R\sum_{j=1}^{i-1}(i-j-1)(i-j-2)$$
$$\cdot [e^{-aT}(1-W)]^{2i-2j-4}\Bigg) \tag{36}$$

Similarly, the expected values of other second derivatives can be obtained as follows:

$$E\left[\frac{\partial^2 \ln\Lambda}{\partial a\partial W}\right] = -\frac{1}{S}\sum_{i=1}^{N}\Bigg(Te^{-aT}\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{i-j-1}$$
$$\cdot G(i,j)Q$$
$$+ Te^{-2aT}(1-3W)\sum_{j=1}^{i-1}(i-j-1)$$
$$\cdot [e^{-aT}(1-W)]^{i-j-2}G(i,j)Q$$
$$- Te^{-3aT}W(1-W)\sum_{j=1}^{i-1}(i-j-1)(i-j-2)$$
$$\cdot [e^{-aT}(1-W)]^{i-j-3}G(i,j)Q$$
$$- 2Te^{-2aT}W\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}[e^{-aT}(1-W)]^{2i-j-k-2}G(j,k)Q$$
$$- 2Te^{-2aT}WR\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{2i-2j-2}$$
$$- Te^{-3aT}W(2-4W)\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}(i-j-1)$$
$$\cdot [e^{-aT}(1-W)]^{2i-j-k-3}G(j,k)Q$$
$$- Te^{-3aT}W(2-4W)R\sum_{j=1}^{i-1}(i-j-1)[e^{-aT}(1-W)]^{2i-2j-3}$$
$$+ Te^{-4aT}W^2(1-W)\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}(i-j-1)(2i-j-k-3)$$
$$\cdot [e^{-aT}(1-W)]^{2i-j-k-4}G(j,k)Q$$
$$+ Te^{-4aT}W^2(1-W)R\sum_{j=1}^{i-1}(i-j-1)(2i-2j-3)$$
$$\cdot [e^{-aT}(1-W)]^{2i-2j-4}\Bigg) \tag{37}$$

$$E\left[\frac{\partial^2 \ln\Lambda}{\partial a\partial S}\right] = -\frac{1}{S}\sum_{i=1}^{N}\Bigg(Te^{-aT}W\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{i-j-1}$$
$$\cdot G(i,j)Q$$
$$+ Te^{-2aT}W(1-W)\sum_{j=1}^{i-1}(i-j-1)$$
$$\cdot [e^{-aT}(1-W)]^{i-j-2}G(i,j)Q$$
$$- Te^{-2aT}W^2\sum_{j=1}^{i-1}\sum_{k=1}^{i-1}[e^{-aT}(1-W)]^{2i-j-k-2}G(j,k)Q$$
$$- Te^{-2aT}W^2R\sum_{j=1}^{i-1}[e^{-aT}(1-W)]^{2i-2j-2}$$

$$
\begin{aligned}
&- \frac{1}{2} T e^{-3aT} W^2 (1-2W) \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} (2i-j-k-2) \\
&\cdot [e^{-aT}(1-W)]^{2i-j-k-3} G(j,k) Q \\
&- T e^{-3aT} W^2 (1-W) R \\
&\left. \cdot \sum_{j=1}^{i-1} (i-j-1) [e^{-aT}(1-W)]^{2i-2j-3} \right) \qquad (38)
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial^2 \Lambda}{\partial W^2}\right] &= -\frac{1}{S} \sum_{i=1}^{N} \left(e^{-2aT} \right. \\
&\cdot \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} [e^{-aT}(1-W)]^{2i-j-k-2} G(j,k) Q \\
&+ e^{-2aT} R \sum_{j=1}^{i-1} [e^{-aT}(1-W)]^{2i-2j-2} \\
&- 4 e^{-3aT} W \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} (i-j-1) [e^{-aT}(1-W)]^{2i-j-k-3} \\
&\cdot G(j,k) Q \\
&- 4 e^{-3aT} W R \sum_{j=1}^{i-1} (i-j-1) [e^{-aT}(1-W)]^{2i-2j-3} \\
&+ e^{-4aT} W^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} (i-j-1)(i-k-1) \\
&\cdot [e^{-aT}(1-W)]^{2i-j-k-4} G(j,k) Q \\
&+ e^{-4aT} W^2 R \sum_{j=1}^{i-1} (i-j-1)^2 [e^{-aT}(1-W)]^{2i-2j-4} \\
&+ 2 e^{-2aT} \sum_{j=1}^{i-1} (i-j-1) [e^{-aT}(1-W)]^{i-j-2} G(i,j) Q \\
&- e^{-3aT} W \sum_{j=1}^{i-1} (i-j-1)(i-j-2) \\
&\cdot [e^{-aT}(1-W)]^{i-j-3} G(i,j) Q \\
&+ e^{-4aT} W^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} (i-j-1)(i-j-2) \\
&\cdot [e^{-aT}(1-W)]^{2i-j-k-4} G(j,k) Q \\
&+ e^{-4aT} W^2 R \sum_{j=1}^{i-1} (i-j-1)(i-j-2) \\
&\left. \cdot [e^{-aT}(1-W)]^{2i-2j-4} \right) \qquad (39)
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial^2 \Lambda}{\partial W \partial S}\right] &= \frac{1}{S^2} \sum_{i=1}^{N} \left(-e^{-aT} \sum_{j=1}^{i-1} [e^{-aT}(1-W)]^{i-j-1} \right. \\
&\cdot G(i,j) Q \\
&+ e^{-2aT} W \sum_{j=1}^{i-1} (i-j-1) [e^{-aT}(1-W)]^{i-j-2} G(i,j) Q \\
&+ e^{-2aT} W \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} [e^{-aT}(1-W)]^{2i-j-k-2} G(j,k) Q \\
&+ e^{-2aT} W R \sum_{j=1}^{i-1} [e^{-aT}(1-W)]^{2i-2j-2} \\
&- e^{-3aT} W^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} (i-j-1) [e^{-aT}(1-W)]^{2i-j-k-3} \\
&\cdot G(j,k) Q \\
&\left. - e^{-3aT} W^2 R \sum_{j=1}^{i-1} (i-j-1) [e^{-aT}(1-W)]^{2i-2j-3} \right) \qquad (40)
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial^2 \Lambda}{\partial S^2}\right] &= \frac{N}{2S^2} - \frac{1}{4S^3} \sum_{i=1}^{N} \left(G(i,i) Q + R \right. \\
&- 2 e^{-aT} W \sum_{j=1}^{i-1} [e^{-aT}(1-W)]^{i-j-1} G(i,j) Q \\
&+ e^{-2aT} W^2 \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} [e^{-aT}(1-W)]^{2i-j-k-2} G(j,k) Q \\
&\left. + e^{-2aT} W^2 R \sum_{j=1}^{i-1} [e^{-aT}(1-W)]^{2i-2j-2} \right) \qquad (41)
\end{aligned}
$$

References

[1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. Hoboken, NJ: Wiley, 2001.

[2] M. El-Diasty and S. Pagiatakis, “A rigorous temperature-dependent stochastic modelling and testing for MEMS-based inertial sensor errors,” *Sensors*, vol. 9, pp. 8473–8489, Oct. 2009.

[3] S. Guerrier, R. Molinari, and J. Balamuta, “Discussion on maximum likelihood-based methods for inertial sensor calibration,” *IEEE Sensors Journal*, vol. 16, no. 14, Jul. 2016.

[4] E. M. Hemerly, “MEMS IMU stochastic error modelling,” *Systems Science and Control Engineering*, Jan. 1, 2017.

[5] *IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Laser Gyros (IEEE Std 647-2006)*. IEEE Aerospace and Electronic Systems Society, 2006.

[6] R. L. Kashyap, “Maximum likelihood identification of stochastic linear systems,” *IEEE Transactions on Automatic Control*, vol. 15, no. 1, pp. 25–34, Feb. 1970.

[7] S. Kay, *Fundamentals of Statistical Processing: Estimation Theory*. Prentice Hall, 1993.

[8] X. R. Li and Y. Bar-Shalom, “A recursive multiple model approach to noise identification,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 671–684, Jul. 1994.

[9] X. R. Li and V. Jilkov, “Survey of maneuvering target tracking part I: Dynamic models,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.

[10] L. Millefiori, P. Braca, and P. Willett, “Modeling vessel kinematics using a stochastic mean-reverting process for long-term prediction,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 5, pp. 2313–2340, Oct. 2016.

[11] J. Nikolic, P. Furgale, A. Melzer, and R. Siegwart, “Maximum likelihood identification of inertial sensor noise model parameters,” *IEEE Sensors Journal*, vol. 16, no. 1, pp. 163–176, Jan. 2016.

[12] J. Nocedal, F. Öztoprak, and R. Waltz, “An interior point method for nonlinear programming with infeasibility detection capabilities,” *Optimization Methods and Software*, vol. 29, no. 4, pp. 837–854, Jul. 2014.

[13] G. Panahandeh, I. Skog, and M. Jansson, “Calibration of the accelerometer triad of an inertial measurement unit, maximum likelihood estimation and Cramér-Rao bound,” in *2010 International Conference on Indoor Positioning and Indoor Navigation*, 2010.

[14] A. Roy and W. Fuller, “Estimation for autoregressive time series with a root near one,” *Journal of Business and Economic Statistics*, vol. 19, no. 4, pp. 482–493, 2001.

[15] P. Smith, “Estimation of the covariance parameters of nonstationary time-discrete linear systems,” Aerospace Corporation, Technical Report TR-0172(2311)-3, 1971.

[16] Y. Stebler, S. Guerrier, J. Skaloud, and M.-P. Victoria-Feser, “Constrained expectation-maximization algorithm for stochastic inertial error modeling: Study of feasibility,” *Measurement Science and Technology*, vol. 22, no. 8, 2011.

[17] Y. Stebler, S. Guerrier, and J. Skaloud, “An Approach for Observing and Modeling Errors in MEMS-Based Inertial Sensors Under Vehicle Dynamic,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 11, pp. 2926–2936, Nov. 2015.

[18] G. Thomas, N. Cooper, C. Venditti, A. Meade, and R. Freckleton, “Bias and measurement error in comparative analyses: A case study with the Ornstein-Uhlenbeck model,” *Biological Journal of the Linnean Society*, vol. 118, no. 1, pp. 64–77, May 2016.

[19] R. Vaccaro and A. Zaki, “Statistical modeling of rate gyros,” *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 3, pp. 673–684, Mar. 2012.

[20] R. J. Vaccaro and A. S. Zaki, “Statistical models of inertial sensors and integral error bounds,” in *Advanced Research in Naval Engineering*, A. A. Ruffa and B. Toni, Eds., Cham, Switzerland: Springer, 2018, pp. 143–160.

[21] S. Ye, Y. Bar-Shalom, P. Willett, and A. Zaki, “Maximum likelihood identification of an Ornstein-Uhlenbeck model and its CRLB,” in *Proc. 2024 27th International Conference on Information Fusion (FUSION)*, Jul. 2024, forthcoming.

[22] S. Ye, Y. Bar-Shalom, P. Willett, and A. Zaki, “Statistically efficient estimation of noise variances for a Wiener process observed with measurement noise,” in *Proc. SPIE 13057, Signal Processing, Sensor/Information Fusion, and and Target Recognition XXXIII*, 1305710, 2024.