JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 23, No. 6, November–December 2000

Online Estimation of Allan Variance Parameters

Jason J. Ford* and Michael E. Evans†
Defence Science Technology Organisation, Salisbury, South Australia 5108, Australia

A new online method is presented for estimation of the angular random walk and rate random walk coefficients of inertial measurement unit gyros and accelerometers. In the online method, a state-space model is proposed, and recursive parameter estimators are proposed for quantities previously measured from offline data techniques such as the Allan variance method. The Allan variance method has large offline computational effort and data storage requirements. The technique proposed here requires no data storage and computational effort of approximately 100 calculations per data sample.

I. Introduction

INERTIAL navigation is the process by which measurements made using gyroscopes and accelerometers are used to determine position.[1,2] Inertial navigation is an essential technology for modern aircraft and an important tool for navigation in general. Accelerometer and gyros in inertial measurement units (IMUs) suffer from particular types of noise interferences that induce IMU navigation errors that grow with time. Because of the importance of systems with these characteristics, a vast amount of analysis of this type of noise has been performed, including frequency fluctuations in atomic clocks,[3] noise from ring laser gyros,[4] and gyro sensors in general.[5–7]

Initially, the noise produced by IMU sensors was characterized by a single rms number, but this was an overly conservative characterization of the short-term behavior and a single rms number can not adequately model the longer term error growth. To overcome this conserative characterization, the Allan variance method was developed and is now the standard method of analysis.[6,7] The Allan variance $\sigma_A^2(N)$ of a sensor output is defined as the variance of the difference of the means of successive subsets (of size N) of the data.

When the Allan deviation $\sigma_A(N)$ is plotted against the sample size, particular noise features in the gyro output appear as structures in the Allan deviation curve.[4,6–8]

There are several disadvantages of the Allan variance technique. By its very nature, it is an offline process and requires a large amount of data to be stored. Additionally, the technique requires that lines of best fit be manually placed on the graph and intercepts be read off to obtain estimates of the noise contributions. Recently, a new technique for calculating the Allan variance has been proposed.[9] This technique has the advantage of providing estimates of the Allan variance as the data arrive, but still requires lines of best fit to be placed on the graph. Nonlinear estimation techniques, such as given in Ref. 10, can be used to estimate lines of best fit automatically from Allan variance data; however, the Allan variance still needs to be calculated separately.

In this paper we consider an approach that estimates the size of two key gyro noise sources (the angular random walk and the rate random walk) directly from the data set without the need to calculate the Allan variance of the gyro output. A key motivation for considering only angular random walk and random rate walk noise sources is that the stochastic model of gyros presented in the literature[5–7] reduces to a simple state-space representation when only these two noise terms are present.

Using the state-space representation, global and local convergence results are established in several stages using the law of large numbers and an ordinary differential equation approach. Unlike the Allan variance technique, the proposed parameter estimators do not require large amounts of data to be stored and only require approximately 100 calculations per time step. For larger data sets, greater than 10^4, the proposed online technique is computationally superior to the Allan variance technique (a typical data set might contain 24 h of 1-Hz measurements, that is, $\sim 10^7$ data points).

Although the proposed algorithm is computationally attractive, it relies on the simplifying assumption that the noise is a combination of only angular random walk and rate random walk noise sources. This assumption needs to be verified prior to use. Hence, we suggest that the proposed algorithm be used in conjunction with the Allan variance method. For example, the Allan variance method could be used to establish that noise characteristics of a particular device are dominated by angular random walk and rate random walk noises, and then the proposed algorithm could be used to verify the performance of the device at later times. Alternatively, a batch of similar devices could be tested using the proposed algorithm once the noise characteristics of one of the devices has been established using the Allan variance method.

The performance of the algorithm is demonstrated against both computer-generated data and data from a real device.

This paper is organized as follows. In Sec. II.A, we introduce the stochastic noise model for gyros and the Allan variance analysis technique. Then the Allan variance method is applied to a low-grade IMU that has error characteristics typical of many IMUs. The error characteristics of the low-grade IMU are then used to motivate our assumption that angular random walk and rate random walk are the key noise sources. In Sec. II.D, by the use of this simplifying assumption, a state-space model for angular and rate random walk noises is introduced. In Sec. II.E, parameter estimators based on several simplifying assumptions are introduced, and a partial convergence result is established (as a necessary step toward full convergence results). In Sec. II.F, conditional mean estimates are defined, and an estimation algorithm for realistic problems is introduced. Local convergence results are established using an ordinary differential equation approach. In Sec. III, implementation issues are discussed including the initialization of the algorithm, the computational effort required, techniques for improving convergence rates from poor initializations, the effect of other noise sources being present, and the need to handle numerical overflow. Some results are presented in Sec. IV. This includes results from the algorithm applied to computer-generated data and results from analysis of data from a low-grade IMU. Finally, in Sec. V, some conclusions are presented.

II. Preliminaries

A. Stochastic Model and the Allan Variance Method

There are two Institute of Electrical and Electronics Engineers standards related to specification and testing of single-axis gyros.[6,7] These two standards provide an overview of dynamic and stochastic modeling of single-axis gyros, as well as detailing the two preferred methods of analysis: the power spectral density and the Allan variance method. This paper uses the stochastic model for single-axis

Received 21 June 1999; revision received 17 February 2000; accepted for publication 28 February 2000.

*Research Scientist, Weapons Systems Division, P.O. Box 1500; Jason.Ford@dsto.defence.gov.au.

†Senior Research Scientist, Weapons Systems Division, P.O. Box 1500.

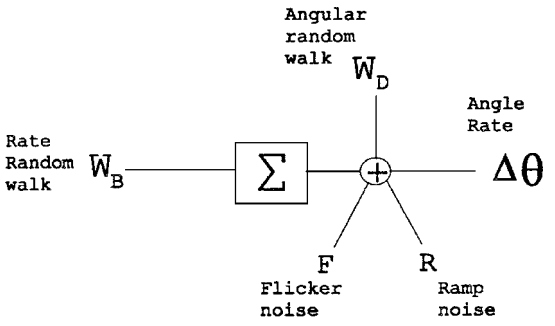

Fig. 1 Discrete-time stochastic model.

gyros and the Allan variance method of analysis described in the standards.

According to the standards,[6,7] the random component of gyro sensor error dynamics can be modeled by a discrete-time, time-invariant, linear stochastic model. The stochastic model in the standards allows for a multitude of error sources of a random nature. Noise of a nonrandom nature (including scale factor errors, fixed-bias errors, and misalignment errors) are considered separately from the stochastic model, and for the purposes of this paper, we assume that issues about these deterministic errors have already been resolved.

In this paper we consider gyros whose measurements are angular rates (or delta angles), and, hence, some of the noise sources allowed in the gyro dynamical stochastic model are not relevant. Figure 1 shows a discrete-time stochastic model of gyro error dynamics including several possible noise sources. This model includes angular random walk (or white rate noise), rate random walk, flicker noise (or bias instability), and random-bias (or fixed drift) noise sources. Noise terms such as rate quantization noise, Markov noise or correlated random drift, sinusoidal errors, ramp instability, and others are possible, but are not shown.

The Allan variance is a method of analysis that determines the error characteristics of a gyro using sampled measurements of the gyro outputs. Let y_k denote the measurement of angular rate (or delta angle) at time k. Let T be the number of data points available for analysis, and let $\bar{y}_\ell(N)$ be the average of the N samples beginning at time $k = (\ell - 1)N + 1$. That is,

$$\bar{y}_\ell(N) := \frac{1}{N}\sum_{i=1}^{N} y_{(\ell-1)N+i} \qquad \text{for} \qquad N \le R_N \tag{1}$$

where R_N is the largest integer not greater than T/N. Then the Allan variance $\sigma_A^2(N)$ is defined as

$$\sigma_A^2(N) := \frac{1}{2(R_N - 1)}\sum_{i=1}^{R_N-1}[\bar{y}_{i+1}(N) - \bar{y}_i(N)]^2 \tag{2}$$

The attractive feature of the Allan variance technique is the ability to determine the contribution of various error sources from features of the log-log plot of Allan deviation, $\sigma_A(N)$ vs N. Details of the features of various error contributions may be found in Refs. 4, 6, and 7 and other papers. In brief, the most common and significant noise sources are angular random walk, rate random walk, and flicker noise (or bias instability), which can be determined from regions in the log-log plot of Allan deviation with slopes of $-\frac{1}{2}$, $\frac{1}{2}$, and 0, respectively. Full details of the procedure may be found in Refs. 4, 6, and 7 and other papers.

A final comment about the literature describing the Allan variance method is that some confusion can result from the terms commonly used to describe the noise characteristics of single-axis gyros. For example, the quantization noise described in much of the literature (see Refs. 4 and 8) and the standards (see Refs. 6 and 7) is noise resulting from quantization of angular measurements. Modern gyros provide angular rate measurements directly, and hence, angular quantization is meaningless. Quantization of rate measurements is termed rate quantization in the standards.[6,7] Rate quantization is not even mentioned in much of the literature (see Refs. 4, 8, and 9). In fact, much of the literature is quite misleading and seems to suggest mistakenly that results for angle quantization are actually results for rate quantization (for example, see Ref. 9, which considers angular quantization noise even though the paper considers a device that measures angular rate).

In the following section we use the Allan variance technique to analyze a low-grade IMU.

B. Testing of a Low-Grade IMU

Over a 45-h period, angular rate measurements from a low-grade gyro were sampled at 1 Hz. The rate sensors of this IMU consist of micromachined vibrating ceramic plates that utilize Coriolis forces to output angular rate measurements. To allow quick processing, the data were preprocessed by averaging over 10-s periods to give measurements in units of degrees per 10 s. This preprocessing does not change the basic nature of the data's error characteristics, but it does scale the size of their contribution. All results presented in this paper refer to the error characteristics of the 10-s averaged data.

The error characteristics of this device can be analyzed using the Allan variance method. Equations (1) and (2) are used to calculate the Allan variance of the data set recorded from the device. Figure 2 shows the log-log plot of Allan deviation against N. The estimated coefficients for this device are given in Sec. IV.

The Allan deviation plot of this device has two distinct features. These features are asymptotes with slopes of $+\frac{1}{2}$ and $-\frac{1}{2}$, which correspond to angular random walk and rate random walk noise sources, respectively. This plot suggests that the dominant noise sources are angular random walk and rate random walk. Other noises such as bias instability or Markov noise may be present in the data, but the Allan deviation plot suggests that even if these additional noises are present their effect is dominated by the two main noise sources.

The Allan deviation plot for this device is similar in nature to many other gyros examined including tactical grade ring laser gyros and fiber optic gyros. In a practical sense, the dominant noise sources in all of these devices appear to be angular random walk and rate random walk. For this reason, in the following section we consider a simplified stochastic model that assumes these are the only noise sources.

C. Simplified Model

In this section, we propose a simplified stochastic model (see Fig. 3). This simplification seems appropriate for many of the sensors that we have examined via the Allan variance method.

In the following section we show this simplified stochastic model can be represented as a state-space model. The unknown angular random walk and rate random walk coefficients are parameters of this representation. This state-space representation is convenient because estimation of linear system parameters is a well-researched problem and there are many well-developed techniques.[11–13] This paper uses a recently developed technique that is particularly well suited to this particular parameter estimation problem.[11]

One key advantage of formulating the noise characterization problem as a parameter estimation problem is that the angular random walk and rate random walk coefficients can be estimated directly without the need to calculate the Allan variance (if angular random walk and rate random walk are the only noise terms present). Of course, the standard way to establish what noise contributions are present is to perform the Allan variance analysis. We suggest that the proposed algorithm could be useful in situations where the types of noise contributions are known with some confidence a priori. Two situations where the proposed algorithm could be used are retesting of a gyro previously tested by the Allan variance method and testing a batch of devices of the same type.

D. State-Space Model

This section gives the formal mathematical framework for the parameter estimation problem. Consider a probability measure space $(\Omega, \mathcal{F}, P)$. Here Ω is an arbitrary space or set of points ω; also, $\mathcal{F}$ is σ algebra in Ω (a class of subsets containing Ω and closed under the formation of complements and countable unions) and P is a probability measure on $\mathcal{F}$. If $x(\omega)$ is a random variable on this probability space, then a particular realization of the random variable is $x(\omega = \omega_1)$, where ω_1 is a particular value of ω taken from

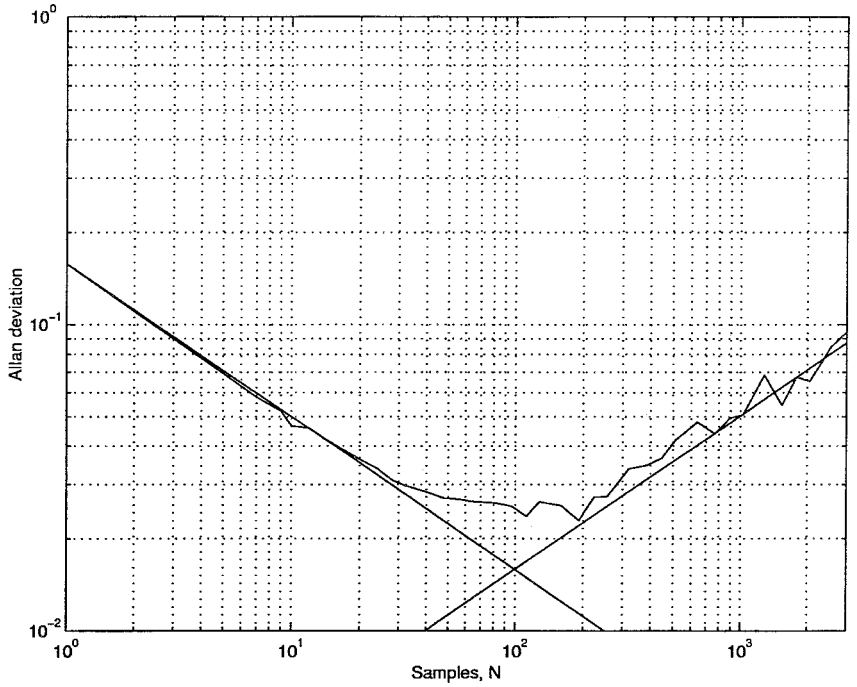

Fig. 2 Low-grade IMU: Allan deviation vs sample size; lines corresponding the parameters estimate from the proposed algorithm added.

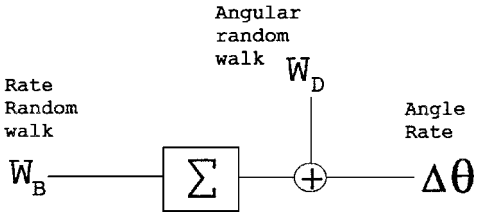

Fig. 3 Simplified discrete-time stochastic model.

Ω. For simplicity, we assume the ω dependence is understood and will write simply x for the random variable (except where the ω dependence aids understanding) (see Ref. 14 for more details on probability theory).

In this paper, we have several random variables on the same probability space and, hence, ω parameterizes a set of random variables $[\{w_k(\omega)\}, \{v_k(\omega)\}, x_0(\omega)]$, where w_k, v_k, and x_0 are defined as follows. Suppose $\{x_\ell\}$, $\ell \in \mathcal{Z}^+$ ($\mathcal{Z}^+$ is the set of positive integers), is a discrete-time linear stochastic process, taking values in R, with dynamics given by

$$x_{k+1} = x_k + \sqrt{B} w_{k+1}, \qquad x_0 \in R \tag{3}$$

where $k \in \mathcal{Z}^+$, $\sqrt{B}$ is the rate random walk coefficient, and $\{w_\ell\}$, $\ell \in \mathcal{Z}^+$, is a sequence of independent and identically distributed $N(0, 1)$ scalar random variables. The notation $N(0, 1)$ is shorthand for a random variable whose density is Gaussian with zero mean and variance 1. Let x_0 be a random variable whose density is Gaussian with known mean and variance.

The state process x_ℓ, $\ell \in \mathcal{Z}^+$, is observed indirectly via the scalar observation process $\{y_\ell\}$, $\ell \in \mathcal{Z}^+$, given by

$$y_k = x_k + \sqrt{D} v_k \in R \tag{4}$$

where $k \in \mathcal{Z}^+$, $\sqrt{D}$ is the angular random walk coefficient, and $\{v_\ell\}$, $\ell \in \mathcal{Z}^+$, is a sequence of independent and identically distributed $N(0, 1)$ scalar random variables. We assume that x_0, $\{w_\ell\}$, and $\{v_\ell\}$ are mutually independent of each other. The observations $\{y_k\}$ are a sequence of angle rates measured by the IMU (or $\Delta\theta$ in Fig. 3). The variance of noise W_B is B, and the variance of noise W_D is D. Let $K^{\text{av}} = \sqrt{B}$ and $N^{\text{av}} = \sqrt{D}$ be the rate random walk and angular random walk coefficients estimated by the Allan variance method.[6,7] We have considered here only single-axis devices, but in principle this technique can be extended to higher dimensions with B and D becoming covariance matrices.

Finally, let $\mathcal{Y}_k := \{y_0, \ldots, y_k\}$ and $\mathcal{G}_k := \{x_0, \ldots, x_k\}$ be the history of the observations and the state until time k. The symbol $:=$ denotes defined as.

The model described by Eqs. (3) and (4) is denoted by

$$\lambda = \lambda(B, D, x_0) \tag{5}$$

We will present convergence results that hold almost surely (a.s.). We will say that in a probability space $(\Omega, \mathcal{F}, P)$ a result holds a.s. if it holds with probability 1, or equivalently that it holds for all ω in an $\mathcal{F}$ set of probability 1 where $\omega \in \Omega$. Consider the following illustration. Let $z, z_1, z_2, \ldots,$ be random variables. If

$$\lim_{k \to \infty} z_k(\omega) = z(\omega) \tag{6}$$

for all ω in an $\mathcal{F}$ set of probability 1, then

$$\lim_{k \to \infty} z_k = z \quad \text{a.s.} \tag{7}$$

A weaker convergence condition is that

$$\lim_{k \to \infty} P[|z_k(\omega) - z(\omega)| \geq \epsilon] = 0 \tag{8}$$

which is called convergence in probability and is equivalent to mean square convergence. Note that almost sure convergence implies convergence in probability but convergence in probability does not imply almost sure convergence.

E. Parameter Estimation: Full Observations

In this section we assume that both $\{x_k\}$ and $\{y_k\}$ are fully observed. The results for the full observation case are presented as a stepping stone to the more interesting and general results that follow.

From Eq. (3), with simple manipulation then squaring both sides and summing over k, we obtain

$$B \sum_{i=1}^{k} w_i^2 = \sum_{i=1}^{k} (x_i - x_{i-1})^2 \tag{9}$$

Now consider the scalars

$$J_k = \sum_{i=1}^{k} x_i x_{i-1}, \qquad O_k = \sum_{i=1}^{k} x_i^2$$

$$H_k = \sum_{i=1}^{k} x_{i-1}^2, \qquad M_k = \sum_{i=1}^{k} w_i^2 \tag{10}$$

From Eq. (9), we see that

$$BM_k = O_k + H_k - 2J_k \tag{11}$$

Similarly, from Eq. (4) with simple manipulation then squaring both sides and summing over k, we obtain

$$D\sum_{i=1}^{k} v_i^2 = \sum_{i=1}^{k} (y_i - x_i)^2 \tag{12}$$

Now consider the scalars

$$T_k = \sum_{i=1}^{k} y_i x_i, \qquad Q_k = \sum_{i=1}^{k} y_i^2, \qquad V_k = \sum_{i=1}^{k} v_i^2 \tag{13}$$

From Eq. (12), we see that

$$DV_k = Q_k + O_k - 2T_k \tag{14}$$

1. Almost Sure Convergence

Lemma 1: Consider the linear system (3) and (4) denoted λ. Suppose $\{x_k\}$ and $\{y_k\}$ exist and are measured, then

$$\lim_{k\to\infty} \hat{B}_k, \hat{D}_k = B, D \quad \text{a.s.} \tag{15}$$

where

$$\hat{B}_k = (O_k + H_k - 2J_k)/k, \qquad \hat{D}_k = (Q_k + O_k - 2T_k)/k$$

Proof: Consider the estimation of B first. By the use of Eq. (11), rewrite the estimate as

$$\hat{B}_k = BM_k/k \tag{16}$$

By the strong law of large numbers,[14] $\lim_{k\to\infty}(1/k)M_k = 1$ a.s. The lemma results follow from Eq. (16). The lemma result for $\hat{D}_k$ follows similarly. □

2. Finite Data Sets

Lemma 2: Consider the linear system (3) and (4) denoted λ. Suppose finite date sets $\{x_0, x_1, \ldots, x_T\}$ and $\{y_0, y_1, \ldots, y_T\}$ are measured, then

$$P(|\hat{B}_T - B| \geq \epsilon) \leq B/T\epsilon^2, \qquad P(|\hat{D}_T - D| \geq \epsilon) \leq D/T\epsilon^2 \tag{17}$$

Proof: The proof follows from an argument similar to Lemma 1 by using the weak law of large number.[14] □

F. Conditional Mean Estimates

We now consider parameter estimation based on conditional mean estimates in lieu of the true states. We define a model set $\Lambda := \{\lambda(B, D, x_0) \mid B, D, x_0 \in R\}$ of allowable model estimates $\hat{\lambda}_k \in \Lambda$ and assume the true model λ lies in the model set Λ. We denote the history or sequence of estimates as $\hat{\Lambda}_k := \{\hat{\lambda}_1, \ldots, \hat{\lambda}_k\}$. Let us denote the associated conditional mean estimates based on the model estimates $\hat{\Lambda}_k$ as

$$\hat{J}_{k|k,\hat{\Lambda}_k} = E[J_k \mid \mathcal{Y}_k, \hat{\Lambda}_k], \qquad \hat{O}_{k|k,\hat{\Lambda}_k} = E[O_k \mid \mathcal{Y}_k, \hat{\Lambda}_k]$$

$$\hat{H}_{k|k,\hat{\Lambda}_k} = E[H_k \mid \mathcal{Y}_k, \hat{\Lambda}_k], \qquad \hat{T}_{k|k,\hat{\Lambda}_k} = E[T_k \mid \mathcal{Y}_k, \hat{\Lambda}_k] \tag{18}$$

Optimal finite-dimensional recursions for these conditional mean estimates can be found in Ref. 15. Hence, we propose estimators for B and D based on observations $\{y_k\}$ as

$$\hat{B}_{k|\hat{\Lambda}_k} = \text{proj}\left[\left(\hat{O}_{k|k,\hat{\Lambda}_k} + \hat{H}_{k|k,\hat{\Lambda}_k} - 2\hat{J}_{k|k,\hat{\Lambda}_k}\right)/k\right] \tag{19}$$

$$\hat{D}_{k|k,\hat{\Lambda}_k} = \text{proj}\left[\left(Q_k + \hat{O}_{k|k,\hat{\Lambda}_k} - 2\hat{T}_{k|k,\hat{\Lambda}_k}\right)/k\right] \tag{20}$$

where

$$\hat{\lambda}_{k+1} = \lambda\left(\hat{B}_{k|\hat{\Lambda}_k}, \hat{D}_{k|\hat{\Lambda}_k}, \hat{x}_0\right), \qquad \hat{\Lambda}_{k+1} = \{\hat{\lambda}_1, \ldots, \hat{\lambda}_{k+1}\} \tag{21}$$

and $\text{proj}(x) = \max(\delta, x)$, where $\delta (0 < \delta < B, D)$ is a small constant that ensures $\hat{B}_{k|\hat{\Lambda}_k}$ and $\hat{D}_{k|\hat{\Lambda}_k}$ remain positive.

We consider first the situation in which conditional mean estimates based on the true model λ are available.

Lemma 3: Consider the linear system (3) and (4) denoted by λ. Suppose $\{y_k\}$ exists and is measured. Also assume that conditional mean estimates based on the correct model are available. Assume the true model λ satisfies $\lambda \in \Lambda$. Then

$$\lim_{k\to\infty} \hat{B}_{k|k,\lambda}, \hat{D}_{k|k,\lambda} = B, D \quad \text{a.s.} \tag{22}$$

Proof: See the Appendix.

Remarks:

1) The projection step is only a technical point to ensure the estimates remain positive.

2) A finite data set result that extends Lemma 3 can be established using the weak law of large numbers.

The following theorem holds.

Theorem 1: Consider the linear system (3) and (4) denoted by λ. Consider a sequence of estimated models $\hat{\Lambda}_k$ adaptively updated by previous parameter estimates as shown in Eq. (21). Then the recursion converges a.s. to the set S (or possibly converges to the projection boundaries $\bar{B} = \delta$ and $\bar{D} = \delta$), where

$$S := \left\{\bar{B}, \bar{D}: \frac{\mathrm{d}V(\bar{B}, \bar{D})}{\mathrm{d}\bar{B}} = 0 \text{ and } \frac{\mathrm{d}V(\bar{B}, \bar{D})}{\mathrm{d}\bar{D}} = 0\right\} \tag{23}$$

where $\bar{\lambda} = \lambda(\bar{B}, \bar{D}, \bar{x}_0)$. That is, S contains the local minimum of the cost function

$$V(\bar{B}, \bar{D}) := E\left[\left((x_k - x_{k-1})^2 - \bar{B}\right)^2 + \left((y_k - x_k)^2 - \bar{D}\right)^2 \mid \bar{\lambda}\right]$$

Proof: See the Appendix.

Remarks:

1) This establishes local convergence to the true model because it follows from Lemma 3 that $\lambda \in S$.

2) Convergence rates have not been established but can be shown using an approach similar to that in Ref. 11.

3) Local convergence results for estimation from finite data sets can be established in the offline situation when the recursions (19) and (20) are passed over the data set multiple times and the model estimate is updated after each pass. In this multipass offline situation, the estimators are an example of the expectation maximization algorithm and local convergence results are available.[16] Convergence results in the situation of a single pass through a finite data set that are analogous to Lemma 2 have not yet been established.

III. Implementation Issues

In this section we discuss some implementation issues relating to the algorithm just given.

A. Filters

Implementation of the estimator defined by Eqs. (19–21) requires the quantities $\hat{J}_{k|k,\hat{\Lambda}_k}$, $\hat{O}_{k|k,\hat{\Lambda}_k}$, $\hat{H}_{k|k,\hat{\Lambda}_k}$, and $\hat{T}_{k|k,\hat{\Lambda}_k}$ as defined in Eq. (18) to be calculated. Optimal finite-dimensional recursions for $\hat{J}_{k|k,\hat{\Lambda}_k}$, $\hat{O}_{k|k,\hat{\Lambda}_k}$, $\hat{H}_{k|k,\hat{\Lambda}_k}$, and $\hat{T}_{k|k,\hat{\Lambda}_k}$ are given in Ref. 15. These estimators are not repeated here to save space.

B. Online or Batch Processing

It is possible to use the algorithm presented in either an online or batch manner. The recursion (19) and (20) can be iterated as each new data point arrives to produce a new estimate. Alternatively, if only a finite data set is available, then the algorithm can be iterated through the data set multiple times with the parameter estimates improving on each successive pass through the data.

In a multipass situation, convergence results follow by considering the data set as the concatenation of copies of the finite data set and convergence occurs to a minimum of Eq. (A10) (see the Appendix). It has been shown that resolution errors result from working on a finite date set.[17] These resolution errors will be apparent on this concatenated data set, and estimation bias may result.

C. Initialization

The algorithm requires an initial value for the angular random walk and rate random walk coefficients. In our experience the algorithm will converge to the true angular random walk and rate random walk coefficients from a reasonable initialization. In particular, for the studies to be presented, convergence to the true values occurs whenever $\hat{B}_0$ is significantly smaller that $\hat{D}_0$. For typical applications (IMU measurements), this requirement is not restrictive.

The filters from Ref. 15 should be allowed some data to initialize, and we suggest that these filters be iterated for at least 500 data points before parameter estimation is started.

D. Improved Convergence Rates

The rate of convergence of adaptive algorithm (19–21) from initial parameter estimates can be improved. Using the difference form of the estimators (A3), convergence rates can be improved using the Polyak and Juditsky technique.[18]

Consider the following estimation algorithm for B (the estimator for D can be modified in a similar way):

$$\hat{B}_{k+1|k+1,\hat{\Lambda}_{k+1}} = \hat{B}_{k|k,\hat{\Lambda}_k} + \gamma_k\left(\Delta\hat{O}_{k|k,\hat{\Lambda}_k} + \Delta\hat{H}_{k|k,\hat{\Lambda}_k} - 2\Delta\hat{J}_{k|k,\hat{\Lambda}_k} - \hat{B}_{k|k,\hat{\Lambda}_k}\right) \tag{24}$$

where $\Delta\hat{O}_{k|k,\hat{\Lambda}_k} := \hat{O}_{k|k,\hat{\Lambda}_k} - \hat{O}_{k-1|k-1,\hat{\Lambda}_{k-1}}$ and $\Delta\hat{H}_{k|k,\hat{\Lambda}_k}$ and $\Delta\hat{J}_{k|k,\hat{\Lambda}_k}$ are defined in a similar way. If $\gamma_k = 1/k$, then Eq. (24) is algebraically equivalent to Eq. (19). To improve the convergence rate of the algorithm, a Polyak and Juditsky[18] step is used instead, $\gamma_k = 1/k^p$, where $0 < p \leq 1$, and the estimates from Eq. (24) are averaged to reduce their variation (see Ref. 18 for more details).

Good choices for p depend on the relative power of the w and v noises. In the examples examined in the Sec. IV, the gyro's major source of noise is angular random walk. Increasing the step size for the B recursion allows the recursion longer time to forget initial estimates (relative to the D recursion), and this facilitates convergence from poor initializations.

Other techniques, such as forgetting factors, can also be useful in improving convergence rates from poor initial parameter estimates.[19]

E. Other Noise Sources

The algorithm presented is proposed under the assumption that the gyro output error has contributions only from angular random walk and rate random walk. For devices where the gyro error is dominated by angular random walk and rate random walk noises but there are small contributions from other noise sources, the bias introduced will be small (except perhaps if there is a pathological flicker noise that has infinite variance). For devices where there are significant contributions from other noise sources, then the estimates from the proposed algorithm will be significantly biased away from the actual B and D.

However, note that although biased estimated B and D will still minimize the cost function $V(B, D)$ given in Theorem 1, this minimum simply no longer occurs at actual values of B and D. The proposed algorithm estimates the B and D that best fit the data in the sense of $V(B, D)$. In situations where only a rough feeling for the device's error behavior is required, this information still may be quite useful.

F. Computational Effort

The computational effort required to implement the recursions (19–21) is approximately 100 floating point calculations per data point (and does not depend on the length of the data). This includes implementation of the optimal filters for $\hat{J}_{k|k,\hat{\Lambda}_k}$, etc. Issues of computational effort for the Allan variance technique (which will depend on implementation choices) are not discussed in the many references. If the data set has T points, then to calculate the Allan variance of size N requires $\mathcal{O}[T(1 + 3/N)]$ calculations. Because the Allan deviation is plotted on a log-log graph, the Allan variance needs at least $K\log(T)$ points to be plotted, where K is the number of points per order of magnitude. The computational effort is, hence, at least $\mathcal{O}[KT\log(T)]$ compared to the $\mathcal{O}(100T)$ required to implement Eqs. (19–21). For large data sets the estimators presented here are computationally superior. In addition, the recursions (19–21) always have the advantage that the data do not need to be stored.

G. Numerical Overflow

When implemented in an online manner, the algorithm needs to be modified to ensure overflow of $\hat{O}_{k|k,\hat{\Lambda}_k}$, etc., does not occur on a very large data set, that is, for $k > 2^{M_p}$, where M_p is the precision of the floating point numbers. Scaling or forgetting factors could be introduced into the filters of Ref. 15 to overcome this overflow issue. Alternatively, the filters could be reset after a reasonable amount of data and then reinitialized on data to avoid numerical overflow.

IV. Results

In this section we examine the proposed technique for estimating the angular random walk and rate random walk coefficients. First, we compare the proposed technique with the Allan variance method using computer-generated data with known noise characteristics. Then, we apply the proposed technique on a data set taken from a low-grade IMU (the same data set analyzed using the Allan variance method earlier). This low-grade IMU, contains ring laser gyros and accelerometers and produces angular rate and linear acceleration signals in the X, Y, and Z directions. The analysis presented is of angular rate data about the Z axis. The device was tested statically.

A. Computer-Generated Data

These computer simulations were performed using the MATLAB® package.

A 40,000 point sequence was generated for linear system (3) and (4) with $B = 0.001$ and $D = 1$.

The Allan variance was calculated, and Fig. 4 shows the log-log Allan deviation plot. Lines of worst and best fit were manually placed on the graph (the best-fit lines were visually based and the worst-fit lines used the data points farthest from the best-fit lines). The lines of best fit are shown in Fig. 4. From intercepts of the lines with $\pm\frac{1}{2}$ slopes, the angular random walk and rate random walk coefficients K^{av} and N^{av} were estimated. The results are given in Table 1.

From the results in Table 1, it appears that the characteristics of the generated noise do not correspond exactly to the system parameters $B = 0.001$ and $D = 1$, that is, $K^{\text{av}} = 0.0316$ and $N^{\text{av}} = 1$. This is because the data set is finite and the random noise sources in the model are approximated by pseudorandom noises. The Allan deviation graph reflects the actual characteristics of the data, and hence, the coefficients estimated from this graph are used as a basis for testing our proposed algorithm (rather than the parameters used to generate the data).

Table 1 Estimated coefficients/parameters

Allan variance method	Proposed method
Simulated data	
$\hat{K}^{\text{av}} = 0.0361\,(-0.0022, +0.0014)$	$\hat{K}^{\text{est}} = 0.0346 \Leftarrow \hat{B} = 0.0012$
$\hat{N}^{\text{av}} = 0.9894\,(-0.0324, +0.1060)$	$\hat{N}^{\text{est}} = 1.0011 \Leftarrow \hat{D} = 1.0023$
IMU data	
$\hat{K}^{\text{av}} = 0.0028\,(-0.0005, +0.0005)$	$\hat{K}^{\text{est}} = 0.00275 \Leftarrow \hat{B} = 7.6 \times 10^{-6}$
$\hat{N}^{\text{av}} = 0.1550\,(-0.0120, +0.0090)$	$\hat{N}^{\text{est}} = 0.1582 \Leftarrow \hat{D} = 0.0250$

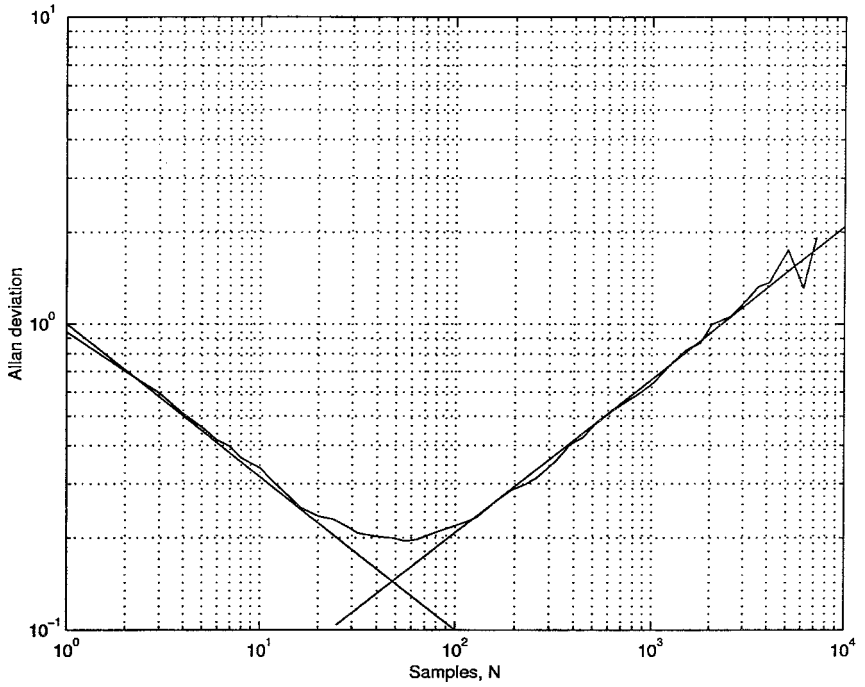

Fig. 4 Generated data: Allan deviation vs sample size; lines of best fits added.

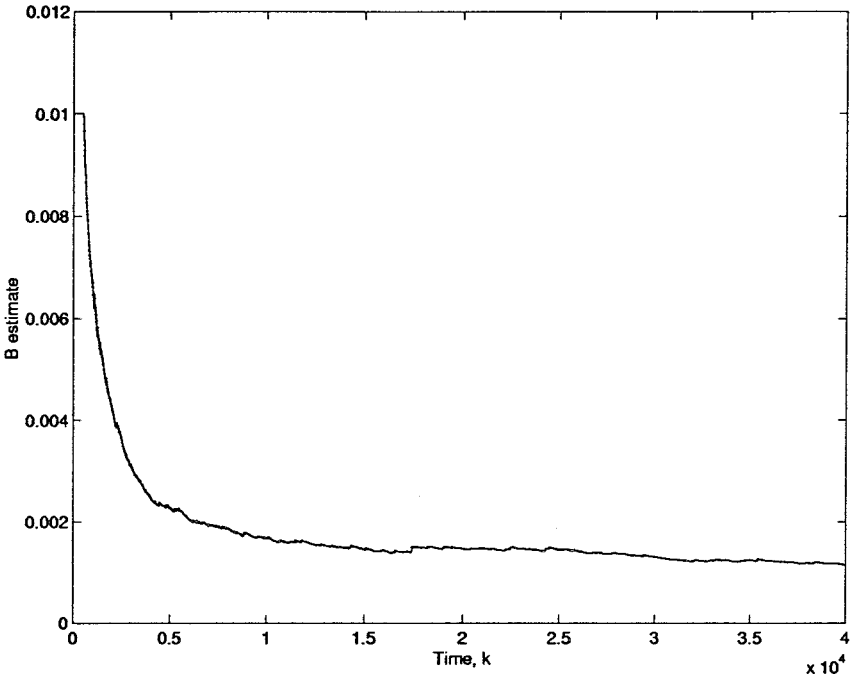

Fig. 5 Estimates of B vs time.

From initial estimates, $\hat{B}_0 = 0.1$ and $\hat{D}_0 = 1.5$, algorithm (19) was used to estimate the B and D coefficients. The final estimates after 40,000 points are shown in Table 1 (the values $\hat{K}^{\text{est}} = \sqrt{\hat{B}}$ and $\hat{N}^{\text{est}} = \sqrt{\hat{D}}$ are the parameter estimates in the Allan variance technique form). The estimated values are well within the error bounds of the estimates obtained from the Allan deviation graph.

Convergence of the estimates $\hat{B}$, $\hat{D}$ is shown in Figs. 5 and 6, respectively. The proposed algorithm was tested using other initial estimates, and convergence occurred from all initial estimates for which $\hat{B}_0 \ll \hat{D}_0$. If the algorithm was initialized with $\hat{B}_0 > \hat{D}_0$, then convergence was not to the values estimated by the Allan variance method (but to a local minimum). This type of local convergence can be avoided by using initial estimates that satisty $\hat{B}_0 \ll \hat{D}_0$.

B. Low-Grade IMU

In this section, our proposed algorithm and the Allan variance method are compared on data from a low-grade IMU. From the Allan deviation plot (Fig. 2) the angular random walk and rate random walk coefficients can be estimated using lines of best fit. The parameter estimates $\hat{K}^{\text{av}}$ and $\hat{N}^{\text{av}}$ obtained from this analysis are given in Table 1.

Our proposed algorithm was then applied to the same measurements set. The algorithm was initialized with parameter values $\hat{B}_0 = 1 \times 10^{-5}$ and $\hat{D}_0 = 0.01$. After several passes of the data, the algorithm had converged to the values $\hat{B}$ and $\hat{D}$ shown in Table 1. The estimated values correspond well to the coefficients estimated from the Allan deviation plot. In Fig. 2, lines based on the estimated values have been added to demonstrate the results of the proposed

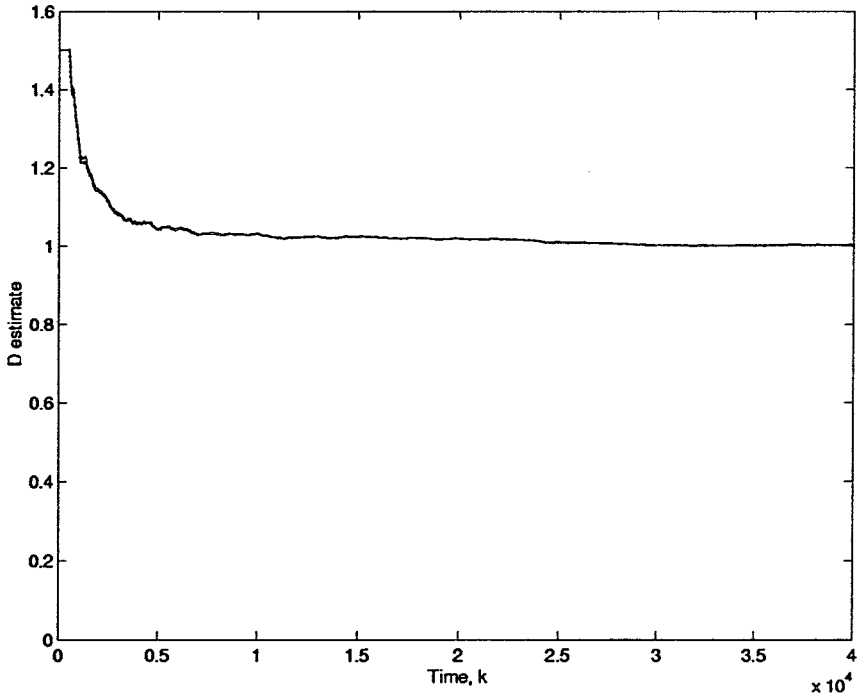

Fig. 6 Estimates of D vs time.

algorithm with respect to the Allan deviation graph for measured data.

V. Conclusions

A new technique is presented for estimating the parameters in the noise model of IMU gyros. The existing techniques for analysis of IMU gyros require large amounts of data to be stored and to be processed offline to produce an Allan deviation graph. The technique proposed here can analyze data as it arrives from the gyro and can produce immediate estimates of the noise model in the sense that it does not require the user to manually analyze an Allan variance graph.

Appendix: Proofs of Lemma 3 and Theorem 1

Proof of Lemma 3: The proof follows Lemma 1. We consider estimation of B first. We rewrite estimate (19) without the projection step as

$$\hat{B}_{k\,|\,\lambda} = E[BM_k \mid \mathcal{Y}_k, \lambda]/k = B\{E[M_k \mid \mathcal{Y}_k, \lambda]/k\} \tag{A1}$$

Hence, the estimate is

$$\begin{aligned}\hat{B}_{k\,|\,\lambda} &= B\{E[M_k \mid \mathcal{Y}_k, \lambda]/k\}\\ &= BE[M_k/k \mid \mathcal{Y}_k, \lambda]\\ &= BE\{E[M_k/k \mid \mathcal{G}_k\mathcal{Y}_k, \lambda] \mid \mathcal{Y}_k, \lambda\}\end{aligned}$$

Hence,

$$\lim_{k\to\infty} \hat{B}_{k|\lambda} = B \quad \text{a.s.} \tag{A2}$$

We have used the property that $E[X|A_1] = E\{E[X|A_2]|A_1\}$, when $A_1 \subset A_2$, and the strong law of large number result, $\lim_{k\to\infty} (1/k)E[M_k \mid \mathcal{G}_k] = 1$ a.s. Because $\phi_k = (1/k)E[M_k \mid \mathcal{G}_k, \mathcal{Y}_k, \lambda]$ is uniformly integrable, that is, $E[\phi_k\phi_k']$ is bounded, and $\lim_{k\to\infty} \phi_k = 1$ a.s., then there is convergence in the conditional mean:

$$\lim_{k\to\infty} E[\phi_k \mid \mathcal{Y}_k, \lambda] = 1 \quad \text{a.s.}$$

The lemma result follows for $\hat{D}_{k\,|\,\lambda}$ similarly. □

Proof of Theorem 1: First we consider estimation of B only. Simple manipulation of Eq. (19) gives a recursion for the estimate $\hat{B}_{k\,|\,k,\hat{\Lambda}_k}$ as follows:

$$\hat{B}_k = \hat{B}_{k-1} + (1/k)(\Delta\hat{O}_k + \Delta\hat{H}_k - 2\Delta\hat{J}_k - \hat{B}_{k-1}) \tag{A3}$$

where we denote $\hat{B}_{k\,|\,k,\hat{\Lambda}_k}$ by the shorthand notation $\hat{B}_k$, we define $\Delta\hat{O}_k := \hat{O}_{k\,|\,k,\hat{\Lambda}_k} - \hat{O}_{k-1\,|\,k-1,\hat{\Lambda}_{k-1}}$, and we define $\Delta\hat{H}_k$ and $\Delta\hat{J}_k$ similarly. To simplify the presentation, we will initially ignore the projection step. Convergence results will still hold if projection is performed (see Ljung[20]).

We follow the technique presented in Ref. 11. This proof relies on the ordinary differential approach described in Refs. 19 and 21–22 (see Ref. 22 for detailed proofs regarding algorithms with projection steps). In the ordinary differential equation (ODE) approach, convergence of a difference equation is established by considering the convergence properties of an associated ODE.

Convergence of Eq. (A3) can be analyzed by considering the following ODE:

$$\frac{\mathrm{d}\bar{B}(\tau, k)}{\mathrm{d}\tau} = \frac{1}{k} f[\bar{B}(\tau, k), k] \tag{A4}$$

where, for the purposes of establishing convergence of the ODE, k is considered a fixed parameter. With $\bar{B}(\tau, k)$ abbreviated as $\bar{B}$, we define $f(\bar{B}, k)$ as follows:

$$f(\bar{B}, k) := E[\Delta\hat{O}_{k\,|\,\bar{B}} + \Delta\hat{H}_{k\,|\,\bar{B}} - 2\Delta\hat{J}_{k\,|\,\bar{B}} - \bar{B}] \tag{A5}$$

We have explicitly shown in the notation that $\Delta\hat{O}_k$, etc., depend on the parameter estimate $\bar{B}$.

In Ref. 20, it is stated that Eq. (A3) will converge to the set $\bar{S}$ or a boundary, where $\bar{S} := \{\bar{B} \mid \lim_{k\to\infty} f(\bar{B}, k) = 0\}$, if the following hold: 1) certain regularity and exponential stability conditions hold as listed in Ref. 20 and 2) ODE (A4) (which is parameterized by k) is asymptotically stable in the limit as $k \to \infty$.

For this system, the regularity conditions required to satisfy condition 1 are shown and discussed in Refs. 11 and 19. Condition 1 also requires that the conditional mean filters for $\hat{J}_{k\,|\,k,\hat{\Lambda}_k}$, $\hat{O}_{k\,|\,k,\hat{\Lambda}_k}$, $\hat{H}_{k\,|\,k,\hat{\Lambda}_k}$, and $\hat{T}_{k\,|\,k,\hat{\Lambda}_k}$ be exponentially forgetting. The exponential forgetting property has not yet been established for these filters, but we will assume that the exponential forgetting property holds.

To establish condition 2, we use a Lyapunov function approach.[19] Consider the function

$$W(\bar{B}, k) = \tfrac{1}{2} E\left\{\left[(x_k - x_{k-1})^2 - \bar{B}\right]^2\right\} \tag{A6}$$

It follows from classical expectation results, including that $E\{E[X|A_2]|A_1\} = E[X|A_1]$ when $A_1 \subset A_2$, that $W(\bar{B}, k) \geq 0$. Under ergodicity and certain smoothness conditions, the differentiation with respect to $\bar{B}$ and the expectation operations can be interchanged. Hence,

$$\frac{\mathrm{d}W[\bar{B}(\tau, k), k]}{\mathrm{d}\bar{B}(\tau, k)} = -f[\bar{B}(\tau, k), k] \tag{A7}$$

and it then follows that

$$\frac{\mathrm{d}W[\bar{B}(\tau, k), k]}{\mathrm{d}\tau} = \frac{\mathrm{d}W[\bar{B}(\tau, k), k]}{\mathrm{d}\bar{B}(\tau, k)} \frac{\mathrm{d}\bar{B}(\tau, k)}{\mathrm{d}\tau}$$

$$= -f[\bar{B}(\tau, k), k](1/k) f[\bar{B}(\tau, k), k] \tag{A8}$$

$$< \quad \text{if} \quad f[\bar{B}(\tau, k), k] \neq 0 \tag{A9}$$

It follows that $W(\bar{B}, k)$ is a Lyapunov function.

It follows from Lynapov's direct method and Eq. (A7) that $\bar{B}(\tau, k)$ converges to the set $\{\bar{B} \mid \lim_{k \to \infty} f(\bar{B}, k) = 0\}$ (or possibly to the boundary $\bar{B} = \delta$ if a projection step is implemented). Convergence of the difference equation (A3) follows from Ljung.[20]

We note that under asymptotic ergodicity (and other regularity conditions given by Ljung[20] and shown in Ref. 11) and using the results by Ljung[20] the difference equation (A3) and the ODE (A4) converge to the same set. That is, in the limit $k \to \infty$ convergence to the set $\{\bar{B} \mid \lim_{k \to \infty} f(\bar{B}, k) = 0\}$ is equivalent to the set of local minima of $E\{[(x_k - x_{k-1})^2 - \bar{B}]^2 | \bar{B}, \mathcal{Y}_k\}$.

The result for simultaneous estimation follows using a similar approach by writing the estimate $\hat{D}_{k|k,\hat{\Lambda}_k}$ as a recursion and using the Lyapunov function

$$W(\bar{B}, \bar{D}, k) = \tfrac{1}{2}\left(E\left\{\left[(x_k - x_{k-1})^2 - \bar{B}\right]^2\right\}\right.$$

$$\left. + E\left\{\left[(y_k - x_k)^2 - \bar{D}\right]^2\right\}\right) \tag{A10}$$

Note that only a local convergence result is established when simultaneous estimation is required (which is normally the case) because the set S may contain more than one point. □

References

[1]Titterton, D., and Weston, J., *Strapdown Inertial Navigation Technology*, Peter Peregrinus, London, 1997, pp. 1–54.

[2]Lawrence, A., *Modern Inertial Technology: Navigation, Guidance, and Control*, Springer–Verlag, New York, 1993, pp. 4–19.

[3]Allan, D., "Statistics of Atomic Frequency Standards," *Proceedings of the IEEE*, Vol. 54, No. 2, 1996, pp. 221–230.

[4]Tehrani, M., "Cluster Sampling Technique for RLG Noise Analysis," *Seventeenth Biennial Guidance Test Symposium*, Central Inertial Guidance Test Facility, Holloman AFB, NM, May 1995, pp. 66–76.

[5]Erickson, G., "An Overview of Dynamic and Stochastic Modeling of Gyros," *Proceedings of the 1993 National Technical Meeting of the ION*, Inst. of Navigation, 1993, pp. 339–351.

[6]*IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Laser Gyros: IEEE STD 647-1995*, Inst. of Electrical and Electronics Engineers, New York, 1996.

[7]*IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros: IEEE STD 952-1997*, Inst. of Electrical and Electronics Engineers, New York, 1998.

[8]Bielas, M., "Stochastic and Dynamic Modeling of Fiber Gyros," *Proceedings of the SPIE—Fiber Optic and Laser Sensors XII*, Vol. 2292, 1994, pp. 240–254.

[9]Ng, L., and Pines, D., "Characterization of Ring Laser Gyro Performance Using the Allan Variance Method," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 1, 1997, pp. 211–214.

[10]Sargent, D., and Wyman, B., "Extraction of Stability Statistics from Integrated Rate Data," AIAA Guidance and Control Conf., Denvers, MA, Aug. 1980, pp. 88–94.

[11]Ford, J., "Adaptive Hidden Markov Model Estimation and Applications," Ph.D. Dissertation, Dept. of Mechanical Engineering, Australian National Univ., Jan. 1998, Chap. 5.

[12]Kumar, P., and Varaiya, P., *Stochastic Systems: Estimation, Identification, and Adaptive Control*, Prentice–Hall, Englewood Cliffs, NJ, 1986.

[13]Söderström, T., and Stoica, P., *System Identification*, Prentice–Hall, New York, 1989.

[14]Billingsley, P., *Probability and Measure*, 3rd ed., Wiley, New York, 1995, pp. 17–32, 85, 86.

[15]Elliott, R., and Krishnamurthy, V., "Finite Dimensional Filters for ML Estimation of Discrete-Time Gauss–Markov Models," *Proceedings of IEEE Conference on Decision and Control CDC97*, Vol. 2, 1997, pp. 1637–1642.

[16]McLachlan, G., and Krishnan, T., *The EM Algorithm and Extensions*, Wiley, New York, 1997, pp. 82–109.

[17]Tehrani, M., "Ring Laser Gyro Data Analysis with Cluster Sampling Technique," *Proceedings of SPIE—Fibre Optic and Laser Sensors*, Vol. 412, Society of Photo-Optical Instrumentation Engineers, 1983, pp. 207–210.

[18]Polyak, B., and Juditsky, A., "Acceleration of Stochastic Approximation by Averaging," *SIAM Journal on Control and Optimization*, Vol. 30, No. 4, 1992, pp. 838–855.

[19]Ljung, L., and Söderström, T., *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1993, pp. 271–289.

[20]Ljung, L., and Söderström, T., *Theory and Practice of Recursive Identification*, MIT Press, 1993, Chap. 4 and pp. 158–170.

[21]Gerencsér, L., "Rate of Convergence of Recursive Estimators," *SIAM Journal on Control and Optimization*, Vol. 30, No. 5, 1992, pp. 1200–1227.

[22]Kushner, H., and Yin, G. G., *Stochastic Approximation Algorithms and Applications*, Springer–Verlag, New York, 1997, Chap. 5.