On-line Estimation of Allan Variance Parameters

J.J. Ford and M.E. Evans

DSTO-RR-0141

APPROVED FOR PUBLIC RELEASE

© Commonwealth of Australia

DEPARTMENT OF DEFENCE
DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
ON-LINE ESTIMATION OF ALLAN VARIANCE PARAMETERS

J.J. Ford and M.E. Evans

Weapons Systems Division
Aeronautical and Maritime Research Laboratory

DSTO-RR-0141

ABSTRACT

A new on-line method is presented for estimation of the angular random walk and rate random walk coefficients of IMU (Inertial Measurement Unit) gyros and accelerometers. The on-line method proposes a state space model and proposes parameter estimators for quantities previously measured from off-line data techniques such as the Allan variance graph. Allan variance graphs have large off-line computational effort and data storage requirements. The technique proposed here requires no data storage and computational effort of $O(100)$ calculations per data sample.
On-line Estimation of Allan Variance Parameters

EXECUTIVE SUMMARY

Inertial navigation is an important technology in which measurements provided by gyroscopes and accelerometers are used to determine position for modern aircraft. Because of the importance of these devices in modern navigation, a full understanding of their noise characteristics is important.

The most commonly used tool for analyzing the noise characteristics of gyros and accelerometers is the Allan variance graph method which suffers from several deficiencies. The Allan variance graph method requires the data to be collected off-line before processing can be performed. It also requires the user to read slopes and intercepts manually off the Allan variance graph so that noise characteristics can be calculated.

The key contribution of this paper is a new technique that can analyze data as it is received and calculates on-line the contribution from various noise sources. The proposed method, unlike the Allan variance graph method, does not require large amounts of data to be stored. The proposed method also has the additional advantage, if the data set is large, of requiring less computational effort overall.

Convergence results are presented for the proposed technique based on the law of large numbers and an ordinary differential equation approach. The technique is verified on both simulated data and data obtained from a HG1700-AE11 Honeywell IMU (Inertial Measurement Unit) by comparison with the results from an Allan variance graph analysis.

The new methodology represents a fundamental improvement in the testing procedures for inertial navigation systems and will enable a more efficient response to ADF requirements for IMU modelling in support of weapon system performance assessments.
Authors

J.J. Ford
*Weapons Systems Division*

J.J. Ford joined the Guidance and Control group in the Weapons Systems Division in February 1998. He received B.Sc. (majoring in Mathematics) and B.E. degrees from the Australian National University in 1995. He also holds the PhD (1998) degree from the Australian National University. His thesis presents several new on-line locally and globally convergent parameter estimation algorithms for hidden Markov models and linear systems.

His research interests include: adaptive signal processing, adaptive parameter estimation and risk-sensitive techniques.

M.E. Evans
*Weapons Systems Division*

M.E. Evans received the B.E. and M.Eng.Sc. degrees in Mechanical Engineering from the University of Queensland in 1974 and 1976 respectively. He also holds the M.S., M.Phil. and PhD (1981) degrees in control theory from Yale University, and pursued part-time studies for a time in the Faculty of Law at the University of Western Australia. He lectured in Mechanical Engineering at The University of Western Australia before joining the DSTO in 1991 as a Senior Research Scientist, where he has been working in the area of guidance and control of guided weapons.
THIS PAGE IS INTENTIONALLY BLANK
## Contents

1 Introduction .................................................. 1

2 Stochastic Model .............................................. 2
   2.1 State Space Model ........................................ 3

3 Parameter Estimation - Full Observations .................. 4
   3.1 Almost Sure Convergence .................................. 5
   3.2 Finite Data Sets ......................................... 5

4 Conditional Mean Estimates .................................. 6

5 Implementation ................................................ 9

6 Simulation studies ........................................... 11
   6.1 Computer Generated data .................................. 11
   6.2 The Honeywell IMU (HG1700-AE11) ....................... 12
   6.3 Discussion ................................................ 12

7 Conclusions .................................................. 13

References ...................................................... 13
1 Introduction

Inertial navigation is the process by which measurements made using gyroscopes and accelerometers are used to determine position\cite{1, 2}. Inertial navigation is an essential technology for modern aircraft and an important tool for navigation in general. Accelerometer and gyro IMUs (Inertial Measurement units) suffer from particular types of noise interferences which induce IMU navigation errors which 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\cite{3}, noise from ring laser gyros\cite{4} and gyro sensors in general\cite{5, 6}.

Characterizing noise produced by IMU sensors by a single RMS number was overly conservative in the short term and did not adequately model the longer term error growth. The Allan variance method was developed to better characterize the noise model and is now the standard method of analysis\cite{6}. The Allan variance, $A(N)$, of a sensor output is the variance of the means of successive subsets of the data of size $N$.

When the Allan variance, $A(N)$, is plotted against the sample size, $N$, particular noise features in the gyro output appear as structures in the Allan variance curve, as illustrated in \cite{4, 6, 7} and Figure 5 in this paper.

There are several disadvantages of the Allan variance technique. Obviously, by its very nature, it is an off-line 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 to be read off to obtain estimates of the noise contributions. Unfortunately, the Allan variance technique does not characterize the noise well when contributions from various sources overlap on the Allan variance graph.

This paper works in the stochastic model framework proposed in \cite{5}. From this stochastic model, we develop a state space model that models only the noise contributions from the angular random walk (ARW) and rate random walk (RRW).

The key contribution of this paper is the proposal of on-line parameter estimation algorithms for the ARW and RRW constants. 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 10000, the proposed on-line technique is computationally superior to the Allan variance technique (a typical data set might contain 24hrs of 1Hz measurements, ie. $\sim 10^7$ data points).

Simulation studies are presented both for computer generated data and for data from a Honeywell IMU. Results from the on-line estimation algorithm are compared to estimates from the Allan variance technique.

This paper is organized as follows: In Section 2, we introduce a state space model for angular and rate random walks. In Section 3, parameter estimators are introduced and a partial convergence result is established in the situation, admittedly artificial, when the error model state variable is measurable. In Section 4, we introduce a conditional mean estimate and an estimation algorithm for the more realistic situation where only output measurements are available. Local convergence results are established using an ordinary
differential equation approach. In Section 5, implementation issues are discussed including initialization, computational effort and techniques for improving convergence rates from poor initializations. Simulation studies are presented in Section 6. Finally, in Section 7 some conclusions are presented.

2 Stochastic Model

Gyro sensor error dynamics can be modelled by the discrete-time stochastic model presented in Figure 1, see also [7, 4]. In this model there are contributions from angular random walk noise ($W_D$), rate random walk noise ($W_B$), flicker noise ($F$) and ramp noise ($R$). Not shown on the figure are contributions from quantization noise, Markov noise, or sinusoidal errors.

The terms “angular random walk” and “rate random walk” are slightly misleading because angle rate $\Delta \theta$ is often measured directly and these terms would normally be seen from a linear system’s viewpoint as measurement noise and process noise (which produces a classical random walk in the output).

In this paper, we propose a simplified stochastic model in which the only non-zero contributions are the angular random walk and rate random walk terms. This simplification seems appropriate for some of the sensors that have been examined via the alternative Allan variance graph technique. Consider Figure 5 in this paper which shows the Allan Variance plot for the output of the gyro examined in this paper. The dominate noise sources appear to be an angular random walk component (the $-\frac{1}{2}$ slope part of the curve) and a rate random walk component (the $+\frac{1}{2}$ slope curve). A discussion of Allan variance plots can be found in [4, 6, 7].

The simplified stochastic model shown in Figure 2 can be represented as a state space model. The unknown angular random walk and rate random walk coefficients are parameters of this representation. Estimation of these quantities can then be examined from a linear system’s viewpoint [8, 9, 10]. In the following section we define our mathematical framework for the parameter estimation problem.
2.1 State Space Model

Consider a probability measure space \((\Omega, \mathcal{F}, P)\). Here \(\Omega\) is an arbitrary space or set of points \(\omega\), \(\mathcal{F}\) is \(\sigma\)-algebra in \(\Omega\) (a class of subsets containing \(\Omega\) and closed under the formation of complements and countable unions) and \(P\) is a probability measure on \(\mathcal{F}\). See [11, Pages 17-23] for more details. In some sense, each point \(\omega\) corresponds to a triple \(\{w_\ell\}, \{v_\ell\}, x_0\), where \(w_\ell\), \(v_\ell\) and \(x_0\) are defined below. Suppose \(\{x_\ell\}, \ell \in \mathbb{Z}^+\) (\(\mathbb{Z}^+\) is the set of positive integers) is a discrete-time linear stochastic process, taking values in \(\mathbb{R}\), with dynamics given by

\[
x_{k+1} = x_k + \sqrt{B} \, w_{k+1}, \quad x_0 \in \mathbb{R}
\]

Here \(k \in \mathbb{Z}^+\), \(\sqrt{B}\) is the rate random walk coefficient and \(\{w_\ell\}, \ell \in \mathbb{Z}^+,\) is a sequence of independent and identically distributed \(N(0, 1)\) scalar random variables. The notation \(N(0, 1)\) is short hand for a random variable whose density is Gaussian with zero mean and variance 1.

The state process \(x_\ell, \ell \in \mathbb{Z}^+\), is observed indirectly via the scalar observation process \(\{y_\ell\}, \ell \in \mathbb{Z}^+,\) given by

\[
y_k = x_k + \sqrt{D} \, v_k \in \mathbb{R}
\]

Here \(k \in \mathbb{Z}^+, \sqrt{D}\) is the angular random walk coefficient and \(\{v_\ell\}, \ell \in \mathbb{Z}^+,\) 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. The sequence \(\{y_\ell\}\) is the sequence of angle rates measured by the IMU (or \(\Delta \theta\) in Figure 2). The variance of noise \(W_B\) is \(B\) and the variance of noise \(W_D\) is \(D\). We have considered here only single axis devices but in principle this technique can be extended to 3 dimensions with \(B\) and \(D\) becoming covariance matrices.

Let \(\mathcal{G}_k\) and \(\mathcal{Y}_k\) denote the complete filtrations generated by \(\{x_\ell\}\) and \(\{y_\ell\}\) respectively. For example, \(\mathcal{G}_k\) is the \(\sigma\)-algebra generated by \(x_0, x_1, \ldots, x_k\), denoted \(\sigma\{x_0, x_1, \ldots, x_k\}\), augmented by including all subsets of events of probability zero. See [12] for more details about filtrations.

The model described by (2.1),(2.2) is denoted by

\[
\lambda = \lambda(B, D, x_0)
\]

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

\[
\lim_{k \to \infty} z_k(\omega) = z(\omega) \quad (2.4)
\]

for all \( \omega \) in an \( \mathcal{F} \)-set of probability 1, then

\[
\lim_{k \to \infty} z_k = z \text{ almost surely or a.s.} \quad (2.5)
\]

A weaker convergence condition is that

\[
\lim_{k \to \infty} P[|z_k(\omega) - z(\omega)| \geq \varepsilon] = 0 \quad (2.6)
\]

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.

### 3 Parameter Estimation - Full Observations

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

From (2.1), with simple manipulation then squaring both sides and summing over \( k \) we obtain,

\[
B \sum_{i=1}^{k} w_i^2 = \sum_{i=1}^{k} (x_i - x_{i-1})^2. \quad (3.1)
\]

Now consider the matrices

\[
J_k = \sum_{i=1}^{k} x_i x_{i-1}, \quad O_k = \sum_{i=1}^{k} x_i^2, \quad H_k = \sum_{i=1}^{k} x_{i-1}^2 \quad \text{and} \quad M_k = \sum_{i=1}^{k} w_i^2. \quad (3.2)
\]

From (3.1), we see that

\[
BM_k = O_k + H_k - 2J_k \quad (3.3)
\]

Similarly, from (2.2) with simple manipulation then squaring both sides and summing over \( k \) we obtain,

\[
D \sum_{i=1}^{k} v_i^2 = \sum_{i=1}^{k} (y_i - x_i)^2. \quad (3.4)
\]

Now consider the matrices

\[
T_k = \sum_{i=1}^{k} y_i x_i, \quad Q_k = \sum_{i=1}^{k} y_i^2 \quad \text{and} \quad V_k = \sum_{i=1}^{k} v_i^2. \quad (3.5)
\]
From (3.4), we see that
\[ DV_k = Q_k + O_k - 2T_k. \]  \hfill (3.6)

### 3.1 Almost Sure Convergence

**Lemma 1** Consider the linear system (2.1),(2.2) denoted $\lambda$. Suppose $\{x_k\}$ and $\{y_k\}$ exist and are measured, then
\[ \lim_{k \to \infty} \hat{B}_k, \hat{D}_k = B, D \text{ a.s.} \]  \hfill (3.7)

where
\[ \hat{B}_k = \frac{O_k + H_k - 2J_k}{k} \quad \text{and} \quad \hat{D}_k = \frac{Q_k + O_k - 2T_k}{k}. \]

**Proof:** Consider estimate $\hat{B}_k$ first.

Rewrite the estimate as
\[ \hat{B}_k = \frac{BM_k}{k}. \]  \hfill (3.8)

By the strong law of large numbers [11, Page 85], \( \lim_{k \to \infty} \frac{1}{k} M_k = 1 \text{ a.s.} \). The lemma results follows from (3.8). The lemma result for $\hat{D}_k$ follows similarly.

### 3.2 Finite Data Sets

**Lemma 2** Consider the linear system (2.1),(2.2) denoted $\lambda$. Suppose finite data sets $\{x_0, x_1, \ldots, x_T\}$ and $\{y_0, y_1, \ldots, y_T\}$ are measured, then
\[ P(|\hat{B}_T - B| \geq \epsilon) \leq \frac{B}{T\epsilon^2} \]
\[ P(|\hat{D}_T - D| \geq \epsilon) \leq \frac{D}{T\epsilon^2} \]  \hfill (3.9)

**Proof:** Follows from a similar argument as Lemma 1 by using the weak law of large number[11, Page 86].

**Remarks**

1. Lemma 2 is a weaker convergence result than Lemma 1 in the sense that in the limit as $T \to \infty$ a convergence in probability result rather than an almost sure convergence result is established.
4 Conditional Mean Estimates

In this section we consider parameter estimation based on conditional mean estimates in lieu of the true states. We define a model set, \( \Lambda := \{ \lambda(B,D,x_0) | B,D,x_0 \in R \} \), of allowable model estimates \( \hat{\lambda}_k \in \Lambda \) and assume the true model, \( \lambda \), lies in the model set \( \Lambda \). 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

\[
\begin{align*}
\hat{J}_{k|k,\hat{\lambda}_k} &= E[J_k|Y_k,\hat{\lambda}_k], \quad \hat{O}_{k|k,\hat{\lambda}_k} = E[O_k|Y_k,\hat{\lambda}_k], \\
\hat{H}_{k|k,\hat{\lambda}_k} &= E[H_k|Y_k,\hat{\lambda}_k] \quad \text{and} \quad \hat{T}_{k|k,\hat{\lambda}_k} = E[T_k|Y_k,\hat{\lambda}_k],
\end{align*}
\]

(4.1)

Optimal finite dimensional recursions for these conditional mean estimates can be found in [13]. Hence, we propose estimators for \( B \) and \( D \) based on observations \( \{Y_k\} \) as:

\[
\begin{align*}
\hat{B}_{k|\hat{\lambda}_k} &= \text{proj}(\frac{\hat{O}_{k|k,\hat{\lambda}_k} + \hat{H}_{k|k,\hat{\lambda}_k} - 2\hat{J}_{k|k,\hat{\lambda}_k}}{k}) \\
\hat{D}_{k|\hat{\lambda}_k} &= \text{proj}(\frac{Q_k + \hat{O}_{k|k,\hat{\lambda}_k} - 2\hat{T}_{k|k,\hat{\lambda}_k}}{k}),
\end{align*}
\]

(4.2, 4.3)

where

\[
\begin{align*}
\hat{\lambda}_{k+1} &= \lambda(\hat{B}_{k|\hat{\lambda}_k}, \hat{D}_{k|\hat{\lambda}_k}, \hat{x}_0) \quad \text{and} \\
\hat{\lambda}_{k+1} &= \{ \hat{\lambda}_1, \ldots, \hat{\lambda}_{k+1} \},
\end{align*}
\]

(4.4)

and \( \text{proj}(x) = \max(\delta, x) \) where \( 0 < \delta < B, D \) is a small constant, and 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 \( \lambda \) are available.

**Lemma 3** Consider the linear system (2.1), (2.2) denoted by \( \lambda \). Suppose \( \{y_k\} \) exists and is measured. Also assume that conditional mean estimates based on the "correct model" are available. Assume the true model \( \lambda \) satisfies \( \lambda \in \Lambda \). Then

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

(4.5)

**Proof:** The proof follows Lemma 1. We consider estimation of \( B \) first. We rewrite the estimate (4.2) without the projection step as

\[
\hat{B}_{k|\lambda} = \frac{E[B M_k|Y_k,\lambda]}{k} = B \frac{E[M_k|Y_k,\lambda]}{k}.
\]

(4.6)

Hence, the estimate is

\[
\begin{align*}
\hat{B}_{k|\lambda} &= B \left( \frac{E[M_k|Y_k,\lambda]}{k} \right) \\
&= B E \left[ \frac{M_k}{k} \big| Y_k, \lambda \right] \\
&= B E \left[ E \left[ \frac{M_k}{k} \big| Y_k, \lambda \right] \big| Y_k, \lambda \right].
\end{align*}
\]
Hence,
\[
\lim_{k \to \infty} \hat{B}_{k|\lambda} = B \text{ a.s..} \tag{4.7}
\]
Here, we have used 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 \(\lim_{k \to \infty} \frac{1}{k} E[M_k|G_k] = 1 \text{ a.s..}\) Also, because \(\phi_k = \frac{1}{k} E[M_k|G_k, Y_k, \lambda]\) is uniformly integrable (that is \(E[\phi_k \phi_k']\) is bounded) and \(\lim_{k \to \infty} \phi_k = 1 \text{ a.s.}\) then there is convergence in conditional mean:
\[
\lim_{k \to \infty} E[\phi_k|Y_k, \lambda] = 1 \text{ a.s..}
\]
The lemma result follows for \(\hat{D}_{k|\lambda}\) similarly.

\[\square\]

Remarks

1. A finite date set result which extends the above Lemma can be established using the weak law of large numbers.

The following theorem holds.

**Theorem 1** Consider the linear system (2.1), (2.2) denoted by \(\lambda\). Consider a sequence of estimated models \(\hat{\lambda}_k\) adaptively updated by previous parameter estimates as shown in (4.4). Then the recursion converges almost surely to the set \(S\) (or possibly the projection boundaries \(\tilde{B} = \delta\) and \(\tilde{D} = \delta\)), where
\[
S := \text{local arg min}_{\tilde{B}, \tilde{D}} E \left[ \left( (x_k - x_{k-1})^2 - \tilde{B} \right)^2 + \left( (y_k - x_k)^2 - \tilde{D} \right)^2 \right] \tag{4.8}
\]
where \(\tilde{\lambda} = \lambda(\tilde{B}, \tilde{D}, \tilde{x}_0)\). That is, \(S\) contains the local minimum of the cost function
\[
E \left[ \left( (x_k - x_{k-1})^2 - \tilde{B} \right)^2 + \left( (y_k - x_k)^2 - \tilde{D} \right)^2 \right] \tag{4.8}
\]

**Proof** First we consider estimation of \(B\) only. Simple manipulation of (4.2) gives a recursion for the estimate \(\hat{B}_{k|k, \tilde{\lambda}_k}\) as follows:
\[
\hat{B}_k = \hat{B}_{k-1} + \frac{1}{k} \left( \Delta \hat{O}_k + \Delta \hat{H}_k - 2 \Delta \hat{J}_k - \hat{B}_{k-1} \right) \tag{4.9}
\]
where we denote \(\hat{B}_{k|k, \tilde{\lambda}_k}\) by the shorthand notation \(\hat{B}_k\) and define \(\Delta \hat{O}_k := \hat{O}_{k|k, \tilde{\lambda}_k} - \hat{O}_{k-1|k-1, \tilde{\lambda}_{k-1}}\) and similarly define \(\Delta \hat{H}_k\) and \(\Delta \hat{J}_k\). To simplify the presentation we will ignore the projection step. Convergence results will still hold if projection is performed, see Ljung [14].

We follow the technique presented in [8]. This proof relies on the ordinary differential approach described in [14, 15, 16]. In the ODE approach, convergence of a difference equation is established by considering the convergence properties of an associated ordinary differential equation.
Convergence of (4.9) can be analyzed by considering the following ordinary differential equation,

$$\frac{d\tilde{B}(\tau, k)}{d\tau} = \frac{1}{k} f(\tilde{B}(\tau, k), k)$$  \hspace{1cm} (4.10)

where $k$ is here a fixed parameter. Also, with $\tilde{B}(\tau, k)$ abbreviated as $\tilde{B}$, we define $f(\tilde{B}, k)$ as follows:

$$f(\tilde{B}, k) := E[\Delta \hat{O}_{k|\tilde{B}} + \Delta \hat{H}_{k|\tilde{B}} - 2\Delta \dot{J}_{k|\tilde{B}} - \tilde{B}]$$  \hspace{1cm} (4.11)

Here we have explicitly shown in the notation used that $\Delta \hat{O}_{k}$ etc. depend on the parameter estimate $\tilde{B}$.

In Ljung [14], it is stated that (4.9) will converge to the set $\mathcal{S}$ or a boundary, where $\mathcal{S} := \{\tilde{B} | \lim_{k \to \infty} f(\tilde{B}, k) = 0\}$, if the following hold:

1. Certain regularity and exponential stability conditions hold as listed in [14].
2. The ODE (4.10) (which is parameterized by $k$) is asymptotically stable in the limit as $k \to \infty$.

The satisfy condition 1 we require the conditional mean filters for $\dot{J}_{k|k, \lambda_k}$, $\dot{O}_{k|k, \lambda_k}$, $\dot{H}_{k|k, \lambda_k}$ and $\ddot{J}_{k|k, \lambda_k}$ to be exponentially forgetting (which has not been shown yet but which we will assume to be the case) and regularity conditions for this system are shown and discussed in [8, 14].

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

$$W(\tilde{B}, k) = \frac{1}{2} E \left[ \left( (x_k - x_{k-1})^2 - \tilde{B} \right)^2 \right]$$  \hspace{1cm} (4.12)

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 $W(\tilde{B}, k) \geq 0$. Under asymptotic ergodicity and certain smoothness conditions the differentiation w.r.t. $\tilde{B}$ and the expectation operations can be interchanged. Hence,

$$\frac{dW(\tilde{B}(\tau, k), k)}{d\tilde{B}(\tau, k)} = -f(\tilde{B}(\tau, k), k)$$  \hspace{1cm} (4.13)

and it then follows that

$$\frac{dW(\tilde{B}(\tau, k), k)}{d\tau} = \frac{dW(\tilde{B}(\tau, k), k)}{d\tilde{B}(\tau, k)} \frac{d\tilde{B}(\tau, k)}{d\tau} = -f(\tilde{B}(\tau, k), k) \frac{1}{k} f(\tilde{B}(\tau, k), k).$$  \hspace{1cm} (4.14)

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

It follows from Ljapunov's direct method and equation (4.13) that $\tilde{B}(\tau, k)$ converges to the set $\{\tilde{B} | \lim_{k \to \infty} f(\tilde{B}, k) = 0\}$ (or possibly to the boundary $\tilde{B} = \delta$ if a projection step is implemented). Convergence of the difference equation (4.9) follows from Ljung[14].

We note that under asymptotic ergodicity (and other regularity conditions given in Ljung [14] and shown in [8]) and using the results in Ljung [14] that the difference equation
(4.9) and the ODE (4.10) converge to the same set. That is, in the limit $k \to \infty$
convergence to the set $\{B | \lim_{k \to \infty} f(\tilde{B}, k) = 0\}$ is equivalent to the local minimum of
$E \left[ ((x_k - x_{k-1})^2 - \tilde{B})^2 | B, Y_k \right]$. The result for simultaneous estimation follows using a similar approach by writing the
estimate $\hat{D}_{k|k, \lambda_k}$ as a recursion and using the Lyapunov function,

$$W(\tilde{B}, \tilde{D}, k) = \frac{1}{2} \left( E \left[ ((x_k - x_{k-1})^2 - \tilde{B})^2 \right] + E \left[ ((y_k - x_k)^2 - \tilde{D})^2 \right] \right). \quad (4.15)$$

Note that only a local convergence result is established in the simultaneous estimation
case because the set $S$ may contain more than one point.

\[ \square \]

**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 [8].

3. Local convergence results for estimation from finite date sets can be established in the
   off-line situation when the recursions (4.2)-(4.3) are passed over the data set multiple
   times and the model estimate is updated after each pass. In this multi-pass off-
   line situation the estimators are an example of the EM (expectation-maximization)
   algorithm and local convergence results are available [17]. Convergence results in the
   situation of a single pass through a finite data set which are analogous to Lemma 2
   have not yet been established.

5 Implementation

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

**Filters**

To implement the estimator defined by (4.2)-(4.4) requires the quantities $\hat{J}_{k|k, \lambda_k}$, $\hat{O}_{k|k, \lambda_k}$,
$\hat{H}_{k|k, \lambda_k}$, and $\hat{T}_{k|k, \lambda_k}$ as defined in (4.1) to be calculated. Optimal finite dimensional recurrences
for $\hat{J}_{k|k, \lambda_k}$, $\hat{O}_{k|k, \lambda_k}$, $\hat{H}_{k|k, \lambda_k}$ and $\hat{T}_{k|k, \lambda_k}$ are given in [13].

**On-line or batch processing**

It is possible to use the algorithm presented in this paper in both an on-line and batch
manner. The recursion (4.2)-(4.3) 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 multi-pass, situation convergence results follow by considering the data set as the concatenation of copies of the finite data set (sometimes known as the periodic extension of a finite signal) and then convergence occurs to a minimum of (4.15). It has been shown that resolution errors result from working on a finite data set\[18\]. These resolution errors will be apparent on this concatenated data set and estimation bias will result. The estimation error will consist of two components as follows:

\[(B_{k+1|k} - B)^2 = B^2_L + B^2_V\]  

(5.1)

The error $B_L$ is a bias due to using a finite data set and $B_V$ is the estimation variance of the estimator. The estimation variance $B_V$ can be reduced by increasing the number of passes through the data while $B_L$ depends only on the finite data set.

**Initialization**

The algorithm requires an initial guess for the ARW and RRW coefficients. In our experience the algorithm will converge to the true ARW and RRW coefficients from a reasonable initialization. In particular, for the studies presented below, convergence to the true values occurs whenever $B_0$ is significantly smaller than $D_0$. For typical applications (IMU measurements) this requirement is not restrictive.

The filters from [13] 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.

**Improved Convergence Rates**

The rate of convergence of the adaptive algorithm (4.2)-(4.4) from initial parameter estimates can be improved. Using the difference form of the estimators, i.e. (4.9) etc., convergence rates can be improved using the Polyak technique [19]. That is,

\[\hat{B}_{k+1|k+1,\hat{\lambda}_k+1} = \hat{B}_{k|k,\hat{\lambda}_k} + \gamma_k \left( \Delta O_{k|k,\hat{\lambda}_k} + \Delta H_{k|k,\hat{\lambda}_k} - 2\Delta J_{k|k,\hat{\lambda}_k} - \hat{B}_{k|k,\hat{\lambda}_k} \right)\]  

(5.2)

where $\gamma_k = \frac{1}{k^p}$ and $0 < p \leq 1$ and similarly for the D estimate. The estimates are then averaged. The convergence results still hold.

Other techniques such as forgetting factors can also be useful[14].

**Computational effort**

The computational effort required to implement the recursions (4.2)-(4.4) 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 are not discussed in the given
references and will depend on implementation choices. If the data set has $T$ points, then
to calculate the Allan variance of size $N$ requires $O(T(1 + 3/N))$ (or order$(T(1 + 3/N))$)
calculations. Because Allan variance 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 $O(KT \log(T))$ compared to the
$O(100T)$ required to implement (4.2)-(4.4).

For large data sets the estimators presented here are clearly computationally superior. The
exact number of points at which our on-line algorithm is computationally superior depends
on implementation choices. For example, if the Allan variance is calculated at log spaced
points and $K = 10$ starting at $N = 2$ then for $T > 5000$ (4.2)-(4.4) is computationally
superior. However, if calculated at log spaced points and $K = 5$ then the Allan variance
technique is computationally superior until $T > 4 \times 10^6$. A typical data set might contain
24hrs of 1Hz measurements, i.e. $\approx 10^7$ data points. The recursions (4.2)-(4.4) always have
the advantage that the data does not need to be stored.

**Numerical Overflow**

If implemented in an on-line manner the algorithm will need to be modified to ensure
overflow of $\dot{\theta}_{k|k|,\dot{\lambda}_k}$ etc. does not occur. Scaling or forgetting factors could be introduced
into the filters of [13]. Alternatively, the filters could be reset and then reinitialized on
data to avoid numerical overflow.

### 6 Simulation studies

In this section we examine the proposed technique for estimating the angular random walk
and rate random walk coefficients. First, we test the technique on computer generated
data with known noise characteristics. Secondly, we test the technique on data taken from
a real device, namely a Honeywell HG1700-AE11 IMU. This device, contains ring laser
gyros and produces delta angle and delta velocity signals in the $X, Y$ and $Z$ directions.
We have used data from only the $Y$ direction and have assumed that it is independent of
the other axes. The device was static.

#### 6.1 Computer Generated data

These computer simulations were performed using the MATLAB package and matlab's
*randn* normal distributed noise sources.

A 40000 point sequence was generated for the linear system (2.1),(2.2) with $B = 0.1$
and $D = 1$. From initial estimates, $\hat{B}_0 = 0.3$ and $\hat{D}_0 = 1.5$, the algorithm (4.2) was
implemented to produce new estimates of the $B$ and $D$ coefficients. The final estimates
after one pass through the data were $\hat{B} = 0.1008$ and $\hat{D} = 0.9991$. 


6.2 The Honeywell IMU (HG1700-AE11)

Navigational data arrives from the Honeywell IMU at 600Hz and 100Hz. This data was averaged to 1Hz before any estimation was performed to allow recording of reasonable time lengths of data. Other transformations were performed on the data to scale the output but these are not important in this discussion.

The recorded data were analyzed by the standard Allan variance graph techniques. Figure 5 shows the Allan variance graph obtained from analysis of a 80000 data point set. It appears from the Allan variance plot that the dominate noise sources are angular random walk and rate random walk which justifies our noise model assumption for this device.

From the Figure 5 the angular random walk and rate random walk coefficients can be estimated, using the technique in [4]. To do this, lines of best fit are drawn on the parts of log-log plot with slope $-1/2$ and $+1/2$. Then the coefficients can be calculated as $\text{ARW} = \sqrt{A_-(2)} = 5 \pm 1$ and $\text{RRW} = \sqrt{A_+(3)/\sqrt{2}} = 0.08(+0.02 - 0.04)$ where $A_-(N)$ is the value of the $-1/2$ line at $N$ and $A_+(N)$ is the value of the $+1/2$ line at sample size $N$.

The technique proposed in this paper is an alternative method for calculating these two coefficients without producing the Allan variance graph.

The recursions with Polyak step, that is (5.2) with $p = 0.5$ for the $B$ recursion and $p = 1$ for the $D$ recursion and the model updated adaptively according to (4.4), were used to estimate the parameters from the 80000 point set. The algorithm was initiated with parameter estimates, $B_0 = 0.1$ and $D_0 = 20$.

The parameter estimates after one pass through the data were $\hat{B} = 0.0045$ and $\hat{D} = 31.09$ (or equivalently $\text{ARW} = \sqrt{\hat{B}} = 5.6$ and $\text{RRW} = \sqrt{\hat{B}} = 0.0668$). In Figures 3 and 4 the evolution of parameter estimates versus time is shown. These final estimates are within the error bounds of the estimates from the Allan variance graph as shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1 (U): Comparison of Estimated Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{ARW}$</td>
</tr>
<tr>
<td>$\text{RRW}$</td>
</tr>
</tbody>
</table>

Various other initializations ($\hat{B}_0$ and $\hat{D}_0$) were tried including: $\hat{B}_0 = 0.1$ and $\hat{D}_0 = 1$; $\hat{B}_0 = 0.3$ and $\hat{D}_0 = 1.5$; amongst others. Convergence occurs when $\hat{B}_0 < \hat{D}_0$. Note that both the Allan variance graph method and the method proposed here require large data sets to enable estimation of $B$. Estimation of $D$ can be done on smaller data sets.

6.3 Discussion

Good choices for $p$ in the step modified estimation algorithm will depend of the relative power of the $w$ and $v$ noises. The Honeywell IMU'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.

7 Conclusions

This paper presented a new technique for estimating 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 off-line to produce an Allan variance graph. The technique proposed here can analyze data on-line as it arrives from the gyro and can produce immediately estimates of the noise model in the sense that it does not require the user to manually analyze an Allan variance graph.

Acknowledgments

The authors wish to acknowledge the helpful suggestions made by the reviewer.

References


Figure 3 (U): Estimates of ARW versus time.

Figure 4 (U): Estimates of RRW versus time.
Figure 5 (U): Allan variance (standard derivation) versus sample size with error bars and lines of worst fit.
DISTRIBUTION LIST

On-line Estimation of Allan Variance Parameters

Jason J. Ford and Michael E. Evans

AUSTRALIA

DEFENCE ORGANISATION

S&T Program
Chief Defence Scientist
FAS Science Policy
AS Science Corporate Management
Director General Science Policy Development
Counsellor Defence Science, London (Doc Data Sheet only)
Counsellor Defence Science, Washington (Doc Data Sheet only)
Scientific Adviser to MRDC Thailand (Doc Data Sheet)
Director General Scientific Advisers and Trials (shared copy)
Scientific Adviser Policy and Command
Navy Scientific Adviser (Doc Data Sheet and distribution list only)
Scientific Adviser - Army (Doc Data Sheet and distribution list only)
Air Force Scientific Adviser (Doc Data Sheet and distribution list only)
Director Trials (Doc Data Sheet and distribution list only)

Aeronautical and Maritime Research Laboratory
Director, AMRL
Chief, Weapons Systems Division
Research Leader Air Weapons Systems
Head, Guidance and Control
Task Manager, Dr Peter Simms WSD
Author(s): Dr J. Ford, Dr M. Evans

DSTO Library
Library Fishermens Bend
Library Maribyrnong
Library Salisbury (2 copies)
Australian Archives
Library, MOD, Pyrmont (Doc Data sheet only)

Capability Development Division
Director General Maritime Development (Doc Data Sheet only)
Director General Land Development (Doc Data Sheet only)
Director General C3I Development (Doc Data Sheet only)

Intelligence Program
DGSTA Defence Intelligence Organisation
Corporate Support Program (libraries)
OIC TRS, Defence Regional Library, Canberra 17
Officer in Charge, Document Exchange Centre
(Doc Data Sheet and distribution list only)
*US Defence Technical Information Centre, 2 copies 18,19
*UK Defence Research Information Center, 2 copies 20,21
*Canada Defence Scientific Information Service, 1 copy 22
*NZ Defence Information Centre, 1 copy 23
National Library of Australia, 1 copy 24

UNIVERSITIES AND COLLEGES

Australian Defence Force Academy
Library 25
Head of Aerospace and Mechanical Engineering 26
Deakin University, Serials Section (M list), Deakin University Library 27
Senior Librarian, Hargrave Library, Monash University 28
Librarian, Flinders University 29

OTHER ORGANISATIONS

NASA (Canberra) 30
AGPS 31

OUTSIDE AUSTRALIA

ABSTRACTING AND INFORMATION ORGANISATIONS
INSPEC: Acquisitions Section Institution of Electrical Engineers 32
Library, Chemical Abstracts Reference Service 33
Engineering Societies Library, US 34

INFORMATION EXCHANGE AGREEMENT PARTNERS
Acquisitions Unit, Science Reference and Information Service, UK 37
Library - Exchange Desk, National Institute of Standards and Technology 38

SPARES (5 copies)

Total number of copies: 43
A new on-line method is presented for estimation of the angular random walk and rate random walk coefficients of IMU (Inertial Measurement Unit) gyros and accelerometers. The on-line method proposes a state space model and proposes parameter estimators for quantities previously measured from off-line data techniques such as the Allan variance graph. Allan variance graphs have large off-line computational effort and data storage requirements. The technique proposed here requires no data storage and computational effort of $O(100)$ calculations per data sample.