Constrained Expectation-Maximization Algorithm for Stochastic Inertial Error Modeling: Study of Feasibility

STEBLER, Yannick, et al.

Abstract

Stochastic modeling is a challenging task for low-cost sensors which errors can have complex spectral structures. This makes the tuning process of the INS/GNSS Kalman filter often sensitive and difficult. For example, first-order Gauss-Markov processes are very often used in inertial sensor models. But the estimation of their parameters is a non-trivial task if the error structure is mixed with other types of noises. Such estimation is often attempted by computing and analyzing Allan variance plots. This contribution demonstrates solving situations when the estimation of error parameters by graphical interpretation is rather difficult. The novel strategy performs direct estimation of these parameters by means of the Expectation-Maximization (EM) algorithm. The algorithm results are first analyzed with a critical and practical point of view using simulations with typically encountered error signals. These simulations show that the EM algorithm seems to perform better than the Allan variance and offers a procedure to estimate first-order Gauss-Markov processes mixed with other types of noises. At the same time, the conducted [...]
Constrained Expectation-Maximization Algorithm for Stochastic Inertial Error Modeling: Study of Feasibility

Yannick Stebler¹, Stéphane Guerrier², Jan Skaloud¹, Maria-Pia Victoria-Feser²
¹ Geodetic Engineering Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland
² Research Center for Statistics, University of Geneva, Switzerland
E-mail: yannick.stebler@epfl.ch, Stephane.Guerrier@unige.ch, jan.skaloud@epfl.ch,Maria-Pia.VictoriaFeser@unige.ch

Abstract. Stochastic modeling is a challenging task for low-cost sensors which errors can have complex spectral structures. This makes the tuning process of the INS/GNSS Kalman filter often sensitive and difficult. For example, first-order Gauss-Markov processes are very often used in inertial sensor models. But the estimation of their parameters is a non-trivial task if the error structure is mixed with other types of noises. Such estimation is often attempted by computing and analyzing Allan variance plots. This contribution demonstrates solving situations when the estimation of error parameters by graphical interpretation is rather difficult. The novel strategy performs direct estimation of these parameters by means of the Expectation-Maximization (EM) algorithm. The algorithm results are first analyzed with a critical and practical point of view using simulations with typically encountered error signals. These simulations show that the EM algorithm seems to perform better than the Allan variance and offers a procedure to estimate first-order Gauss-Markov processes mixed with other types of noises. At the same time, the conducted tests revealed limits of this approach that are related to the convergence and stability issues. Suggestions are given to circumvent or mitigate these problems when complexity of error structure is “reasonable”. This work also highlights the fact that the suggested approach via EM algorithm and the Allan variance may not be able to estimate reasonably well parameters of complex error models, and shows the need for new estimation procedures to be developed in this context. Finally, an empirical scenario is presented to support the former findings. There, the positive effect of using the more sophisticated EM-based error modeling on a filtered trajectory is highlighted.

Keywords: Allan variance, EM algorithm, IMU, error modeling
1. Introduction

Satellite-based navigation (GNSS) is nowadays a standard approach for performing localization in outdoor environment. Situations where GNSS signals are partially or completely unavailable severely degrade the performance of such systems. Furthermore, GNSS sensor bandwidth may be too low for applications with high dynamics. A well accepted and largely proven approach for improving navigation in such situations is to integrate GNSS with inertial sensors such as gyroscopes and accelerometers. A conventional Inertial Measurement Unit (IMU) is composed of a triad of orthogonally mounted sensors yielding specific force and angular rate signals. After initialization, these quantities are then integrated with respect to time to give instantaneous body velocity, position and attitude. This procedure is the core of an Inertial Navigation System (INS). The inertial and GNSS data are combined through Bayesian techniques like Extended Kalman Filtering (EKF). During periods affected by gaps in GNSS signal reception, the navigation quality is mainly driven by the errors affecting the inertial sensor outputs. These errors are integrated in the strapdown inertial navigation and their impact consequently grow with time. Correct modeling and estimation of these errors is thus very important for improving the quality of the inertial navigation.

IMU errors have deterministic and stochastic nature. On the one hand, deterministic errors (e.g. axes misalignment) are compensated through physical models during calibration procedures and will not be treated in this article; see [1, 2] for more details. On the other hand, stochastic errors contain components which have random behavior (e.g. dynamic dependent errors) or are too complicated to model deterministically (e.g. environmental changes). The error structure is often described by stochastic processes among which Gaussian White Noise (WN), Random Walk (RW), first-order Gauss-Markov (GM), Bias Instability (BI), Rate Ramp (RR) or Quantization Noise (QN) are part of a non-exhaustive list of commonly used processes in the navigation community. With the exception of QN, these processes can be described by stochastic differential equations owning some parameter(s). These are included in the EKF State-Space Model (SSM) through state vector augmentation technique [3, 4]. This leads to the following questions that often arise during the filter tuning: Which processes should be considered to best describe the stochastic part of the error signal? Once the processes are selected, which parameter values should be used in the SSM/EKF?

Classical techniques which may be exploited for answering these questions are the autocorrelation function (ACF) and the well known variance analysis techniques based on Allan Variance (AV) [5], Hadamard Variance (HV) or Total Variance (TV). The first method can be used to validate the employed error model by computing the ACF of the residuals in the EKF. However, this doesn’t allow the direct identification of processes and their parameter(s) when the signal is composed of mixtures of processes [6]. On the other side, the AV, HD, TV are well established techniques for identifying processes and estimating their parameter(s). However, they only provide good results for processes which are clearly separable in the spectral domain and not subjected to
spectral ambiguity [7]. With the exception of the GM, the parameters of the processes mentioned previously can be estimated through variance analysis. However, the often encountered issue is to estimate the inverse correlation time $\beta$ and the driving noise variance $\sigma^2_{GM}$ of a GM process when mixed with some other processes. Large values of $\beta$ make the GM process approach to WN, while small values to RW. In practice, these values are estimated through ad-hoc tuning, by using available sensor specifications, or by experience [8]. The work in [6] proposes a methodology in which GM processes are used to overbound the sensor error but the success of this methodology is quite limited.

The estimation of the parameters that specify a SSM (see Eq. 4 and 5) is in general quite challenging. Indeed, the (log) likelihood function $\log L(\theta | y_t, x_t)$, where $\theta$ is the vector of unknown parameters that defines the SSM, $y_t$ is the measurement vector, and $x_t$ is the system state vector, is a highly nonlinear and complicated function [9]. Historically, a Newton-Raphson algorithm was employed to successively update $\theta$ until the log likelihood was maximized (for more details on this approach see for example [10]). However, a conceptually simpler estimation procedure was proposed in [11] which relies on the Expectation-Maximization (EM) algorithm originally developed in [12]. The EM is a procedure that is guaranteed to converge to the Maximum Likelihood Estimate (MLE) and is therefore often employed when dealing with difficult likelihood maximization. It is based on the idea of replacing a complex likelihood maximization by a sequence of easier maximizations whose limit is the answer to the original problem [13]. The EM is particularly suited to problems with missing data which often render calculations cumbersome. With such approach, two different likelihood problems are considered. The first is the problem we are interested to solve, namely the “incomplete-data” problem, while the second is the problem that is actually solved by the EM, the so called “complete-data” problem [13]. Suppose for the moment that we could observe the state vector $x_t$ in addition to the observations $y_t$. Then the log likelihood function of this problem $\log L(\theta | x_t, y_t)$ could be easily maximized using results from the multivariate normal theory [9]. This maximization can be considered as the “complete-data” problem. However, we do not have the complete data since the state vector $x_t$ is unobserved and thus we aim to solve the “incomplete-data” problem. In this context, the EM algorithm offers an iterative method for finding the MLE $\hat{\theta}$ by successively maximizing the conditional expectation of the complete data likelihood. A more formal treatment of this method is given in section 3.

In the context of integrated navigation, the EM algorithm has mainly been used for obtaining system process noise and measurement noise parameters where all parameter elements are estimated [14, 15]. However, when using EM for more complex SSMs describing stochastic error behaviour of inertial sensors, some elements must remain fixed in the parameter matrices. An example can be provided by the often encountered stochastic error model which is composed of WN, GM, RW and RR processes where elements of the state-space transition matrix must remain fixed (e.g. the one element of RW, and the null off-diagonal elements), and some others can be freely estimated (e.g. the $\beta$ of GM).
In this article, we show the adaptation of the classical EM algorithm to constrain the SSM parameters. Such constraints allow estimating more complex stochastic error models as those used in INS/GNSS filters/smoothers. Realistic examples will be first simulated for illustrating the algorithm performance and for providing a critical analysis of its practical use in inertial navigation. The results obtained by the EM algorithm will be compared with the benchmark methods, i.e. the AV, HV and TV. Later, an example with real data will be included to demonstrate the positive impact of the derived stochastic model on INS/GNSS trajectory.

This article is organized as follows: in the Sec. 2, we briefly introduce the AV, HV and TV and show how these techniques can be used in the context of sensor error modeling. In Sec. 3 we explain the theory related to the constrained EM algorithm applied to SSMs. In Sec. 4 we present a simulation study which allows comparing the performances of the estimation based on the EM algorithm with that based on the AV, HV and TV. Finally, in Sec. 5 we illustrate the EM approach with real inertial signals and we conclude the paper with a discussion.

2. Allan, Hadamard and Total Variance Techniques

Several techniques have been developed for the analysis of the noise altering the output signal of sensors, such as the AV, the HV or the TV [16, 17, 18, 19]. The AV was invented in 1966 by David Allan when he criticized the use of the sample variance estimator in the context of non \( \text{iid} \) time series. He proposed the AV as an alternative theoretical measure of variability [5]. Although this method was originally intended to study the stability of oscillators, it has been successfully applied to problems dealing with a large number of different types of sensors, among which stands the modeling of inertial sensor errors [20, 21, 22, 23, 24, 6]. In 1998, the IEEE standard put forward this technique as a noise identification method to determine the characteristics of the underlying random processes that perturb data [25].

Suppose that \( y_t, t = 1, 2, ..., N \) is the realization of some (univariate) stochastic process, say the observed signal of an IMU. This could be the constant signal of an accelerometer or gyroscope acquired during static conditions where the mean has been removed. The remaining signal represents therefore the error. Let \( \bar{y}_t(\tau) \) be the sample average of \( \tau \) consecutive observations, i.e.

\[
\bar{y}_t(\tau) = \frac{1}{\tau} \sum_{j=0}^{\tau-1} \bar{y}_{t-j}
\]

Intuitively, the AV at scale \( \tau \) (noted as \( \sigma_{y}^2(\tau) \)) aims to measure how much the sample average \( \bar{y}_t(\tau) \) changes from one period of time to another. More rigorously, this quantity is defined as half the expectation of squared differences between adjacent nonoverlapping \( \bar{y}_t(\tau) \):

\[
\sigma_{y}^2(\tau) = \frac{1}{2} \mathbb{E} \left[ (\bar{y}_t(\tau) - \bar{y}_{t-\tau}(\tau))^2 \right]
\]  (1)
Constrained Expectation-Maximization Algorithm for Stochastic Inertial Error Modeling: Study of Feasibility

Despite there exist several estimators of the AV, we solely focus on the maximal-overlap estimator proposed in [26] which present some superior statistical properties. Such an estimator is defined as:

$$\hat{\sigma}_y^2(\tau) = \frac{1}{2(N - 2\tau + 1)} \sum_{t=2\tau}^{N} (\bar{y}_t(\tau) - \bar{y}_{t-\tau}(\tau))^2$$

The AV can be expressed in the frequency domain through the unique relationship between $\sigma^2_y(\tau)$ and the Power Spectral Density (PSD) $S_{yy}(f)$ of the intrinsic processes [23]:

$$\sigma^2_y(\tau) = 4 \lim_{a \to \infty} \int_0^a S_{yy}(f) \frac{\sin^4(\pi f \tau)}{(\pi f \tau)^2} df$$

where $f$ is the frequency. This last equation enables to link the parameter vector $\theta$ of some SSM to $\sigma^2_y(\tau)$. Indeed, this relationship is due to the known form of the PSD function characterizing different noise processes which enables to express $\theta$ as a function of $\sigma^2_y(\tau)$ (a detailed discussion on how to express this link between $\theta$ and $\sigma^2_y(\tau)$ can be found in [22]). In general, only five basic processes are considered with the AV: QN, WN, BI, RW, RR. These processes correspond to linear regions in a “$\sigma^2_y(\tau)$ v.s. $\tau$” log-log plot as schematically depicted in Fig. 1. Therefore, $\theta$ is usually estimated by performing linear regression of (visually) identified linear regions in such log-log plots. This method is only well defined for a few type of processes and it is not clear how inference on $\theta$ can be made with this approach.

Several extensions of the AV have been proposed in the literature such as the HV or the TV. The HV was proposed in [27] as a generalization of the sample variance weighted with binomial coefficients. This approach has, compared to the AV, a higher-resolution spectral analysis and reduces the uncertainty of long-term estimates of frequency stability without increasing the length of a data run. The HV is defined
as:
\[ \sigma^2_H = \frac{1}{6} \mathbb{E} \left[ (\bar{y}_t(\tau) - 2\bar{y}_{t-\tau}(\tau) + \bar{y}_{t-2\tau}(\tau))^2 \right] \]

An estimator of \( \sigma^2_H \) can simply be found by replacing the theoretical by the empirical expectation:
\[ \hat{\sigma}^2_H = \frac{1}{6(N - 3\tau + 1)} \sum_{t=3\tau}^{N} (\bar{y}_t(\tau) - 2\bar{y}_{t-\tau}(\tau) + \bar{y}_{t-2\tau}(\tau))^2 \]

A more detailed description of the HV can, for example, be found in [16] which also introduced a modified version of this method.

The TV was proposed in [28] as an estimator of the AV (Eq. 1) that has lesser mean square error than the standard unbiased estimator (Eq. 2). The TV can be defined as:
\[ \hat{\sigma}^2_T = \frac{1}{N - 1} \sum_{k=1}^{N-1} \left[ \frac{1}{2(N - 2\tau + 1)} \sum_{t=2\tau}^{N} (\bar{y}_{t,k}(\tau) - \bar{y}_{t-\tau,k}(\tau))^2 \right] \]

where \( \{\bar{y}_{t,k}(\tau)\} = \bar{y}_{k+1}(\tau), \bar{y}_{k+2}(\tau), ..., \bar{y}_1(\tau), \bar{y}_2(\tau), ..., \bar{y}_k(\tau) \). Note that \( \hat{\sigma}^2_T \) can be computationally very intensive and, therefore, several faster methods have been proposed [29, 30].

3. Constrained EM Algorithm

3.1. State-Space Model

A linear dynamical system can be described by the following differential equation:
\[ \dot{x}(t) = F(t)x(t) + G(t)w(t) + L(t)u(t) \] (3)
where \( x(t) \) is the \( p \times 1 \) system state vector at time \( t \), \( F(t) \) is the \( p \times p \) time-varying dynamic coefficient matrix, \( G(t) \) is the \( p \times r \) time-varying process noise coupling matrix, \( w(t) \) is a \( r \times 1 \) random forcing function such that \( w \sim \mathcal{N}(0, Q) \), \( L(t) \) is a \( p \times r \) time-varying input coupling matrix, and \( u \) is a \( r \times 1 \) deterministic input vector. Integration and discretization of Eq. 3 yields the discrete SSM:
\[ x_{t+1} = \Phi_t x_t + w_t + u_t \] (4)
with measurements
\[ y_{t+1} = H_{t+1} x_{t+1} + v_{t+1} \] (5)
where \( \Phi_t \) is the discrete form of \( F(t) \), \( H_t \) is the \( l \times p \) design matrix which maps the true state space \( x_t \) into the observed space, and \( v_t \) is a \( l \times 1 \) noise vector such that \( v \sim \mathcal{N}(0, R) \). The initial state \( x_1 \) is assumed to be a normal random vector with mean vector \( \mu \) and \( p \times p \) initial covariance matrix \( \Sigma \) (i.e. \( P_1 \)).

As mentioned earlier, the vector \( \theta \in \Theta \) is the set of SSM parameters such that:
\[ \theta = \{u, \Phi, Q, R, H, \mu, \Sigma\} \] (6)
In the context of stochastic error modeling, some elements of the SSM parameters are not known and have to be estimated by maximizing the likelihood function. Unfortunately, the direct maximization is in general a very difficult task. However, the EM algorithm provides an iterative means for solving this problem quite effectively.

### 3.2. The Likelihood Function

The log-likelihood of $\theta$ given $y_t$ and $x_t$ for $t = 1, \ldots, N$ is:

$$
\log L(\theta|y_t, x_t) = -\frac{1}{2} (x_1 - \mu)^T \Sigma^{-1} (x_1 - \mu) - \frac{1}{2} \sum_{t=2}^{N} (x_t - \Phi x_{t-1} - u_t)^T Q^{-1} (x_t - \Phi x_{t-1} - u_t) - \frac{1}{2} \sum_{t=1}^{N} (y_t - H x_t)^T R^{-1} (y_t - H x_t) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} N \log |Q| - \frac{1}{2} N \log |R| - \frac{N}{2} \log 2\pi
$$

The objective of the EM is to find $\theta$ that maximizes $\log L(\theta|y_t, x_t)$. As $x_t$ is unobservable in our case, it is replaced by the “complete-data” likelihood $\Psi = \mathbb{E} [\log L(\theta|y_t, x_t)]$ which entire expression can be found in [31, 11].

### 3.3. The Algorithm

The EM algorithm switches iteratively between an Expectation (E-)step and a Maximization (M-)step [32]. On the $(j+1)$th iteration, the E- and M-steps are defined as follows:

**E-Step**. Calculate $Q(\theta|\theta^{(j)})$, where

$$
Q(\theta|\theta^{(j)}) = \mathbb{E} [\log L(\theta^{(j)}|y_t, x_t)] = \Psi^{(j)}
$$

**M-Step**. Choose $\theta^{(j+1)}$ to be any value of $\theta \in \Theta$ which belongs to:

$$
\theta^{(j+1)} = \arg\max_{\theta \in \Theta} Q(\theta|\theta^{(j)})
$$

The E- and M-steps are iteratively repeated until some convergence criterion is fulfilled (e.g. until $|L(\theta^{(j+1)}|y_t, x_t) - L(\theta^{(j)}|y_t, x_t)| < \epsilon$ for some arbitrarily small amount $\epsilon$) [32]. In the sequel, these two steps are detailed.

In the E-step, the expected states $x_t^{N}$ (as well as the associated covariance matrices) are computed such that they could be considered fixed in the M-step where $\Psi^{(j)}$ will be maximized. Therefore, the E-step requires computing:

$$
x_t^{N} = \mathbb{E} [x_t|y_1, \ldots, y_N, \theta^{(j)}]
$$

$$
P_t^{N} = \text{cov} [x_t|y_1, \ldots, y_N, \theta^{(j)}]
$$

$$
P_{t,t-1}^{N} = \text{cov} [x_t, x_{t-1}|y_1, \ldots, y_N, \theta^{(j)}]
$$
which can be calculated using a Kalman smoother provided for completeness in the Appendix [11].

In the M-step, the parameter vector is updated to $\theta^{(j+1)}$ by finding the parameters that maximize $\Psi^{(j)}$ considering the values $x_N^{(j)}$, $P_N^{(j)}$, and $P_{t,t-1}^{(j)}$ obtained in the E-step as fixed. For doing that, the expression yielded by $\Psi^{(j)}$ is minimized by computing the partial derivatives with respect to $\theta^{(j)}$ and setting them to zero. The results of these derivatives for the classical unconstrained case can be found in many articles like [11, 31]. The work in [33, 31] provides the way of constraining elements in the matrices of $\theta$. Let $M$ be any matrix contained in $\theta$ (e.g. $\Phi$, $Q$, $H$, $R$) with fixed and $p$ free elements (i.e. to be estimated): This matrix can be decomposed in matrices containing the fixed and free elements:

$$M = M_{\text{fixed}} + M_{\text{free}}$$  \hspace{1cm} (7)

Let the vec() operator create a column vector from $M$ by stacking the column vectors of

$$M = \begin{bmatrix} m_{*,1} & m_{*,2} & m_{*,3} \end{bmatrix}$$

below one another:

$$\text{vec}(M) = \begin{bmatrix} m_{*,1} \\ m_{*,2} \\ m_{*,3} \end{bmatrix}$$  \hspace{1cm} (8)

Eq. 8 can be rewritten as the following linear combination:

$$\text{vec}(M) = f + D \cdot m$$  \hspace{1cm} (9)

where $f = \text{vec}(M_{\text{fixed}})$ and $D \cdot m = \text{vec}(M_{\text{free}})$ with $m$ a $(p \times 1)$ vector containing the $p$ free values, and $D$ a design matrix transforming $m$ into vec($M_{\text{free}}$). To derive the update equations, the likelihood has to be rewritten as a function of vec($M$) where $M$ is whatever parameter matrix for which the update equation is derived [31]. Then, this result is rewritten as a function of $m$ using Eq. 9. The $m$ is found by setting $\frac{\partial \Psi}{\partial m} = 0$. As detailed derivations can be found in [31], only the final results are given here. If we set the following quantities:

$$\tilde{P}_t = P_N^t + x_N^t (x_N^t)^T$$
$$\tilde{P}_{t,t-1} = P_{N,t-1}^t + x_N^t (x_N^{t-1})^T$$

we can write the update equations for estimating the individual parameters of $\theta$ in Eq. 6 like: $u$ update (unknown parameter in Eq. 4):

$$u^{(j+1)} = f_u + D_u \cdot m_u$$

with

$$m_u = \frac{1}{N-1} (D_u^T Q^{-1} D_u)^{-1}$$
$$D_u^T Q^{-1} \sum_{t=2}^{N} (x_t^N - \Phi x_{t-1}^N - f_u)$$
\( \mu \) update (unknown parameter in Eq. 4):
\[
\mu^{(j+1)} = f_\mu + D_\mu \cdot m_\mu
\]
with
\[
m_\mu = (D_\mu^T \Sigma^{-1} D_\mu)^{-1} D_\mu^T (\Sigma^{-1}(x_t^N - f_\mu))
\]
\( \Phi \) update (unknown parameter in Eq. 4):
\[
\text{vec}(\Phi^{(j+1)}) = f_\Phi + D_\Phi \cdot m_\Phi
\]
with
\[
m_\Phi = \left( \sum_{t=2}^N D_\Phi^T (\tilde{P}_{t-1} \otimes Q^{-1}) D_\Phi \right)^{-1}
\]
\[
D_\Phi^T \left( \sum_{t=2}^N \left[ \text{vec}(Q^{-1} \tilde{P}_{t,t-1})
- (\tilde{P}_{t-1} \otimes Q^{-1}) f_\Phi - \text{vec}(Q^{-1} u(x_{t-1}^N)^T) \right] \right)
\]
\( H \) update (unknown parameter in Eq. 5):
\[
\text{vec}(H^{(j+1)}) = f_H + D_H \cdot m_H
\]
with
\[
m_H = \left( \sum_{t=1}^N D_H^T (\tilde{P}_t \otimes R^{-1}) D_H \right)^{-1}
\]
\[
D_H^T \left( \sum_{t=1}^N \left[ \text{vec}(R^{-1} y_t(x_t^N)^T) - (\tilde{P}_t \otimes R^{-1}) f_H \right] \right)
\]
\( Q \) update (unknown parameter in Eq. 4):
\[
\text{vec}(Q^{(j+1)}) = f_Q + D_Q \cdot m_Q
\]
with
\[
m_Q = \frac{1}{N-1} (D_Q^T D_Q)^{-1} D_Q^T \text{vec}(S)
\]
and
\[
S = \sum_{t=2}^N \left( \tilde{P}_t - \tilde{P}_{t,t-1} \Phi^T - \Phi \tilde{P}_{t-1,t} - x_t^N u^T
- u(x_t^N)^T + \Phi \tilde{P}_{t-1} \Phi^T + \Phi x_{t-1}^N u^T \Phi^T
+ u(x_{t-1}^N)^T \Phi^T + uu^T \right)
\]
\( R \) update (unknown parameter in Eq. 5):
\[
\text{vec}(R^{(j+1)}) = f_R + R_R \cdot m_R
\]
with
\[
m_R = \frac{1}{N} (D_R^T D_R)^{-1} D_R^T
\]
\[
\text{vec} \left( \sum_{t=1}^{N} (y_t - Hx_t^N)(y_t - Hx_t^N)^T + HP_t^N H^T \right)
\]
The tensor symbol stands for the Kronecker product. Note that as stated in [34, 31], simultaneous estimation of \( \mu \) and \( \Sigma \) makes the algorithm fail in practice. Thus, as proposed in [34], we kept \( \Sigma \) fixed at a small value.

3.4. Asymptotic Distribution of the MLEs

In the context of SSMs, the MLE \( \hat{\theta} \) is consistent and has an asymptotic normal distribution given by:
\[
\sqrt{N} \left( \hat{\theta} - \theta \right) \xrightarrow{D} N \left( 0, J(\theta)^{-1} \right)
\]
where \( J(\theta) \) is the asymptotic information matrix given by
\[
J(\theta) = \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \left[ -\frac{\partial^2 \log L(\theta|y_t)}{\partial \theta \partial \theta^T} \right]
\]

There are (mainly) 3 conditions for the results stated above to hold. A full treatment of the necessary conditions for the consistency and the asymptotic normality of MLEs can be found in [35]. In this article, we only provide some crucial elements to establish these results, similarly as it is done in [9]. Indeed, it is necessary to assume that the absolute eigenvalues of the matrix \( \Phi \) are less than one. This assumption guarantees that the filter is stable. Moreover, the SSM must be observable and controllable to ensure that the results given in Eq. 10 and 11 hold.

3.5. Practical Issues

In general, the maximization of the likelihood of SSMs is an uneasy task due to the highly nonlinear form of this function. Therefore, the employment of the EM algorithm implies three main issues:

(i) The EM algorithm is a “hill-climbing” algorithm and can therefore converge to a local maxima. This makes the method sensitive to initial conditions. In practice, diffusion priors could be employed to perform some kind of pre-search of the parameter space. Alternatively, the AV, HV and TV methods could be used to obtain an initial guess of some of the parameter values.

(ii) In numerical applications of the EM algorithm, setting a proper stopping criterion is crucial. The work of [12] showed that the EM algorithm is guaranteed to converge to at least a local maxima. Thus, the convergence criterion mentioned in Sec. 3 is in theory correct. In practice however, this is not feasible due to numerical imprecision
and the large number of iteration that is often required to reach the maxima. A typical convergence criterion could be given by:

\[
\frac{L(\theta^{(j+1)} | y_t, x_t) - L(\theta^{(j)} | y_t, x_t)}{0.5 \cdot |L(\theta^{(j+1)} | y_t, x_t) + L(\theta^{(j)} | y_t, x_t) | + c} < \epsilon
\]

where the averaging in the denominator increases the stability of the criterion, and the \(c\) value is used to keep the criterion well behaved in the case where a fixed point is reached.

(iii) Joint estimation of \(\Phi\), \(Q\) and \(u\) can sometimes lead to instabilities. In such situations, the EM algorithm is very likely to diverge. For example, this situation could arise from ridges in the likelihood surface of \(\Phi\) vs. \(u\) and \(\Phi\) vs. \(Q\). In practice, it has been found that eliminating the estimation of the \(u\) parameter in the EM by using classical Least-Square (LS) estimation largely improves the performance of the EM algorithm. However, this renders the inference on \(\theta\) improper since \(u\) is fixed (and thus viewed as true), which may bias the estimation of the remaining parameters.

4. Simulations

To analyze the potential of employing the constrained EM algorithm for estimating parameters of stochastic random processes, we simulate different mixtures of processes described in the introduction where \(y_t\), \(t = 1, \ldots, N\) is the compound random error signal affecting a gyroscope or an accelerometer.

4.1. Random Walk, White Noise and Rate Ramp

We first assume that \(y_t\) is a mixture of WN, RW and RR processes and compare their identification by EM, AV, HV and TV techniques. Such process is modeled as a discrete time-invariant SSM:

\[
x_{t+1} = x_t + w_t + (\omega \Delta t)
y_{t+1} = x_{t+1} + v_{t+1}
\]

such that \(w_t \sim \mathcal{N}(0, q)\) with \(q = \sigma_{RW}^2\) and \(v_t \sim \mathcal{N}(0, r)\) with \(r = \sigma_{WN}^2\). The goal is to estimate the following parameter set:

\[
\theta = \{\sigma_{WN}^2, \sigma_{RW}^2, \omega\}
\]

This specific problem doesn’t require any constraints on the SSM parameters. Thus, the \(M_{\text{fixed}}\) matrices will be null, which leads to the classical unconstrained EM algorithm of [11]. Furthermore, this example allows to yield a complete comparison with AV since all processes are clearly identifiable (see Fig. 1). We simulated 200 signals \(y_t\) with \(N = 6000\) issued from the following true parameter set:

\[
\theta = \{0.04, 4 \cdot 10^{-4}, 0.003\}
\]
To highlight the importance of the initialization, we ran the EM algorithm by starting at two different initial values:

\[
\theta^{(0)} = \{0.25, 10^{-8}, 0.0\} \quad \text{(12)}
\]

\[
\theta^{*(0)} = \{25, 100, 10\} \quad \text{(13)}
\]

where the second set \(\theta^{*(0)}\) contains values which are far away from \(\theta\). The EM estimates are compared to what the AV, HV and TV techniques would provide by fitting lines on the linear regions of the log-log plot which correspond to the WN (slope is \(-1/2\)), RW (slope is \(1/2\)) and RR (slope is \(1\)) processes. Comparison of estimation is shown in Fig. 2 where the \(EM\) and \(EM^{*}\) columns correspond to the EM results when started at \(\theta^{(0)}\) and \(\theta^{*(0)}\), respectively, and the horizontal lines represent the true parameter set \(\theta\).

In the first case \((EM)\), the solution clearly converged to the global maxima of the likelihood function. In the second case \((EM^{*})\), the rate ramp parameter was also correctly estimated since it is not dependent on the estimation of \(\Phi\) (which is fixed). However, the estimation of \(\sigma_{WN}^2\) and \(\sigma_{RW}^2\) (defining \(q\) and \(r\)) is clearly affected by a convergence to a wrong local maxima. The performances of the AV, HV and TV for estimating \(\sigma_{WN}^2\) are relatively similar to the EM algorithm. However, the RW and RR are often not well separable by these method, because the fitted line slopes do not perfectly correspond to the correct values for the respective processes. The TV has a smaller variance than the the AV and HV. But the later has a smaller bias than the two other approaches. Note that the drift \(\omega\) was estimated by LS with the HV.

Figure 2. Performance comparison between the EM algorithm started at good \((EM)\) and bad \((EM^{*})\) initial values, the AV technique \((AV)\), the HV technique with least-squares estimation of the drift \((HV)\) and the TV technique \((TV)\) for 200 simulations of a mixture of WN, RW and RR processes. The true parameters are marked by horizontal lines.
Constrained Expectation-Maximization Algorithm for Stochastic Inertial Error Modeling: Study of Feasibility

4.2. Gauss-Markov, White Noise and Rate Ramp

In this second example, we assume that $y_t$ is a mixture of WN, GM and RR processes. Such combined process is modeled as a discrete time-invariant SSM:

$$
x_{t+1} = (1 - \beta \Delta t) x_t + w_t + (\omega \Delta t)
$$

$$
y_{t+1} = x_{t+1} + v_{t+1}
$$

such that $w_t \sim \mathcal{N}(0, q)$ with $q = 2\beta\sigma_{GM}^2$ and $v_t \sim \mathcal{N}(0, r)$ with $r = \sigma_{WN}^2$. The goal is to estimate the parameter set:

$$\theta = \{\beta, \sigma_{GM}^2, \sigma_{WN}^2, \omega\}$$

defining $\Phi$, $u$, $Q$ and $R$ from the signal $y_t$. Again, this problem does not require any constraint on the SSM parameters, but this time, the $\Phi$ matrix including the $\beta$ value has to be estimated. This makes the global maxima search task in the likelihood “surface” more difficult. To analyse the performance of an EM algorithm in this scenario, EM algorithm was applied on 200 realizations of $y_t$ with $N = 6000$ issued from the following parameters:

$$\theta = \{0.008, 0.25, 0.64, 10^{-3}\}$$

We started the algorithm with the following initial parameter values:

$$\theta^{(0)} = \{10^{-3}, 1.0, 1.0, 0.0\}$$

To highlight the advantage of eliminating some parameters by other estimation techniques, we ran the EM algorithm with estimating $u$ ($EM^*$), and with eliminating it through LS adjustment prior to EM estimation ($EM$). The results are shown in Fig.

![Figure 3](image.png)

**Figure 3.** Performance comparison between the EM algorithm with prior estimation of $u$ by LS ($EM$) and without ($EM^*$) for 200 simulations of a mixture containing WN, GM and RR processes. The true values of the parameters are marked by horizontal lines.
3. They are much better if \( u \) is correctly eliminated from the EM estimation, since the \( \Phi \) and \( Q \) updates depend on \( u \).

We now study the restitution of AV plots by the parameter set \( \hat{\theta} \) estimated via EM. We selected randomly 3 from the 200 estimated parameters \( \hat{\theta} \) for a case where \( u \) was removed by LS. We then computed the AV and PSD for 20 realizations issued from these 3 solutions \( \hat{\theta} \) (thin curves in Fig. 4) and compared them to the respective 3 true signals \( y_t \) (thick curves in Fig. 4). It can be seen that the resulting AV sequences are fairly well contained in the 95% confidence interval associated with AV estimation of the true signals.

4.3. Gauss-Markov, Random Walk, White Noise and Rate Ramp

In this example, we assume that \( y_t \) is a mixture of WN, RW, GM and RR processes. This is modeled as a discrete time-invariant SSM:

\[
\begin{align*}
    x_{t+1} &= \begin{bmatrix} 1 - \beta \Delta t & 0 \\ 0 & 1 \end{bmatrix} x_t + w_t + \begin{bmatrix} 0 \\ \omega \Delta t \end{bmatrix} \\
    y_{t+1} &= \begin{bmatrix} 1 & 1 \end{bmatrix} x_{t+1} + v_{t+1}
\end{align*}
\]
such that $w_t \sim N(0, Q)$ with

$$Q = \begin{bmatrix} 2\beta \sigma_{GM}^2 & 0 \\ 0 & \sigma_{RW}^2 \end{bmatrix}$$

and $v_t \sim N(0, R)$ with

$$R = \begin{bmatrix} \sigma_{WN}^2 \end{bmatrix}$$

The goal is to estimate the parameter set:

$$\theta = \{\beta, \sigma_{GM}^2, \sigma_{RW}^2, \sigma_{WN}^2, \omega\}$$

from the signal $y_t$. Such problem typically requires that some elements in the involved matrices must remain fixed while others are estimated. For example, all elements in $\Phi$ excepting $1 - \beta \Delta t$ must remain fixed. In $u$, the first element must stay null, while only the diagonal of $Q$ contains free elements. Since all the elements of $H$ are fixed, this matrix does not need to be updated. We illustrate the performance of EM by the same procedure as for the previous simulation scenario. The $y_t$ signal is issued from the following parameters:

$$\theta = \{0.008, 0.25, 10^{-8}, 0.09, 10^{-4}\}$$

The initial parameters where set to:

$$\theta^{(0)} = \{10^{-4}, 10^{-6}, 10^{-10}, 2.5 \cdot 10^{-7}, 0\}$$

The results of the 200 runs are shown in Fig. 5. Note that $u$ has been estimated by LS for improving the estimation of the remaining parameters in EM. The estimation appears to be seriously biased, specially for the inverse correlation time of GM and
Figure 6. Results for the signals containing GM, RW, WN and RR processes. Each panel shows the AV of one realization issued from \( \theta \) (thick curve) and 20 simulations driven from the corresponding estimated parameter sets \( \hat{\theta} \) (thin curves).

RW strength which are difficult to separate in the spectral space. As for the previous scenario, we computed the AV for 20 realizations issued from 3 solutions \( \hat{\theta} \) (see Fig. 6). The effect of the bias in some parameters is visible through the systematic overbounding in the middle part of the AV sequences.

5. Application on Real Data Set

We apply the constrained EM algorithm on signals issued from a tactical-grade IMU (IMAR-FSAS [36]). Three hours long static data were collected in constant temperature conditions at a sampling frequency of 100Hz. The AV plots revealed that the gyroscope error signals are mainly composed of a white noise and thus present no need for more sophisticated modeling. However, the AV plot of accelerometer errors (thick curve in Fig. 7 for Y-axis accelerometer) show a more complex structure. The analyses are similar for the X- and Z-axis sensors and are therefore not shown here. Since the slopes of the linear parts in the thick AV curve do not correspond to any of the theoretical processes depicted in Fig. 1, we choose to model this error by superposing two GM
Constrained Expectation-Maximization Algorithm for Stochastic Inertial Error Modeling: Study of Feasibility

processes and a WN. Such model can be written as:

\[
x_{t+1} = \begin{bmatrix} 1 - \beta_1 \Delta t & 0 \\ 0 & 1 - \beta_2 \Delta t \end{bmatrix} x_t + w_t
\]

\[
y_{t+1} = \begin{bmatrix} 1 & 1 \end{bmatrix} x_{t+1} + v_{t+1}
\]

such that \( w_t \sim \mathcal{N}(0, Q) \) with

\[
Q = \begin{bmatrix} 2\beta_1 \sigma_{GM,1}^2 & 0 \\ 0 & 2\beta_2 \sigma_{GM,2}^2 \end{bmatrix}
\]

and \( v_t \sim \mathcal{N}(0, R) \) with

\[
R = \begin{bmatrix} \sigma_{WN}^2 \end{bmatrix}
\]

The goal is to estimate the parameter set:

\[
\theta = \{\beta_1, \beta_2, \sigma_{GM,1}^2, \sigma_{GM,2}^2, \sigma_{WN}^2\}
\]

from the signal \( y_t \). The estimated values for the parameters obtained with the EM algorithm are:

\[
\hat{\theta} = \{0.0004, 0.10, 4 \cdot 10^{-8}, 10^{-8}, 3.6 \cdot 10^{-5}\}
\]

where the units of the \( \beta \) and variances are \([1/s]\) and \([m/s^2]^2\), respectively. The quality of the estimation is illustrated in Fig. 7 in which AV plots of 100 realizations issued from the estimated \( \hat{\theta} \) (thin curves) are compared with that of sensor signal (thick curve). The estimated power level of WN appears to fit well the real signal (left part of AV curve). However, long-term errors modeled by the two GM processes match the signal AV sequence only approximately (right part of AV curve). This can be explained by several reasons, which highlight the limitations of the constrained EM on inertial sensors. First,
the task of identifying the GM parameters within a process containing much higher power of WN is difficult and induce very long convergence time. Second, accumulation of numerical imprecision in many iteration may influence the results if the parameters are of small magnitudes (which is the case). Third, using longer time series would most likely improve the uncertainty of parameter estimation; however, this was not feasible due to memory limitations of the computing hardware. Indeed, increasing the length of the analyzed signal may improve the observation of the underlaying long-term processes (right part in the AV plot), in other words, decreasing the 95% confidence intervals in this region, while improving the estimation of the GM process parameters by the EM algorithm. We will address this problem in future experiments.

In the sequel, we analyze the impact of the estimated model on the INS/GNSS integration via optimal forward Kalman filtering and backward smoothing. For that, the IMAR-FSAS IMU was mounted together with a high-grade dual-frequency GPS receiver (JAVAD) on a car, and the motion was sampled at 100 Hz and 10 Hz, respectively. The carrier-phase GPS observation were double-differenced in post-processing to yield high-precision (cm-level) GPS positioning. These have been combined with the inertial observation in an EKF. To highlight the impact of proper stochastic modeling, we introduced artificially two outages in GPS solutions of different duration, at times where good and reliable GPS solutions were available as reference. During these outages, the navigation solution is solely dependent on inertial navigation, meaning that the residual systematic errors affecting these signals are integrated with time. We then recomputed the INS/GPS trajectory using the traditional IMAR-FSAS stochastic error model provided by the manufacturer, and compared it to the EKF/smoothed solution using the EM-estimated model. In both cases we compare the positioning differences (N-E-D) with respect to the reference. This allows to compute the positioning error along each direction in the local-level frame (North, East and Down axes) by comparing both solutions with the reference trajectory (the one without gap).

The first 20s-long outage has been introduced in a time during which the car was turning in a roundabout. Fig. 8 depicts the processed position differences along each axis when using the traditional IMAR-FSAS model (full curves) and the new model (dotted curves). Except for the East component, the new model significantly decreased the trajectory errors based on inertial coasting during this period. The second outage was longer (about 130s) and affected a period in which the car was moving on a straight road, and its acceleration varied. As shown in Fig. 9, the filtered trajectory errors were better bounded at the end of this outage (by a factor of 2-4) when using the new model. Indeed, the maximum observed difference could be decreased from 23 m to 10 m along the North component, from -6 m to -1.2 m along the East component, and from 16 m to 4.2 m along the vertical component.
6. Conclusions and Perspectives

The potential of using both, an unconstrained and constrained EM algorithm for estimating mixtures of stochastic processes for inertial sensor error modeling has been analyzed. In particular, the often encountered problem of selecting the inverse correlation time and driving noise variance of a first-order GM process in an EKF used in INS/GNSS integration has been treated. The focus was specially given to cases where classical AV, HV or TV techniques cannot be used. It was demonstrated that an EM algorithm can successfully estimate such parameters if preceded by some a priori data processing like drift removing. However, it was also shown that this method is likely to converge to a local maxima if the initial values are “far” from the “true” parameter values. Therefore, using diffusion priors or search of parameter initial values is recommended when applying this method in practice. Finally, the algorithm has been applied to real inertial sensors for estimating an error structure whose complexity would make its estimation impossible using classical (i.e. AV, HV or TV) techniques. Its positive effects on Kalman-filtered trajectory were demonstrated by showing the mitigation of the trajectory errors during the artificial absence of external measurements. The performed studies also revealed an important limit of the proposed method. Indeed,

![Figure 8](image)

**Figure 8.** Position errors along North (first panel), East (second panel) and Down (third panel) component occurring during the 20s-long GPS outage when using the traditional error model (full curves) and the new model (dotted curves) in the EKF.
as the likelihood “surfaces” encountered with SSMs are often very complex and highly nonlinear, the (EM) optimization of this function may become consequently unstable and can diverge. However, in contrast to the AV, HV and TV methods, the EM based approach enables computing confidence intervals for the estimated parameters. Nevertheless, as shown in the lastly simulated example and the real application, there are at the moment no other estimation methods that can determine successfully very complex error structures.

Figure 9. Position errors along North (first panel), East (second panel) and Down (third panel) component occurring during the 130s-long GPS outage when using the traditional error model (full curves) and the new model (dotted curves) in the EKF.
References

Appendix

The Kalman smoother estimator can be computed by first performing forward Kalman filtering. For $t = 2, \ldots, N$, compute:

\[
\begin{align*}
    x_{t-1}^t &= \Phi x_{t-1}^{t-1} + u_t \\
    P_{t-1}^t &= \Phi P_{t-1}^{t-1} \Phi^T + Q \\
    K_t &= P_{t-1}^t H^T (HP_{t-1}^{t-1} H^T + R)^{-1} \\
    x_t' &= x_{t-1}^t + K_t (y_t - Hx_{t-1}^t) \\
    P_t' &= P_{t-1}^t - K_t H P_{t-1}^{t-1}
\end{align*}
\]
where the initial conditions are given by $x_1 = \mu$ and $P_1 = \Sigma$. The backward recursion can be calculated for $t = N, \ldots, 1$:

\[
J_{t-1} = P_{t-1}^{t-1} \Phi^T (P_t^{-1})^{-1}
\]

\[
x_{t-1}^N = x_{t-1}^{t-1} + J_{t-1} (x_t^N - \Phi x_{t-1}^{t-1})
\]

\[
P_{t-1}^N = P_{t-1}^{t-1} + J_{t-1} (P_t^N - P_t^{t-1}) J_{t-1}^T
\]

The covariance matrix $P_{t,t-1}^N$ can be computed for $t = N, \ldots, 2$ using the following relation:

\[
P_{t-1,t-2} = P_{t-1}^{t-1} J_{t-2}^T + J_{t-1} (P_{t,t-1}^N - \Phi P_{t-1}^{t-1}) J_{t-2}^T
\]

where

\[
P_{N,N-1}^N = (I - K_N H) \Phi P_{N-1}^{N-1}
\]