Meas. Sci. Technol. **22** (2011) 085204 (12pp) doi:10.1088/0957-0233/22/8/085204

Constrained expectation-maximization algorithm for stochastic inertial error modeling: study of feasibility

Yannick Stebler[1], Stéphane Guerrier[2], Jan Skaloud[1] and Maria-Pia Victoria-Feser[2]

[1] Geodetic Engineering Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland
[2] Research Center for Statistics, University of Geneva, Geneva, Switzerland

E-mail: yannick.stebler@epfl.ch, Stephane.Guerrier@unige.ch, jan.skaloud@epfl.ch and Maria-Pia.VictoriaFeser@unige.ch

Received 17 March 2011, in final form 27 May 2011
Published 15 July 2011
Online at stacks.iop.org/MST/22/085204

Abstract
Stochastic modeling is a challenging task for low-cost sensors whose 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 an 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 the parameters of complex error models reasonably well 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

(Some figures in this article are in colour only in the electronic version)

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 grows 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 paper; see [1, 2] for more details. On the other hand, stochastic errors contain components which have random behavior (e.g. dynamics-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 the 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 that 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 does not allow the direct identification of processes and their parameter(s) when the signal is composed of mixtures of processes [6]. On the other hand, the AV, HD and 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 subject 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 β and the driving noise variance σ^2_{GM} of a GM process when mixed with some other processes. Large values of β make the GM process approach WN, while small values give 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 equations (4) and (5)) is in general quite challenging. Indeed, the (log) likelihood function $\log L(\boldsymbol{\theta}|\mathbf{y}_t, \mathbf{x}_t)$, where $\boldsymbol{\theta}$ is the vector of unknown parameters that defines the SSM, $\mathbf{y}_t$ is the measurement vector and $\mathbf{x}_t$ is the system state vector, is a highly nonlinear and complicated function [9]. Historically, a Newton–Raphson algorithm was employed to successively update $\boldsymbol{\theta}$ 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 an approach, two different likelihood problems are considered. The first is the problem we are interested in solving, 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 $\mathbf{x}_t$ in addition to the observations $\mathbf{y}_t$. Then the log likelihood function of this problem $\log L(\boldsymbol{\theta}|\mathbf{x}_t, \mathbf{y}_t)$ could easily be 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 $\mathbf{x}_t$ 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 $\hat{\boldsymbol{\theta}}$ 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 behavior 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 β of GM).

In this paper, we show the adaptation of the classical EM algorithm to constrain the SSM parameters. Such constraints allow estimating more complex stochastic error models such as those used in INS/GNSS filters/smoothers. Realistic examples are 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 are compared with the benchmark methods, i.e. the AV, HV and TV. Later, an example with real data is included to demonstrate the positive impact of the derived stochastic model on the INS/GNSS trajectory.

This paper is organized as follows: in section 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 section 3, we explain the theory related to the constrained EM algorithm applied to SSMs. In section 4, we present a simulation study which allows comparison of the performances of the estimation based on the EM algorithm with those based on the AV, HV and TV. Finally, in section 5, we illustrate the EM approach with real inertial signals and 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, HV or TV [16–19]. The AV was invented in 1966 by David Allan when he criticized the use of the sample variance estimator in the context of non-*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 [6, 20–24]. 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, $t = 1, 2, \ldots, N$, 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 therefore represents the error. Let $\bar{y}_t(\tau)$ be the sample average of τ 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 τ (noted as $\sigma^2_{\bar{y}}(\tau)$) 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 $\bar{y}_t(\tau)$:

$$\sigma^2_{\bar{y}}(\tau) = \frac{1}{2}\mathbb{E}[(\bar{y}_t(\tau) - \bar{y}_{t-\tau}(\tau))^2]. \tag{1}$$

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

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

The AV can be expressed in the frequency domain through the unique relationship between $\sigma^2_{\bar{y}}(\tau)$ and the power spectral density (PSD) $S_{yy}(f)$ of the intrinsic processes [23]:

$$\sigma^2_{\bar{y}}(\tau) = 4\lim_{a\to\infty}\int_0^a S_{yy}(f)\frac{\sin^4(\pi f\tau)}{(\pi f\tau)^2}\,\mathrm{d}f,$$

where f is the frequency. This equation enables the parameter vector $\boldsymbol{\theta}$ of some SSM to be linked to $\sigma^2_{\bar{y}}(\tau)$. Indeed, this relationship is due to the known form of the PSD function characterizing different noise processes which enables $\boldsymbol{\theta}$ to be expressed as a function of $\sigma^2_{\bar{y}}(\tau)$ (a detailed discussion on how to express this link between $\boldsymbol{\theta}$ and $\sigma^2_{\bar{y}}(\tau)$ 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 the linear regions in a '$\sigma_{\bar{y}}(\tau)$ versus τ' log–log plot as schematically depicted in figure 1. Therefore, $\boldsymbol{\theta}$ is usually estimated by performing linear regression of (visually) identified linear regions in such log–log plots. This method is well defined only for a few type of processes, and it is not clear how inference on $\boldsymbol{\theta}$ 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

$$\sigma^2_H = \tfrac{1}{6}\mathbb{E}[(\bar{y}_t(\tau) - 2\bar{y}_{t-\tau}(\tau) + \bar{y}_{t-2\tau}(\tau))^2].$$

An estimator of σ^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 (equation (1)) that has less mean square error than the standard unbiased estimator (equation (2)). The TV can be defined as

$$\hat{\sigma}^2_T = \frac{1}{N-1}\sum_{k=1}^{N-1}\left[\frac{1}{2(N-2\tau+1)} \times \sum_{t=2\tau}^{N}(\bar{y}_{t,k}(\tau) - \bar{y}_{t-\tau,k}(\tau))^2\right],$$

where $\{\bar{y}_{t,k}(\tau)\} = \bar{y}_{k+1}(\tau), \bar{y}_{k+2}(\tau), \ldots, \bar{y}_1(\tau), \bar{y}_2(\tau), \ldots, \bar{y}_k(\tau)$. 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:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{w}(t) + \mathbf{L}(t)\mathbf{u}(t), \tag{3}$$

where $\mathbf{x}(t)$ is the $p \times 1$ system state vector at time t, $\mathbf{F}(t)$ is the $p \times p$ time-varying dynamic coefficient matrix, $\mathbf{G}(t)$ is the $p \times r$ time-varying process noise coupling matrix, $\mathbf{w}(t)$ is a $r \times 1$ random forcing function such that $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$,

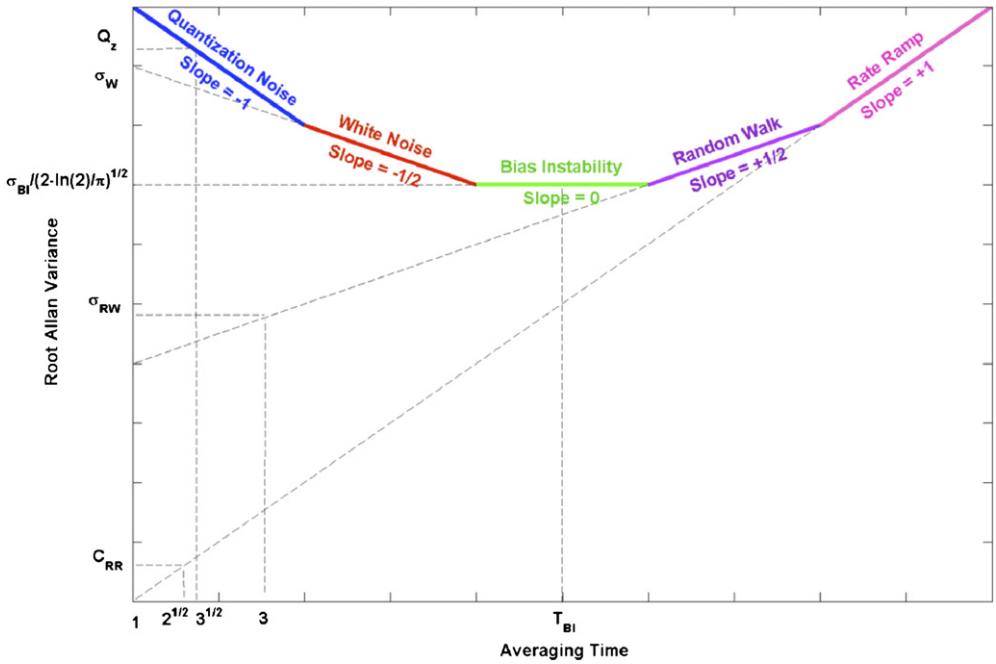

Figure 1. Schematic sample representation of the root Allan variance using analytical results.

$\mathbf{L}(t)$ is a $p \times r$ time-varying input coupling matrix and $\mathbf{u}$ is a $r \times 1$ deterministic input vector. Integration and discretization of equation (3) yield the discrete SSM:

$$\mathbf{x}_{t+1} = \mathbf{\Phi}_t \mathbf{x}_t + \mathbf{w}_t + \boldsymbol{u}_t \tag{4}$$

with measurements

$$\mathbf{y}_{t+1} = \mathbf{H}_{t+1}\mathbf{x}_{t+1} + \mathbf{v}_{t+1}, \tag{5}$$

where $\mathbf{\Phi}_t$ is the discrete form of $\mathbf{F}(t)$, $\mathbf{H}_t$ is the $l \times p$ design matrix which maps the true state space $\mathbf{x}_t$ into the observed space and $\mathbf{v}_t$ is a $l \times 1$ noise vector such that $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. The initial state $\mathbf{x}_1$ is assumed to be a normal random vector with mean vector $\boldsymbol{\mu}$ and $p \times p$ initial covariance matrix $\mathbf{\Sigma}$ (i.e. $\mathbf{P}_1^1$).

As mentioned earlier, the vector $\boldsymbol{\theta} \in \mathbf{\Theta}$ is the set of SSM parameters such that

$$\boldsymbol{\theta} = \{\mathbf{u}, \mathbf{\Phi}, \mathbf{Q}, \mathbf{R}, \mathbf{H}, \boldsymbol{\mu}, \mathbf{\Sigma}\}. \tag{6}$$

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 $\boldsymbol{\theta}$ given $\mathbf{y}_t$ and $\mathbf{x}_t$ for $t = 1, \ldots, N$ is

$$\log L(\boldsymbol{\theta}|\mathbf{y}_t, \mathbf{x}_t) = -\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu})$$
$$-\frac{1}{2}\sum_{t=2}^{N}(\mathbf{x}_t - \mathbf{\Phi}\mathbf{x}_{t-1} - \mathbf{u}_t)^T \mathbf{Q}^{-1}(\mathbf{x}_t - \mathbf{\Phi}\mathbf{x}_{t-1} - \mathbf{u}_t)$$
$$-\frac{1}{2}\sum_{t=1}^{N}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t)^T \mathbf{R}^{-1}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t) - \frac{1}{2}\log|\mathbf{\Sigma}|$$
$$-\frac{1}{2}N\log|\mathbf{Q}| - \frac{1}{2}N\log|\mathbf{R}| - \frac{N}{2}\log 2\pi.$$

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

3.3. The algorithm

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

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

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}) = \mathbb{E}[\log L(\boldsymbol{\theta}^{(j)}|\mathbf{y}_t, \mathbf{x}_t)] = \mathbf{\Psi}^{(j)}.$$

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

$$\boldsymbol{\theta}^{(j+1)} = \underset{\boldsymbol{\theta}\in\mathbf{\Theta}}{\operatorname{argmax}}\, Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(j)}).$$

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

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

$$\mathbf{x}_t^N = \mathbb{E}[\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_N, \boldsymbol{\theta}^{(j)}]$$
$$\mathbf{P}_t^N = \mathrm{cov}[\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_N, \boldsymbol{\theta}^{(j)}]$$
$$\mathbf{P}_{t,t-1}^N = \mathrm{cov}[\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{y}_1, \ldots, \mathbf{y}_N, \boldsymbol{\theta}^{(j)}],$$

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 $\boldsymbol{\theta}^{(j+1)}$ by finding the parameters that maximize $\boldsymbol{\Psi}^{(j)}$ considering the values $\mathbf{x}_t^N$, $\mathbf{P}_t^N$ and $\mathbf{P}_{t,t-1}^N$ obtained in the E-step as fixed. For doing that, the expression yielded by $\boldsymbol{\Psi}^{(j)}$ is minimized by computing the partial derivatives with respect to $\boldsymbol{\theta}^{(j)}$ 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 [31, 33] provides the way of constraining the elements in the matrices of $\boldsymbol{\theta}$. Let $\mathbf{M}$ be any matrix contained in $\boldsymbol{\theta}$ (e.g. $\boldsymbol{\Phi}$, $\mathbf{Q}$, $\mathbf{H}$, $\mathbf{R}$) with fixed and p free elements (i.e. to be estimated). This matrix can be decomposed into matrices containing the fixed and free elements:

$$\mathbf{M} = \mathbf{M}_{\text{fixed}} + \mathbf{M}_{\text{free}}. \tag{7}$$

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

$$\mathbf{M} = [\mathbf{m}_{\bullet,1} \quad \mathbf{m}_{\bullet,2} \quad \mathbf{m}_{\bullet,3}]$$

below one another:

$$\text{vec}(\mathbf{M}) = \begin{bmatrix} \mathbf{m}_{\bullet,1} \\ \mathbf{m}_{\bullet,2} \\ \mathbf{m}_{\bullet,3} \end{bmatrix}. \tag{8}$$

Equation (8) can be rewritten as the following linear combination:

$$\text{vec}(\mathbf{M}) = \mathbf{f} + \mathbf{D} \cdot \mathbf{m}, \tag{9}$$

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

$$\widetilde{\mathbf{P}}_t = \mathbf{P}_t^N + \mathbf{x}_t^N \left(\mathbf{x}_t^N\right)^T$$
$$\widetilde{\mathbf{P}}_{t,t-1} = \mathbf{P}_{t,t-1}^N + \mathbf{x}_t^N \left(\mathbf{x}_{t-1}^N\right)^T,$$

we can write the update equations for estimating the individual parameters of $\boldsymbol{\theta}$ in equation (6) such as $\mathbf{u}$ update (unknown parameter in equation (4)):

$$\mathbf{u}^{(j+1)} = \mathbf{f}_{\mathbf{u}} + \mathbf{D}_{\mathbf{u}} \cdot \mathbf{m}_{\mathbf{u}},$$

with

$$\mathbf{m}_{\mathbf{u}} = \frac{1}{N-1} \left(\mathbf{D}_{\mathbf{u}}^T \mathbf{Q}^{-1} \mathbf{D}_{\mathbf{u}}\right)^{-1} \mathbf{D}_{\mathbf{u}}^T \mathbf{Q}^{-1} \sum_{t=2}^{N} \left(\mathbf{x}_t^N - \boldsymbol{\Phi} \mathbf{x}_{t-1}^N - \mathbf{f}_{\mathbf{u}}\right);$$

$\boldsymbol{\mu}$ update (unknown parameter in equation (4)):

$$\boldsymbol{\mu}^{(j+1)} = \mathbf{f}_{\boldsymbol{\mu}} + \mathbf{D}_{\boldsymbol{\mu}} \cdot \mathbf{m}_{\boldsymbol{\mu}}$$

with

$$\mathbf{m}_{\boldsymbol{\mu}} = \left(\mathbf{D}_{\boldsymbol{\mu}}^T (\boldsymbol{\Sigma})^{-1} \mathbf{D}_{\boldsymbol{\mu}}\right)^{-1} \mathbf{D}_{\boldsymbol{\mu}}^T (\boldsymbol{\Sigma})^{-1} \left(\mathbf{x}_1^N - \mathbf{f}_{\boldsymbol{\mu}}\right);$$

$\boldsymbol{\Phi}$ update (unknown parameter in equation (4)):

$$\text{vec}(\boldsymbol{\Phi}^{(j+1)}) = \mathbf{f}_{\boldsymbol{\Phi}} + \mathbf{D}_{\boldsymbol{\Phi}} \cdot \mathbf{m}_{\boldsymbol{\Phi}}$$

with

$$\mathbf{m}_{\boldsymbol{\Phi}} = \left(\sum_{t=2}^{N} \mathbf{D}_{\boldsymbol{\Phi}}^T (\widetilde{\mathbf{P}}_{t-1} \otimes \mathbf{Q}^{-1}) \mathbf{D}_{\boldsymbol{\Phi}} \right)^{-1} \times \mathbf{D}_{\boldsymbol{\Phi}}^T \left(\sum_{t=2}^{N} \left[\text{vec}(\mathbf{Q}^{-1} \widetilde{\mathbf{P}}_{t,t-1}) - (\widetilde{\mathbf{P}}_{t-1} \otimes \mathbf{Q}^{-1}) \mathbf{f}_{\boldsymbol{\Phi}} - \text{vec}\left(\mathbf{Q}^{-1} \mathbf{u} \left(\mathbf{x}_{t-1}^N\right)^T\right) \right] \right);$$

$\mathbf{H}$ update (unknown parameter in equation (5)):

$$\text{vec}(\mathbf{H}^{(j+1)}) = \mathbf{f}_{\mathbf{H}} + \mathbf{D}_{\mathbf{H}} \cdot \mathbf{m}_{\mathbf{H}}$$

with

$$\mathbf{m}_{\mathbf{H}} = \left(\sum_{t=1}^{N} \mathbf{D}_{\mathbf{H}}^T (\widetilde{\mathbf{P}}_t \otimes \mathbf{R}^{-1}) \mathbf{D}_{\mathbf{H}} \right)^{-1} \times \mathbf{D}_{\mathbf{H}}^T \left(\sum_{t=1}^{N} \left[\text{vec}\left(\mathbf{R}^{-1} \mathbf{y}_t \left(\mathbf{x}_t^N\right)^T\right) - (\widetilde{\mathbf{P}}_t \otimes \mathbf{R}^{-1}) \mathbf{f}_{\mathbf{H}} \right] \right);$$

$\mathbf{Q}$ update (unknown parameter in equation (4)):

$$\text{vec}(\mathbf{Q}^{(j+1)}) = \mathbf{f}_{\mathbf{Q}} + \mathbf{D}_{\mathbf{Q}} \cdot \mathbf{m}_{\mathbf{Q}}$$

with

$$\mathbf{m}_{\mathbf{Q}} = \frac{1}{N-1} \left(\mathbf{D}_{\mathbf{Q}}^T \mathbf{D}_{\mathbf{Q}}\right)^{-1} \mathbf{D}_{\mathbf{Q}}^T \text{vec}(\mathbf{S})$$

and

$$\mathbf{S} = \sum_{t=2}^{N} \left(\widetilde{\mathbf{P}}_t - \widetilde{\mathbf{P}}_{t,t-1} \boldsymbol{\Phi}^T - \boldsymbol{\Phi} \widetilde{\mathbf{P}}_{t-1,t} - \mathbf{x}_t^N \mathbf{u}^T - \mathbf{u} \left(\mathbf{x}_t^N\right)^T + \boldsymbol{\Phi} \widetilde{\mathbf{P}}_{t-1} \boldsymbol{\Phi}^T + \boldsymbol{\Phi} \mathbf{x}_{t-1}^N \mathbf{u}^T \boldsymbol{\Phi}^T + \mathbf{u} \left(\mathbf{x}_{t-1}^N\right)^T \boldsymbol{\Phi}^T + \mathbf{u}\mathbf{u}^T \right)$$

$\mathbf{R}$ update (unknown parameter in equation (5)):

$$\text{vec}(\mathbf{R}^{(j+1)}) = \mathbf{f}_{\mathbf{R}} + \mathbf{R}_{\mathbf{R}} \cdot \mathbf{m}_{\mathbf{R}}$$

with

$$\mathbf{m}_{\mathbf{R}} = \frac{1}{N} \left(\mathbf{D}_{\mathbf{R}}^T \mathbf{D}_{\mathbf{R}}\right)^{-1} \mathbf{D}_{\mathbf{R}}^T \times \text{vec}\left(\sum_{t=1}^{N} \left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^N\right)\left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^N\right)^T + \mathbf{H}\mathbf{P}_t^N \mathbf{H}^T \right).$$

The $\otimes$ symbol stands for the Kronecker product. Note that as stated in [31, 34], simultaneous estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ makes the algorithm fail in practice. Thus, as proposed in [34], we kept $\boldsymbol{\Sigma}$ fixed at a small value.

3.4. Asymptotic distribution of the MLEs

In the context of SSMs, the MLE $\hat{\boldsymbol{\theta}}$ is consistent and has an asymptotic normal distribution given by

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \boldsymbol{J}(\boldsymbol{\theta})^{-1}), \tag{10}$$

where $\boldsymbol{J}(\boldsymbol{\theta})$ is the asymptotic information matrix given by

$$\boldsymbol{J}(\boldsymbol{\theta}) = \lim_{N \to \infty} \frac{1}{N} \mathbb{E}\left[-\frac{\partial^2 \log L(\boldsymbol{\theta}|\mathbf{y}_t)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]. \tag{11}$$

There are (mainly) three 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 paper, we provide only some crucial elements to establish these results, similarly as done in [9]. Indeed, it is necessary to assume that the absolute eigenvalues of the matrix $\mathbf{\Phi}$ are less than 1. This assumption guarantees that the filter is stable. Moreover, the SSM must be observable and controllable to ensure that the results given in equations (10) and (11) hold.

3.5. Practical issues

In general, the maximization of the likelihood of SSMs is not an easy 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 maximum. 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 [12] showed that the EM algorithm is guaranteed to converge to at least a local maximum. Thus, the convergence criterion mentioned in section 3 is in theory correct. In practice, however, this is not feasible due to numerical imprecision and the large number of iterations that is often required to reach the maximum. A typical convergence criterion could be given by

$$\frac{L(\boldsymbol{\theta}^{(j+1)}|\mathbf{y}_t, \mathbf{x}_t) - L(\boldsymbol{\theta}^{(j)}|\mathbf{y}_t, \mathbf{x}_t)}{0.5 \cdot |L(\boldsymbol{\theta}^{(j+1)}|\mathbf{y}_t, \mathbf{x}_t) + L(\boldsymbol{\theta}^{(j)}|\mathbf{y}_t, \mathbf{x}_t)| + c} < \epsilon,$$

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 $\mathbf{\Phi}$, $\mathbf{Q}$ and $\mathbf{u}$ 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 $\mathbf{\Phi}$ versus $\mathbf{u}$ and $\mathbf{\Phi}$ versus $\mathbf{Q}$. In practice, it has been found that eliminating the estimation of the $\mathbf{u}$ parameter in the EM by using classical least-squares (LS) estimation largely improves the performance of the EM algorithm. However, this renders the inference on $\boldsymbol{\theta}$ improper since $\mathbf{u}$ 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 a 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 $q = \sigma^2_{RW}$ and $v_t \sim \mathcal{N}(0, r)$ with $r = \sigma^2_{WN}$. The goal is to estimate the following parameter set:

$$\boldsymbol{\theta} = \left\{\sigma^2_{WN}, \sigma^2_{RW}, \omega\right\}.$$

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

$$\boldsymbol{\theta} = \{0.04, 4 \times 10^{-4}, 0.003\}.$$

To highlight the importance of the initialization, we ran the EM algorithm by starting at two different initial values:

$$\boldsymbol{\theta}^{(0)} = \{0.25, 10^{-8}, 0.0\} \tag{12}$$

$$\boldsymbol{\theta}^{*(0)} = \{25, 100, 10\}, \tag{13}$$

where the second set $\boldsymbol{\theta}^{*(0)}$ contains values which are far away from $\boldsymbol{\theta}$. 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. A comparison of estimation is shown in figure 2 where the *EM* and *EM** columns correspond to the EM results when started at $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{\theta}^{*(0)}$, respectively, and the horizontal lines represent the true parameter set $\boldsymbol{\theta}$.

In the first case (*EM*), the solution clearly converged to the global maxima of the likelihood function. In the second case (*EM**), the RR parameter was also correctly estimated since it is not dependent on the estimation of $\mathbf{\Phi}$ (which is fixed). However, the estimation of σ^2_{WN} and σ^2_{RW} (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 σ^2_{WN} are relatively similar to the EM algorithm. However, the RW and RR are often not well separable by these methods 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 latter has a smaller bias than the two other approaches. Note that the drift ω was estimated by LS with the HV.

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 a combined process is modeled as a discrete time-invariant SSM:

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

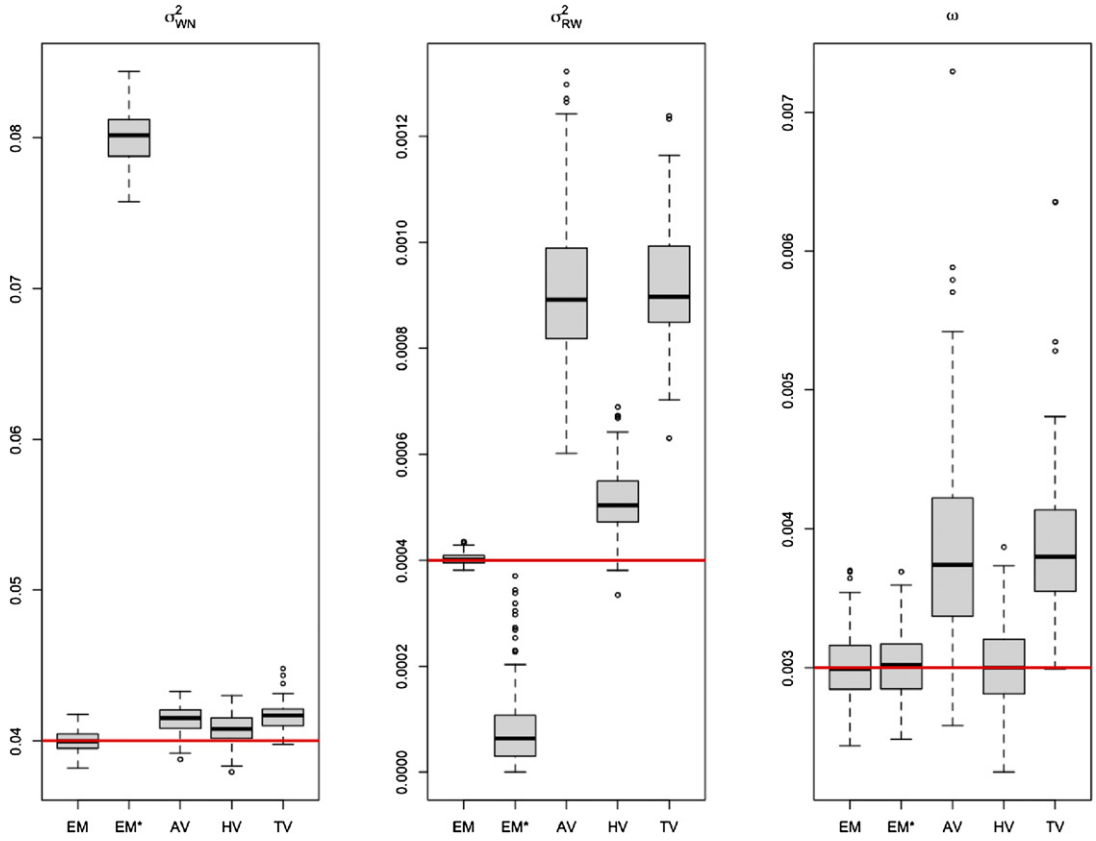

Figure 2. Performance comparison between the EM algorithm started at good (*EM*) and bad (*EM**) initial values, the AV technique (*AV*), the HV technique with least-squares estimation of the drift (*HV*) and the TV technique (*TV*) for 200 simulations of a mixture of WN, RW and RR processes. The true parameters are marked by horizontal lines.

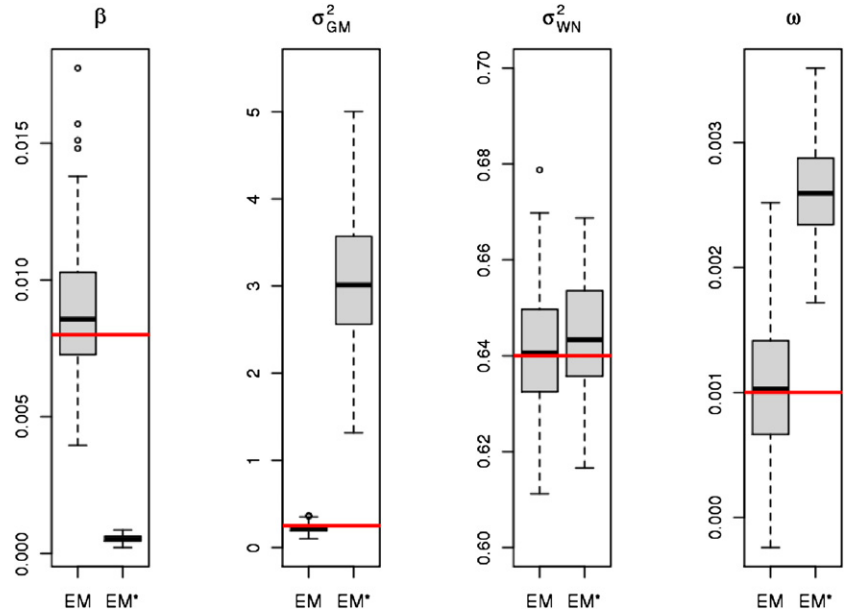

Figure 3. Performance comparison between the EM algorithm with prior estimation of **u** by LS (*EM*) and without (*EM**) for 200 simulations of a mixture containing WN, GM and RR processes. The true values of the parameters are marked by horizontal lines.

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

$$\boldsymbol{\theta} = \left\{\beta, \sigma^2_{GM}, \sigma^2_{WN}, \omega\right\}$$

defining $\boldsymbol{\Phi}$, **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 $\boldsymbol{\Phi}$ matrix including the β value has to be estimated.

This makes the global maxima search task in the likelihood 'surface' more difficult. To analyze the performance of an EM algorithm in this scenario, the EM algorithm was applied on 200 realizations of y_t with $N = 6000$ issued from the following parameters:

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

We started the algorithm with the following initial parameter values:

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

To highlight the advantage of eliminating some parameters by other estimation techniques, we ran the EM algorithm both estimating $\mathbf{u}$ (*EM**) and eliminating it through LS adjustment prior to EM estimation (*EM*). The results are shown in figure 3. They are much better if $\mathbf{u}$ is correctly eliminated from the EM estimation, since the $\boldsymbol{\Phi}$ and $\mathbf{Q}$ updates depend on $\mathbf{u}$.

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

$$\mathbf{x}_{t+1} = \begin{bmatrix} 1 - \beta\Delta t & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_t + \mathbf{w}_t + \begin{bmatrix} 0 \\ \omega\Delta t \end{bmatrix}$$

$$y_{t+1} = [1 \quad 1\,]\mathbf{x}_{t+1} + v_{t+1}$$

such that $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ with

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

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

$$\mathbf{R} = \left[\sigma_{WN}^2\right].$$

The goal is to estimate the parameter set

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

from the signal y_t. Such a problem typically requires that some elements in the involved matrices must remain fixed while others are estimated. For example, all elements in $\boldsymbol{\Phi}$ excepting $1 - \beta\Delta t$ must remain fixed. In $\mathbf{u}$, the first element must stay null, while only the diagonal of $\mathbf{Q}$ contains free elements. Since all the elements of $\mathbf{H}$ 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:

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

The initial parameters were set to

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

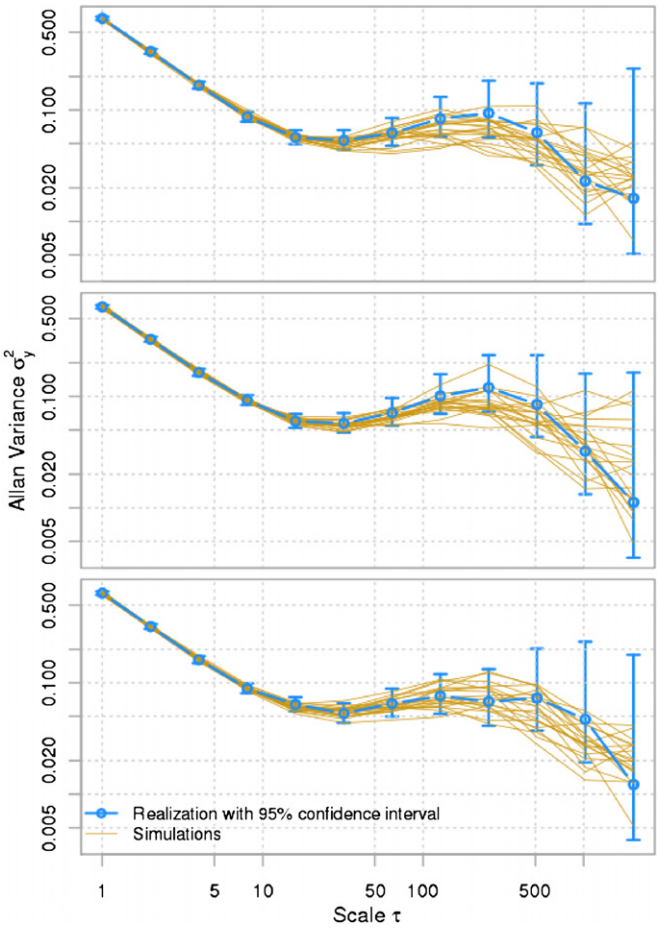

Figure 4. Results for the signals containing GM, WN and RR processes. Each panel shows the AV of one realization issued from $\boldsymbol{\theta}$ (thick curve) and 20 simulations driven from the corresponding estimated parameter sets $\hat{\boldsymbol{\theta}}$ (thin curves).

The results of the 200 runs are shown in figure 5. Note that $\mathbf{u}$ 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 RW strength, which are difficult to separate in the spectral space. As for the previous scenario, we computed the AV for 20 realizations of 3 solutions $\hat{\boldsymbol{\theta}}$ (see figure 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 to the real data set

We apply the constrained EM algorithm to signals from a tactical-grade IMU (*IMAR-FSAS* [36]). Three hours of static data were collected under constant temperature conditions at a sampling frequency of 100 Hz. The AV plots revealed that the gyroscope error signals are mainly composed of WN and thus present no need for more sophisticated modeling. However, the AV plot of accelerometer errors (thick curve in figure 7 for the *Y*-axis accelerometer) shows a more complex structure. The analyses are similar for the *X*- and *Z*-axes sensors and are therefore not shown here. Since

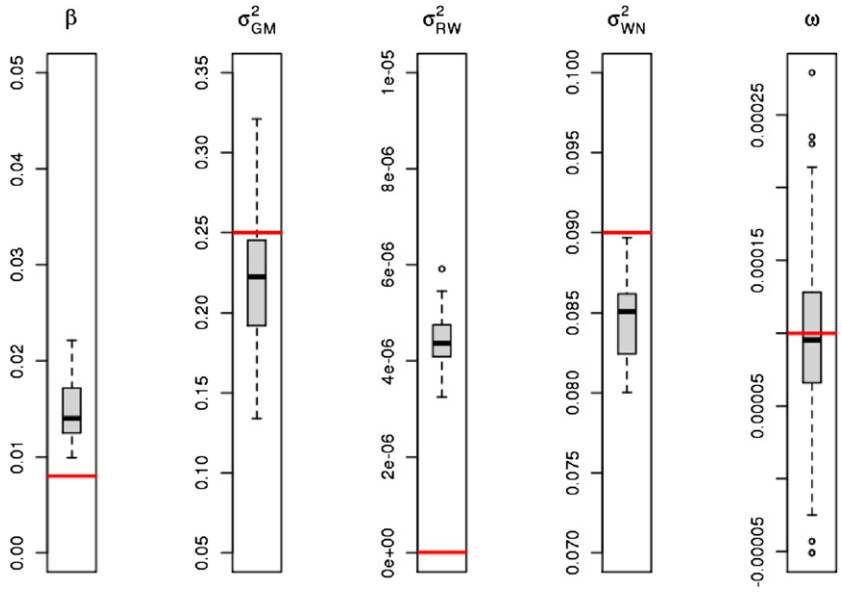

Figure 5. Performance of the EM algorithm where **u** has been removed by LS for 200 simulations of a mixture containing WN, GM, RW and RR processes. The true parameters are marked by horizontal lines.

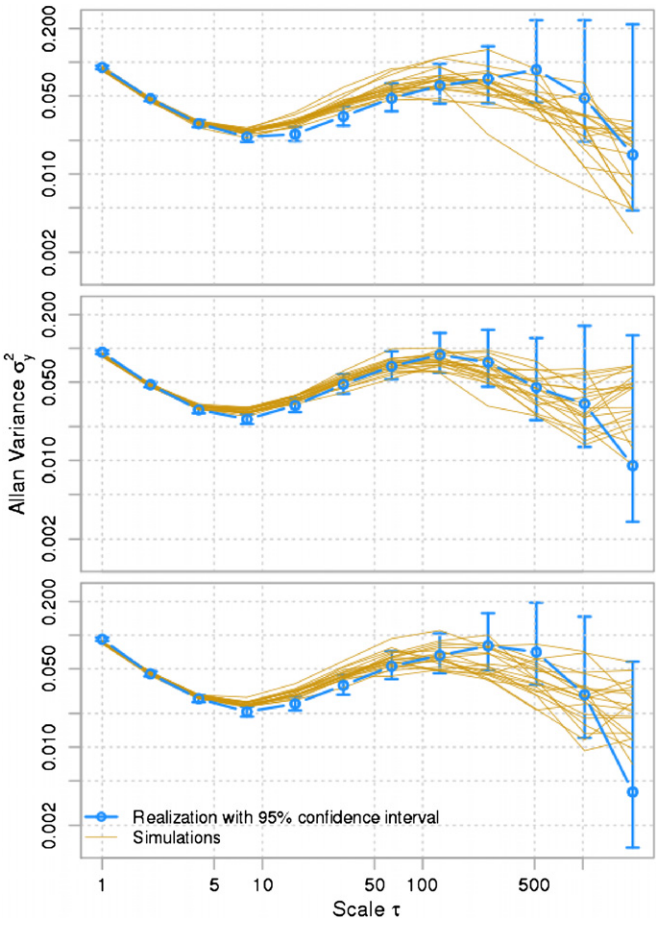

Figure 6. Results for the signals containing GM, RW, WN and RR processes. Each panel shows the AV of one realization issued from $\boldsymbol{\theta}$ (thick curve) and 20 simulations driven from the corresponding estimated parameter sets $\hat{\boldsymbol{\theta}}$ (thin curves).

the slopes of the linear parts in the thick AV curve do not correspond to any of the theoretical processes depicted in figure 1, we choose to model this error by superposing two GM processes and a WN. Such a model can be written as follows:

$$\mathbf{x}_{t+1} = \begin{bmatrix} 1-\beta_1\Delta t & 0 \\ 0 & 1-\beta_2\Delta t \end{bmatrix}\mathbf{x}_t + \mathbf{w}_t$$

$$y_{t+1} = [1 \quad 1]\mathbf{x}_{t+1} + v_{t+1}$$

such that $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ with

$$\mathbf{Q} = \begin{bmatrix} 2\beta_1\sigma^2_{GM,1} & 0 \\ 0 & 2\beta_2\sigma^2_{GM,2} \end{bmatrix}$$

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

$$\mathbf{R} = \left[\sigma^2_{WN}\right].$$

The goal is to estimate the parameter set

$$\boldsymbol{\theta} = \left\{\beta_1, \beta_2, \sigma^2_{GM,1}, \sigma^2_{GM,2}, \sigma^2_{WN}\right\}$$

from the signal y_t. The estimated values for the parameters obtained with the EM algorithm are

$$\hat{\boldsymbol{\theta}} = \{0.0004, 0.10, 4\times10^{-8}, 10^{-8}, 3.6\times10^{-5}\},$$

where the units of the β and variances are (s^{-1}) and $((m\,s^{-2})^2)$, respectively. The quality of the estimation is illustrated in figure 7 in which the AV plots of 100 realizations of the estimated $\hat{\boldsymbol{\theta}}$ (thin curves) are compared with that of the sensor signal (thick curve). The estimated power level of WN appears to fit well with the real signal (left part of the AV curve). However, long-term errors modeled by the two GM processes match the signal AV sequence only approximately (right part of the 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 induces very long convergence time. Second, the

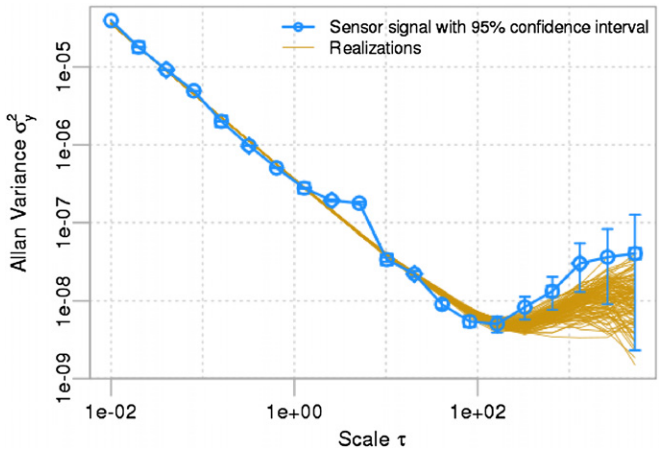

Figure 7. Results of the estimation of two GM processes and a WN process applied to the *IMAR-FSAS Y*-axis accelerometer error signal. The figure shows the AV of 100 realizations issued from $\hat{\theta}$ (thin curves) and the AV of the sensor signal (thick curve).

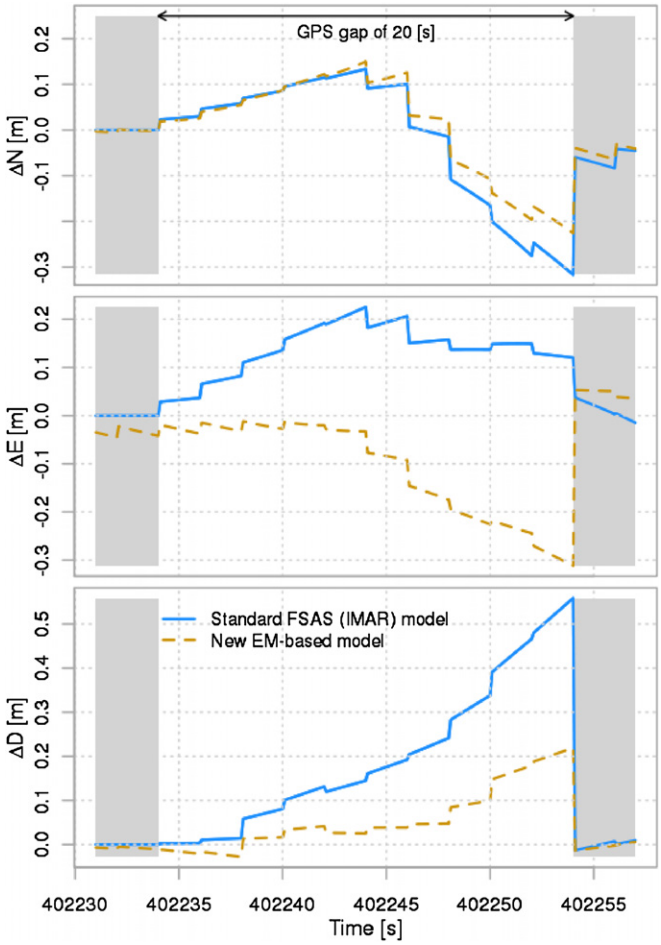

Figure 8. Position errors along the North (first panel), East (second panel) and Down (third panel) components occurring during the 20 s long GPS outage when using the traditional error model (full curves) and the new model (dotted curves) in the EKF.

accumulation of numerical imprecision in many iterations 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.

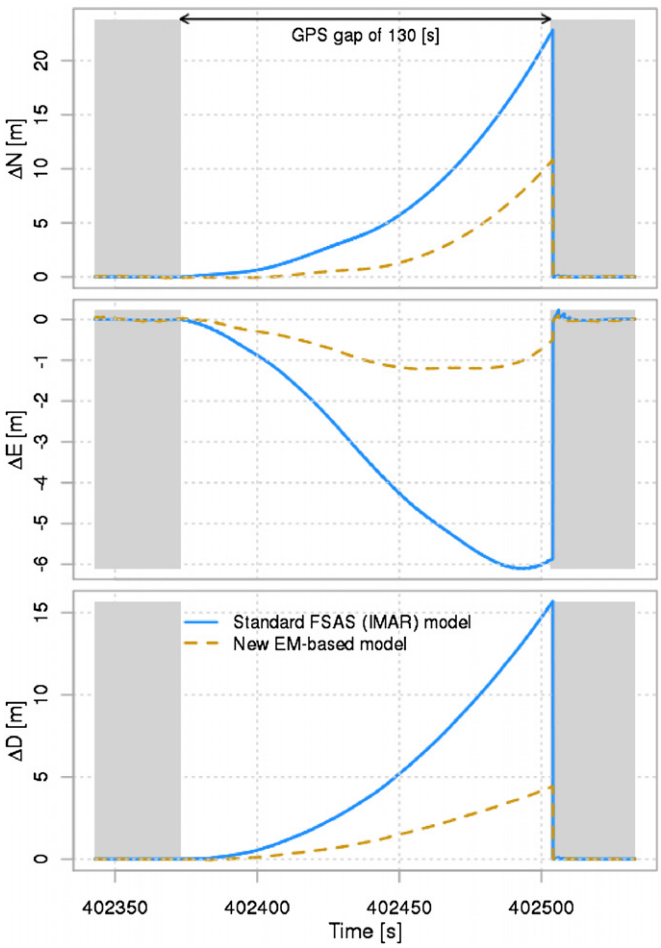

Figure 9. Position errors along the North (first panel), East (second panel) and Down (third panel) components occurring during the 130 s long GPS outage when using the traditional error model (full curves) and the new model (dotted curves) in the EKF.

In the following, 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 and 10 Hz, respectively. The carrier-phase GPS observation was double-differenced in post-processing to yield high-precision (cm-level) GPS positioning. This has 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 durations, at times when 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 us 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 20 s long outage has been introduced in a time period during which the car was turning a roundabout. Figure 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 130 s) and affected a period in which the car was moving on a straight road, and its acceleration varied. As shown in figure 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 to 10 m along the North component, from −6 to −1.2 m along the East component and from 16 to 4.2 m along the vertical component.

6. Conclusions and perspectives

The potential of using both an unconstrained and a 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 on the cases where the 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 maximum if the initial values are 'far' from the 'true' parameter values. Therefore, using diffusion priors or search for 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 the 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, 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 last simulated example and the real application, there are at the moment no other estimation methods that can successfully determine very complex error structures.

Appendix

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

$$\begin{aligned}
\mathbf{x}_t^{t-1} &= \boldsymbol{\Phi}\mathbf{x}_{t-1}^{t-1} + \mathbf{u}_t \\
\mathbf{P}_t^{t-1} &= \boldsymbol{\Phi}\mathbf{P}_{t-1}^{t-1}\boldsymbol{\Phi}^T + \mathbf{Q} \\
\mathbf{K}_t &= \mathbf{P}_t^{t-1}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_t^{t-1}\mathbf{H}^T + \mathbf{R}\right)^{-1} \\
\mathbf{x}_t^t &= \mathbf{x}_t^{t-1} + \mathbf{K}_t\left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^{t-1}\right) \\
\mathbf{P}_t^t &= \mathbf{P}_t^{t-1} - \mathbf{K}_t\mathbf{H}\mathbf{P}_t^{t-1},
\end{aligned}$$

where the initial conditions are given by $\mathbf{x}_1^1 = \boldsymbol{\mu}$ and $\mathbf{P}_1^1 = \boldsymbol{\Sigma}$. The backward recursion can be calculated for $t = N, \ldots, 1$:

$$\begin{aligned}
\mathbf{J}_{t-1} &= \mathbf{P}_{t-1}^{t-1}\boldsymbol{\Phi}^T\left(\mathbf{P}_t^{t-1}\right)^{-1} \\
\mathbf{x}_{t-1}^N &= \mathbf{x}_{t-1}^{t-1} + \mathbf{J}_{t-1}\left(\mathbf{x}_t^N - \boldsymbol{\Phi}\mathbf{x}_{t-1}^{t-1}\right) \\
\mathbf{P}_{t-1}^N &= \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-1}\left(\mathbf{P}_t^N - \mathbf{P}_t^{t-1}\right)\mathbf{J}_{t-1}^T.
\end{aligned}$$

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

$$\mathbf{P}_{t-1,t-2} = \mathbf{P}_{t-1}^{t-1}\mathbf{J}_{t-2}^T + \mathbf{J}_{t-1}\left(\mathbf{P}_{t,t-1}^N - \boldsymbol{\Phi}\mathbf{P}_{t-1}^{t-1}\right)\mathbf{J}_{t-2}^T,$$

where

$$\mathbf{P}_{N,N-1}^N = \left(\mathbf{I} - \mathbf{K}_N\mathbf{H}\right)\boldsymbol{\Phi}\mathbf{P}_{N-1}^{N-1}.$$

References

[1] Titterton D and Weston J 1997 *Strapdown Inertial Navigation Technology* (London: Peter Peregrinus)

[2] Fong W, Ong S and Nee A 2008 Methods for in-field user calibration of an inertial measurement unit without external equipment *Meas. Sci. Technol.* **19** 085202

[3] Gelb A 1974 *Applied Optimal Estimation* (Cambridge, MA: MIT Press)

[4] Grewal M and Andrews A 2001 *Kalman Filtering: Theory and Practice Using MATLAB* 2nd edn (New York: Wiley Interscience)

[5] Allan D 1966 Statistics of atomic frequency standards *Proc. IEEE* **54** 221–30

[6] Xing Z and Gebre-Egziabher D 2008 Modeling and bounding low cost inertial sensor errors *Position, Location and Navigation Symp., 2008 IEEE/ION* pp 1122–32

[7] Greenhall C 1998 Spectral ambiguity of Allan variance *IEEE Trans. Instrum. Meas.* **47** 623–7

[8] Waegli A and Skaloud J 2007 Assessment of GPS/MEMS IMU integration performance in ski racing *Proc. ENC-GNSS 2007 (TimeNav'07) (Geneva, Switzerland)*

[9] Shumway R and Stoffer D 2000 *Time Series Analysis and Its Applications* (Berlin: Springer)

[10] Gupta N and Mehra R 1974 Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations *IEEE Trans. Autom. Control* **19** 774–83

[11] Shumway R and Stoffer D 1982 An approach to time series smoothing and forecasting using the EM algorithm *J. Time Ser. Anal.* **3** 253–64

[12] Dempster A, Laird N and Rubing D 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc.* B **39** 1–38