Fault Detection by Desynchronized Kalman Filtering, Introduction to Robust Estimation

Olivier Bilenne
Multitel A.S.B.L.
Parc Initialis - Avenue N. Copernic, 1
B 7000 Mons
Belgium
bilenne@multitel.be

Abstract – *This paper deals with the state estimation of dynamic systems. A recursive linear MMSE estimator is presented as an alternative to Kalman filtering . This estimator has the ability to cope with asynchronous measurements, and to process the data by sets of undefined sizes. It is particularly suitable for fault detection, because the decisions can be based on more data. This is an open door to robust estimation. A mixed estimator robust to various failure scenarios is then derived by using the Bayesian approach. This mixed estimator is originally thought for applications requiring high integrity estimations. It is next tested on a rail navigation problem.*

Keywords: Dynamic state estimation, fault detection, robust estimation, sensor data fusion, biased sensor, Kalman filter, Bayesian approach, asynchronous observations, information fusion.

1 Introduction

The object of this paper is the problem of robust state estimation of dynamic systems. Linear estimators are optimum when dealing with Gaussian variables. Unfortunately, real sensors are often biased and may provide erroneous measurements (*outliers*), causing important biases in the linear estimation. As a consequence, the integrity of the estimation may drop considerably. This problem could have dramatic consequences in applications where integrity is of first importance. Unlike linear estimators, robust estimators present a reduced sensitivity to erroneous measurements, and tend to be more powerful in practical applications. The work of reference on robust estimation is owed to P.J. Huber [1]. Huber assumes that the actual measurement noise distribution lies in a certain neighborhood of the assumed noise distribution. A necessary condition for an estimator to be robust is that it guarantees good performance all over this neighborhood. Huber's approach of the problem of seeking robust estimators relies on the *minimax* strategy. The selected estimator is the one with the best performance under the least favorable element of the neighborhood. The performance criterion is the minimum asymptotic covariance of the estimation error, when the amount of measurements grows to infinity. In dynamic estimation however, the asymptotic performance criterion makes little sense because the variable to be estimated is randomly changing over time. Moreover, an accuracy criterion based on the expected variance of the estimation error is highly insufficient in the context of applications demanding very high levels of confidence on the state estimates, such as safe navigation systems. In this case, it is preferable to focus on on the *integrity* of the estimates rather than on their accuracy. The robust estimation literature is quite vast, though few papers are really integrity-directed. This paper is mainly based on the work of P.B. Ober, who proposed a complete solution to the static robust estimation problem for navigation systems [2]. Ober's approach consists roughly in using fault detection techniques to isolate the faulty sensors, providing a nonlinear estimation from the unbiased measurements only and attaching to the state estimate a safe confidence interval which certifies a very high level of integrity. The aim of this paper is to show how those static robust estimation techniques can be adapted to the dynamic problem. We use, as a basic tool, a recursive linear estimator, close to the Kalman filter, which has the characteristic of being able to deal recursively and asynchronously with sets of measurements collected at various times and in various numbers. This recursive estimator can not only handle asynchronous observations but also presents an increased flexibility in fault detection because the fault detection algorithm can be fed with an undefined number of measurements of various natures and provide decisions of increased reliability. We will see that the robustness of the linear algorithm can be improved by using the Bayesian approach, which leads to a nonlinear recursive estimation algorithm. In Sec. 4, the algorithms are tested on a rail navigation problem.

2 Recursive linear estimation of systems by means of discrete-time observations

We consider continuous-time systems of known linear dynamics. The problem consists of estimating, in real time, the state of the system from discrete-time observations. We make no assumption on the observation rhythm and allow the observations to be collected asynchronously.

We will make a distinction between the *estimation times*

$$t_1, t_2, \ldots, t_k, \ldots \ , \tag{1}$$

which delimit the *steps* of the estimation algorithm, and the *observation times*. We call x_k the n-dimentional state of the system at estimation time t_k. Step k ($k = 1, 2, \ldots$) is the part the algorithm taking place in time interval $]t_{k-1}; t_k]$.

Step k leads to the estimation of x_k from all the measurements prior to t_k. During step k, a set of n_k observations

$$y_1^k, y_2^k, \dots, y_{n_k}^k \tag{2}$$

are collected at *observation times*

$$o_1^k, o_2^k, \dots, o_{n_k}^k \ , \tag{3}$$

included in $]t_{k-1}; t_k]$. We call x_k^i the state of the system at observation time o_i^k.

We suppose the dynamics of the system is linear and can be written under the following discrete-time state equations : for all $k \geq 1$,

$$\left| \begin{array}{l} x_1^k = F_1^k x_{k-1} + G_1^k u_1^k + \omega_1^k \ , \\ x_i^k = F_i^k x_{i-1}^k + G_i^k u_i^k + \omega_i^k \ , \quad 1 < i \leq n_k \ , \\ x_k = F_{n_k+1}^k x_{n_k}^k + G_{n_k+1}^k u_{n_k+1}^k + \omega_{n_k+1}^k \ , \\ y_i^k = H_i^k x_i^k + J_i^k u_i^k + \upsilon_i^k \ , \ i = 1, \dots, n_k \ , \end{array} \right. \tag{4}$$

where F_i^k is the full-rank *state transition matrix*, H_i^k the *observation matrix*, u_i^k a known constant control signal, random vector ω_i^k is the *plant* or *process noise* of covariance matrix Ω_i^k, and random vector υ_i^k is the *observation* or *measurement noise* of covariance matrix Υ_i^k. Moreover, we suppose the process and measurement noise to be "memory-less", i.e.

$$E\{\omega_i^k {\omega_j^l}^T\} = 0 \quad \text{if } (k,i) \neq (l,j) \ , \tag{5}$$

$$E\{\upsilon_i^k {\upsilon_j^l}^T\} = 0 \quad \text{if } (k,i) \neq (l,j) \ . \tag{6}$$

We suppose all those random variables to be Gaussian, or that they can be conservatively substituted by Gaussian variables by the technique of *Overbounding*[1]. [2] [3]

In this paper, observations y_i^k of states x_k^i are seen as *shifted observations* of x_k. All the observations y_i^k ($i = 1, \dots, n_k$) collected at step k are gathered in a composite vector z_k, which is considered as the observation vector of the system at estimation time t_k.

By ignoring the observation times, and solving the propagation model (4) for variables x_k, we obtain the state equations for the estimation times :

$$\begin{array}{l} x_k = F_k x_{k-1} + G_k u_k + W_k \omega_k \\ z_k = H_k x_k - \Theta_k \omega_k + \tilde{J}_k u_k + \upsilon_k \ . \end{array} \tag{7}$$

Definitions of matrices F_k, H_k, G_k, $\tilde{J}_k$, W_k, Θ_k can be found in Appendix (A.1). Matrix H_k propagates x_k from estimation time t_k back to the observation times of step k. Matrix Θ_k can be seen as a *correction matrix* which compensates for the extra process noise contained in x_k and posterior to the observations. One characteristic of those equations is that the observations z_k and the process noise ω_k are conditionally dependent given state x_k.

[1] Overbounding consists of using Gaussian distributions that garantee the conservation of the performance assessment. This technique is beyond the scope of this paper.

2.1 State estimation

Our goal now is to deduce from Eqs. (7) an estimator of state x_k given all the observations $z_1, z_2, \dots, z_k$ prior to t_k. We call z_1^k the vector resulting from the concatenation of those observations[2] :

$$z_1^k = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{pmatrix} = \textit{vect}\{z_1, z_2, \dots, z_k\} \ . \tag{8}$$

We have opted for the commonly-used *minimum least-square error criterion* (MMSE), which prevails under the hypothesis of Gaussian variables. The *minimum-covariance estimator* $\hat{x}_k$ minimizes the mean square estimation error

$$E\{(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T | z_1^k\} \ . \tag{9}$$

The MMSE criterion has the property that estimator $\hat{x}_k$ coincides with the expected value of x_k conditioned by the past observations :

$$\hat{x}_k = E\{x_k | z_1^k\} \ . \tag{10}$$

The solution to this estimation problem verifies the Wiener-Hopf equation ([4]) and is given by

$$\hat{x}_{k|k} = E\{x_k {z_1^k}^T\} E\{z_1^k {z_1^k}^T\}^{-1} z_1^k \ , \tag{11}$$

provided that covariance matrix

$$R_k = E\{z_1^k {z_1^k}^T\} \tag{12}$$

is strictly positive definite, which is generally the case in well-posed problems. When the algorithm is applied to time signals, the ever-growing dimension of the covariance matrix R_k makes the online application of the Wiener filter impossible. We then resort to the *Levinson Recursion*, which turns the Wiener filter into a time-recursive linear estimator.

Levinson recursion. The Levinson recursion consists in estimating $\hat{x}_k$ by updating the previous estimation $\hat{x}_{k-1}$ according to the new observation collected at step k. In particular, this recursion scheme allows to derivate the Kalman filter algorithm from the Wiener-Hopf equation and a dynamic model of the system ([4],[5]). A similar development applied to estimator (11) together with model (7) leads to a time-recursive version of the linear estimator. The Levinson recursion is shortly explained in Appendix (A.2). One must however keep in mind that the observations are analyzed by sets of several measurements which possibly happened at different times but are all taken as *shifted observations* of the state at time t_k.

Recursive estimation is based on the concepts of *prediction* and *innovation*. The procedure followed at step k of the recursion can be divided into three phases.

[2] The z_1^k notation must not be confused with entities such as x_1^k, y_1^k, or H_1^k, which are variables referring to observation time o_1^k.

Firstly, the algorithm predicts an *a priori* estimate $\hat{x}_{k|k-1}$ of x_k by propagating through the dynamic model the last estimation computed at step $k-1$:

$$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + G_k u_k \ . \tag{13}$$

We call $e_{k|k-1}$ the estimation error on the predicted estimation at step k, and $P_{k|k-1}$ the covariance matrix of this error :

$$e_{k|k-1} = x_k - \hat{x}_{k|k-1} \ , \tag{14}$$
$$P_{k|k-1} = E\{e_{k|k-1} e_{k|k-1}^T\} \ . \tag{15}$$

Next, the prediction is confronted with the new observation z_k. The resulting information is usually called *innovation*. The innovation γ_k at step k is defined as the difference between the observation z_k collected at step k and the predicted observation $\hat{z}_{k|k-1}$. We call Γ_k the covariance matrix of innovation γ_k at step k :

$$\hat{z}_{k|k-1} = H_k \hat{x}_{k|k-1} + J_k u_k \ , \tag{16}$$
$$\gamma_k = z_k - \hat{z}_{k|k-1} \ , \tag{17}$$
$$\Gamma_k = E\{\gamma_k \gamma_k^T\} \ . \tag{18}$$

Finally, the *a posteriori* estimate $\hat{x}_{k|k}$ is obtained by upgrading the prediction $\hat{x}_{k|k-1}$ according to equation

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + E\{x_k \gamma_k^T\} E\{\gamma_k \gamma_k^T\}^{-1} \gamma_k \ , \tag{19}$$

which can be derived easily by introducing the *Inverse matrix partitioning lemma* presented in Appendix (A.2) into the Wiener filter equation (11). Similarly, we call $e_{k|k}$ the estimation error at step k and $P_{k|k}$ its covariance matrix :

$$e_{k|k} = x_k - \hat{x}_{k|k} \ , \tag{20}$$
$$P_{k|k} = E\{e_{k|k} e_{k|k}^T\} \ . \tag{21}$$

Straightforward computations lead to the recursive algorithm summarized in Table 1. One can see that our algorithm is very close to the basic equations of the Kalman filter, except for a few additional terms issued because of the dependency between the observations and the process noise.

Remark. The originality here is that the recursive version of the algorithm divides the estimation problem into a sequel of smaller Wiener-Hopf estimation algorithms. The difference from classical Kalman filtering lies in the fact that it processes the data by sets of several consecutive observations. This algorithm presents a degree of freedom in the choice of the estimation times although those estimation times have no influence on the estimates themselves. In a sense, it can be seen as a generalization of the Kalman filter, which analyzes the data synchronously, the estimation times being imposed by the observations.

Robust estimation. This linear estimator gives satisfying results when the sensors are perfectly unbiased. Unfortunately, when a failure occurs, the linear estimator stops working properly, and a nonlinear estimator is needed. The class of nonlinear estimators called *robust estimators* have the property to be less sensitive to the measurements containing abnormally large errors. This goal can be achieved for example by dropping the measurements that are likely to be erroneous and performing the state estimation from the other valid measurements (we will come back to that strategy in Sec. 3). However, the corrupted measurements still have to be uncovered. The decision is generally based on *a priori* information on the nature of the sensors and their likelihood to be faulty, and on the collected measurements themselves. One way of checking for faults consists of estimating the measurement noise υ_k of the sensors. This is the object of the next section.

Table 1: Recursive linear estimator.

Initialisation :
$k \leftarrow 0$
$\hat{x}_{0|0} = \hat{x}_0$
$\hat{P}_{0|0} = \hat{P}_0$
Repeat :
$k \leftarrow k+1$
Prediction :
$\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + G_k u_k$
$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + W_k \Omega_k W_k^T$
$\hat{z}_{k|k-1} = H_k \hat{x}_{k|k-1} + J_k u_k$
Innovation :
$\gamma_k = z_k - \hat{z}_{k|k-1}$
$\Gamma_k = H_k F_k P_{k-1|k-1} F_k^T H_k^T + \Upsilon_k$
$+(H_k W_k - \Theta_k)\Omega_k (H_k W_k - \Theta_k)^T$
Update :
$K_k = (P_{k|k-1} H_k^T - W_k \Omega_k \Theta_k^T)\Gamma_k^{-1}$
$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \gamma_k$
$P_{k|k} = (I - K_k H_k) P_{k|k-1} - K_k H_k \Omega_k W_k^T$
End of loop.

2.2 Measurement noise estimation

The purpose of this section is to show that fault detection can be coupled with the execution of the recursive estimation algorithm of Table 1, since a reliable fault detection algorithm would make robust estimation possible.

The first idea consists in inspecting the size of innovation γ_k at step k. The bigger the innovation, the more chances the measurements have to be erroneous. Unfortunately, this method is dangerous because the innovation γ_k is corrupted by the process noise ω_k. The same conclusion can be made about the use, as a fault detector, of the *residual* of the recursive estimation algorithm :

$$r_k = z_k - (H_k \hat{x}_{k|k-1} + \tilde{J}_k u_k) = (I - K_k H_k)\gamma_k \ . \tag{22}$$

The residual depends on the process noise as well. A better solution consists of estimating straightforwardly the measurement noise υ_k according to some optimization criterion such as the *minimum-covariance criterion*. The optimum estimation of the measurement noise then verifies the Wiener-Hopf equation :

$$\hat{\upsilon}_k = E\{\upsilon_k {z_1^k}^T\} E\{z_1^k {z_1^k}^T\}^{-1} z_1^k \ . \tag{23}$$

Once again, the introduction of the *Levinson recursion* scheme simplifies the estimation procedure :

$$\hat{\upsilon}_{k|k} = \hat{\upsilon}_{k|k-1} + E\{\upsilon_k \gamma_k^T\} E\{\gamma_k \gamma_k^T\}^{-1} \gamma_k \ . \tag{24}$$

Moreover, the last developments of Appendix (A.2) show that the measurement noise υ_k can be estimated straight from the innovations γ_k of the recursive linear algorithm and Eq. (24) reduces to :

$$\hat{\upsilon}_{k|k} = \Upsilon_k \Gamma_k^{-1} \gamma_k \ . \tag{25}$$

This is a direct consequence of hypothesis (6). Like the estimator of Sec. 2.1, this estimator of the measurement noise has the characteristic of being biased when one of the sensors is faulty. This property is the basis of our fault detection strategy. On the other hand, Eq. (25) implies that, when no failure occurs and under the Gaussian hypothesis, the actual measurement noise follows a centered normal density function of covariance matrix

$$\hat{\Upsilon}_{k|k} = \Upsilon_k \Gamma_k^{-1} \Upsilon_k^T \ . \tag{26}$$

We introduce a squared norm of the estimation of the measurement errors called the Sum of the Squared (estimated) Errors :

$$SSE_k = \hat{\upsilon}_{k|k}^T \hat{\Upsilon}_{k|k}^{-1} \hat{\upsilon}_{k|k} \ , \tag{27}$$

This norm, which was used in [2] also, intervenes directly in the estimation of the probability of realization of $\hat{\Upsilon}_{k|k}$ as measurement noise under the no-failure hypothesis. The posterior likelihood of the no-failure hypothesis can next be deduced from this probability by applying the *Bayes Rule* and compared with the posterior probabilities of various failure scenarios. This constitutes the final step of the fault detection algorithm.

3 Robust estimation

3.1 The Bayesian approach

In this section we present a nonlinear state estimator which proves to be more robust than the linear estimator of Table 1. This nonlinear estimation algorithm relies on the *Bayesian approach*. We first begin by making a series of hypotheses h_i on the validity of the measurements and indexing all the possible sets of correct measurements. The mean and variance of the conditional distribution of the actual state x_k under each hypothesis h_i are given by the recursive linear estimation algorithm of Table 1 fed with the set of measurements corresponding to h_i only :

$$\begin{aligned} \hat{x}_{k|k,i} &= E\{x_k | h_i, z_1^k\} \ , & (28)\\ P_{k|k,i} &= E\{(x_k - \hat{x}_{k|k,i})^2 | h_i, z_1^k\} \ . & (29) \end{aligned}$$

At each step k, the recursive linear estimator splits into several parallel estimators starting from the same initial estimate $\hat{x}_{k-1|k-1}$ but operating under different hypotheses h_i. The global posterior expectation $\hat{x}_{k|k}$ of state x_k can then be derived by marginalizing over all the failure hypotheses :

$$\hat{x}_{k|k} = \sum_i p(h_i | z_1^k) \ \hat{x}_{k|k,i} \ . \tag{30}$$

The posterior expectation is the estimate of x_k offering the minimum mean square error

$$P_{k|k} = \sum_i p(h_i | z_1^k) \ (P_{k|k,i} + \Delta_{k,i}) \ , \tag{31}$$

where the extra term

$$\Delta_{k,i} = (\hat{x}_{k|k,i} - \hat{x}_{k|k})(\hat{x}_{k|k,i} - \hat{x}_{k|k})^T \tag{32}$$

results from the divergence between the failure hypotheses. The posterior distribution of x_k is no longer Gaussian and not necessarily symmetric or unimodal if the measurements of the sensors do not match exactly. In those conditions, the posterior expectation is not an accurate estimator of the state x_k, despite its property of minimum mean square error. Other estimators such as the *maximum a posteriori* (MAP) estimator prove to perform better outside the context of symmetric and unimodal distributions. However, this discussion is slightly out of the scope of the applications requiring high levels of safety, for which one would like to provide confidence intervals rather than accurate state estimates. This issue is the object of Sec. 3.2.
The *mixture* distribution of $\hat{x}_{k|k}$ represents a problem for the fault detection algorithm of Sec. 2.2 which requires normally distributed variables only, or failing this, a full knowledge of the distribution of the estimate of the measurement noise. Unfortunately, the normal structure is lost in the recursion as soon as the posterior expectation of x_k, averaged over all hypotheses, is considered as the past estimate of step $k+1$. We may note already that this phenomenon will also tend to weaken the reliability of the integrity certification procedure based directly on the posterior distribution of the state and presented in Sec. 3.2. In the application of Sec. 4, we assimilate, at step $k+1$, the posterior distribution of x_k to a Gaussian distribution of same mean and variance. By doing so, we hope that this loss of precision will be compensated to some extent by the additional uncertainty brought by the $\Delta_{k,i}$ terms. Of course such a delicate operation can not be valid without taking further precautions so as to guarantee the conservativeness of the estimated posterior distributions and the relevance of the fault detection algorithm at future steps. This problem is currently under investigation.

3.2 Integrity certification

In the introduction, we have emphasized the importance of the integrity certification of the estimates. *Integrity* measures the probability of misleading information. *Misleading information* (MI) is said to occur when the state estimate lies outside its (multi-dimensional) *precision interval* and no major failure has been detected. It is a capital concept in the fields where estimation mistakes can not be tolerated, such as railway and air navigation. The integrity of the estimator is compromised whenever the system provides misleading information at a rate greater than the maximum tolerated MI rate.

The integrity of the state estimator can be estimated directly from the distribution of the state estimate, provided that this distribution is known, by integrating the density function inside bounds, that delimit the confidence interval. Those bounds are generally called *alert limits*, and can be determined from the precision and continuity requirements of the problem. They are safe if the system was correctly modeled and if conservative probabilities were attributed to the failure hypotheses. The role of the robust estimator, according to Ober's approach, is to place the estimate and its confidence interval in the state space in such a way that the actual state has the highest posterior probability to lie inside the confidence interval. As pointed in Sec. 3.1, the posterior distribution of the mixed estimator is non-normal and may be difficult to estimate. In the following application, the assimilation at step k of the posterior estimation of x_{k-1} to a normally distributed variable leads to a Gaussian mixture distribution for estimate $\hat{x}_{k|k}$ on which alert bounds are easy to fix.

4 Fault detection and robust estimation : an application

In this section, we consider the problem of real-time estimation of the speed of a train from the measurements of various speed sensors. Our tests were realized on a database recorded in Italy during an Arezzo-Pontassieve journey. The train speed estimation was based on the data collected by three devices : a wheel sensor, a Doppler radar, and the Doppler speed output of a GPS receiver. The measurements given by the sensors were the projections, in the direction of motion of the train, of the three-dimensional speed vectors, restricting the problem to a scalar unknown. In perfect conditions, the devices provided measurements normally distributed around the train speed with a standard deviation of $\sigma_{WS} = 1m/s$, $\sigma_{DR} = 2m/s$, and $\sigma_{GPS} = 0.5m/s$ respectively. Furthermore, we assigned to each sensor heuristic *a priori* probabilities of not being in their normal operating mode. The fault probability of each sensor was set to 0.1. We considered that those *a priori* probabilities were constant over time, and did not depend on past measurements or the other sensors diagnoses. In the failure case, we supposed that the measurements followed a uniform distribution in a large interval around the actual train speed.

Modeling of the system's dynamics. We arbitrarily chose to use a constant acceleration model to describe the dynamics of the train. The speed x_1 and acceleration x_2 of the train form the two-dimensional system state. The dynamics of the train follows the differential equations

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \,, \\ \dot{x}_2(t) &= \tau \,, \end{aligned} \qquad (33)$$

where τ is an additive white noise of constant standard deviation $\sigma_\tau = 0.08ms^{-4}$. Perturbation τ represents the jerk endured by the train. By integrating Eqs. (33) on the time intervals defined in Eqs. (1,3), we obtain discrete time state equations which fit Eqs. (7).

The choice of a constant acceleration model is quite popular in the literature and proves to be particularly suitable for this speed estimation problem. It was motivated by the physical limitations and security constraints of the rail transport mainly, but remains of course arbitrary.

Linear state estimation. Firstly, we have tested the recursive linear filter presented in Sec. 2.1. Since this filter can be viewed as an online version of the Wiener filter (11), the properties of which are well-known, we did not elaborate on the behaviour of the filter throughout the whole experiment. Rather, we focused on the property of invariance of the state estimation process towards the frequency of estimation. Table 2 gathers the first estimates of the speed (m/s) and the standard error on speed estimation (m/s) collected at various frequencies in geometric progression f_0, $\frac{1}{2}f_0$, $\frac{1}{4}f_0$ ($f_0 = 5Hz$). Indeed, we observe the perfect matching of the three estimators.

Table 2: Invariance of the estimator to estimation frequency.

T	T_0	$2T_0$	$4T_0$
f	f_0	$\frac{1}{2}f_0$	$\frac{1}{4}f_0$
T_0	$8.5211, 0.2862$	–	–
$2T_0$	$8.5482, 0.3295$	$8.5482, 0.3295$	–
$3T_0$	$8.6938, 0.3465$	–	–
$4T_0$	$8.7135, 0.3481$	$8.7135, 0.3481$	$8.7135, 0.3481$
$5T_0$	$8.9498, 0.3444$	–	–
$6T_0$	$8.9741, 0.2850$	$8.9741, 0.2850$	–
$7T_0$	$8.9562, 0.3090$	–	–
$8T_0$	$9.0269, 0.3254$	$9.0269, 0.3254$	$9.0269, 0.3254$

Robust state estimation. This section highlights the limitations of linear estimators in presence of failures. Linear estimation is optimal in the Gaussian case, but is quite sensitive to erroneous measurements and becomes largely biased and suboptimal when one or more sensors stop working properly. In our tests, the linear filter of Sec. 2.1 is compared to the bias-desensitized estimator presented in Sec. 3.

Again, our tests were based on the Arezzo-Pontassieve ride introduced above. The database did not provide the evolution over time of the actual speed of the train and prevented us to evaluate the performances of the estimation algorithms in a quantitative fashion. We will rather adopt a qualitative approach by analyzing how the linear and nonlinear estimators cope with classical ambiguous situations of the railway navigation such as isolated failures, permanently-biased sensors, slipping wheels at sudden acceleration or braking and on wet or icy rails, freezing of the GPS output inside tunnels, etc.

Fig. 1 shows the measurements collected by the three sensors during time interval $1590s - 1630s$. The thick continuous line represents the evolution of the state estimate computed by the recursive linear estimator on the top graph, and by the mixed estimator below. The crosses represent the measurements of the sensors that were taken as faulty by the most likely hypothesis, i.e. the hypothesis with the highest a posteriori probably. The thin dotted lines represent the alert limits corresponding to an integrity of $1 - 10^{-11}$. Fig. 1 highlights the problem of the unavailability of satellite signals, inside tunnels for example. The GPS signal loses track of the train speed and gives constant meaningless measurements for a while. Fig. 1 shows

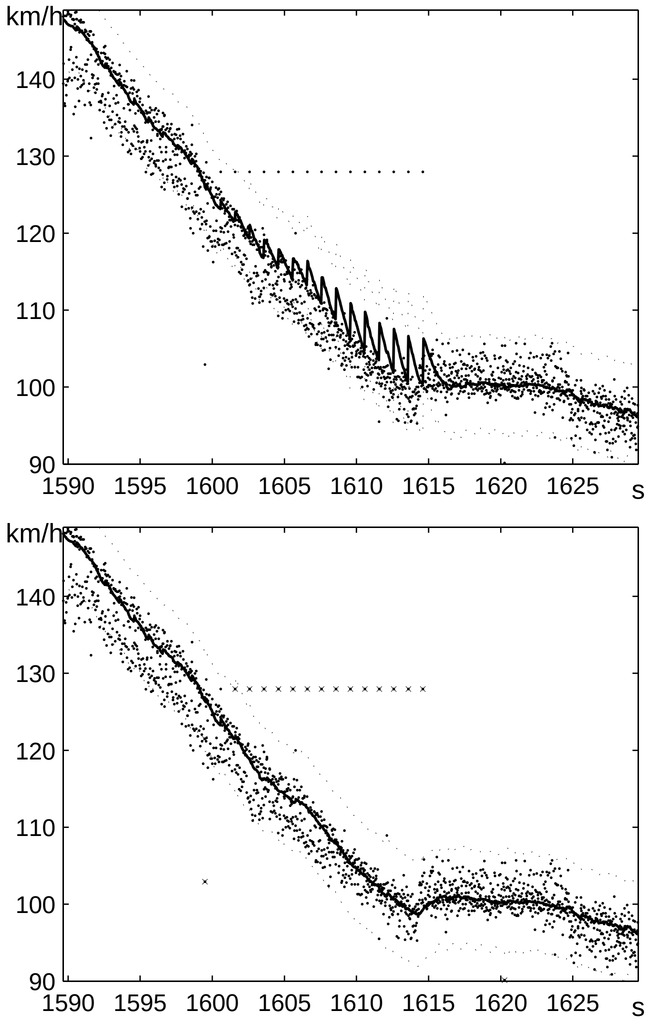

Fig. 1: Linear and $10Hz$ mixed estimators.

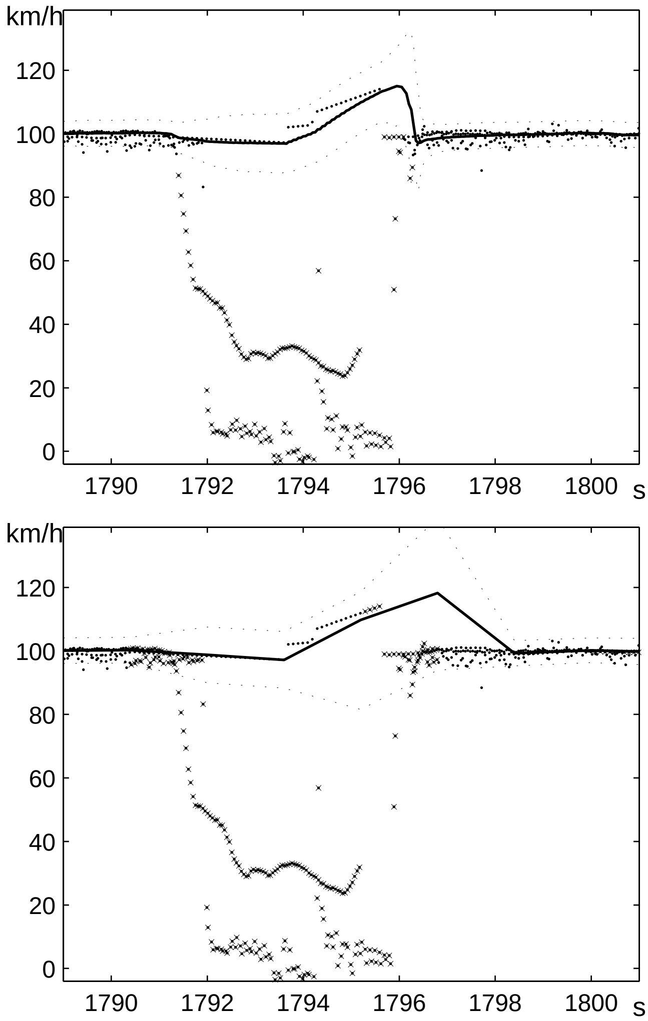

Fig. 2: $10Hz$ and $1.25Hz$ mixed estimators.

how the mixed estimator detects the abnormal behaviour of the GPS signal and ignores the wrong measurements until the GPS provides reliable measurements again. The erroneous radar measurement $(1600s, 103km/h)$ has also been rejected, which confirms that the mixed estimator detects the isolated failures as well.

The speed estimation depicted in Fig. 2 is based on the measurements of three different Doppler radars of respective standard deviations $\sigma_{DR1} = 2m/s$, $\sigma_{DR2} = 1m/s$, $\sigma_{DR3} = 0.5m/s$. All the sensors seem to lose track of the speed between times $1792s$ and $1796s$. The measurements of sensors 1 and 3 collapse abruptly while the speed of the train remains constant. On the other hand, the measurements provided by sensor 3 deviate only progressively from their last correct value. The mixed estimator was tested with two different frequencies of estimation. The $10Hz$ estimator detects the failure of sensors 1 and 3 and follows naively the measurements of sensor 2 until the true speed of the train crosses dangerously the alert bounds of the confidence interval. This confusion can be avoided by lowering the frequency of fault detection and estimation and basing the decisions on a greater amount of measurements. The $1.25Hz$ estimator detects the suspicious sudden convergence of the measurements of the three sensors at time $1796s$ and considers all the sensors as faulty on a larger time interval. Reducing the frequency of estimation increases the amount of data fed into the fault decision algorithm and may improve the reliability of the decisions. The drawback of a lowered frequency of estimation is that delaying the estimation time increases the complexity of the matrices and may lead to excessive computation times.

An unrealistic model. In this application, the train was modeled by a constant acceleration model, which is in fact a simplification of the real system's dynamics. As a consequence, the model slightly differs from the reality. A strictly rigorous approach would require the introduction in the model's matrices of an uncertainty component, say Δ, modeling the deviation from the real system. Such uncertainty variable may also be used in the case of nonlinear systems to model the inaccuracy resulting from the linearization of the dynamics. Moreover, we considered that each sensor could provide either unbiased or completely erroneous measurements containing no relevant information. This reduction to two distinct cases is often unsatisfying in current navigation problems. Real sensors are seldom perfectly calibrated, and often subject to perturbations which cause the measurements to be permanently biased to a certain extent. Though we hope those biases to be small enough so that they can be detected by robust estimation techniques and their impact on the estimate can be limited. A more realistic modeling of this speed estimation problem would consider the hypotheses of slightly biased sensors and complement the measurement noise distribution with a degree of uncertainty. Perturbations of the model's dynamics and the measurement noise will cause the state estimation to be permanently biased. The procedure of integrity certification of the estimated confidence interval must imperatively take this bias into account. However, when recursive estimation algorithms are used, evaluating the bias on the estimate is complicated by the fact that the bias presents a dynamic profile. For information, Mangoubi studied in [6] the stability and the performance of dynamic estimators with respect to uncertain perturbations of the model's dynamics and the measurement noise. Mangoubi derives the dynamic state estimators that minimize the impact of those perturbations on the estimate. These robust estimators guarantee a minimum bound on the resulting bias in the case of bounded-norm perturbations, or minimize the expected value of an arbitrary cost function penalizing the bias when the perturbations are supposed to follow a Gaussian distribution. Mangoubi's developments rest on minimax techniques in agreement with Huber's theory of robust statistics. It could be interesting to investigate up to what point those results apply to the present problem. Another solution proposed in [2] would adopt the following scheme. In a first step, the probability densities of those biases would be estimated from conservative a priori knowledge and the present and past measurements. The structure of the recursive Wiener estimator is such that the estimation at step $k+1$ is based on the previous estimate at step k. Consequently, the measurement biases at step k have the property to corrupt not only the current estimation $\hat{x}_{k|k}$ but the following estimations $\hat{x}_{k+1|k+1}, \hat{x}_{k+2|k+2}, \ldots$ as well. A complete description of the signature on the future estimates and innovations of the bias of a sensor can be found in [7]. Because we can estimate the density function of the noise vector $\upsilon_{k|k-1}$ with respect to a given bias scenario, the density function of the current noise vector can be obtained by marginalization with respect to the estimated bias vectors. However, this problem is out of the scope of this paper and research on that subject is under progress. For information, [2] studies the static case, while the dynamic case is tackled in [8].

The mixed estimator (30) results from the averaging of the distributions of conditional estimates weighted according to the posterior likelihood of various failure hypotheses. When evaluating the likelihood of each hypothesis, we supposed that the measurement noise of a faulty sensor is uniformly distributed in an large interval around the true value. Unfortunately, in real applications the *contaminating* distribution of outliers is generally unknown and this involves the necessity to reconsider the concept of robustness and derive estimators that can cope with various contaminated distributions. By instance in [9], distributions similar to Huber's *ϵ-contaminated neighborhood* are used :

$$\mathcal{P}_\epsilon(F_0) = \{F | F = (1-\epsilon)F_0 + \epsilon H, H \in \mathcal{S}\} \,, \qquad (34)$$

where F_0 is the no-failure distribution, ϵ is the *fraction of contamination*, and $\mathcal{S}$ the set of all possible contaminating distributions. Schick and Mitter present a dynamic state-estimator close to our posterior mean estimator (30) which is robust in the presence of rare and isolated outliers. To justify the robustness of their estimator they propose a minimax approach which ensures that the estimator performs well in the worst conditions. They consider the least favorable measurement noise distribution, i.e. the distribution H that maximizes a given cost function (the estimation error variance) and then propose to derive the dynamic estimator which minimizes that expected cost. A complete solution for finding the least favorable distribution in practice is unknown to date. The context of this paper is different since in safety applications the integriy of the state estimation prevails over its precision and the objective is no longer to minimize any cost function but to provide safe confidence intervals. The problem of certifying the integrity of confidence intervals in the least favourable noise conditions using a minimax approach remains open to further research.

5 Conclusion

In this paper we have developed a recursive MMSE linear estimator that has the property to be able to deal with data collected asynchronously. We have proven this estimator to be equivalent to a Kalman estimator which would update its estimate at the exact times of the measurements. In everyday applications, sensors sometimes provide erroneous measurements which induce a bias in the MMSE estimator. As a result, the precision and integrity of the state estimator are not guaranteed anymore. Therefore, a fault detection algorithm is necessary. We took a starting point from the methods of fault detection in static systems developed in [2] for state estimation applications requiring very high levels of integrity, and adapted to the dynamic problem an algorithm for detecting sensor failures by estimation of the measurement noise. The interest of our recursive linear estimator lies in the fact that it allows measurements to be analyzed by sets of undefined sizes, offering the fault detection algorithm the possibility to provide more reliable decisions based on large sets of measurements.

Finally, a recursive estimator robust to faulty sensors was deduced from the averaging of linear recursive estimators operating under various hypotheses. The likelihood of those hypotheses is estimated by the fault detection algorithm. The recursive linear estimator and the fault detection algorithm were tested on a rail navigation problem. The results obtained on a real database of measurements from various speed sensors show the recursive estimator to behave particularly well even in presence of multiple simultaneous failures. Though many questions remain open regarding the application of robust dynamic estimation techniques to problems demanding very high levels of safety, namely the integrity monitoring of the estimates in the presence of modeling errors or when the sensors are permanently biased, and the determination of the least favorable noise distribution that would ensure the conservativeness of the estimation process. We leave those issues to future work.

Acknowledgements

This work was supported by Multitel A.S.B.L.. Data are property of ALSTOM Transport Belgium. Acknowledgements to Andreea I. Carnu, Rocío Moraga, Marc Massar, François Meers, Stuart Murdoch and Silvio Ferreira.

A Appendix

A.1 Matrix definitions

Definition of composite vectors z_k, υ_k, ω_k, u_k and matrices Υ_k, Ω_k, F_k, G_k, W_k, H_k, Θ_k, and $\tilde{J}_k$ corresponding to model (4) and appearing in Eq. (7) :

$$z_k = vect\{y_1^k, y_2^k, \ldots, y_{n_k}^k\} , \quad (35)$$

$$\upsilon_k = vect\{\upsilon_1^k, \upsilon_2^k, \ldots, \upsilon_{n_k}^k\} , \quad (36)$$

$$\Upsilon_k = diag\{\Upsilon_1^k, \Upsilon_2^k, \ldots, \Upsilon_{n_k}^k\} , \quad (37)$$

$$\omega_k = vect\{\omega_1^k, \omega_2^k, \ldots, \omega_{n_k+1}^k\} , \quad (38)$$

$$\Omega_k = diag\{\Omega_1^k, \Omega_2^k, \ldots, \Omega_{n_k+1}^k\} , \quad (39)$$

$$u_k = vect\{u_1^k, u_2^k, \ldots, u_{n_k+1}^k\} , \quad (40)$$

$$F_k = F_{n_k+1}^k \ldots F_2^k F_1^k \overset{d}{=} \prod_{l=1}^{n_k+1} F_l^k , \quad (41)$$

$$\begin{array}{l} G_k = (\bar{G}_1^k \; \bar{G}_2^k \; \cdots \; \bar{G}_{n_k+1}^k) , \\ \bar{G}_j^k = \prod_{l=j+1}^{n_k+1} F_l^k G_j^k , \end{array} \quad (42)$$

$$\begin{array}{l} W_k = (\bar{W}_1^k \; \bar{W}_2^k \; \cdots \; \bar{W}_{n_k+1}^k) , \\ \bar{W}_j^k = \prod_{l=j+1}^{n_k+1} F_l^k , \end{array} \quad (43)$$

$$\begin{array}{l} H_k = (\bar{H}_1^k \; \bar{H}_2^k \; \cdots \; \bar{H}_{n_k}^k)^T , \\ \bar{H}_i^{k\,T} = H_i^k \prod_{l=n_k+1}^{i+1} F_l^{k\,-1} , \end{array} \quad (44)$$

$$\begin{array}{l} \Theta_k = \begin{pmatrix} \bar{\Theta}_{1,1}^k & \cdots & \bar{\Theta}_{1,n_k+1}^k \\ \vdots & \ddots & \vdots \\ \bar{\Theta}_{n_k,1}^k & \cdots & \bar{\Theta}_{n_k,n_k+1}^k \end{pmatrix} , \\ \bar{\Theta}_{i,j}^k = \left| \begin{array}{ll} H_i^k \prod_{l=j}^{i+1} F_l^{k\,-1} , & \text{if } j > i , \\ 0 , & \text{if } j \le i , \end{array} \right. \end{array} \quad (45)$$

$$\tilde{J}_k = J_k - \Xi_k \; , \; J_k = vect\{J_1^k, J_2^k, \ldots, J_{n_k}^k\} , \quad (46)$$

$$\begin{array}{l} \Xi_k = \begin{pmatrix} \bar{\Xi}_{1,1}^k & \cdots & \bar{\Xi}_{1,n_k+1}^k \\ \vdots & \ddots & \vdots \\ \bar{\Xi}_{n_k,1}^k & \cdots & \bar{\Xi}_{n_k,n_k+1}^k \end{pmatrix} , \\ \bar{\Xi}_{i,j}^k = \left| \begin{array}{ll} H_i^k \prod_{l=j}^{i+1} F_l^{k\,-1} G_j^k & \text{if } j > i , \\ 0 , & \text{if } j \le i . \end{array} \right. \end{array} \quad (47)$$

A.2 Levinson recursion

The Hermitian matrix R_k defined in Eq. (12) can be partitioned as follows :

$$R_k = \begin{pmatrix} E\{z_1^{k-1} z_1^{k-1\,T}\} & E\{z_1^{k-1} z_k{}^T\} \\ E\{z_k z_1^{k-1\,T}\} & E\{z_k z_k{}^T\} \end{pmatrix} \quad (48)$$

$$= \begin{pmatrix} R_{k-1} & r_k \\ r_k{}^T & \rho_k \end{pmatrix} . \quad (49)$$

The *matrix inversion by partitioning lemma* ([4],[5]) states that

$$R_k^{-1} = \begin{pmatrix} R_{k-1}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} R_{k-1}^{-1} r_k^T \\ -I \end{pmatrix} \tilde{\Gamma}_k^{-1} \begin{pmatrix} R_{k-1}^{-1} r_k^T \\ -I \end{pmatrix}^T , \quad (50)$$

where we have introduced expression $\tilde{\Gamma}_k$ which proves to be equal to innovation γ_k of the recursive linear estimation algorithm of Table 1 :

$$\tilde{\Gamma}_k = \rho - r_k^T R_{k-1}^{-1} r_k = \Gamma_k . \quad (51)$$

We will now compute two important results of Secs. 2.1 and 2.2. We first partition the following composite vectors :

$$z_1^k = \begin{pmatrix} z_1^{k-1} \\ z_k \end{pmatrix} , \quad (52)$$

$$E\{x_k z_1^{k\,T}\} = (\; E\{x_k z_1^{k-1\,T}\} \;\; E\{x_k z_k^T\} \;) , \quad (53)$$

$$E\{\upsilon_k z_1^{k\,T}\} = (\; 0 \;\; E\{\upsilon_k z_k^T\} \;) . \quad (54)$$

The introduction of Eqs. (50,53) into the Wiener estimator (11) leads to the recursive estimator (19). Similarly, Eq. (24) can be found by introducing Eq. (50,54) into estimator (23).

References

[1] P.J. Huber. *Robust Statistics*. John Wiley, New York, USA, 1981.

[2] P.B. Ober. *Integrity Prediction and Monitoring of Navigation Systems*. PhD thesis, Technische Universiteit Delft, The Netherlands, 2003.

[3] B. Decleene. Defining pseudorange integrity. In *Proc. ION-GPS*, Salt Lake, City, Utah, 2000.

[4] S.M. Kogon D.G. Manolakis, V.K. Ingle. *Statistical and Adaptive Signal Processing*. McGraw-Hill, Boston, MA., 2000.

[5] B. Noble and J.W. Daniel. *Applied Linear Algebra*. Prentice Hall, Englewood Cliffs, NJ., 3rd edition, 1988.

[6] R. Mangoubi. *Robust Estimation and Failure Detection, A Concise Treatment*. Springer-Verlag, Berlin, Germany, 1998.

[7] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Information and system science series. Prentice Hall, Englewood Cliffs, NJ., 1993.

[8] G. Brown and P.Y.C. Hwang. GPS failure detection by autonomous means within the cockpit. *Journal of Institute of Navigation*, 33(4):335–353, 1986.

[9] I.C. Schick and S.K. Mitter. Robust recursive estimation in the presence of heavy-tailed observation noise. *Annals of Statistics*, 22(1):1045–1080, 1994.