Moving-horizon Estimation With Guaranteed Robustness for Discrete-time Linear Systems and Measurements Subject to Outliers

A. Alessandri [a], M. Awawdeh [a]

[a] *Department of Mechanical Engineering – University of Genoa, P.le Kennedy Pad. D, 16129 Genova*

Abstract

An approach to state estimation for discrete-time linear time-invariant systems with measurements that may be affected by outliers is presented by using only a batch of most recent inputs and outputs according to a moving-horizon strategy. The approach consists in minimizing a set of least-squares cost functions in which each measure possibly contaminated by outlier is left out in turn. The estimate that corresponds to the lowest cost is retained and propagated to the next time instant, where the procedure is repeated with the new information batch. The stability of the estimation error for the proposed moving-horizon estimator is proved under mild conditions concerning the observability of the free-noise state equation and the selection of a tuning parameter in the cost function. Robustness is guaranteed with sufficiently large outliers. The effectiveness of the proposed method as compared with the Kalman filter is shown by means of a numerical example.

Key words: Moving horizon, State estimation, Outlier

1 Introduction

In numerous applications there exists the problem of dealing with large deviations in the measurements because of sensor malfunctions, wrong replacement of measures, or large non-Gaussian noises. These abnormal signals are usually called outliers in many different fields such as process control [22], heart surgery [21], intrusion detection [28], environmental monitoring [14], positioning [9], cloud management [19], and fault detection [11]. Various filtering methods have been proposed to attenuate or detect outliers (see, e.g., [13] and the references therein). In this paper, a more general problem is addressed that consists in estimating the state variables of a linear system by means of measures possibly corrupted by outliers. The estimation is performed by using a moving-horizon approach, which will be set in such a way to make it robust to outliers.

⋆ The material in this paper was partially presented at the 53rd IEEE Conference on Decision and Control, December 1517, 2014, Los Angeles, CA, USA. This paper was recommended for publication in revised form by Associate Editor Giancarlo Ferrari-Trecate under the direction of Editor Ian R. Petersen.

Email addresses: `alessandri@dime.unige.it` (A. Alessandri), `awawdeh@dime.unige.it` (M. Awawdeh).

The problem of estimating the state variable of a linear system with output contaminated by outliers can be treated by using the Kalman filter with some convenient adjustment. As is well-known, under the assumption that initial state and disturbances are white Gaussian stochastic processes, the best estimator in the sense of the minimization of the expected quadratic estimation error is the Kalman filter. Such an estimator is recursive in that the new output is processed by iterating the estimate update based on the current residual, i.e., the output error given by the difference between the measure and its prediction obtained from the last state estimate. Thus, one may check abnormal residuals via a threshold test to skip the Kalman estimate update with such residuals. This procedure can be motivated from a theoretical point of view by using the maximum likelihood criterion [2].

The first ideas about what is currently denoted as moving-horizon estimation (MHE) are presented in [16]. MHE consists in performing state estimation by using a limited amount of most recent information. The state estimates are obtained by minimizing a least-squares cost function with a batch of the inputs and outputs according to a sliding-horizon strategy. Constraints on the state variables may be easily taken into account since the optimization is carried on line. The first results on MHE for linear systems [3,24] have been extended to nonlinear (see, e.g., [5,7,8,15,25]) and large-scale systems [10].

Outliers are particular type of uncertainty that prevent an estimator from ensuring guaranteed performances [18]. Robustness is thus a fundamental requirement in the design of filters for uncertain systems. A method to enhance the robustness of the Kalman filter in the presence of outliers is presented in [13]. In [1,26] statistical tests are proposed that are less sensitive to abnormal noises. For the same reasons, an l_1 loss function is more suitable for the the purpose of identification with measures affected by outliers [17,27]. The reader is referred to [23] for a complete review of the most important methods of regression that account for robustness to outliers.

Based on the preliminary results of [2], here we focus on MHE for linear discrete-time systems with measurements contaminated by outliers. Toward this end, first we will prove the stability of the estimation error and, second, the robustness to outliers. Conditions for the stability of moving-horizon estimators for uncertain linear systems are reported in [6], where explicit bounding sequences are provided thanks to the adoption of worst-case cost functions. Unfortunately, such cost functions are not helpful in case of measurements affected by outliers, thus a different criterion is proposed here. More specifically, at each time instant we separately minimize a set of least-squares cost functions, where the measurements that can be affected by outliers are left out in turn. Then, we choose the minimizer associated with the lowest cost, and this estimate is propagated ahead to the next time instant according to the usual moving-horizon strategy. Such an estimation criterion ensures robustness to outliers of sufficiently large amplitude.

The paper is organized as follows. In Section 2, the proposed MHE method is described. Stability and robustness properties are illustrated in Sections 3 and 4, respectively. In Section 5, simulation results are presented and discussed. Finally, the conclusions are drawn in Section 6.

Let $\mathbb{N} := \{0,1,2,\ldots\}$. The minimum and maximum eigenvalues of a real, symmetric matrix P are denoted by $\lambda_{\min}(P)$ and $\lambda_{\max}(P)$, respectively; in addition, $P > 0$ means that it is positive definite. Given a generic matrix M, $\|M\| := \left(\lambda_{\max}(M^\top M)\right)^{1/2} = \left(\lambda_{\max}(M M^\top)\right)^{1/2}$. For a vector v, $\|v\| := (v^\top v)^{1/2}$ denotes its Euclidean norm, and $B(r) := \{v \in \mathbb{R}^n : \|v\| \leq r\}$ for $r > 0$. Given a sequence of vectors $v_i, v_{i+1}, \ldots, v_j$ for $i < j$, let us define $v_i^j := \mathrm{col}\,(v_i, v_{i+1}, \ldots, v_j)$. Moreover, $v_i^j\Big|_k$ denotes v_i^j without the k-th element, with $k = 1, 2, \ldots, j-i+1$. In other words, $v_i^j\Big|_1 := v_{i+1}^j$, $v_i^j\Big|_k := \mathrm{col}\,(v_i, v_{i+1}, \ldots, v_{i+k-2}, v_{i+k}, \ldots, v_j)$ for $k = 2, 3, \ldots, j-i$, and $v_i^j\Big|_{j-i+1} := v_i^{j-1}$. For the sake of simplicity, let $v_i^j\Big|_0 := v_i^j$. Finally, recall the square sum bound, namely, given m scalars $s_1, s_2, \ldots, s_m \in \mathbb{R}$, we have

$$\left(\sum_{i=1}^m s_i\right)^2 \leq m \sum_{i=1}^m s_i^2 \,.$$

2 MHE With Measures Corrupted by Outliers

Let us consider the discrete-time linear system

$$x_{t+1} = A\,x_t + B\,u_t + w_t \tag{1a}$$
$$y_t = C\,x_t + v_t \tag{1b}$$

where $t = 0, 1, \ldots$ is the time instant, $x_t \in \mathbb{R}^n$ is the state vector, $u_t \in \mathbb{R}^m$ is the control vector, $w_t \in \mathbb{R}^n$ is the system noise vector, $y_t \in \mathbb{R}$ is the measure, and $v_t \in \mathbb{R}$ is the measurement noise.

As to the system disturbance, w_t is supposed to be "small" as compared with the dynamics (i.e., bounded and taking zero or around zero values). In other words, we assume the following.

Assumption 1 *There exists $r_w \in (0, \infty)$ such that, for all $t = 0, 1, \ldots$, $\|w_t\| \leq r_w$.*

The measurement noise, instead, is "small" except on rare occurrences. More specifically, we assume the following.

Assumption 2 *There exist $r_v \in (0, \infty)$, $M > r_v$, and a nonnegative, strictly increasing sequence $\{\bar{t}_i\}$ such that, for all $t = 0, 1, \ldots$ and $i = 0, 1, \ldots$, (a) $|v_t| \leq r_v$ for $t \notin \{\bar{t}_i\}$, (b) $|v_{\bar{t}_i}| \in (r_v, M]$.*

The assumption above means that the measurement noises may take abnormal but bounded values at certain instants $\bar{t}_i$ since, of course, M is much larger than r_v. Such time instants correspond to the outliers and they are unknown. Indeed, we suppose to know r_w and r_v, and the reader is referred to [20] for an overview of the methods to estimate such parameters together with the underlying model. As will be clearer later, the knowledge of M is not required since it would be sufficient to assume that the outliers, though large, are bounded. In Section 4, a lower bound on the absolute value of outliers will be provided in such way that robustness is ensured for the proposed MHE method.

The moving-horizon approach consists in deriving a state estimate at the current time t by using the information given by $y_{t-N}, y_{t-N+1}, \ldots, y_t, u_{t-N}, u_{t-N+1}, \ldots, u_{t-1}$ with the

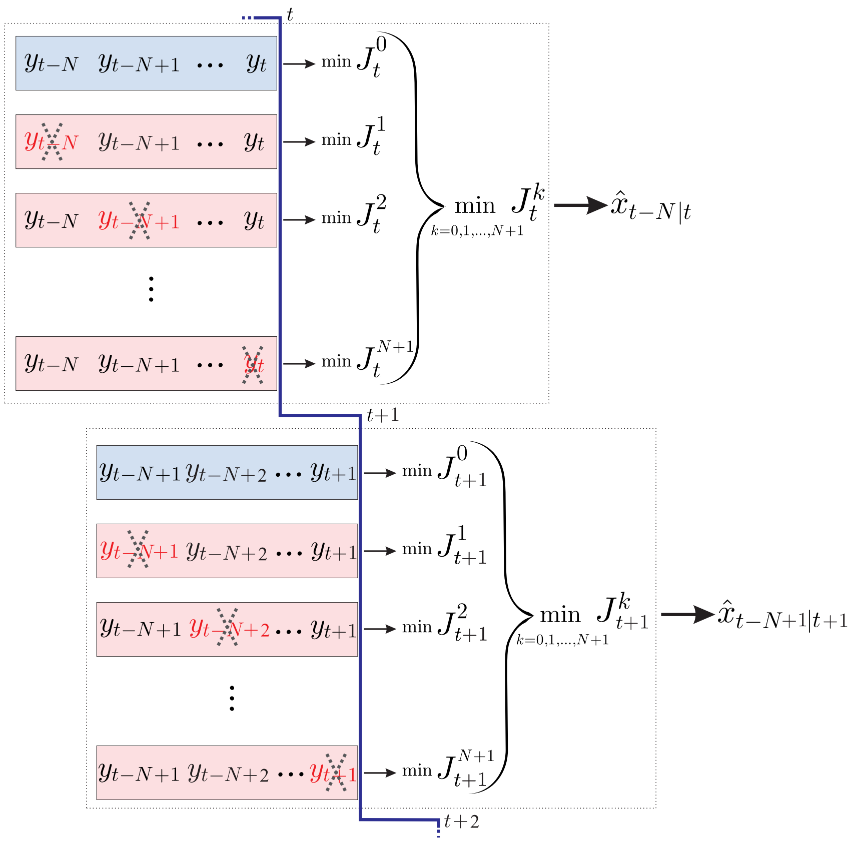

Fig. 1. Pictorial sketch of the MHE algorithm when moving from time t to $t+1$.

integer $N \geq 1$. More specifically, we aim to estimate $x_{t-N}, \ldots, x_t$ on the basis of such information and of a "prediction" $\bar{x}_{t-N}$ of the state x_{t-N} at the beginning of the moving window. We denote the estimates of $x_{t-N}, \ldots, x_t$ at time t by $\hat{x}_{t-N|t}, \hat{x}_{t-N+1|t}, \ldots, \hat{x}_{t|t}$, respectively.

As compared with the previous literature on MHE, here we consider explicitly the occurrence of outliers in the measures. In such a setting, a natural criterion to derive the estimator consists in resorting to a least-squares approach by explicitly trying to reduce the effect of the outliers. Though in principle we can deal with an arbitrary number of outliers, we restrict our attention to the case of at most only one measurement affected by outlier in the batch of measures included in the sliding window, thus assuming what follows.

Assumption 3 *The sequence* $\{\bar{t}_i\}$ *is such that* $\inf_{i\in\mathbb{N}} (\bar{t}_{i+1} - \bar{t}_i) > N + 1$.

If an outlier corrupts the k-th measure of the batch $1, 2, ..., N + 1$, a least-squares cost function that leaves out such a measure is

$$J_t^k(\hat{x}_{t-N}) = \mu \,\|\, \hat{x}_{t-N} - \bar{x}_{t-N} \,\|^2 + \alpha_k \sum_{\substack{i=t-N \\ i\neq t-N+k-1}}^{t} (y_i - C\hat{x}_i)^2 \tag{2}$$

for $k = 1, 2, \ldots, N+1$, where $\mu \geq 0$ and $\alpha_k > 0$. The cost (2) is to be minimized together with the constraints

$$\hat{x}_{i+1} = A\hat{x}_i + Bu_i \,,\; i = t-N, \ldots, t-1 \,. \tag{3}$$

If no outlier affects the measures of the batch, we may use all of them:

$$J_t^0(\hat{x}_{t-N}) = \mu \,\|\, \hat{x}_{t-N} - \bar{x}_{t-N} \,\|^2 + \alpha_0 \sum_{i=t-N}^{t} (y_i - C\hat{x}_i)^2 \tag{4}$$

where $\alpha_0 > 0$. Of course, also the minimization of (4) has to be performed with the constraints (3). In practice, at each time $t = N, N+1, \ldots$ we have to solve $N+2$ problems given by

$$\min_{\substack{\hat{x}\in\mathbb{R}^n \text{ s.t.} \\ (3) \text{ holds}}} J_t^k(\hat{x}) \,,\; k = 0, 1, \ldots, N+1$$

and compare the optimal costs (2) and (4): the best of such costs is associated with the estimate (see Fig. 1). Such a strategy is thus given by

$$\min_{k=0,1,\ldots,N+1} \min_{\substack{\hat{x}\in\mathbb{R}^n \text{ s.t.}\\ (3) \text{ holds}}} J_t^k(\hat{x}) \tag{5}$$

and, at each time $t = N, N+1, \ldots$, it can be summarized as follows.

MHE Algorithm

Input: $\bar{x}_{t-N}, y_{t-N}, y_{t-N+1}, \ldots, y_t, u_{t-N}, u_{t-N+1}, \ldots, u_{t-1}$

Output: $\hat{x}_{t-N|t}$

1: `for` $k = 0$ `to` $N+1$ `do`

compute $\hat{x}^k_{t-N|t} := \operatorname*{argmin}_{\substack{\hat{x}\in\mathbb{R}^n \text{ s.t.}\\ (3) \text{ holds}}} J_t^k(\hat{x})$

2: `endfor`

3: choose $k_t^* \in \operatorname*{argmin}_{k=0,1,\ldots,N+1} J_t^k(\hat{x}^k_{t-N|t})$

4: set $\hat{x}_{t-N|t} = \hat{x}^{k_t^*}_{t-N|t}$

Note that the estimate $\hat{x}_{t-N|t}$ is not unique in general. To complete the estimation, we have to determine the remaining estimates at time t by using (3) as follows: $\hat{x}_{t-N+i+1|t} = A\hat{x}_{t-N+i|t} + Bu_{t-N+i}$, $i = 0, \ldots, N-1$. If x_t belongs to $X \subset \mathbb{R}^n$ compact, one may take into account this information by performing the cost minimization with such additional constraints, namely $\hat{x}_{t-N+i|t} \in X$, $i = 0, 1, \ldots, N$.

Before solving the minimization problems, we need to assign $\bar{x}_{t-N}$, for which various choices can be made. For example, we can choose the result of the corresponding estimate at previous step, i.e., $\bar{x}_{t-N} = \hat{x}_{t-N|t-1}$. Another possibility consists in propagating the value of $\hat{x}_{t-N-1|t-1}$ as follows: $\bar{x}_{t-N} = A\,\hat{x}_{t-N-1|t-1} + Bu_{t-N-1}$. To simplify the stability analysis, we will adopt this last choice. Of course, it is necessary to select an a-priori prediction of x_0, which will be denoted by $\bar{x}_0$.

Remark 1 *In case of multiple outliers in the batch, one has to deal in general with $l = 2, 3, \ldots, N+1$ outliers affecting the batch of measures $y_{t-N}, y_{t-N+1}, \ldots, y_t$ by considering all permutations of l measures over the entire set. The number of costs to evaluate is equal to*

$$n_l = \binom{N+1}{l}.$$

In such a general case, one may resort to a mixed-integer formulation as follows:

$$\begin{aligned}
&\min_{\substack{\hat{x}_{t-N}\in\mathbb{R}^n,\beta^k_{t-N}\in\{0,1\},\\ k=0,1,\ldots,N+1}} \mu \,\|\, \hat{x}_{t-N} - \bar{x}_{t-N} \,\|^2 \\
&+ \beta^0_{t-N}\,\alpha_0 \sum_{i=t-N}^{t} (y_i - C\hat{x}_i)^2 \\
&+ \sum_{k=1}^{N+1} \beta^k_{t-N}\,\alpha_k \sum_{\substack{i=t-N\\ i\neq t-N+k-1}}^{t} (y_i - C\hat{x}_i)^2 \\
&\text{s.t.} \quad \sum_{k=0}^{N+1} \beta^k_{t-N} = l \text{ and } (3) \text{ hold}.
\end{aligned}$$

where the binary variables β^k_{t-N}, $k = 0, 1, \ldots, N+1$ are introduced to the scope [29]. The problem with $l = 1$ reduces to (5) and there is no computational advantage as compared with the solution method based on the MHE Algorithm. Clearly, the problem for $l \in \{2, 3, \ldots, N+1\}$ is computationally challenging. The use of branch-and-bound methods may result in some computational savings as compared with a pure zero-one enumeration technique, but in general mixed-integer optimization problems with quadratic cost function are pretty well-known to be NP-hard [12].

The stability properties of the proposed method are presented in the next section.

3 Stability of MHE

The collections of measures at time steps $t-N, t-N+1, \ldots, t$, with the k-th measure left out are given by

$$y^t_{t-N}\big|_k = F_k x_{t-N} + H_k w^{t-1}_{t-N} + v^t_{t-N}\big|_k\,,\quad k = 0, 1, \ldots, N+1$$

with F_k and H_k for $k \neq 0$ obtained from F_0 and H_0 by deleting the k-th block row, respectively, where

$$F_0 := \begin{pmatrix} C \\ C\,A \\ \vdots \\ C\,A^N \end{pmatrix}$$

$$H_0 := \begin{pmatrix} 0 & 0 & \ldots & 0 \\ C & 0 & \ldots & 0 \\ CA & C & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-1} & CA^{N-2} & \ldots & C \end{pmatrix}.$$

In the following, we will denote $(\alpha_0, \alpha_1, \ldots, \alpha_{N+1})$ by α and let

$$\alpha_{\min} := \min_{k=0,1,\ldots,N+1} \alpha_k \quad \alpha_{\max} := \max_{k=0,1,\ldots,N+1} \alpha_k$$

and

$$h := \max_{k=0,1,\ldots,N+1} \|H_k\| \,.$$

Based on the aforesaid, we can state the following result on the stability of the estimation error $e_{t-N} := x_{t-N} - \hat{x}_{t-N|t}$.

Theorem 1 *Suppose that Assumptions 1, 2, and 3 hold, F_k is of full rank for all $k = 0, 1, \ldots, N+1$, and let*

$$\delta := \min_{k=0,1,\ldots,N+1} \lambda_{\min}\left(F_k^\top F_k\right) > 0 \,. \tag{6}$$

Then the sequence $\{\zeta_t\}$ given by

$$\zeta_0 = \kappa\,(\mu, \alpha) \tag{7a}$$

$$\zeta_{t+1} = a\,(\mu, \alpha)\ \zeta_t + b\,(\mu, \alpha)\ ,\ t = 0, 1, \ldots \tag{7b}$$

is such that $\|e_{t-N}\|^2 \le \zeta_t$ for $t = N, N+1, \ldots,$ where

$$\begin{aligned}
\kappa\,(\mu, \alpha)\ &:= \frac{2}{\mu + \delta\,\alpha_{\min}}\ \left(2\mu\ \|x_0 - \bar{x}_0\|^2 + c\right) \\
a\,(\mu, \alpha)\ &:= \frac{8\mu\,\|A\|^2}{\mu + \delta\,\alpha_{\min}} \\
b\,(\mu, \alpha)\ &:= \frac{8\mu r_w^2 + 2c}{\mu + \delta\,\alpha_{\min}} \\
c &:= 2\,(\alpha_{\min} + \alpha_{\max})\ \left(N^2 r_w^2 h^2 + (N r_v + M)^2\right)\ .
\end{aligned}$$

If μ is chosen such that $a\,(\mu, \alpha) < 1$, the sequence $\{\zeta_t\}$ converges to

$$\frac{b\,(\mu, \alpha)}{1 - a\,(\mu, \alpha)}$$

and is strictly decreasing if

$$\zeta_0 > \frac{b\,(\mu, \alpha)}{1 - a\,(\mu, \alpha)}\,.$$

Proof. First, we will derive a useful bound, which will be used later on. Since $\left\|H_k w_{t-N}^{t-1}\right\| \le \|H_k\|\ \left\|w_{t-N}^{t-1}\right\| \le \|H_k\|\, N r_w$ and $\left\|\,v_{t-N}^t\big|_k\right\| \le N r_v + M$, from the bound

$$\begin{aligned}
&\left\|\,y_{t-N}^t\big|_k - F_k x_{t-N}\right\| = \left\|H_k w_{t-N}^{t-1} + v_{t-N}^t\big|_k\right\| \\
&\le \left\|H_k w_{t-N}^{t-1}\right\| + \left\|\,v_{t-N}^t\big|_k\right\|
\end{aligned}$$

we obtain

$$\left\|\,y_{t-N}^t\big|_k - F_k x_{t-N}\right\|^2 \le 2\ \left(N^2 r_w^2 h^2 + (N r_v + M)^2\right)\ . \tag{8}$$

Consider the optimal cost at time t:

$$\begin{aligned}
J_t^*(\hat{x}_{t-N|t}) = \mu\,\|\,\hat{x}_{t-N|t} - \bar{x}_{t-N}\,\|^2 \\
+ \alpha_{k_t^*}\ \left\|\,y_{t-N}^t\big|_{k_t^*} - F_{k_t^*}\,\hat{x}_{t-N|t}\right\|^2
\end{aligned}$$

and hence

$$\begin{aligned}
J_t^*(\hat{x}_{t-N|t}) \ge \mu\,\|\,\hat{x}_{t-N|t} - \bar{x}_{t-N}\,\|^2 \\
+ \alpha_{\min}\ \left\|\,y_{t-N}^t\big|_{k_t^*} - F_{k_t^*}\,\hat{x}_{t-N|t}\right\|^2
\end{aligned} \tag{9}$$

where, from now on, J_t^* stands for $J_t^{k_t^*}$. Since

$$\begin{aligned}
&\|F_k x_{t-N} - F_k \hat{x}_{t-N|t}\| = \left\|F_k x_{t-N} - y_{t-N}^t\big|_k + y_{t-N}^t\big|_k\right. \\
&\left. - F_k \hat{x}_{t-N|t}\right\| \le \left\|F_k x_{t-N} - y_{t-N}^t\big|_k\right\| \\
&+ \left\|\,y_{t-N}^t\big|_k - F_k \hat{x}_{t-N|t}\right\|\ ,
\end{aligned}$$

we have

$$\begin{aligned}
&\|F_k x_{t-N} - F_k \hat{x}_{t-N|t}\|^2 \le 2\left\|F_k x_{t-N} - y_{t-N}^t\big|_k\right\|^2 \\
&+ 2\left\|\,y_{t-N}^t\big|_k - F_k \hat{x}_{t-N|t}\right\|^2
\end{aligned}$$

and hence

$$\begin{aligned}
&\left\|\,y_{t-N}^t\big|_k - F_k \hat{x}_{t-N|t}\right\|^2 \ge \frac{1}{2}\|F_k x_{t-N} - F_k \hat{x}_{t-N|t}\|^2 \\
&- \left\|\,y_{t-N}^t\big|_k - F_k x_{t-N}\right\|^2
\end{aligned} \tag{10}$$

for $k = 0, 1, \ldots, N+1$. Using (8) and $\|F_k x_{t-N} - F_k \hat{x}_{t-N|t}\|^2 \ge \delta\,\|x_{t-N} - \hat{x}_{t-N|t}\|^2$ with $\delta > 0$ since all the matrices F_k are of full rank by assumption as pointed out in (6), it follows from (10) for $k = k_t^*$ that

$$\begin{aligned}
&\alpha_{\min}\left\|\,y_{t-N}^t\big|_{k_t^*} - F_{k_t^*}\hat{x}_{t-N|t}\right\|^2 \ge \frac{\delta\,\alpha_{\min}}{2}\|x_{t-N} \\
&- \hat{x}_{t-N|t}\|^2 - c_1
\end{aligned}$$

where

$$c_1 := 2\,\alpha_{\min}\left(N^2 r_w^2 h^2 + (N r_v + M)^2\right)$$

and, using this inequality in (9), that

$$\begin{aligned}
J_t^*(\hat{x}_{t-N|t}) \ge \mu\,\|\,\hat{x}_{t-N|t} - \bar{x}_{t-N}\,\|^2 + \frac{\delta\,\alpha_{\min}}{2}\|x_{t-N} \\
- \hat{x}_{t-N|t}\|^2 - c_1\,.
\end{aligned} \tag{11}$$

Since

$$\|x_{t-N} - \hat{x}_{t-N|t}\| \le \|x_{t-N} - \bar{x}_{t-N}\| + \|\bar{x}_{t-N} - \hat{x}_{t-N|t}\| \tag{12}$$

and hence $\|x_{t-N} - \hat{x}_{t-N|t}\|^2 \leq 2\|x_{t-N} - \bar{x}_{t-N}\|^2 + 2\|\bar{x}_{t-N} - \hat{x}_{t-N|t}\|^2$, we have

$$\|\hat{x}_{t-N|t} - \bar{x}_{t-N}\|^2 \geq \frac{1}{2}\|x_{t-N} - \hat{x}_{t-N|t}\|^2 - \|x_{t-N} - \bar{x}_{t-N}\|^2$$

and, using this inequality in (11), we finally obtain

$$J_t^*(\hat{x}_{t-N|t}) \geq \left(\frac{\mu}{2} + \frac{\delta\,\alpha_{\min}}{2}\right) \|x_{t-N} - \hat{x}_{t-N|t}\|^2 - \mu\|x_{t-N} - \bar{x}_{t-N}\|^2 - c_1\,. \tag{13}$$

Consider the following inequalities, which hold for the various definitions we have introduced so far:

$$J_t^*(\hat{x}_{t-N|t}) \leq J_t^k(\hat{x}^k_{t-N|t}) \leq J_t^k(x_{t-N}) = \mu\,\|x_{t-N} - \bar{x}_{t-N}\|^2 + \alpha_k\,\left\| y^t_{t-N}\big|_k - F_k\,x_{t-N}\right\|^2 \leq \mu\,\|x_{t-N} - \bar{x}_{t-N}\|^2 + \alpha_{\max}\,\left\| y^t_{t-N}\big|_k - F_k\,x_{t-N}\right\|^2$$

for $k = 0, 1, \ldots, N+1$. Using (8), the previous inequality yields

$$J_t^*(\hat{x}_{t-N|t}) \leq \mu\,\|x_{t-N} - \bar{x}_{t-N}\|^2 + c_2 \tag{14}$$

where

$$c_2 := 2\alpha_{\max}\left(N^2 r_w^2 h^2 + (N r_v + M)^2\right)\,.$$

It follows from (13) and (14) that

$$\left(\frac{\mu}{2} + \frac{\delta\,\alpha_{\min}}{2}\right) \|e_{t-N}\|^2 \leq 2\,\mu\,\|x_{t-N} - \bar{x}_{t-N}\|^2 + c \tag{15}$$

where $c := c_1 + c_2$. Since $x_{t-N} = A x_{t-N-1|t-1} + B u_{t-N-1} + w_{t-N-1}$ and $\bar{x}_{t-N} = A\hat{x}_{t-N-1} + B u_{t-N-1}$, we obtain

$$\|x_{t-N} - \bar{x}_{t-N}\|^2 \leq \left\|A(x_{t-N-1|t-1} - \hat{x}_{t-N-1}) + w_{t-N-1}\right\|^2 \leq 2\,\|A\|^2\,\left\|x_{t-N-1} - \hat{x}_{t-N-1|t-1}\right\|^2 + 2\,r_w^2\,. \tag{16}$$

Summing up, if we combine (15) and (16), it follows that

$$\left(\frac{\mu}{2} + \frac{\delta\,\alpha_{\min}}{2}\right) \|e_{t-N}\|^2 \leq 4\,\mu\,\|A\|^2\,\|e_{t-N-1}\|^2 + 4\mu r_w^2 + c$$

and hence that

$$\|e_{t-N}\|^2 \leq \frac{8\,\mu\,\|A\|^2}{\mu + \delta\,\alpha_{\min}}\,\|e_{t-N-1}\|^2 + \frac{8\mu r_w^2 + 2c}{\mu + \delta\,\alpha_{\min}}$$

for all $t = N, N+1, \ldots$. After fixing the initial condition for $t = N$ in (7a) by using (15), from the above inequality it is straightforward to set the sequence $\{\zeta_t\}$ according to (7b), thus concluding the proof. □

Note that Theorem 1 with the choice of a sufficiently small μ ensures that the estimation error e_{t-N} belongs to the compact set $B(r_e)$, where

$$r_e := \max\left(\kappa(\mu,\alpha), \frac{b(\mu,\alpha)}{1 - a(\mu,\alpha)}\right)$$

from now on.

Remark 2 *The stability condition $\delta > 0$ in (6) is to be ascribed to observability. More precisely, such a condition corresponds to have full rank for the observability matrix F_0 as well for all the matrices obtained from F_0 by deleting each row in turn. In other words, the system should be observable also by excluding each of the measures in the batch in accordance with what one can expect because of the leave-one-out moving-horizon strategy proposed here. The bounding sequence (7) depends on the various system parameters such as $\|A\|$, r_w, r_v, and M. The smaller such parameters, the tighter the bound. The choice of μ and N is more involving and deserves a careful analysis with some specific tools such as quadratic boundedness [4].*

In the next section, we will address the robustness property of the moving-horizon estimators.

4 Robustness of MHE

In this section, we will discuss the selection of the design parameters $\alpha_0, \alpha_1, \ldots, \alpha_{N+1}$ in such a way to ensure the rejection of the outliers to improve the performance of the moving-horizon estimator.

The MHE strategy (5) guarantees outlier rejection if

$$J_t^k\left(\hat{x}^k_{t-N|t}\right) < J_t^i\left(\hat{x}^i_{t-N|t}\right)\,, \quad k = 1, 2, \ldots, N+1, \quad i = 0, 1, \ldots, N+1, i \neq k \tag{17}$$

with the k-th measure in the information batch affected by outlier, i.e.,

$$y_{t-N+i-1} = \bar{F}_i x_{t-N} + \bar{H}_i w^{t-1}_{t-N} + v_{t-N+i-1}\,, \quad i = 1, 2, \ldots, N+1 \tag{18a}$$

$$|v_{t-N+i-1}| \leq r_v\,, i = 1, 2, \ldots, N+1, \quad i \neq k \tag{18b}$$

$$\bar{r}_v < |v_{t-N+k-1}| \tag{18c}$$

for some $\bar{r}_v > 0$ with $\bar{F}_i$ and $\bar{H}_i$ denote the i-th rows of F_0 and H_0, respectively. More specifically, the following result holds.

Theorem 2 *Under the assumptions of Theorem 1 and having chosen $\alpha_0, \alpha_1, \ldots, \alpha_{N+1}$ all equal to any strictly positive constant and $\mu > 0$ such that $a(\mu, \alpha) < 1$, (18) implies (17) if the absolute values of outliers are strictly larger than*

$$\bar{r}_v = \sqrt{3\big(\bar{F}_{i^*} e^* + \bar{H}_{i^*} w^* + v^*\big)^2 + 3\bar{f}^2 r_e^2 + 3\bar{h}^2 N r_w^2} \tag{19}$$

where

$$\bar{f} := \max_{k=1,2,\ldots,N+1} \|\bar{F}_k\| \quad \bar{h} := \max_{k=1,2,\ldots,N+1} \|\bar{H}_k\|$$

$$(e^*, w^*, v^*, i^*) \in \operatorname*{argmax}_{e \in B(r_e), w \in B(\sqrt{N} r_w), |v| \le r_v, i \in \{1,\ldots,N+1\}} \big(\bar{F}_i e + \bar{H}_i w + v\big)^2 \tag{20}$$

with

$$\bar{r}_v > r_v \, . \tag{21}$$

Proof. Instead of proving (17) directly, we will consider the stronger but more easily tractable condition

$$J_t^k(\hat{x}_{t-N|t}) < J_t^i(\hat{x}_{t-N|t}) \, , \quad k = 1, 2, \ldots, N+1, \; i = 0, 1, \ldots, N+1, i \neq k \, , \quad \forall \hat{x}_{t-N|t} \in \mathbb{R}^n \tag{22}$$

since it is straightforward to verify that (22) implies (17). We will consider first (22) for $i = 0$ and later on the remaining cases, namely, for $i = 1, 2, \ldots, N+1$.

Case analysis for $i = 0$

To ensure $J_t^k(\hat{x}_{t-N|t}) < J_t^0(\hat{x}_{t-N|t})$ for $k = 1, 2, \ldots, N+1$ and all $\hat{x}_{t-N|t} \in \mathbb{R}^n$, consider the following:

$$\begin{aligned} v_{t-N+k-1}^2 &= \big(\bar{F}_k e_{t-N} + \bar{H}_k w_{t-N}^{t-1} + v_{t-N+k-1} \\ &\quad - \bar{F}_k e_{t-N} - \bar{H}_k w_{t-N}^{t-1}\big)^2 \le 3\big(y_{t-N+k-1} \\ &\quad - \bar{F}_k \hat{x}_{t-N|t}\big)^2 + 3\left(\bar{F}_k e_{t-N}\right)^2 + 3\left(\bar{H}_k w_{t-N}^{t-1}\right)^2 \\ &\le 3\big(y_{t-N+k-1} - \bar{F}_k \hat{x}_{t-N|t}\big)^2 + 3\left\|\bar{F}_k\right\|^2 \left\|e_{t-N}\right\|^2 \\ &\quad + 3\left\|\bar{H}_k\right\|^2 \left\|w_{t-N}^{t-1}\right\|^2 \le 3\big(y_{t-N+k-1} - \bar{F}_k \hat{x}_{t-N|t}\big)^2 \\ &\quad + 3\bar{f}^2 r_e^2 + 3\bar{h}^2 N r_w^2 \end{aligned} \tag{23}$$

where the first inequality was obtained by using the square sum bound. Thus, the term $y_{t-N+k-1} - \bar{F}_k \hat{x}_{t-N|t}$ is nonnull and hence

$$\big(y_{t-N+k-1} - \bar{F}_k \hat{x}_{t-N|t}\big)^2 > 0 \tag{24}$$

if

$$v_{t-N+k-1}^2 > 3\bar{f}^2 r_e^2 + 3\bar{h}^2 N r_w^2 \, . \tag{25}$$

Using (24), we easily derive

$$\alpha_k \sum_{\substack{i=t-N \\ i \neq t-N+k-1}}^{t} \big(y_i - C\hat{x}_{t-N|t}\big)^2 < \alpha_0 \sum_{i=t-N}^{t} \big(y_i - C\hat{x}_{t-N|t}\big)^2$$

as, by assumption, $\alpha_0 = \alpha_k$. After adding $\mu\|\hat{x}_{t-N|t} - \bar{x}_{t-N}\|^2$ to both sides of such an inequality, we obtain (22) for $i = 0$.

Case analysis for $i = 1, 2, \ldots, N+1$

We will adopt a reasoning quite similar to the previous case. First, note that, since we assume that all the cost parameters $\alpha_1, \alpha_2, \ldots, \alpha_{N+1}$ are equal to some strictly positive value, (22) restricted to $i = 1, 2, \ldots, N+1$ holds if

$$\big(\bar{F}_i e_{t-N} + \bar{H}_i w_{t-N}^{t-1} + v_{t-N+i-1}\big)^2 < \big(\bar{F}_k e_{t-N} + \bar{H}_k w_{t-N}^{t-1} + v_{t-N+k-1}\big)^2 \tag{26}$$

for $i = 1, 2, \ldots, N+1$. Toward this end, we will get an upper bound on the l.h.s. and a lower bound on the r.h.s. of (26) and then combine these bounds in such way to satisfy the inequality. The former is simply given by

$$\big(\bar{F}_k e_{t-N} + \bar{H}_k w_{t-N}^{t-1} + v_{t-N+k-1}\big) \le \big(\bar{F}_{i^*} e^* + \bar{H}_{i^*} w^* + v^*\big)^2 \tag{27}$$

where the maximizer defined in (20) exists for the Weierstrass theorem. The latter stems from (23), which yields

$$\frac{v_{t-N+k-1}^2}{3} - \bar{f}^2 r_e^2 - \bar{h}^2 N r_w^2 \le \big(\bar{F}_k e_{t-N} + \bar{H}_k w_{t-N}^{t-1} + v_{t-N+k-1}\big)^2 \, . \tag{28}$$

Using (27) and (28), it follows that (26) holds if

$$\big(\bar{F}_{i^*} e^* + \bar{H}_{i^*} w^* + v^*\big)^2 < \frac{v_{t-N+k-1}^2}{3} - \bar{f}^2 r_e^2 - \bar{h}^2 N r_w^2$$

or, after a little algebra,

$$v_{t-N+k-1}^2 > 3\big(\bar{F}_{i^*} e^* + \bar{H}_{i^*} w^* + v^*\big)^2 + 3\bar{f}^2 r_e^2 + 3\bar{h}^2 N r_w^2 \, . \tag{29}$$

Let us now combine the two cases. Since $\big(\bar{F}_{i^*} e^* + \bar{H}_{i^*} w^* + v^*\big)^2 \ge 0$, the r.h.s. of (25) is less than or equal to the r.h.s. of (29), it is straightforward to conclude that (22) holds by choosing $\bar{r}_v$ as in (19).

Proof of (21)

We proceed by a contradiction argument and hence let us suppose $\bar{r}_v \leq r_v$. Thus, (20) yields

$$r_v^2 \geq \bar{r}_v^2 \geq 3\big(\bar{F}_i e + \bar{H}_i w + v\big)^2 + 3\bar{f}^2 r_e^2 + 3\bar{h}^2 N r_w^2 \quad (30)$$

for all $e \in B(r_e)$, $w \in B(\sqrt{N} r_w)$, $v : |v| \leq r_v$, and $i = 1, 2, \ldots, N+1$. If we choose $e = 0$, $w = 0$, and v such that $|v| = r_v$, (30) results in a contradiction and hence we conclude that (21) holds. □

Remark 3 *Theorem 2 deserves a special comment since, in principle, one may deduce a less conservative condition for robustness by means of the direct use of (17) instead of (22). The condition (17) may facilitate the selection of the parameters* $\alpha_0, \alpha_1, \ldots, \alpha_{N+1}$ *but it requires the introduction of additional assumptions on the boundedness of the state trajectories. Indeed, the stronger condition (22) enables to deal only with estimation error and disturbances, thus taking advantage of the results of Theorem 1 and allowing for a simple choice of such parameters, as detailed later on.*

Remark 4 *The robustness of the proposed method in case of multiple outliers is difficult to be proved. Despite the single outlier case, the combinatorial nature of the problem prevents one to find suitable upper and lower bounds on the l.h.s. and r.h.s. of (26), respectively, and hence to fix the minimum amplitude of the absolute value of the outliers for which rejection is ensured. The proof of robustness in the sense of Theorem 2 is thus nontrivial under the assumption of multiple outliers, though the stability property of Theorem 1 holds.*

Based on the results presented so far and without loss of generality, we adopt the following cost functions:

$$J_t^0\,(\hat{x}_{t-N}) = \rho \,\|\, \hat{x}_{t-N} - \bar{x}_{t-N} \,\|^2 + \sum_{i=t-N}^{t} (y_i - C\hat{x}_i)^2 \quad (31a)$$

$$J_t^k\,(\hat{x}_{t-N}) = \rho \,\|\, \hat{x}_{t-N} - \bar{x}_{t-N} \,\|^2 + \sum_{\substack{i=t-N \\ i \neq t-N+k-1}}^{t} (y_i - C\hat{x}_i)^2\,, \quad k = 1, 2, \ldots, N+1 \quad (31b)$$

where the scalar $\rho = \mu/\alpha \geq 0$ is to be suitably chosen, as the minimization of (31a) and (31b) depends just only on the value of the ratio μ/α. Likewise, the stability properties of Theorem 1 depend only on μ/α and not separately on μ and α.

Theorem 2 claims that robustness is guaranteed in case of occurrence of outliers with absolute value larger than $\bar{r}_v$. Clearly, the dependence of $\bar{r}_v$ on M is such that an increase of M determines a larger $B(r_e)$ and hence a larger $\bar{r}_v$. Thus, we may reduce $\bar{r}_v$ by decreasing M as much as possible, which can be accomplished by solving an optimization problem that takes into account the various relations between M and $\bar{r}_v$. Toward this end, we explicitly point out the dependence on M in some previous definitions by redefining them with a little abuse of notation as follows:

$$\kappa(\rho, M) := \frac{4\left(\rho\, d_0^2 + N^2 h^2 r_w^2 + (N r_v + M)^2\right)}{\rho + \delta} \quad (32)$$

$$a(\rho) := \frac{8\rho\,\|A\|^2}{\rho + \delta} \quad (33)$$

$$b(\rho, M) := \frac{8\left(\left(\rho + N^2 h^2\right) r_w^2 + (N r_v + M)^2\right)}{\rho + \delta} \quad (34)$$

where d_0 is an upper bound on $\|x_0 - \bar{x}_0\|$. Note that in (33) there is no dependence on M, which indeed affects both (32) and (34). Thus, let us choose ρ such that $a(\rho) < 1$ (i.e., the stability result of Theorem 1 holds) and consider the following *minmax* problem:

$$\min \max_{i=0,1,\ldots,N+1} f_i(e, w, v, r_e, M) \quad (35a)$$

$$e^\top e - r_e^2 \leq 0 \quad (35b)$$

$$w^\top w - N r_w^2 \leq 0 \quad (35c)$$

$$v^2 - r_v^2 \leq 0 \quad (35d)$$

$$r_v - M \leq 0 \quad (35e)$$

$$r_e - \max\left(\kappa(\rho, M), \frac{b(\rho, M)}{1 - a(\rho)}\right) = 0 \quad (35f)$$

with unknowns $e \in \mathbb{R}^n$, $w \in \mathbb{R}^{n \times N}$, $v \in \mathbb{R}$, $r_e \geq 0$, and $M \geq 0$ and where

$$f_i(e, w, v, r_e, M) := \begin{cases} M^2\,, \text{ for } i = 0 \\ 3\big(\bar{F}_i e + \bar{H}_i w + v\big)^2 + 3\bar{f}^2 r_e^2 \\ +3\bar{h}^2 N r_w^2\,, \text{ for } i = 1, 2, \ldots, N+1. \end{cases}$$

Because of the continuity of the functions f_i, the problem (35) admits a solution.

In the next section, simulation results will be presented and discussed.

5 Numerical Example

We compared the Kalman filter with estimate update enabled on residual check (KF) and the proposed method (denoted by MHF, moving-horizon filter) about the estimation of the state variables of a second-order oscillating system by using only the measures of the first variable, possibly corrupted by outliers. Such an autonomous system with damping ratio ξ and (undamped) natural pulsation ω is described by a discrete-time linear equation with

$$A = \begin{bmatrix} 1 & T \\ -T\omega^2 & -2\omega\xi T + 1 \end{bmatrix} \quad C = [1 \quad 0]$$

with $\xi = 0.2$, $\omega = 0.5$ rad/s, and sampling period $T = 0.5$ s. The initial states were generated as white Gaussian noises with mean $[1 \quad 1]^\top$ and covariance $P_0 = \mathrm{diag}(2,2)$. White Gaussian processes with means equal to zero and covariances $Q = \mathrm{diag}(1,1)$, and $r = 0.01, 0.1, 1.0$ (except in case of outlier) were chosen as system and measurement disturbances, respectively. For each simulation run, we computed the corresponding values of r_v and r_w and finally of $\bar{r}_v$ by solving (35). The number of time steps between one outlier occurrence and the next one was randomly generated between $N+1$ and $2\,(N+1)$. The outlier amplitudes were chosen according to a zero-mean, white Gaussian distribution with dispersion equal to $\bar{r}_v$ and thus with a probability of having an outlier amplitude for which the rejection of the MHF is not ensured of about 68% (see Theorem 2).

The KF was designed by using P_0, Q, and r according to the simulation setting and different thresholds σ_t to enable the estimate update based on the current output error. (i.e., only in case the absolute value of such an error is less than σ_t). We considered different choices of σ_t by scaling the covariance of the output error, namely, $s_t = r + CP_{t|t-1}C^\top$, where $P_{t|t-1}$ is the "a priori" covariance of the KF. Specifically, we choose σ_t equal to $\sqrt{s_t}$, $2\sqrt{s_t}$, $5\sqrt{s_t}$, and $10\sqrt{s_t}$. The estimates of the MHF were obtained according to (5) by using the cost functions (31) with different choices of ρ such that the stability condition $\alpha(\rho) < 1$ is satisfied.

To compare the performances of all the estimators, we will show the mean computational time (MCT, in s) and the root mean square error (RMSE). The MCT is just the mean duration of the computation that is required to generate an estimate of MHF or KF at each time instant. The RMSE is defined as follows:

$$RMSE(t) = \left(\sum_{i=1}^{L} \frac{\|e_{t,i}\|^2}{L} \right)^{1/2}$$

where $e_{t,i}$ is the estimation error at time t in the i-th simulation run, and L is the number of simulation runs. The initial states of KF and MHF were initialized with $[1 \quad 1]^\top$ in all the simulation runs.

The result of a simulation run is presented in Fig. 2. Fig. 3 shows the boxplots of the RMSEs for both MHFs and KFs with different choices of ρ and σ_t, respectively. Tables 1-4 illustrate the results concerning tests over 100 simulation runs with zero-mean, Gaussian measurement noises having different dispersions. More specifically, Table 1 and 2 report the RMSE medians separately for the first and second state variable, respectively. As compared with the KF, the MHF performs much better in terms of RMSE with an MCT that is about three times that of the KF on average, as shown in Table 3. Moreover, the performances of the MHF turn out to be quite similar over large variations of ρ, whereas, by contrast, the KF performs badly in case of wrong choice of the threshold. Table 4 shows that the MCT of the MHF grows almost linearly with the increase of N (i.e., with the number of measures of the batch).

6 Conclusions

We have addressed the problem of state estimation for linear systems with measurements affected by outliers by devising a novel approach based on a moving-horizon strategy, for which stability and robustness have been established. We have verified the effectiveness of the proposed approach via simulations, where the higher estimation precision and enhanced robustness to outliers is paid in terms of a moderate increase of computational burden as compared with the Kalman filter with estimate update driven by a threshold test on the output error.

Future work will concern the extension of the proposed method to estimation in the presence of multiple outliers and for nonlinear systems. The use of heuristics to reduce the computational effort will be another topic to investigate.

References

[1] A.D. Akkaya and M.L. Tiku. Robust estimation in multiple linear regression model with non-gaussian noise. *Automatica*, 44(2):407–417, 2008.

[2] A. Alessandri and M. Awawdeh. Moving-horizon estimation for discrete-time linear systems with measurements subject to outliers. In *Proc. 53rd IEEE Conference on Decision and Control*, pages 2591–2596, Los Angeles, California, USA, 2014.

[3] A. Alessandri, M. Baglietto, and G. Battistelli. Receding-horizon estimation for discrete-time linear systems. *IEEE Trans. on Automatic Control*, 48(3):473–478, 2003.

[4] A. Alessandri, M. Baglietto, and G. Battistelli. On estimation error bounds for receding-horizon filters using quadratic boundedness. *IEEE Trans. on Automatic Control*, 49(8):1350–1355, 2004.

[5] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*, 44(7):1753 – 1765, 2008.

[6] A. Alessandri, M. Baglietto, and G. Battistelli. Min-max moving-horizon estimation for uncertain discrete-time systems. *SIAM J. Control and Optimization*, 50(3):1439–1465, 2012.

[7] A. Alessandri, M. Baglietto, G. Battistelli, and M. Gaggero. Moving-horizon state estimation for nonlinear systems using neural networks. *IEEE Trans. Neural Networks*, 22(5):768–780, 2011.

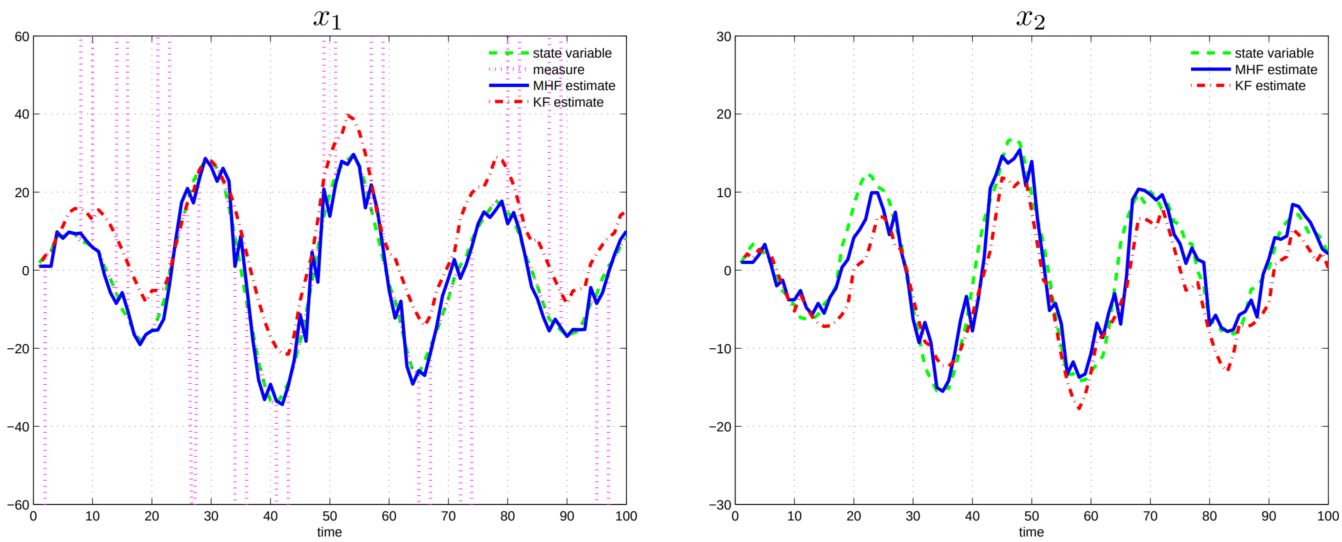

Fig. 2. True state, measures, and estimates of x_1 and x_2 in a simulation run with zero-mean, Gaussian measurement noise having $r = 0.01$ and using a MHF with $\rho = 10^{-4}$ and $N = 3$, and a KF with $\sigma_t = 2\sqrt{s_t}$.

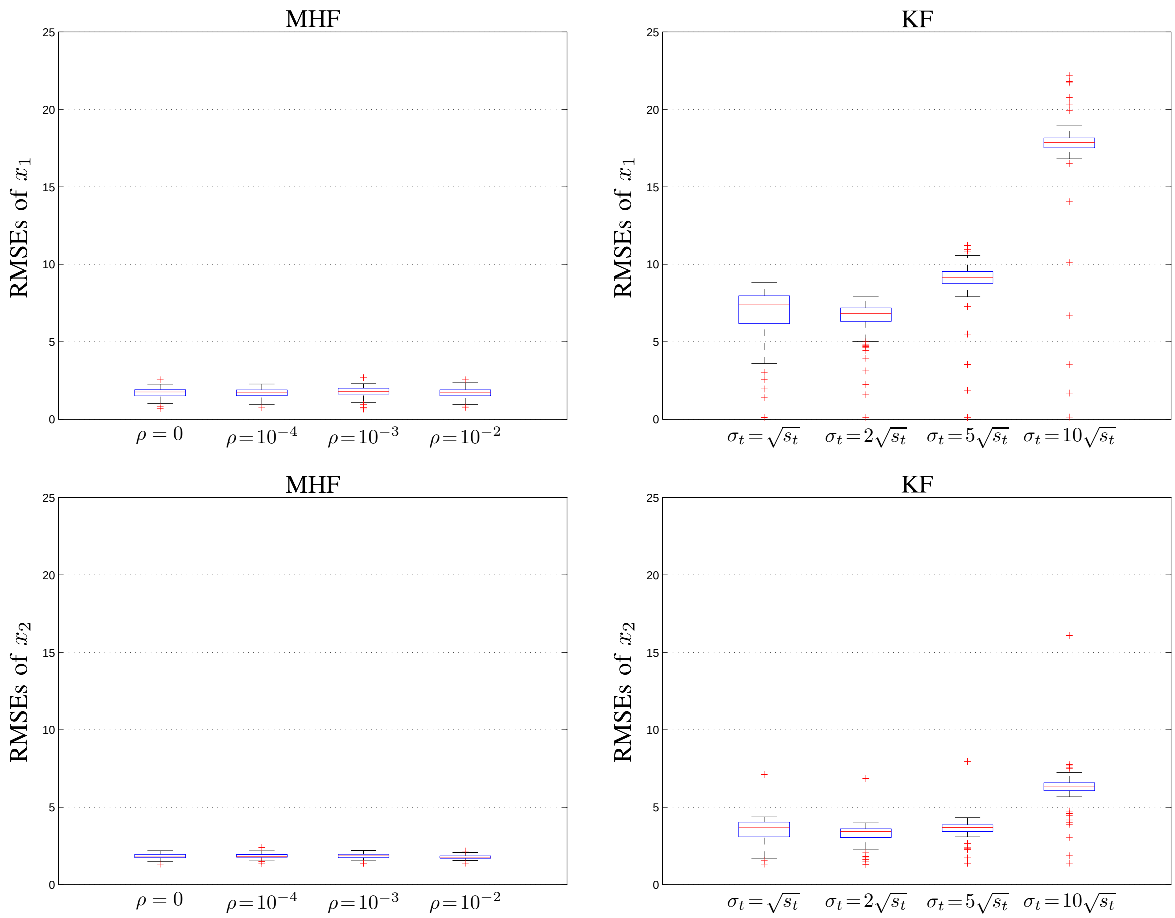

Fig. 3. RMSEs over 100 simulation runs with zero-mean, Gaussian measurement noise having variance $r = 0.01$ for different ρ (MHF with $N = 3$) and σ_t (KF).

Table 1
RMSE medians of x_1 for different ρ (MHF with $N = 3$) and σ_t (KF)

r	MHF				KF			
	$\rho = 0$	$\rho = 10^{-4}$	$\rho = 10^{-3}$	$\rho = 10^{-2}$	$\sigma_t = \sqrt{s_t}$	$\sigma_t = 2\sqrt{s_t}$	$\sigma_t = 5\sqrt{s_t}$	$\sigma_t = 10\sqrt{s_t}$
0.01	1.7593	1.6997	1.8002	1.7427	7.3796	6.8145	9.1603	17.8522
0.1	1.8679	1.7730	1.8191	1.7617	7.3707	6.6347	9.3632	18.2327
1.0	2.1133	2.0334	1.9819	1.9987	7.3584	6.6622	11.1109	22.9324

Table 2
RMSE medians of x_2 for different ρ (MHF with $N = 3$) and σ_t (KF)

r	MHF				KF			
	$\rho = 0$	$\rho = 10^{-4}$	$\rho = 10^{-3}$	$\rho = 10^{-2}$	$\sigma_t = \sqrt{s_t}$	$\sigma_t = 2\sqrt{s_t}$	$\sigma_t = 5\sqrt{s_t}$	$\sigma_t = 10\sqrt{s_t}$
0.01	1.8457	1.8356	1.8591	1.7833	3.6778	3.4288	3.6901	6.3668
0.1	1.8649	1.8692	1.8801	1.8026	3.7273	3.3492	3.7557	6.5189
1.0	2.1249	2.1199	2.0805	2.0212	3.7446	3.3314	4.5664	8.8081

Table 3
MCT (in s) for different ρ (MHF with $N = 3$) and σ_t (KF)

r	MHF				KF			
	$\rho = 0$	$\rho = 10^{-4}$	$\rho = 10^{-3}$	$\rho = 10^{-2}$	$\sigma_t = \sqrt{s_t}$	$\sigma_t = 2\sqrt{s_t}$	$\sigma_t = 5\sqrt{s_t}$	$\sigma_t = 10\sqrt{s_t}$
0.01	0.0281	0.0270	0.0277	0.0247	0.0070	0.0081	0.0070	0.0081
0.1	0.0262	0.0245	0.0252	0.0242	0.0073	0.0086	0.0077	0.0108
1.0	0.0248	0.0278	0.0259	0.0275	0.0089	0.0095	0.0081	0.0073

Table 4
MCT (in s) for different ρ and N (MHF)

ρ	$r = 0.01$				$r = 0.1$				$r = 1.0$			
	N=3	N=4	N=5	N=6	N=3	N=4	N=5	N=6	N=3	N=4	N=5	N=6
0	0.0281	0.0297	0.0372	0.0406	0.0262	0.0322	0.0344	0.0406	0.0248	0.0309	0.0347	0.0422
10^{-4}	0.0270	0.0303	0.0342	0.0386	0.0245	0.0309	0.0345	0.0409	0.0278	0.0288	0.0344	0.0408
10^{-3}	0.0277	0.0322	0.0388	0.0466	0.0252	0.0323	0.0375	0.0417	0.0259	0.0392	0.0422	0.0428
10^{-2}	0.0247	0.0308	0.0364	0.0388	0.0242	0.0291	0.0353	0.0411	0.0275	0.0294	0.0366	0.0439

[8] L. Fagiano and C. Novara. A combined moving horizon and direct virtual sensor approach for constrained nonlinear estimation. *Automatica*, 49(1):193–199, 2013.

[9] K. Fallahi, C.-T. Cheng, and M. Fattouche. Robust positioning systems in the presence of outliers under weak GPS signal conditions. *IEEE Systems Journal*, 6(3):401–413, 2012.

[10] M. Farina, G. Ferrari-Trecate, and R. Scattolini. Moving-horizon partition-based state estimation of large-scale systems. *Automatica*, 46(5):910–918, 2010.

[11] H. Ferdowsi, S. Jagannathan, and M. Zawodniok. An online outlier identification and removal scheme for improving fault detection performance. *IEEE Trans. on Neural Networks and Learning Systems*, 25(5):908–919, 2014.

[12] C.A. Floudas. *Nonlinear and Mixed-Integer Optimization. Fundamentals and Applications.* Oxford University Press, Oxford, UK, 1995.

[13] M.A. Gandhi and L. Mili. Robust Kalman filter based on a generalized maximum-likelihood-type estimator. *IEEE Trans. on Signal Processing*, 58(5):2509–2520, 2010.

[14] H. Garces and D. Sbarbaro. Outliers detection in environmental monitoring databases. *Eng. Appl. of Artificial Intelligence*, 24(2):341–349, 2011.

[15] Y. Guo and B. Huang. Moving horizon estimation for switching nonlinear systems. *Automatica*, 49(11):3270–3281, 2013.

[16] A. H. Jazwinski. Limited memory optimal filtering. *IEEE Trans. on Automatic Control*, 13(5):558–563, 1968.

[17] F. Lauer, G. Bloch, and R. Vidal. A continuous optimization framework for hybrid system identification. *Automatica*, 47(3):608–613, 2011.

[18] A. Matasov and V. Samokhvalov. Guaranteeing parameter estimation with anomalous measurement errors. *Automatica*, 32(9):1317–1322, 1996.

[19] S. Meng and L. Liu. Enhanced monitoring-as-a-service for effective cloud management. *IEEE Trans. on Computers*, 62(9):1705–1720, 2013.

[20] M. Milanese and C. Novara. Set membership identification of nonlinear systems. *Automatica*, 40(6):957975, 2004.

[21] T. Ortmaier, M. Groger, D.H. Boehm, V. Falk, and G. Hirzinger. Motion estimation in beating heart surgery.

IEEE Trans. on Biomedical Engineering, 52(10):1729–1740, 2005.

[22] R.K. Pearson. Outliers in process modeling and identification. *IEEE Trans. on Control Systems Technolgy*, 10(1):55–63, 2002.

[23] P.J., Rousseeuw, and A.M. Leroy. *Robust regression and outlier detection*. Wiley, New York, 1987.

[24] C. V. Rao, J. B. Rawlings, and J. H. Lee. Constrained linear estimation–a moving horizon approach. *Automatica*, 37(10):1619–1628, 2001.

[25] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. *IEEE Trans. on Automatic Control*, 48(2):246–257, 2003.

[26] D. Shi, T. Chen, and L. Shi. New Kalman filter and smoother consistency tests. *Automatica*, 49(10):3141–3144, 2013.

[27] W. Xu, E.-W. Bai, and M. Cho. System identification in the presence of outliers and random noises: A compressed sensing approach. *Automatica*, 50(11):2905–2911, 2014.

[28] J. Zhang, M. Zulkernine, and A. Haque. Random-forests-based network intrusion detection systems. *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(5):649–659, 2008.

[29] G. Zioutas, L. Pitsoulis, and A. Avramidis. Quadratic mixed integer programming and support vectors for deleting outliers in robust regression. *Annals of Operations Research*, 166(1):339–353, 2009.