Regularised estimation for ARMAX process with measurements subject to outliers

Yin, Le; Liu, Shuo; Gao, Hui

2018

Yin, L., Liu, S., & Gao, H. (2018). Regularised estimation for ARMAX process with measurements subject to outliers. *IET Control Theory & Applications*, 12(7), 865-874.

<https://hdl.handle.net/10356/87744>

<https://doi.org/10.1049/iet-cta.2017.1204>

© 2018 Institution of Engineering and Technology. This paper was published in *IET Control Theory and Applications* and is made available as an electronic reprint (preprint) with permission of Institution of Engineering and Technology. The published version is available at: [<http://dx.doi.org/10.1049/iet-cta.2017.1204>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 26 Apr 2023 18:58:57 SGT

Regularised estimation for ARMAX process with measurements subject to outliers

ISSN 1751-8644

Received on 29th October 2017

Accepted on 9th January 2018

E-First on 14th February 2018

doi: 10.1049/iet-cta.2017.1204

www.ietdl.org

Le Yin^{1,2}, Shuo Liu^{1,3}, Hui Gao⁴ ✉¹Berkeley Education Alliance for Research in Singapore, 1 Create Way, Singapore²School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore³Department of Electrical and Computer Engineering, National University of Singapore, 21 Lower Kent Ridge Road, Singapore⁴School of Automation, Chongqing University, Chongqing, People's Republic of China

✉ E-mail: gaohui_cqu@sina.com

Abstract: ARMAX models are widely used in control engineering for both system description and control design. They can accurately describe a large class of real processes with relatively low complexity, but do not take into account observation errors, which can be particularly important in applications like filtering and fault diagnosis. Due to the intrinsic dependence in ARMAX process output, a single outlier may contaminate multiple data entries and completely spoil the conventional least-squares estimate. In this study, the authors develop a novel Moving Horizon Estimator that is not only robust to outliers but also able to reveal outliers. By utilising the fact that outliers are extreme errors that occur infrequently, the estimation problem is formulated as a least-squares optimisation problem with outliers explicitly modelled and regularised with ℓ_1 -norm to induce sparsity. A coordinate descent-based solver is developed to obtain iterative algorithms with guaranteed convergence and closed-form solution available to each coordinate sub-problem per iteration. Due to the explicit modelling of outlier vectors, the impact of an outlier on multiple time instants can be estimated and mitigated. Simulation tests demonstrate that the proposed algorithm can effectively cope with outliers, even for the case when the commonly used Huber's M -estimation approach breaks down.

1 Introduction

The Auto-Regressive-Moving-Average with eXogenous input (ARMAX) model is a basic model for describing practical systems, which has been attracting much attention from researchers not only in time-series analysis but also in systems and control [1–5]. ARMAX models can accurately describe a large class of real processes with relatively low complexity [6, 7], but do not take into account the observation errors on the output of the process, which can be particularly important in applications like filtering and fault diagnosis [8, 9].

Robustness to uncertainties is one of the major concerns in system state estimation and control [10–12]. Uncertainties can be classified into different categories according to their origin. For example, in networked control systems, the exchange of information is subject to network congestion and latency [13], which gives rise to the uncertainty in measurement. In wireless networked control systems, the mobile sensor nodes may be affected by external environment, which contributes to the uncertainty in system state [14]. In addition, the discrepancy between the mathematical model and the actual dynamical system brings the uncertainty in model parameter [11, 15]. Such uncertainties can lead to instability and poor performance of systems, and various proposals have been developed to address this issue in various contexts. In particular, the system identification and state estimation problems under uncertainties have been studied by many authors, among which we quote Wang and Burnham [16], Guidorzi *et al.* [17], and Diversi *et al.* [8, 9].

Kalman filtering of the ARMAX process has been extensively studied in the literature [2, 3, 18, 19]. The Kalman filter is the optimal linear filter for a large class of problems and has enjoyed a great popularity within the engineering community [20, 21]. It can be derived as the weighted least-squares solution to a regression problem, without regard to distributional assumptions [22–24]. However, the Kalman filter is known to be susceptible to outliers. The squared-error criterion used in Kalman filter is extremely sensitive to outliers [25, 26]; squaring a larger number makes it even larger and hence an outlier entering the cost functional is

likely to dominate all other observations. Therefore, even rare occurrences of outliers can severely degrade the performance of the Kalman filter, resulting in poor estimates [27].

The problem may become even worse with the autoregressive and moving average nature of the ARMAX process. Due to the intrinsic dependence in ARMAX process output, i.e. the current output value depends on the previous output values, a single outlier in observation may contaminate multiple data entries when it comes to estimating the process and completely spoil the conventional least-squares estimate. Therefore, it appears extremely useful to have a robust estimator that performs well under such circumstances.

In this paper, the estimation problem for ARMAX process with output contaminated by outliers is addressed by using a moving-horizon approach. In moving-horizon estimation, only the most recent information is processed and the estimates are obtained by solving a least-squares optimisation problem with a batch of inputs and outputs according to a sliding-window strategy. Since the size of the estimation is fixed by the sliding window, efficient online optimisation can be performed. The intrinsic robustness to uncertainties in the knowledge of system dynamics and capability of accounting constraints have made the approach favourable to both linear systems [28–31] and non-linear systems [32–34].

Robust statistical procedures provide formal methods to spot the outlying data points and reduce their influence [35]. One well-known approach for robustifying the least-squares estimator or Kalman filter resorts to a robust M -estimator with some outlier-resilient objective function such as Huber's cost [36–38]. The objective functions used in M -estimators aim to bound the influence of large residuals and thereby suppressing the impact of outliers. This approach requires independent noise entries so that prewhitening is usually performed to uncorrelate the residual error vectors before processing them in a robust regression framework. However, prewhitening spreads the outliers to non-contaminated data entries making outlier-resilient estimation even more challenging [39, 40].

Efforts have also been made to totally remove the impact of extreme errors. For example, a least-trimmed squares based

approach for robust moving horizon estimation of discrete-time linear systems is reported in [41]. The approach consists in minimising 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 the procedure is repeated when the sliding window moves to the next time instant. Although the principle of deleting one measure at a time is possible to be generalised to multiple measures, it is not obvious at all which subset of measures should be deleted. The computations involved may therefore become intractable because of the large number of subsets that would have to be considered [42].

Recent advances in compressed sensing theory have made the ℓ_1 -norm regularisation a promising approach in both signal recovery [43, 44] and sparse error detection [45, 46]. In state estimation, the ℓ_1 -norm regularisation was also used, e.g. to address the problem of detecting and mitigating the effects of an adversarial corruption of sensory data in cyber-physical systems [47–49]. Besides ℓ_1 -norm regularisation, the sum-of-norms regularisation was applied in the smoothing problem for a generic linear system where impulsive disturbances occur in the process model [50]. Similar to ℓ_1 -norm regularisation, which induces sparsity in the estimates, sum-of-norms regularisation produces sparse estimates in which the zero components come in groups.

In this paper, we focus on handling the uncertainty in measurement and explicitly model the outliers as auxiliary variables. By leveraging the fact that the outlier component could indeed be a sparse vector, the estimation problem is formulated as a least-squares problem with outliers regularised with ℓ_1 -norm to induce sparsity. The resultant optimisation problem is jointly convex over the state and outlier vectors. A coordinate descent-based solver is developed to obtain iterative algorithms with guaranteed convergence and closed-form solution available to each coordinate sub-problem per iteration. Due to the explicit modelling of outlier vectors, the impact of an outlier on multiple time instants can be estimated and mitigated. Therefore, a more robust and accurate estimate can be expected. Some attractive aspects of the proposed approach are summarised as follows:

- The approach achieves a dual-task in both obtaining a robust estimate and detecting the outliers without requiring any assumption about the outlier pattern except that it must be sparse. Indeed, outliers, being the most extreme observations, are events that occur infrequently, and thus sparse.
- Since outlier vectors are explicitly modelled, the impact of an outlier on multiple time instants can be estimated and subsequently reduced. This is unlike the commonly used M-estimation methodology such as Huber-based robust scheme which only implicitly incorporates outliers and may break down due to contamination of multiple data entries induced by a single outlier.
- The resultant optimisation problem is convex, so the global solution can be computed efficiently by using many standard methods. In fact, the coordinate descent solver developed in this paper allows it to be solved in linear time complexity, which makes the approach quite practical even for large estimation horizon length.

The remaining of the paper is organised as follows. Section 2 develops the Moving Horizon Estimator (MHE) for ARMAX processes. The robustification scheme based on outlier-pursuit ℓ_1 -norm regularisation is introduced in Section 3, where a coordinate descent-based solver is also developed. In Section 4, both simulation and experimental results are presented and discussed. Finally, the conclusions are drawn in Section 5.

2 Preliminaries

The ARMAX process and moving horizon estimation were already introduced in the literature [1–3, 18, 28–31]. In this section, we only give the equations necessary for the derivation of the ℓ_1 -regularised MHE in Section 3.

2.1 ARMAX process with aberrant observations

Consider the following ARMAX process:

$$A(q^{-1})z_k = B(q^{-1})u_k + C(q^{-1})\varepsilon_k \quad (1)$$

with aberrant observations

$$y_k = z_k + o_k \quad (2)$$

where

$$\begin{aligned} A(q^{-1}) &= I + A_1q^{-1} + \dots + A_nq^{-n} \\ B(q^{-1}) &= B_1q^{-1} + B_2q^{-2} + \dots + B_nq^{-n} \\ C(q^{-1}) &= I + C_1q^{-1} + \dots + C_nq^{-n} \end{aligned}$$

are matrix polynomials in backward-shift operator q^{-1} , i.e. $q^{-1}z_k = z_{k-1}$, with $A_i \in \mathbb{R}^{m \times m}$, $B_i \in \mathbb{R}^{m \times l}$ and $C_i \in \mathbb{R}^{m \times m}$. The process input and output are given by $u_k \in \mathbb{R}^{l \times 1}$ and $z_k \in \mathbb{R}^{m \times 1}$, respectively. The zero-mean white Gaussian noise $\varepsilon_k \in \mathbb{R}^{m \times 1}$ has a known covariance matrix R_k , and is in general independent of u_k . In the measurement contamination model of (2), y_k is what we really observe and z_k is the process output given by (1). The outlier vector o_k is defined to be deterministically non-zero if y_k corresponds to an aberrant observation, and $\mathbf{0}_m$ otherwise.

The ARMAX model in (1) can be represented in the state-space form given as [2]

$$x_{k+1} = \Phi_A x_k + \Gamma u_k + \Omega \varepsilon_k \quad (3)$$

$$z_k = H x_k + \varepsilon_k \quad (4)$$

where

$$\begin{aligned} \Phi_A &= \begin{bmatrix} -A_1 & I & 0 & \dots & 0 \\ -A_2 & 0 & I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -A_{n-1} & 0 & 0 & \dots & I \\ -A_n & 0 & 0 & \dots & 0 \end{bmatrix} \\ \Gamma &= [B_1 \quad B_2 \quad \dots \quad B_n]^T \\ \Omega &= \begin{bmatrix} C_1 - A_1 \\ C_2 - A_2 \\ \vdots \\ C_n - A_n \end{bmatrix} \\ H &= [I \quad 0 \quad \dots \quad 0] \end{aligned}$$

Substituting (4) into (3) gives

$$x_{k+1} = \Phi x_k + \Gamma u_k + \Omega z_k \quad (5)$$

where

$$\Phi = \Phi_A - \Omega H = \begin{bmatrix} -C_1 & I & 0 & \dots & 0 \\ -C_2 & 0 & I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -C_{n-1} & 0 & 0 & \dots & I \\ -C_n & 0 & 0 & \dots & 0 \end{bmatrix}$$

2.2 Moving-horizon estimation for ARMAX process

The moving-horizon approach consists in deriving an estimate at the current time k by using the most recent information given by $y_{k-N}, \dots, y_k$ and $u_{k-N}, \dots, u_{k-1}$ with the integer $N \geq 1$. Using (4) and (5), $N+1$ equations can be set up as follows:

$$\begin{aligned}
z_{k-N} &= Hx_{k-N} + \varepsilon_{k-N} \\
z_{k-N+1} &= H\Phi x_{k-N} + H\Gamma u_{k-N} + H\Omega z_{k-N} \\
&\quad + \varepsilon_{k-N+1} \\
&\vdots \\
z_k &= H\Phi^N x_{k-N} + \sum_{i=1}^N H\Phi^{i-1} \Gamma u_{k-i} \\
&\quad + \sum_{i=1}^N H\Phi^{i-1} \Omega z_{k-i} + \varepsilon_k
\end{aligned} \tag{6}$$

For convenience, let us define the vectors

$$\begin{aligned}
Z_{k,N} &= [z_{k-N}^T, \dots, z_{k-1}^T, z_k^T]^T \\
U_{k-1,N-1} &= [u_{k-N}^T, \dots, u_{k-2}^T, u_{k-1}^T]^T \\
E_{k,N} &= [\varepsilon_{k-N}^T, \dots, \varepsilon_{k-1}^T, \varepsilon_k^T]^T
\end{aligned}$$

and the matrices

$$\begin{aligned}
T_N &= \begin{bmatrix} I & 0 & 0 & \dots & 0 \\ -H\Omega & I & 0 & \dots & 0 \\ -H\Phi\Omega & -H\Omega & I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -H\Phi^{N-1}\Omega & -H\Phi^{N-2}\Omega & -H\Phi^{N-3}\Omega & \dots & I \end{bmatrix} \\
\bar{\Gamma}_N &= \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ -H\Gamma & 0 & 0 & \dots & 0 \\ -H\Phi\Gamma & -H\Gamma & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -H\Phi^{N-1}\Gamma & -H\Phi^{N-2}\Gamma & -H\Phi^{N-3}\Gamma & \dots & -H\Gamma \end{bmatrix} \\
\Pi_{k,N} &= \begin{bmatrix} R_{k-N} & 0 & \dots & 0 \\ 0 & R_{k-N+1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & R_k \end{bmatrix}
\end{aligned}$$

The equations in (6) can be written in a more compact form given as

$$\bar{Z}_{k,N} = M_N x_{k-N} + E_{k,N} \tag{7}$$

where

$$\bar{Z}_{k,N} = T_N Z_{k,N} + \bar{\Gamma}_N U_{k-1,N-1} \tag{8}$$

Let $\bar{x}_{k-N}$ be the a priori prediction and Σ_{k-N} the associated covariance matrix, the least-squares criteria comes up as a natural choice for deriving the MHE. Therefore, at time instant k , our goal is to solve

$$\begin{aligned}
\hat{x}_{k-N|k} &= \arg \min_{x_{k-N}} \left\{ \|\bar{Z}_{k,N} - M_N x_{k-N}\|_{\Pi_{k,N}^{-1}}^2 \right. \\
&\quad \left. + \|x_{k-N} - \bar{x}_{k-N}\|_{\Sigma_{k-N}^{-1}}^2 \right\}
\end{aligned} \tag{9}$$

where $\hat{x}_{k-N|k}$ denotes the estimate of x_{k-N} at time k . Once $\hat{x}_{k-N|k}$ is obtained, the remaining estimates at time k , $\hat{x}_{k-N+1|k}, \dots, \hat{x}_{k|k}$, can be easily iterated from $\hat{x}_{k-N|k}$ by recursively applying (5), i.e. for $i = k-N, \dots, k-1$

$$\hat{x}_{i+1|k} = \Phi \hat{x}_{i|k} + \Gamma u_i + \Omega z_i \tag{10}$$

The estimated process output is then given by

$$\hat{z}_{k|k} = H \hat{x}_{k|k} \tag{11}$$

The second term on the right hand side of (9) is known as the arrival cost in estimation theory and different choices can be made for $\bar{x}_{k-N}$ and Σ_{k-N} . Let P_{ik} denote the error covariance matrix of $\hat{x}_{i|k}$. For example, one can choose the result of the corresponding estimate at the previous step, i.e. $\bar{x}_{k-N} = \hat{x}_{k-N|k-1}$ and $\Sigma_{k-N} = P_{k-N|k-1}$, as proposed in [28, 41]. To simplify the discussion, we adopt the choice in [29, 51] and assume that the prediction $\bar{x}_{k-N}$ and Σ_{k-N} are determined by propagating the estimated value at time $k-N-1$, i.e.

$$\bar{x}_{k-N} = \Phi \hat{x}_{k-N-1|k-N-1} + \Gamma u_{k-N-1} + \Omega z_{k-N-1} \tag{12}$$

$$\Sigma_{k-N} = \Phi P_{k-N-1|k-N-1} \Phi^T \tag{13}$$

In this case, the estimator in (9) is the standard least-squares MHE [51], and the state estimate $\hat{x}_{k|k}$ computed in (10) coincides with the well-known Kalman filter state estimate at time k . The estimation procedure is initialised with some a priori prediction $\bar{x}_0$ and Σ_0 .

3 Regularised MHE

In this section, we explicitly model the outliers by introducing auxiliary variables o_k as described in (2), and propose a robust version of the MHE in (9). The objective here is to estimate the output of the ARMAX process, z_k , given the outlier-contaminated measurements, y_k .

Let y_k and o_k denote the outlier-contaminated measurements and outlier vectors, respectively, then for the sliding window at time k we have

$$\begin{aligned}
Y_{k,N} &= [y_{k-N}^T, \dots, y_{k-1}^T, y_k^T]^T \\
O_{k,N} &= [o_{k-N}^T, \dots, o_{k-1}^T, o_k^T]^T
\end{aligned}$$

and according to (2)

$$Y_{k,N} = Z_{k,N} + O_{k,N} \tag{14}$$

Substituting (14) into (8) gives

$$\bar{Z}_{k,N} = T_N (Y_{k,N} - O_{k,N}) + \bar{\Gamma}_N U_{k-1,N-1} \tag{15}$$

Substituting (15) into (7) gives

$$\bar{Y}_{k,N} = M_N x_{k-N} + T_N O_{k,N} + E_{k,N} \tag{16}$$

where

$$\bar{Y}_{k,N} = \bar{Z}_{k,N} + T_N O_{k,N} \tag{17}$$

It can be seen that a data entry in $\bar{Y}_{k,N}$ is affected by not only the contemporaneous outlier but also a combination of past outliers with weights depending on the model from which the underlying process z_k is generated. Therefore, a single outlier may contaminate multiple data entries when it comes to estimating the process.

The challenge with introducing $O_{k,N}$ is that estimating x_{k-N} and $O_{k,N}$ based solely on $\bar{Y}_{k,N}$ would be an under-determined problem (fewer equations than unknowns), which generally has an infinite number of solutions. Notice that in practise the outlier vectors, o_k , are associated with occasional large errors caused by disturbances, instrument failures and so on. Hence most entries of $O_{k,N}$ are zero, and the key to addressing the problem of underdeterminacy is recovery of the degree of sparsity of the vector $O_{k,N}$.

For this reason, we consider introducing an appropriate regularisation term to $O_{k,N}$ to induce sparse solutions. It is well known that ℓ_0 -norm regularisation is able to quantify the degree of sparsity, however, the optimisation problem is non-convex and requires an intractable combinatorial search [52]. A natural approximation to ℓ_0 -norm regularisation is ℓ_1 -norm regularisation, which is convex and seems to be ubiquitous throughout many

fields of mathematics and engineering [53]. Good examples such as the much used lasso [54] and compressed sensing [55, 56] have demonstrated the effectiveness of such regularisation.

Substituting (17) into (9) and using an ℓ_1 -norm regularisation, the robust MHE amounts to solving

$$\begin{aligned} & [\hat{x}_{k-N|k}, \hat{O}_{k,N}] \\ &= \arg \min_{x_{k-N}, O_{k,N}} \left\{ \|\bar{Y}_{k,N} - M_N x_{k-N} - T_N O_{k,N}\|_{\Pi_{k,N}^{-1}}^2 \right. \\ & \quad \left. + \|x_{k-N} - \bar{x}_{k-N}\|_{\Sigma_{k-N}^{-1}}^2 + \lambda_k \|O_{k,N}\|_1 \right\} \end{aligned} \quad (18)$$

where $\lambda_k \geq 0$ is a tuning parameter that determines the degree of sparsity in $O_{k,N}$, and $\hat{O}_{k,N} = [\hat{O}_{k-N|k}^T, \dots, \hat{O}_{k-1|k}^T, \hat{O}_{k|k}^T]^T$ is the estimate of $O_{k,N}$ at time k .

Remark 1: The ℓ_1 -norm regularisation term $\|O_{k,N}\|_1$ in (18) enables recovery of outlying data entries. If outliers are present in all entries of $O_{k,N}$, then replacing it by $\|O_{k,N}\|_2$ would enable the estimator to identify whole data vectors as outliers instead of entry-wise outliers. The underlying optimisation problem for ℓ_2 -norm regularisation is still convex [57], and iterative solvers can likewise be developed using the same methodology presented in this paper; due to space limitations only ℓ_1 -norm regularisation is discussed hereinafter.

Remark 2: There is a well-known connexion between the ℓ_1 -regularisation approach and the Huber's M -estimation approach (see, e.g. [39, 40, 58]). In this particular application, the state estimates obtained in (18) is equivalent to the solution of the following Huber's M -estimation problem, when $\Pi_{k,N}$, T_N and Σ_{k-N} are all identity matrices, i.e.

$$\begin{aligned} \hat{x}_{k-N|k} &= \arg \min_{x_{k-N}} \left\{ \sum_{d=1}^D \rho_{\lambda_k}(\bar{Y}_{k,N,d} - M_{N,d}^T x_{k-N}) \right. \\ & \quad \left. + \sum_{d'=1}^{n \times m} (x_{k-N,d'} - \bar{x}_{k-N,d'})^2 \right\} \\ &= \arg \min_{x_{k-N}, O_{k,N}} \left\{ \|\bar{Y}_{k,N} - M_N x_{k-N} - O_{k,N}\|^2 \right. \\ & \quad \left. + \|x_{k-N} - \bar{x}_{k-N}\|^2 + \lambda_k \|O_{k,N}\|_1 \right\} \end{aligned} \quad (19)$$

where $\bar{Y}_{k,N,d}$ is the d th entry of $\bar{Y}_{k,N}$, $M_{N,d}^T$ the d th row of M_N , $x_{k-N,d'}$ the d' th entry of x_{k-N} and ρ_{λ_k} the Huber's cost [35] given by

$$\rho_{\lambda_k}(r) = \begin{cases} \frac{r^2}{2}, & |r| \leq \lambda_k \\ \lambda_k |r| - \frac{\lambda_k^2}{2}, & |r| > \lambda_k \end{cases} \quad (20)$$

The optimisation problem in (18) is actually more general than that in (19) by being able to incorporate the weighting matrix T_N , which essentially reveals how the outliers affect the data entries and results in a more reasonable penalisation on the outlier vectors. This distinguishes the proposed MHE from the conventional ones in which the outliers are not explicitly modelled and the estimators may break down due to contamination of multiple data entries caused by a single outlier.

3.1 Robust MHE via coordinate descent

The optimisation problem in (18) is jointly convex in x_{k-N} and $O_{k,N}$, and a method exploiting the problem structure and offering computational advantages is block coordinate descent [59]. In this method, the optimisation variables are partitioned into blocks and the objective function is minimised iteratively with respect to one block variable at a time: given $O_{k,N}$, the optimal x_{k-N} is the least-squares estimate from regression of $\bar{Y}_{k,N} - T_N O_{k,N}$ on design

matrix M_N ; given x_{k-N} , the optimal $O_{k,N}$ is the solution to the ℓ_1 -minimisation problem. This alternation between updates to x_{k-N} and $O_{k,N}$ closely matches the interplay between detecting outliers and estimating states.

Let j be the index of coordinate descent iterations, $x_{k-N}^{(j)}$ and $O_{k,N}^{(j)}$ are updated cyclically by solving the following sub-problems per iteration j :

$$\begin{aligned} x_{k-N}^{(j)} &= \arg \min_{x_{k-N}} \left\{ \|\bar{Y}_{k,N} - M_N x_{k-N} - T_N O_{k,N}^{(j-1)}\|_{\Pi_{k,N}^{-1}}^2 \right. \\ & \quad \left. + \|x_{k-N} - \bar{x}_{k-N}\|_{\Sigma_{k-N}^{-1}}^2 \right\} \end{aligned} \quad (21)$$

$$\begin{aligned} O_{k,N}^{(j)} &= \arg \min_{O_{k,N}} \left\{ \|\bar{Y}_{k,N} - M_N x_{k-N}^{(j)} - T_N O_{k,N}\|_{\Pi_{k,N}^{-1}}^2 \right. \\ & \quad \left. + \lambda_k \|O_{k,N}\|_1 \right\} \end{aligned} \quad (22)$$

Solving the sub-problem in (21) is equivalent to finding the standard least-squares MHE estimate in (9) with outlier-compensated measurements $\tilde{Z}_{k,N}^{(j-1)} = \bar{Y}_{k,N} - T_N O_{k,N}^{(j-1)}$ per iteration j .

For the sub-problem in (22), it is actually a generalised lasso problem that can be further partitioned into a sequence of one-dimensional (1D) coordinate descent iterations [60, 61], i.e.

$$\begin{aligned} O_{k,N,d}^{(j)} &= \arg \min_{O_{k,N,d}} \left\| \bar{Y}_{k,N} - M_N x_{k-N}^{(j)} - T_N \begin{bmatrix} O_{k,N,1:d-1}^{(j)} \\ O_{k,N,d} \\ O_{k,N,d+1:D}^{(j-1)} \end{bmatrix} \right\|_{\Pi_{k,N}^{-1}}^2 \\ & \quad + \lambda_k |O_{k,N,d}| \end{aligned} \quad (23)$$

where $O_{k,N,d}^{(j)}$ is the d th entry of $O_{k,N}^{(j)}$ and $D = (N+1) \times m$ is the dimension of $O_{k,N}$. To put things into perspective, it is helpful to express the 1D problem in (23) as an equivalent scalar lasso given by

$$O_{k,N,d}^{(j)} = \arg \min_{O_{k,N,d}} (O_{k,N,d} - \gamma_{k,d}^{(j)})^2 + \lambda_{k,d} |O_{k,N,d}| \quad (24)$$

where

$$\begin{aligned} \gamma_{k,d}^{(j)} &= \frac{1}{u_{k,d,d'}} \left[\alpha_{k,d}^{(j)} - \sum_{i=1}^{d-1} u_{k,d,i} O_{k,N,i}^{(j)} - \sum_{i=d+1}^D u_{k,d,i} O_{k,N,i}^{(j-1)} \right] \\ \alpha_k^{(j)} &= T_N^T \Pi_{k,N}^{-1} (\bar{Y}_{k,N} - M_N x_{k-N}^{(j)}) \\ U_k &= T_N^T \Pi_{k,N}^{-1} T_N, \quad \lambda_{k,d} = \frac{\lambda_k}{u_{k,d,d'}} \end{aligned}$$

and $u_{k,d,d'}$ is the (d, d') entry of U . The closed-form solution to scalar lasso problem exists [62], and for the problem in (24) the solution is calculated as

$$O_{k,N,d}^{(j)} = \max \left(|\gamma_{k,d}^{(j)}| - \frac{\lambda_{k,d}}{2}, 0 \right) \text{sign}(\gamma_{k,d}^{(j)}) \quad (25)$$

Remark 3: For any initial values of $O_{k,N}^{(0)}$, global convergence of the iterates in (21) and (22) to the solution of (18) is guaranteed according to the results in [63]. However, there is no analytical results available in the literature regarding the rate of convergence. Nevertheless, we can see that the large part of the total computational time for each iteration j will be spent in solving the least-squares sub-problem in (21) for which the complexity is linear in N .

For considerable reduction in the coordinate decent iterations, it was suggested that one should initialise with $O_{k,N}^{(0)}$ close to its global optimum values [39]. At time k , the best initial guess would be the estimate obtained at the previous time instant. Therefore,

one can start naturally with $o_k^{(0)}$ being zero and $o_{k-N}^{(0)}, \dots, o_{k-2}^{(0)}, o_{k-1}^{(0)}$ being the estimates obtained at time $k-1$. In practise, a very fast convergence with only a few iterations has been noticed. This is due to the efficiency of the coordinate descent when the subproblems have a simple closed-form solution and each coordinate minimisation can be done very quickly. Furthermore, often the minimisers for many of the estimates did not change (indeed, many entries of the estimated outlier vectors remained to be zero) as we cycle through the variables, and hence, the iterations can be very fast.

Remark 4: An improved approximation of the ℓ_0 -norm relative to that offered by the ℓ_1 -norm is provided by the weighted ℓ_1 -norm [52]. A major difference between the ℓ_1 -norm and ℓ_0 -norm regularisations is that larger coefficients are penalised more heavily than smaller coefficients in the ℓ_1 -norm, unlike the more equal penalisation in the ℓ_0 -norm. To address this imbalance, weighted ℓ_1 -norm regularisation was developed to more equally penalize non-zero coefficients. The weights used for the current iteration were chosen as inversely proportional to the magnitude of the solution obtained in the previous iteration, thereby decreasing the weights for large coefficients and increasing the weights for small coefficients. Building on this line of thought, consider surrogating (18) by

$$\begin{aligned} & [\hat{x}_{k-N}^{(d)}, \hat{O}_{k,N}^{(d)}] \\ &= \arg \min_{x_{k-N}, O_{k,N}} \left\{ \|\bar{Y}_{k,N} - M_N x_{k-N} - T_N O_{k,N}\|_{\Pi_{k,N}^{-1}}^2 \right. \\ & \quad \left. + \|x_{k-N} - \bar{x}_{k-N}\|_{\Sigma_{k-N}^{-1}}^2 + \lambda_k \|W_k^{(d)} O_{k,N}\|_1 \right\} \end{aligned} \quad (26)$$

where $W_k^{(d)} = \text{diag}\{w_{k,d}^{(d)}\}$ and for $d = 1, 2, \dots, D$

$$w_{k,d}^{(d)} = (|O_{k,N,d}^{(d-1)}| + \delta)^{-1} \quad (27)$$

The parameter δ is a small positive constant preventing the weight $w_{k,d}$ from tending to infinity. The cost in (26) is actually a weighted version of the cost in (18), where each of the outlier norms is weighted by $w_{k,d}$. When an element of $O_{k,N}$ is small, then this element is weighted more during the minimisation of the next iteration, and vice versa. To start the iteration, all weights can be set to one and (26) reduces to (18), whose solution is then used as the initial condition for the next iteration in (26). The simulation results in Section 4 will show that (26) can produce a better estimate than that of (18). Another attractive fact is that each iteration of this algorithm consists only in solving one ℓ_1 -regularised optimisation problem, for which the solver developed in this paper can readily be used.

One limitation of the proposed approach is that there is no analytic solution available for the ℓ_1 -regularised optimisation problem and hence the stability of the proposed MHE can hardly be analysed. Nevertheless, one can still enjoy the stability of the standard least-squares MHE by selecting an appropriate value for the tuning parameter λ_k , which will be introduced in the ensuing subsection.

3.2 Parameter selection and pathwise coordinate optimisation

A nice property of the proposed regularisation-based approach is that it only has one tuning parameter, λ_k , which controls the amount of regularisation and hence detection of outlier. It can be understood that increasing the value of λ_k causes $O_{k,N}$ to be weighted more in the overall objective function and encourages more zero components in $\hat{O}_{k,N}$. Indeed, all the entries of $\hat{O}_{k,N}$ will be identically zero when λ_k exceeds a certain critical value. In this case, the proposed method will reduce to the standard least-squares method. On the contrary, decreasing the value of λ_k can give rise to increasing numbers of non-zero components in $\hat{O}_{k,N}$, making the

proposed method more conservative than the method of least squares.

In robust batch linear regression problems, λ_k is usually assumed to be constant with respect to k , i.e. $\lambda_k = \lambda_c$ for all $k \geq 1$, and judicious tuning of λ_c has been investigated for this case (see, e.g. [39, 64]). However, these methods are not suitable for online optimisation. As a remedy, it was suggested to select λ_c during an initialisation phase and fix its value for all future time instants [40]. Obviously, this approach is restricted to the assumption that the statistical properties of the outliers and nominal noise would remain unchanged.

For the moving horizon estimation problem in this paper, the lasso-based solver developed in Section 3 allows one to simultaneously perform parameter selection and state estimation tasks. Recall that the procedure described in Section 3 is a successive coordinate-wise descent algorithm for minimising the objective function in (18) with a fixed λ_k . Without much difficulty, the coordinate-wise descent procedure can also be applied to obtain the solutions at a grid of λ values (*regularization path*), using each solution as the initial condition for the next λ value on the grid. The strategy here is that we first obtain a path of solutions before selecting one that is optimal with respect to some criteria. This procedure is repeated as the sliding window moves to the next time instant.

Before searching along the path, we first need to choose a starting point, $\lambda_k^{\max}$, such that the estimated outlier vector in (18) will be set exactly equal to zero when $\lambda_k \geq \lambda_k^{\max}$. This requires computing the standard least-squares MHE solution, i.e. solving

$$\begin{aligned} \hat{x}_{k-N}^{\text{mhe}} = \arg \min_{x_{k-N}} \left\{ \|\bar{Y}_{k,N} - M_N x_{k-N}\|_{\Pi_{k,N}^{-1}}^2 \right. \\ \left. + \|x_{k-N} - \bar{x}_{k-N}\|_{\Sigma_{k-N}^{-1}}^2 \right\} \end{aligned} \quad (28)$$

and the following proposition gives the critical value.

Proposition 1: The robust MHE estimate in (18) coincides with the standard least-squares MHE estimate $\hat{x}_{k-N}^{\text{mhe}}$ if and only if

$$\lambda_k \geq \lambda_k^{\max} = 2\|T_N^T \Pi_{k,N}^{-1} (\bar{Y}_{k,N} - M_N \hat{x}_{k-N}^{\text{mhe}})\|_{\infty} \quad (29)$$

The proof is given in the Appendix.

Remark 5: In the presence of impulsive outliers, the $\lambda_k^{\max}$ obtained in (29) may become unreasonably large, and it is inefficient to start the path from such a point. In this case one can use a set of outlier-free training data and run the standard least-squares MHE in (28) to obtain a surrogate defined as

$$\bar{\lambda}_k^{\max} = \max_{N \leq k \leq N_t} 2\|T_N^T \Pi_{k,N}^{-1} (\bar{Z}_{k,N} - M_N \hat{x}_{k-N}^{\text{mhe}})\|_{\infty} \quad (30)$$

where $\bar{Z}_{k,N}$ is the outlier-free measurements and N_t is the length of training data.

Once the starting point is determined at each time instant, it is clear that the ‘best’ λ_k should always locate in the interval $[0, \lambda_k^{\max}]$. Let $\{\lambda_k^i\}_{i=1}^I$ be a grid of λ values distributed over the interval $[0, \lambda_k^{\max}]$, and $[\hat{x}_{k-N}(\lambda_k^i), \hat{O}_{k,N}(\lambda_k^i)]$ the solution of (18) due to λ_k^i . If the estimate is accurate, then $\bar{Y}_{k,N} - M_N \hat{x}_{k-N}(\lambda_k^i) - T_N \hat{O}_{k,N}(\lambda_k^i)$ should have similar statistics as $E_{k,N}$. Therefore, the standardised sample variance should approach one, and the best λ_k is then chosen as

$$\lambda_k^* = \min_{1 \leq i \leq I} |1 - \hat{\sigma}_e(\lambda_k^i)| \quad (31)$$

where

$$\hat{\sigma}_e(\lambda_k^i) = \frac{\| \bar{Y}_{k,N} - M_N \hat{x}_{k-N}(\lambda_k^i) - T_N \hat{O}_{k,N}(\lambda_k^i) \|_{\Pi_{k,N}^{-1}}^2}{(N+1) \times m}$$

Finding λ_k^* in (31) seems to require solving I times the optimisation problem in (18). This is fortunately not the case for our problem, as efficient path algorithms for lasso-based problems exist.

The lasso solutions at a grid of λ values can be efficiently computed using the pathway coordinate descent algorithm [61, 62]. Specifically, let $\lambda_k^1 > \lambda_k^2 > \dots > \lambda_k^I$ be a grid of decreasing λ values, where $\lambda_k^1 = \lambda_k^{\max}$, and $\lambda_k^I = 0$. Starting at λ_k^1 and proceeding along the grid, the estimate obtained at the previous point, say λ_k^{i-1} , is used as the initial condition for the current point λ_k^i . Examples have shown that it can be more efficient to compute the entire path of solutions than to calculate the solution only at a single fixed λ value, especially for large problem dimensions; see [65]. The number of λ values, I , should be chosen large enough to ensure that a λ value in the vicinity of the global minimum of (31) can be obtained.

4 Examples

In this section, the performance of the proposed approach is tested by using both simulated and experimental data. For easy reference, the proposed MHEs with equal weight and refined weight are summarised in Algorithms 1 and 2, respectively. For real-time implementation, one can follow the steps described in the algorithms and repeat the procedure each time as the next measurement becomes available.

Algorithm 1 (Proposed MHE with equal weight):

- 1: Give the initial values x_0 and Σ_0 , and set $k = 0$.
- 2: Select a suitable value for λ_k according to subsection 3.2.
- 3: Solve the optimisation problem (18) by using the coordinate descent algorithm in Section 3.1.
- 4: Compute the estimated output $\hat{z}_{k|k}$ by using (10) and (11) with z_i being replaced by $y_i - \hat{o}_{i|k}$.
- 5: Compute the prediction $\bar{x}_{k-N}$ and Σ_{k-N} by using (12) and (13).
- 6: Set $k = k + 1$ and go back to step 2.

Algorithm 2 (Proposed MHE with refined weight):

- 1: Give the initial values x_0 and Σ_0 , and set $k = 0$.
- 2: Select a suitable value for λ_k according to Section 3.2.
- 3: **for** $l = 1$ to L **do**
- 4: Solve the optimisation problem (26) by using the coordinate descent algorithm in Section 3.1.
- 5: Update the weight $W_k^{(l+1)}$ by using (27).
- 6: **end for**
- 7: Compute the estimated output $\hat{z}_{k|k}$ by using (10) and (11) with z_i being replaced by $y_i - \hat{o}_{i|k}$.
- 8: Compute the prediction $\bar{x}_{k-N}$ and Σ_{k-N} by using (12) and (13).
- 9: Set $k = k + 1$ and go back to step 2.

4.1 Example 1: single outlier

Consider the ARMAX process in (1) where

$$A(q^{-1}) = 1 - 0.9q^{-1}, \quad B(q^{-1}) = 0.1q^{-1} \\ C(q^{-1}) = 1 - 0.8q^{-1}$$

and the measurement noise ε_k was $\mathcal{N}(0, 0.1)$. The process was initialised with $x_0 = 0$, $\Sigma_0 = 0.1I$ and a step input of size 1 (i.e. $u_k = 1$ for all $k \geq 1$). An outlier of size -10 was introduced at $k = 25$.

The proposed MHE was implemented according to Algorithm 1 with $N = 5$ and $\lambda_k = 5$ (approximately equals to $1/3$ of $\bar{\lambda}_k^{\max}$ in (30)). For the sake of comparison, the Kalman filter and Huber-penalty MHE in [66] were implemented. In the latter, the measurement rows in (28) were prewhitened, and the least-squares cost was replaced by Huber's cost in (20). The Huber's parameter was set to 1.345 for standardised Gaussian noise as suggested by [35]. Moreover, the MHE with one iteration of the refined weight in Algorithm 2 was also implemented.

Fig. 1 shows 1000 simulation runs each for the (a) Kalman filter, (b) Huber-penalty MHE, (c) proposed MHE with equal weight in Algorithm 1, and (d) proposed MHE with refined weight in Algorithm 2. The mean values of the 1000 runs are given by the white curves. First, observe that the estimates from the Kalman filter are severely biased by the presence of the outlier at $k = 25$, while the estimates from the proposed MHE are almost not. Second, the Huber-based scheme failed to give a robust estimate and exhibited similar performance to the Kalman filter. This is not surprising as a single outlier may give rise to multiple contaminated entries in $\bar{Y}_{k,N}$ thereby causing the breakdown of Huber's M-estimator. Third, a clear improvement is observed for the MHE with refined weight as a result of more equal penalisation on outlier vectors. The proposed MHE can effectively cope with outlier and outperforms the non-robust Kalman filter even when λ_k is not systematically tuned.

4.2 Example 2: multiple outliers

Consider the ARMAX process in (1) where

$$A(q^{-1}) = I + A_1 q^{-1} + A_2 q^{-2} \\ B(q^{-1}) = B_1 q^{-1} + B_2 q^{-2} \\ C(q^{-1}) = I + C_1 q^{-1} + C_2 q^{-2}$$

and

$$A_1 = \begin{bmatrix} 0 & 0.5 \\ 1 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1.2 & 0 \\ 0 & 0.5 \end{bmatrix} \\ B_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ C_1 = \begin{bmatrix} 1.2 & 0 \\ 0 & 0.6 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 0.36 & 0 \\ 0 & 0 \end{bmatrix}$$

Both $\{u_k\}$ and $\{\varepsilon_k\}$ are independent sequences given by $u_k \sim \mathcal{N}(0, 1)$ and $\varepsilon_k \sim \mathcal{N}(0, R_k)$ where $R_k = \text{diag}\{2, 1\}$. The proposed estimator in Algorithm 1 is designed with $N = 5$ and λ_k chosen according to the criteria in (31). The initial condition was set as $x_0 = 0$ and $\Sigma_0 = I$.

A total of 50 samples were generated in one simulation run and 1000 such simulation runs were conducted. To compare the performance of the estimators, the root mean square error (RMSE) of the estimates from the 1000 simulation runs is defined as follows:

$$\text{RMSE}(k) = \left(\frac{1}{1000} \sum_{r=1}^{1000} \| \hat{x}_{k|k}^{(r)} - x_k \|^2 \right)^{1/2}$$

where $\hat{x}_{k|k}^{(r)}$ is the estimated state at time k in the r th simulation run. The performance of the MHE with different types of outlier contamination was tested. More specifically, two contamination types were considered: two-point contamination which yields values -10 or 10 with equal probability of $\tau/2$, and contamination with a Gaussian distribution $\mathcal{N}(0, 10^2)$ with probability τ . The tested contamination fractions were $\tau = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5$ and 0.6 .

A snapshot of the estimates obtained in one simulation run for Gaussian contamination and $\tau = 0.1$ is presented in Fig. 2. It can be seen that the estimates (circles) from the Kalman filter are severely affected by the outliers but the estimates (triangles) from the

Fig. 1 Estimates from the

(a) Kalman filter, (b) Huber-penalty MHE, (c) MHE with equal weight in (18), and (d) MHE with refined weight in (26). The mean values of the 1000 runs for the estimate are given by the white curves

Fig. 2 True state, measures and estimates of x_1 and x_2 in a simulation run for Gaussian contamination and $\tau = 0.1$ using MHE and KF are given by multiplication symbol, addition symbol, triangle and open circle, respectively

proposed MHE are almost not. Fig. 3 depicts the boxplot of the RMSE versus percentage of outliers for the MHE and Kalman filter, respectively. As compared with the Kalman filter, the proposed MHE performs much better in terms of RMSE under both types of contamination.

4.3 Example 3: liquid-level estimation

Consider the liquid-level estimation problem studied in [67], where the liquid level of a coupled-tank control apparatus shown in Fig. 4 is to be estimated. The transfer function between the liquid level in Tank 1, z_k , and the control voltage, u_k , is given as

$$z_k = \frac{0.066}{1 - 0.99q^{-1}}u_k + e_k \quad (32)$$

where the measurement noise e_k is $\mathcal{N}(0, 0.1)$. The polynomials A , B , and C in the ARMAX model can be obtained by comparing (32) and (1) and is given by

$$A(q^{-1}) = C(q^{-1}) = 1 - 0.99q^{-1}, \\ B(q^{-1}) = 0.066q^{-1}$$

One thousand measurements of the liquid level y_k were collected as shown in Fig. 5 when the control voltage u_k was held constant at 2 V. To obtain statistical results, we estimated the liquid level z_k for 10 samples using the Kalman filter and the proposed MHE in Algorithm 1 with $N = 5$ and $\lambda_k = 1$. This was repeated 100 times and the results are shown in Figs. 6 and 7, respectively. For comparison purposes, we also tested three other different values of

Fig. 3 RMSEs over 1000 simulation runs for estimates formed by the proposed MHE and Kalman filter under two types of outlier contamination

Fig. 4 Photo of the coupled tank

λ_k : 1.5, 2, and ∞ . The variances of the estimation results at each time instant are summarised in Table 1. It can be seen from Table 1 that the variances from the proposed MHE are less than that from the Kalman filter in all cases except for $\lambda_k = \infty$, in which case the proposed MHE reduces to the well-known Kalman filter and hence gives the same variance. For $\lambda_k = 1$, Table 1 shows that the variance from the proposed MHE is about half of that from the Kalman filter. This example shows that the tuning parameter in the

proposed MHE can be chosen gainfully to reduce the impact of outliers and give a smaller variance.

5 Conclusion

In this paper, a robust MHE for ARMAX process with measurements affected by outliers was developed. Outliers were modelled as auxiliary variables and simultaneously estimated with the state using ℓ_1 -norm regularisation. Block coordinate descent-based iterations for the underlying convex optimisation problem were developed such that each coordinate sub-problem involved only simple closed-form updates per time step. A weighted ℓ_1 -regularisation scheme was proposed to achieve improved performance. Simulation and experimental tests demonstrated that the proposed MHE can effectively cope with outliers, even for the case when the commonly used Huber's M-estimator breaks down.

Sum-of-norms regularisation is able to induce group sparsity and has been applied to the detection and estimation of abrupt changes in state smoothing problems [50, 68]. Considering that when the outlier vector o_k is non-zero, all its components are typically non-zero (group sparsity), we can employ sum-of-norms regularisation in the concerned ARMAX process just like the application in the smoothing problems. Furthermore, the proposed approach is developed for handling measurement outliers, and new challenges may arise if uncertainties or disturbances occur in system state and model parameter. It would be interesting to study the robustness issues under such circumstances in our future work.

6 Acknowledgments

The authors acknowledge the help of Dr. Zhou Dexiang in providing the example of liquid level estimation. L. Yin and S. Liu

Fig. 5 Measurement y_k for the liquid-level estimation experiment

Fig. 6 Estimated liquid level $\hat{z}_{k|k}$ from the Kalman filter

Fig. 7 Estimated liquid level $\hat{z}_{k|k}$ from the proposed MHE in Algorithm 1

Table 1 $\text{Var } \hat{z}_{k|k} \times 10^{-3}$ from the proposed MHE and Kalman filter

k	MHE in Algorithm 1				Kalman filter
	$\lambda_k = 1$	$\lambda_k = 1.5$	$\lambda_k = 2$	$\lambda_k = \infty$	
5	0.88	1.03	1.12	1.67	1.67
6	0.95	1.08	1.16	1.69	1.69
7	0.88	1.05	1.13	1.78	1.78
8	0.76	0.91	1.02	1.59	1.59
9	0.66	0.81	0.91	1.41	1.41
10	0.61	0.72	0.80	1.18	1.18

were supported by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

7 References

- [1] Toshani, H., Farrokhi, M., Alipouri, Y.: 'Constrained generalised minimum variance controller design using projection-based recurrent neural network', *IET Control Theory Appl.*, 2016, **11**, (2), pp. 143–154
- [2] Åström, K.J., Wittenmark, B.: 'Computer-controlled systems: theory and design' (Prentice-Hall, Upper Saddle River, New Jersey, USA, 1997)
- [3] Goodwin, G.C., Sin, K.S.: 'Adaptive filtering prediction and control' (Prentice-Hall, Englewood Cliffs, New Jersey, 1984)
- [4] Jørgensen, J.B., Huusom, J.K., Rawlings, J.B.: 'Finite horizon MPC for systems in innovation form'. IEEE 2011 50th IEEE Conf. Decision and Control and European Control Conf., 2011, pp. 1896–1903
- [5] Hannan, E.J., Deistler, M.: 'The statistical theory of linear systems', vol. 70 (John Wiley & Sons, Inc., New York, 1988)
- [6] Ljung, L.: 'System identification: theory for the user' (Prentice Hall, Englewood Cliffs, New Jersey, 1987)
- [7] Söderström, T., Stoica, P.: 'System identification' (Prentice Hall, Upper Saddle River, New Jersey, 1989)
- [8] Diversi, R., Guidorzi, R., Soverini, U.: 'Identification of ARMAX models with additive output noise', *IFAC Proc. Vol.*, 2009, **42**, (10), pp. 1574–1579
- [9] Diversi, R., Guidorzi, R., Soverini, U.: 'Identification of ARMAX models with noisy input and output', *IFAC Proc. Vol.*, 2011, **44**, (1), pp. 13121–13126
- [10] Li, S., Xiang, Z., Lin, H., et al.: 'State estimation on positive markovian jump systems with time-varying delay and uncertain transition probabilities', *Inf. Sci.*, 2016, **369**, pp. 251–266
- [11] Zhao, X., Zhang, L., Shi, P., et al.: 'Robust control of continuous-time systems with state-dependent uncertainties and its application to electronic circuits', *IEEE Trans. Ind. Electron.*, 2014, **61**, (8), pp. 4161–4170
- [12] Arefi, M.M., Jahed-Motlagh, M.R., Karimi, H.R.: 'Adaptive neural stabilizing controller for a class of mismatched uncertain nonlinear systems by state and output feedback', *IEEE Trans. Cybern.*, 2015, **45**, (8), pp. 1587–1596
- [13] You, J., Yin, S., Karimi, H.R.: 'Robust estimation for discrete markov system with time-varying delay and missing measurements', *Math. Probl. Eng.*, 2013, **2013**
- [14] Guo, P., Zhang, J., Karimi, H.R., et al.: 'State estimation for wireless network control system with stochastic uncertainty and time delay based on sliding mode observer', *Abs. Appl. Anal.*, 2014, **2014**
- [15] Duan, Z., Xiang, Z., Karimi, H.R.: 'Robust stabilisation of 2D state-delayed stochastic systems with randomly occurring uncertainties and nonlinearities', *Int. J. Syst. Sci.*, 2014, **45**, (7), pp. 1402–1415
- [16] Wang, Z., Burnham, K.J.: 'Robust filtering for a class of stochastic uncertain nonlinear time-delay systems via exponential state estimation', *IEEE Trans. Signal Process.*, 2001, **49**, (4), pp. 794–804
- [17] Guidorzi, R., Diversi, R., Vincenzi, L., et al.: 'AR+ noise versus AR and ARMA models in SHM-oriented identification'. IEEE 2015 23th Mediterranean Conf. Control and Automation (MED), 2015, pp. 809–814
- [18] Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: 'Time series analysis: forecasting and control', vol. 734 (John Wiley & Sons, Hoboken, New Jersey, 2008)
- [19] Brockwell, P.J., Davis, R.A.: 'Introduction to time series and forecasting' (Springer Science & Business Media, New York, 2002)
- [20] Anderson, B.D.O., Moore, J.B.: 'Optimal filtering' (Prentice-Hall, Englewood Cliffs, New Jersey, 1979)
- [21] Mendel, J.M.: 'Lessons in estimation theory for signal processing, communications, and control' (Prentice Hall, Upper Saddle River, New Jersey, 1995)
- [22] Sorenson, H.W.: 'Least-squares estimation: from Gauss to Kalman', *IEEE Spectr.*, 1970, **7**, (7), pp. 63–68
- [23] Duncan, D.B., Horn, S.D.: 'Linear dynamic recursive estimation from the viewpoint of regression analysis', *J. Am. Stat. Assoc.*, 1972, **67**, (340), pp. 815–821
- [24] Bryson, A.E.: 'Applied optimal control: optimization, estimation and control' (Hemisphere, New York, 1975)
- [25] Tukey, J.W.: 'A survey of sampling from contaminated distributions', *Contrib. Probab. Stat.*, 1960, **2**, pp. 448–485
- [26] Huber, P.J.: 'Robust estimation of a location parameter', *Ann. Math. Stat.*, 1964, **35**, (1), pp. 73–101
- [27] Schick, I.C., Mitter, S.K.: 'Robust recursive estimation in the presence of heavy-tailed observation noise', in George, E.I., Hsing, T. (Eds.): 'The annals of statistics' (1994), pp. 1045–1080
- [28] Alessandri, A., Baglietto, M., Battistelli, G.: 'Receding-horizon estimation for discrete-time linear systems', *IEEE Trans. Autom. Control*, 2003, **48**, (3), pp. 473–478
- [29] Rao, C.V., Rawlings, J.B., Lee, J.H.: 'Constrained linear state estimation—a moving horizon approach', *Automatica*, 2001, **37**, (10), pp. 1619–1628
- [30] Boulkroune, B., Darouach, M., Zasadzinski, M.: 'Moving horizon state estimation for linear discrete-time singular systems', *IET Control Theory Appl.*, 2010, **4**, (3), pp. 339–350
- [31] Gao, C., Zhao, G., Lu, J., et al.: 'Decentralised moving-horizon state estimation for a class of networked spatial-navigation systems with random parametric uncertainties and communication link failures', *IET Control Theory Appl.*, 2015, **9**, (18), pp. 2666–2677

- [32] Alessandri, A., Baglietto, M., Battistelli, G.: 'Moving-horizon state estimation for nonlinear discrete-time systems: new stability results and approximation schemes', *Automatica*, 2008, **44**, (7), pp. 1753–1765
- [33] Rao, C.V., Rawlings, J.B., Mayne, D.Q.: 'Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations', *IEEE Trans. Signal Process.*, 2003, **48**, (2), pp. 246–258
- [34] Guo, Y., Huang, B.: 'Moving horizon estimation for switching nonlinear systems', *Automatica*, 2013, **49**, (11), pp. 3270–3281
- [35] Huber, P.J.: 'Robust statistics' (John Wiley & Sons, Inc., Hoboken, New Jersey, 2009)
- [36] Gandhi, M.A., Mili, L.: 'Robust Kalman filter based on a generalized maximum-likelihood-type estimator', *IEEE Trans. Signal Process.*, 2010, **58**, (5), pp. 2509–2520
- [37] Durovic, Z.M., Kovacevic, B.D.: 'Robust estimation with unknown noise statistics', *IEEE Trans. Autom. Control*, 1999, **44**, (6), pp. 1292–1296
- [38] Kovacevic, B., Durovic, Z., Glavaški, S.: 'On robust Kalman filtering', *Int. J. Control*, 1992, **56**, (3), pp. 547–562
- [39] Farahmand, S., Giannakis, G.B., Angelosante, D.: 'Doubly robust smoothing of dynamical processes via outlier sparsity constraints', *IEEE Trans. Signal Process.*, 2011, **59**, (10), pp. 4529–4543
- [40] Farahmand, S., Giannakis, G.B.: 'Robust RLS in the presence of correlated noise using outlier sparsity', *IEEE Trans. Signal Process.*, 2012, **60**, (6), pp. 3308–3313
- [41] Alessandri, A., Awawdeh, M.: 'Moving-horizon estimation with guaranteed robustness for discrete-time linear systems and measurements subject to outliers', *Automatica*, 2016, **67**, pp. 85–93
- [42] Rousseeuw, P.J., Leroy, A.M.: 'Robust regression and outlier detection' (John Wiley & Sons Inc., New York, 1987)
- [43] Eldar, Y.C., Mishali, M.: 'Robust recovery of signals from a structured union of subspaces', *IEEE Trans. Inf. Theory*, 2009, **55**, (11), pp. 5302–5316
- [44] Eldar, Y.C., Kuppinger, P., Bolcskei, H.: 'Block-sparse signals: uncertainty relations and efficient recovery', *IEEE Trans. Signal Process.*, 2010, **58**, (6), pp. 3042–3054
- [45] Candes, E.J., Tao, T.: 'Decoding by linear programming', *IEEE Trans. Inf. Theory*, 2005, **51**, (12), pp. 4203–4215
- [46] Candes, E.J., Randall, P.A.: 'Highly robust error correction by convex programming', *IEEE Trans. Inf. Theory*, 2008, **54**, (7), pp. 2829–2840
- [47] Fawzi, H., Tabuada, P., Diggavi, S.: 'Secure estimation and control for cyber-physical systems under adversarial attacks', *IEEE Trans. Autom. Control*, 2014, **59**, (6), pp. 1454–1467
- [48] Pajic, M., Weimer, J., Bezzo, N., et al.: 'Robustness of attack-resilient state estimators'. 5th ACM/IEEE Int. Conf. Cyber-Physical Systems (ICCPs), 2014, pp. 163–174
- [49] Shoukry, Y., Nuzzo, P., Puggelli, A., et al.: 'Secure state estimation for cyber physical systems under sensor attacks: a satisfiability modulo theory approach', *IEEE Trans. Autom. Control*, 2017, **62**, (10), pp. 4917–4932
- [50] Ohlsson, H., Gustafsson, F., Ljung, L., et al.: 'Smoothed state estimates under abrupt changes using sum-of-norms regularization', *Automatica*, 2012, **48**, (4), pp. 595–605
- [51] Muske, K.R., Rawlings, J.B., Lee, J.H.: 'Receding horizon recursive state estimation'. IEEE 1993 American Control Conf., 1993, pp. 900–904
- [52] Candes, E.J., Wakin, M.B., Boyd, S.P.: 'Enhancing sparsity by reweighted ℓ_1 minimization', *J. Fourier Anal. Appl.*, 2008, **14**, (5–6), pp. 877–905
- [53] Tibshirani, R.J., Taylor, J.: 'The solution path of the generalized lasso', *Ann. Stat.*, 2011, **39**, (3), pp. 1335–1371
- [54] Tibshirani, R.: 'Regression shrinkage and selection via the lasso', *J. R. Stat. Soc. B, Methodol.*, 1996, **58**, (1), pp. 267–288
- [55] Candès, E.J., Romberg, J., Tao, T.: 'Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information', *IEEE Trans. Inf. Theory*, 2006, **52**, (2), pp. 489–509
- [56] Donoho, D.L.: 'Compressed sensing', *IEEE Trans. Inf. Theory*, 2006, **52**, (4), pp. 1289–1306
- [57] Yuan, M., Lin, Y.: 'Model selection and estimation in regression with grouped variables', *J. R. Stat. Soc. B, Stat. Methodol.*, 2006, **68**, (1), pp. 49–67
- [58] Fuchs, J.-J.: 'An inverse problem approach to robust regression'. Proc. 1999 IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. **4**, 1999, pp. 1809–1812
- [59] Kekatos, V., Giannakis, G.B.: 'From sparse signals to sparse residuals for robust sensing', *IEEE Trans. Signal Process.*, 2011, **59**, (7), pp. 3355–3368
- [60] Li, Y., Osher, S.: 'Coordinate descent optimization for ℓ_1 minimization with application to compressed sensing: a greedy algorithm', *Inverse Probl. Imag.*, 2009, **3**, (3), pp. 487–503
- [61] Friedman, J., Hastie, T., Höfling, H., et al.: 'Pathwise coordinate optimization', *Ann. Appl. Stat.*, 2007, **1**, (2), pp. 302–332
- [62] Hastie, T., Tibshirani, R., Friedman, J.H.: 'The elements of statistical learning: data mining, inference, and prediction' (Springer Science & Business Media, New York, 2009)
- [63] Tseng, P.: 'Convergence of a block coordinate descent method for nondifferentiable minimization', *J. Optim. Theory Appl.*, 2001, **109**, (3), pp. 475–494
- [64] Giannakis, G.B., Mateos, G., Farahmand, S., et al.: 'USPACOR: universal sparsity-controlling outlier rejection'. 2011 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 1952–1955
- [65] Friedman, J., Hastie, T., Tibshirani, R.: 'Regularization paths for generalized linear models via coordinate descent', *J. Stat. Softw.*, 2010, **33**, (1), p. 1
- [66] Geebelen, K., Wagner, A., Gros, S., et al.: 'Moving horizon estimation with a huber penalty function for robust pose estimation of tethered airplanes'. IEEE 2013 American Control Conf. (ACC), 2013, pp. 6169–6174
- [67] Zhou, D.: 'Model predictive control with tracking error bound and an influence function approach to moving horizon estimation'. PhD thesis, Nanyang Technology University, 2017
- [68] Ohlsson, H., Gustafsson, F., Ljung, L., et al.: 'State smoothing by sum-of-norms regularization'. IEEE 2010 49th IEEE Conf. Decision and Control (CDC), 2010, pp. 2880–2885

8 Appendix

Let $J_k(x_{k-N}, O_{k,N})$ denote the cost in (18) and $\check{\nabla}$ the sub-gradient operator, the optimality condition for the non-differentiable problem in (18) is given by

$$\begin{aligned} 0 &\in \check{\nabla}_{O_{k,N}} J_k(\hat{x}_{k-N}, \hat{O}_{k,N}) \Rightarrow \\ 0 &\in \lambda \check{O}_{k,N} - 2T_N^T \Pi_{k,N}^{-1} (\bar{Y}_{k,N} - M_N \hat{x}_{k-N} - T_N \hat{O}_{k,N}) \end{aligned} \quad (33)$$

where $\check{O}_{k,N} = [\check{O}_{k,N,1}, \check{O}_{k,N,2}, \dots, \check{O}_{k,N,D}]^T$ is the subgradient of $\|O_{k,N}\|_1$ and for $d = 1, 2, \dots, D$

$$\check{O}_{k,N,d} = \begin{cases} \text{sign}(\hat{O}_{k,N,d}), & \hat{O}_{k,N,d} \neq 0 \\ s_{N,d}, & \hat{O}_{k,N,d} = 0, \quad |s_{N,d}| \leq 1. \end{cases}$$

The proposed robust MHE coincides with the standard least-squares one when $\hat{O}_{k,N} = \mathbf{0}_D$ which implies $\hat{x}_{k-N} = \hat{x}_{k-N}^{\text{mhe}}$. For $\hat{O}_{k,N} = \mathbf{0}_D$, (33) is satisfied with $\hat{x}_{k-N} = \hat{x}_{k-N}^{\text{mhe}}$ if and only if (29) holds.