Outlier detection and trimmed-average estimation in network systems

Ujjwal Pratap, Carlos Canudas de Wit, Federica Garin

Outlier detection and trimmed-average estimation in network systems

Ujjwal Pratap[a], Carlos Canudas-de-Wit[b], Federica Garin[a,*]

[a] *Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, Grenoble, France*
[b] *CNRS, GIPSA-Lab, Grenoble, France.*

Abstract

This paper addresses the problem of outlier detection and trimmed-average state estimation in an LTI network system. We consider that only some states are measured and there exists an outlier among the unmeasured states, which is so different from the remaining states that it affects the average value significantly. The goal of this paper is both to detect the outlier and to estimate the average state excluding the outlier (trimmed-average). Moreover, we also investigate the case where the system matrices are partially unknown since the outlier results from an unknown localized fault in the system. Finally, we illustrate the method on a thermal diffusion system.

Keywords: Linear time-invariant systems, Network systems, Average estimation, Reduced-order observers, Outlier detection, Trimmed average

1. Introduction

In recent years, there has been a lot of interest in the problem of estimation in network control systems due to its vast area of application such as power networks, communication networks, and traffic networks to name a few. In particular, in some networks, instead of reconstructing the entire states, some aggregation of the unmeasured states are estimated using a few available measurements obtained from the dedicated sensors placed at some positions. For instance, [1] investigates the problem of average state estimation of the unmeasured nodes in a network system and [2] and [3] propose design of some average state observers. This approach is very beneficial in terms of reducing complexity. However, if the unmeasured part has an outlier, which could be a result of an error or anomaly, the average value so estimated may not lead to the true average value. Instead, one might look for an average value that excludes the outlier. For example, in a power distribution network, one might be interested in estimating the average household consumption of an area using dedicated sensors deployed in some households. In case, there is a major consumer (an outlier) in the area, whose consumption is not measured directly, one might not obtain the desired value through average estimation since its high power consumption can affect the average significantly. Therefore, it is natural to ask some questions: what if there is an outlier in the unmeasured section of the network? How to apply an estimation technique such that it filters the outlier and detects it simultaneously?

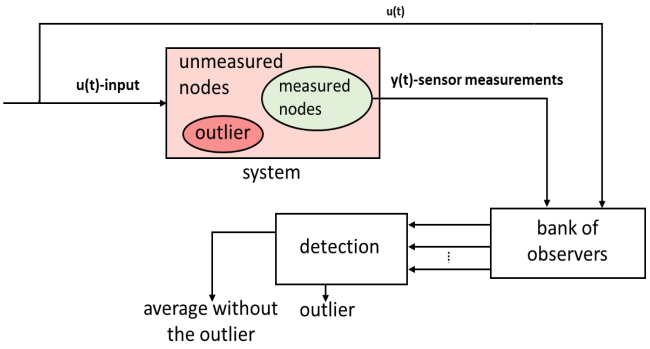

Figure 1: Scenario considered in the paper: An LTI system with dedicated sensors at some positions and an outlier in the unmeasured part.

Outlier detection and analysis is a very well studied problem in statistics and data mining. Some of the classical works are [4, 5]. The former gives a formal accepted definition of the outlier and the latter proposes different regression-based detection methods. Since then, work in this field has flourished. A survey of the state-of-the-art is given in the book [6]. In particular, some of the methods include depth-based methods [7], distance-based methods [8] and k-nearest neighbour methods [9]. However, these techniques apply to available static data points only. In this work, we consider a network system with dynamics and moreover the aim is to detect the outlier present among the unavailable measurements.

In network systems also, there has been focus on out-

*Corresponding author
Email addresses: ujjwal.pratap@gipsa-lab.fr (Ujjwal Pratap), carlos.canudas-de-wit@gipsa-lab.fr (Carlos Canudas-de-Wit), federica.garin@inria.fr (Federica Garin)

lier detection in sensor networks with outliers in the sensed data. The paper [10] is a survey on different techniques used for outlier detection in wireless sensor networks. For example, [11] addresses unsupervised detection in wireless sensor networks which accommodates different unsupervised techniques, [12] focuses on neighbor based detection methods and [13] proposes a robust Kalman filter to detect and exclude the outlier from the sensor measurements which is somehow similar to our goal. However, the main difference is that in this literature, outliers are among sensor measurements, arising due to noise, error, sensor limitations, disturbances etc., while in our case the outlier is in the system itself, and is an unmeasured state which is so different from the other states that it affects the average value significantly.

The scenario in this paper considers a continuous LTI system with dedicated sensors at some positions. There exists an outlier in the set of unmeasured nodes. We propose a centralised method to detect the outlier and estimate the average, excluding it simultaneously. Our approach is to run a bank of observers and compare the estimates so obtained in order to detect the outlier as illustrated in Figure 1. For this, at first, we provide a necessary and sufficient condition under which a bank of scalar and tunable observers can be designed to estimate the average of the unmeasured nodes excluding an element at every possible position. Then, we define a distance-based dissimilarity criterion to differentiate between the average estimates so obtained. Using a simple optimization, we obtain an estimate of the outlier position and of the average excluding it.

A preliminary version of this work was presented in [14]. The current paper significantly extends the results in [14], with two new sections. First, we investigate the case where the system matrices are only partially known, since the outlier results from a fault, and the only available knowledge is the system without the fault. We consider a class of localized faults that result in a single outlier, for which we can extend our method. Then, we illustrate the method on a thermal diffusion system.

2. Problem formulation

Consider a network represented by a weighted directed graph $(G) = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, 3 \dots n\}$ denotes the set of the nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes the set of edges. We follow the convention that the edge $(i, j) \in \mathcal{E}$ is represented as $i \leftarrow j$, since this edge will correspond to the influence of state x_j on the dynamics of state x_i. Let $A = [a_{ij}]$ be the associated weighted adjacency matrix, where a_{ij} is the weight of the edge $(i, j) \in \mathcal{E}$.

The dynamics of the network is described by

$$\begin{cases} \dot{x}(t) = & Ax(t) + Bu(t) \\ y(t) = & Cx(t)\,, \end{cases} \tag{1}$$

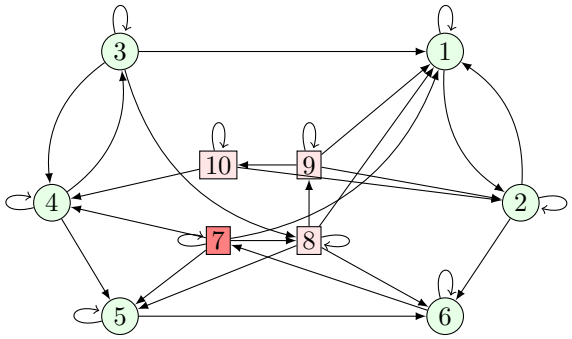

Figure 2: A network system with an outlier at node 7. The nodes in green circles are the measured nodes and the nodes in red squares are the unmeasured nodes with the outlier in a darker red shade.

where $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^{n_1}$ are the state vector, the input vector and the output vector respectively. We assume that the input is bounded and the system is BIBS stable. Therefore, we have bounded input and bounded state trajectories.

We assume to have dedicated sensor measurements at n_1 nodes, i.e., the output y contains the values of n_1 states. Without loss of generality, we order the states starting with the measured ones, so that we have the state partition $x(t) = [x_1^T(t), x_2^T(t)]^T$, where the vector $y = x_1(t) \in \mathbb{R}^{n_1}$ contains the measured states and $x_2(t) \in \mathbb{R}^{n_2}$ contains the unmeasured states. We assume that $n_2 > 1$. Denoting by $I_s \in \mathbb{R}^{s \times s}$ the identity matrix of size s, and by $\mathbf{0}_{s,r} \in \mathbb{R}^{s \times r}$ the zero matrix of size $s \times r$, the block structure of the matrices corresponding to the above-mentioned partition is

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \text{ and } C = \begin{bmatrix} I_{n_1} & \mathbf{0}_{n_1,n_2} \end{bmatrix}. \tag{2}$$

With this partition, the system can be rewritten as

$$\begin{cases} \dot{x}_1(t) &= A_{11}x_1(t) + A_{12}x_2(t) + B_1u(t) \\ \dot{x}_2(t) &= A_{21}x_1(t) + A_{22}x_2(t) + B_2u(t) \\ y(t) &= x_1(t). \end{cases} \tag{3}$$

Having given the system description, we define the outlier in consideration as follows.

Definition 1. *A state is called an outlier if its value differs from all the other states by such a large value that the average value changes significantly.*

In the next subsection we present an example to motivate the reader towards the problem. The example shows how an outlier may affect the average value. Then we use the same example in subsequent sections to illustrate the results and the method for outlier detection proposed in the paper.

2.1. Motivating Example

Example 1. *Consider the network depicted in Figure 2. The dynamics of the network is described as in (1). The*

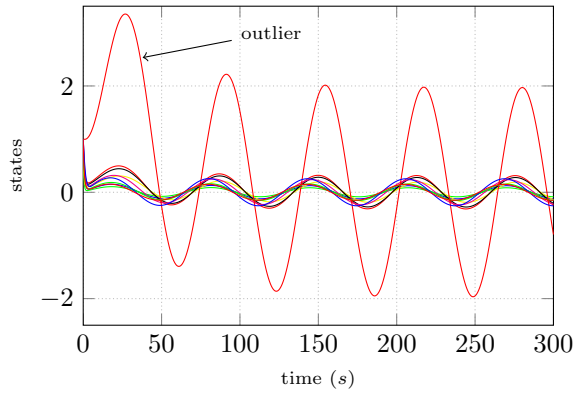

Figure 3: State trajectories of the network in Figure 1 in response to $u(t) = 0.2 * \sin(0.1t)$. Here, it can be seen that the outlier state (in red) is behaving very differently than the other states.

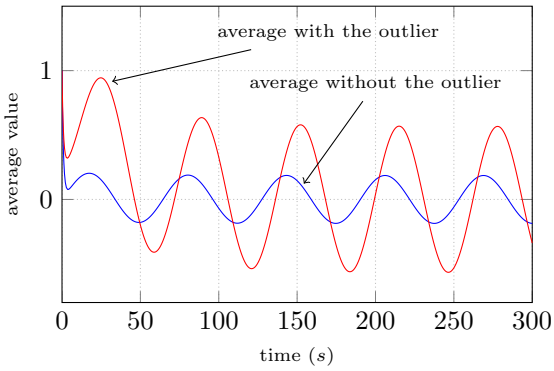

Figure 4: Trajectories of the average of the unmeasured states with and without the outlier, demonstrating the effect of an outlier on the average.

input is given by $u(t) = 0.2 * \sin(0.1t)$, $x(0) = \mathbf{1}_{10}$, *where* $\mathbf{1}_s \in \mathbb{R}^s$ *denotes the vector of all ones. The corresponding system matrices* A, B, C *according to the partition in* (2) *are*

$$A_{11} = \begin{bmatrix} -3.25 & 0.98 & 0.84 & 0 & 0 & 0 \\ 0.61 & -5.33 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3.53 & 0.48 & 0 & 0 \\ 0 & 0 & 0.25 & -3.05 & 0 & 0 \\ 0 & 0 & 0 & 0.13 & -1.69 & 0 \\ 0 & 0.54 & 0 & 0 & 0.85 & -2.18 \end{bmatrix}$$

$$A_{12} = \begin{bmatrix} 0.35 & 0.35 & 0.35 & 0 \\ 0 & 0 & 0.76 & 0.76 \\ 0 & 0.22 & 0 & 0 \\ 0.23 & 0 & 0 & 0.23 \\ 0.16 & 0.16 & 0 & 0 \\ 0 & 0.35 & 0 & 0 \end{bmatrix},$$

$$A_{21} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0.02 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$A_{22} = \begin{bmatrix} -0.03 & 0 & 0 & 0 \\ 0.02 & -1.4 & 0 & 0 \\ 0 & 0.64 & -1.16 & 0 \\ 0 & 0 & 0.4 & -2.0 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T \text{ and}$$

$$C = \begin{bmatrix} I_6 & \mathbf{0}_{6,4} \end{bmatrix}.$$

Here, in this example, we have a network of $n = 10$ *nodes depicted in Figure 2. The sensor measurements are obtained from the nodes* $\{1, \ldots, 6\}$ *denoted by the circle nodes and the nodes* $\{7, \ldots, 10\}$ *are the unmeasured nodes denoted by the square nodes. Here,* $n_1 = 6$ *and* $n_2 = 4$. *We have an outlier at node* 7 *pointed as the shaded node. It can be seen in Figure 4, where the averages of the unmeasured states with and without the outlier have been depicted, that there is a significant difference in the average value because of the outlier. Hence, we must find ways to estimate the average in such a way that it excludes the outlier.*

2.2. Problem statement

Consider a network with the dynamics as (1) with the assumption that there exists an outlier in the unmeasured part of the system. Using only the dedicated sensor measurements $y(t) = x_1(t)$, how is it possible to reconstruct the average of the unmeasured nodes $x_2(t)$ without the outlier? In addition, is it possible to design a scalar observer to estimate such an average with an arbitrary rate of convergence? Moreover, if the position of the outlier is unknown, how is it possible to detect it and compute the average excluding it simultaneously? Finally, can the same result be achieved in the case where the system matrices are only partially known, since the outlier results from a fault, and the only available knowledge is the system without the fault?

3. Outlier at a known position

In this section, we consider the case when the position of the existing outlier is known. Let us define j_o as the true position of the outlier in the set of unmeasured nodes. The position j_o means that the outlier is the node $n_1 + j_o$.

At first, we recall the necessary and sufficient condition to design a scalar and tunable observer to estimate the average of all but one element j in the set of unmeasured nodes. Then, we provide an explicit construction of the observer. In the end, we illustrate the estimation using Example 1.

Let $c_j \in \mathbb{R}^{n_2}$ be a vector of all ones but zero at the jth position. Let us define $q_j = \frac{1}{n_2 - 1} c_j$, so that the average

state of the unmeasured nodes excluding the element at the jth position be denoted as

$$x^{\text{av}}_{2,j}(t) = q_j^T x_2(t), \tag{4}$$

where $x_2(t)$ is defined in (2).

To reconstruct the average $x^{\text{av}}_{2,j}(t)$, we consider a scalar observer, namely a system of the form

$$\begin{cases} \dot{w}_j(t) &= -\alpha w_j(t) + h_j^T y(t) + g_j^T u(t) \\ \hat{x}^{\text{av}}_{2,j}(t) &= w_j(t) + \ell_j^T y(t), \end{cases} \tag{5}$$

where $w_j(t) \in \mathbb{R}$ is the state of the observer, while $\alpha \in \mathbb{R}$, $\ell_j, h_j \in \mathbb{R}^{n_1}$ and $g_j \in \mathbb{R}^m$ will be suitably designed.

Let $\xi_j(t) = x^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,j}(t)$ be the estimation error. We say that (5) is an observer if the parameters α, h_j, g_j and ℓ_j can be chosen such that $\xi_j(t) \to 0$ as $t \to \infty$. Moreover, we are interested in designing an observer such that the error $\xi_j(t) \to 0$ as $t \to \infty$ with a desired rate of convergence. The condition under which this is possible has been studied in [3].

Theorem 1. *[3, Theorem IV.1] A scalar and tunable observer of the form (5) to estimate $x^{\text{av}}_{2,j}(t)$ for a given j exists if and only if*

$$\text{rank} \begin{bmatrix} A_{12} \\ q_j^T A_{22} \\ q_j^T \end{bmatrix} = \text{rank}[A_{12}]. \tag{6}$$

Proof. For proof see [3]. □

Now, assuming that the condition (6) holds, we give the explicit design of the observer:

Choose an arbitrary $\alpha > 0 \in \mathbb{R}$ and compute the parameters ℓ_j^T, g_j^T and h_j^T as

$$\ell_j^T = q_j^T (A_{22} + \alpha I_{n_2}) A_{12}^\dagger, \tag{7a}$$

$$h_j^T = -\ell_j^T (\alpha I_{n_1} + A_{11}) + q_j^T A_{21}, \tag{7b}$$

$$g_j^T = q_j^T B_2 - \ell_j^T B_1, \tag{7c}$$

where $A_{12}^\dagger$ is the Moore-Penrose pseudo-inverse of A_{12}.

In what follows, we show that with the above choice of parameters, the observer (5) has the error dynamics $\dot{\xi}_j(t) = -\alpha \xi_j(t)$. For that, let us consider the error dynamics $\dot{\xi}_j(t)$. From (3), (4) and (5) we have

$$\begin{aligned} \dot{\xi}_j(t) = -\alpha \xi_j(t) &+ (-\alpha \ell_j^T + q_j^T A_{21} - \ell_j^T A_{11} - h_j^T) x_1(t) \\ &+ (q_j^T A_{22} - \ell_j^T A_{12} + \alpha q_j^T) x_2(t) \\ &+ (q_j^T B_2 - g_j^T - \ell_j^T B_1) u(t). \end{aligned} \tag{8}$$

Assuming that condition (6) holds, and with ℓ_j, h_j and g_j as in (7), here, we will show that the following conditions are satisfied:

$$q_j^T A_{21} - \alpha \ell_j^T - \ell_j^T A_{11} - h_j^T = 0, \tag{9a}$$

$$\alpha q_j^T + q_j^T A_{22} - \ell_j^T A_{12} = 0, \text{ and} \tag{9b}$$

$$q_j^T B_2 - g_j^T - \ell_j^T B_1 = 0 \tag{9c}$$

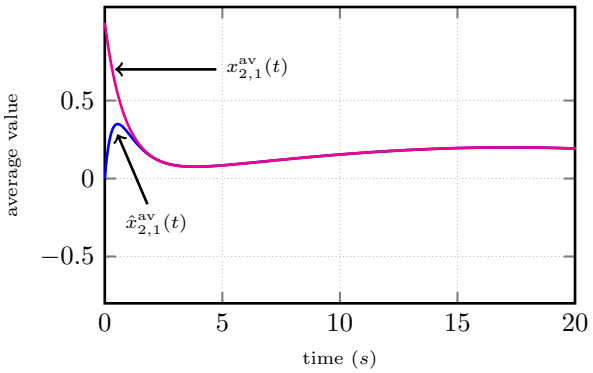

Figure 5: For the network system in Example 1, where the outlier is node 7 (i.e., $j_o = 1$): estimated ($\hat{x}^{\text{av}}_{2,1}(t)$) and original ($x^{\text{av}}_{2,1}(t)$) average of all unmeasured states except the outlier's.

and hence the estimation error ξ_j has stable dynamics

$$\dot{\xi}_j(t) = -\alpha \xi_j(t). \tag{10}$$

It can be seen that if the condition (6) holds, then the row vectors c_j and $c_j^T A_{22}$ lie in the row space of A_{12}. Then, the vector $\alpha q_j^T + q_j^T A_{22}$ also lies in the row space of A_{12} for any $\alpha \in \mathbb{R}$. Therefore, it holds that

$$(\alpha q_j^T + q_j^T A_{22})(I - A_{12}^\dagger A_{12}) = 0.$$

Hence,

$$\alpha q_j^T + q_j^T A_{22} = (\alpha q_j^T + q_j^T A_{22}) A_{12}^\dagger A_{12}. \tag{11}$$

It can be seen that with the choice of ℓ_j^T in (7a), (11) is equivalent to (9b). Further, it can also be seen that with the choice of g_j and h_j in (7b) and (7c), the conditions (9a) and (9c) are satisfied.

We have shown that, under condition (6), the observer (5) with gains as in (7) has error dynamics (10), which is stable for arbitrary $\alpha > 0$. However, the rate of convergence α should be tuned taking into account the discretization scheme. For example, with forward Euler method with fixed time-step δt, α must satisfy $\alpha < 2/\delta t$ in order to ensure stability of the discretized error dynamics $\xi_j(t + \delta t) = (1 - \alpha\, \delta t)\xi_j(t)$. Moreover, the rate of convergence should be adapted to the system properties, with a rule of thumb that suggests α around twice as fast as the system. In the examples of this paper, we will take a smaller α for illustrative purposes, to make the transient behaviours more visible.

For initialization of the observer, in the absence of information on the initial state, it is natural to set $\hat{x}^{\text{av}}_{2,j}(0) = 0$, which can be obtained by choosing $w(0) = -\ell^T y(0)$. The choice $w(0) = 0$, instead, corresponds to $\hat{x}^{\text{av}}_{2,j}(0) = \ell^T y(0)$, which grows with α (see the definition of ℓ in (7a)) and might be used only with small values of α.

Now, we illustrate the estimation using Example 1.

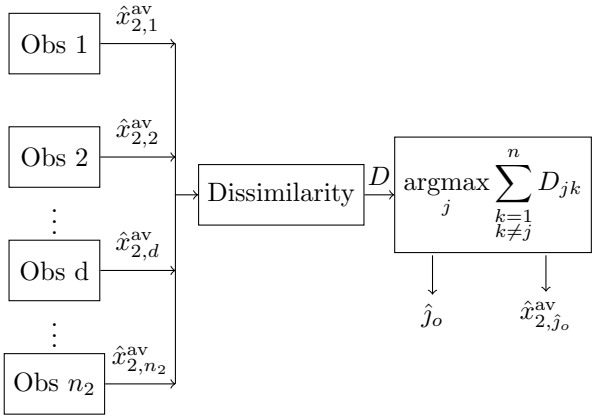

Figure 6: Structure of the estimation and detection algorithm presented in Section 4 for the detection of the outlier and trimmed average estimation. Here, obs stands for the observer and D is the dissimilarity matrix. $\hat{j}_o$ and $\hat{x}^{\text{av}}_{2,\hat{j}_o}$ are the detected position of the outlier and the average estimate excluding the outlier respectively.

Example 1 (continued)**.** *Here, we have the knowledge that the outlier is at node 7, which is the first node in the set of unmeasured nodes, i.e., $j_o = 1$. We want to estimate $x^{\text{av}}_{2,j_o}(t)$, i.e., $x^{\text{av}}_{2,1}(t)$. Note that the condition (6) with $j = 1$ is satisfied, as* $\text{rank}\begin{bmatrix} A_{12} \\ q_1^T A_{22} \\ q_1^T \end{bmatrix} = 4 = \text{rank}[A_{12}]$, *where $q_1^T = \left[0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$. Therefore, we can design an observer of the form (5) to estimate $x^{\text{av}}_{2,1}(t)$ with an arbitrary rate. Let us fix $\alpha = 3$ and use observer (5) with ℓ_1, g_1 and h_1 computed with (7). We initialize the observer with $w_1(0) = -\ell_1^T y(0)$ so as to obtain $\hat{x}^{\text{av}}_{2,1}(0) = 0$.*

It can be seen in Figure 5 that our observer is able to estimate the average $x^{\text{av}}_{2,1}(t)$, i.e., the average of unmeasured states excluding the outlier at node $n_1 + 1 = 7$.

In the next section, we investigate the possibility of estimating the average using an observer of the type (5) in presence of an outlier at an arbitrary position, and we also give a method to detect the outlier.

4. Outlier at an unknown position

In this section, we consider the case when the position of the outlier i.e., j_o is unknown. We propose a centralised method to estimate the average of the unmeasured nodes excluding the outlier and detecting it simultaneously. Our approach is to estimate the averages $x^{\text{av}}_{2,j}(t)$ for all possible $j \in \{1, \ldots, n_2\}$ using a bank of scalar observers and then compare the estimates in order to detect the outlier. For the comparison, we propose a dissimilarity criterion inspired by the distance-based dissimilarity used in signal processing. Figure 6 illustrates the process we use for the detection of the outlier which also gives us the required average. Therefore, at first, we provide a necessary and sufficient condition for the design of observers of the form (5) for every possible $j \in \{1, \ldots, n_2\}$. Then we define a dissimilarity criterion to differentiate between the estimates. After that, we define a very general optimization problem to detect the outlier. In the end, we illustrate the method with the help of Example 1.

4.1. Existence condition for the bank of observers

In order to design observers of the form (5) to reconstruct $x^{\text{av}}_{2,j}(t)$ for all $j \in \{1, \ldots, n_2\}$, the condition (6) must be satisfied for all $j \in \{1, \ldots, n_2\}$. The resulting condition has been stated in the following theorem.

Theorem 2. *A tunable and scalar observer for $x^{\text{av}}_{2,j}(t)$ exists for all $j \in \{1, \ldots, n_2\}$ if and only if*

$$\text{rank}(A_{12}) = n_2. \tag{12}$$

Proof. From Theorem 1, to estimate $x^{\text{av}}_{2,j}(t)$ for all j, a tunable, scalar observer exists if and only if (6) holds for all j.
This is equivalent to

$$\text{rank}\begin{bmatrix} A_{12} \\ q_1^T A_{22} \\ \vdots \\ q_{n_2}^T A_{22} \\ q_1^T \\ \vdots \\ q_{n_2}^T \end{bmatrix} = \text{rank}[A_{12}]. \tag{13}$$

Now define $P = (I_{n_2} - \mathbf{1}_{n_2}\mathbf{1}_{n_2}^T)$, and $Q = -\dfrac{1}{n_2 - 1}P = \begin{bmatrix} q_1^T \\ \vdots \\ q_{n_2}^T \end{bmatrix}$, so that the left-hand side of (13) is equal to $\begin{bmatrix} A_{12} \\ QA_{22} \\ Q \end{bmatrix}$.
We can see that $\text{rank}(Q) = n_2$ by showing that $\text{rank}(P) = n_2$ and we show $\text{rank}(P) = n_2$ by showing that $\det P \neq 0$. For this, notice that

$$\begin{bmatrix} I_{n_2} & \mathbf{0}_{n_2,1} \\ \mathbf{1}_{n_2}^T & 1 \end{bmatrix} \begin{bmatrix} P & -\mathbf{1}_{n_2} \\ \mathbf{0}_{1,n_2} & 1 \end{bmatrix} \begin{bmatrix} I_{n_2} & \mathbf{0}_{n_2,1} \\ -\mathbf{1}_{n_2}^T & 1 \end{bmatrix} = \begin{bmatrix} I_{n_2} & -\mathbf{1}_{n_2} \\ \mathbf{0}_{1,n_2} & 1 - \mathbf{1}_{n_2}^T \mathbf{1}_{n_2} \end{bmatrix}.$$

Taking determinant of matrices on both the sides, we have $\det(P) = 1 - n_2 \neq 0$.
Therefore, $\text{rank}(P) = n_2$ and hence, $\text{rank}(Q) = n_2$.
Finally, since $\text{rank}(Q) = n_2$, we have $\text{rank}\begin{bmatrix} A_{12} \\ QA_{22} \\ Q \end{bmatrix} = n_2$.
Therefore, (13) holds if and only if $\text{rank}(A_{12}) = n_2$. □

Now we present some remarks on how restrictive the conditions in Theorem 2 are.

Remark 1: It can be seen from (2), that $A_{12} \in \mathbb{R}^{n_1 \times n_2}$ and the condition (12) is $\text{rank}(A_{12}) = n_2$. It implies $n_1 \geq n_2$, that is the number of measured nodes must

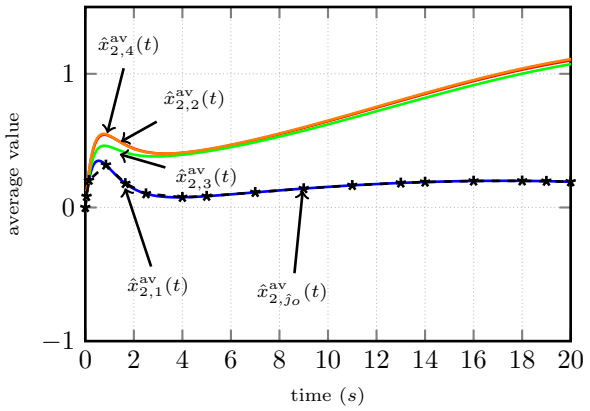

Figure 7: Estimated trajectories of the averages of unmeasured nodes in Example 1 excluding one node at a time i.e. $\hat{x}^{\text{av}}_{2,j}(t)\ \forall j \in \{1,\dots,n_2\}$, the output of the bank of observers. The dash-starred line is $\hat{x}^{\text{av}}_{2,\hat{j}_o}(t)$, the estimated average without the detected outlier at time t. It can be seen that $\hat{x}^{\text{av}}_{2,\hat{j}_o}(t)$ converges to $\hat{x}^{\text{av}}_{2,1}(t)$ very quickly as the outlier is at $j_o = 1$.

be greater than or equal to the number of unmeasured nodes.

Remark 2: From the condition (12), A_{12} is full column rank so it cannot have an all zero column. Therefore, Theorem 2 requires that for every unmeasured node, there exists an edge pointing to some measured node, i.e., for every unmeasured node j there is an edge (i,j) with i a measured node (recall that the edge (i,j) is depicted as an arrow $j \to i$, representing an influence of state x_j on the dynamics of state x_i).

Now, we proceed towards the problem of detection of the outlier in the next section. For this, we propose to run a bank of observers and then compare the estimates we obtain.

4.2. Outlier detection

In this subsection, we define a dissimilarity matrix and an optimization problem in order to detect the outlier.

4.2.1. Dissimilarity criterion

Here, we define a dissimilarity criterion in order to differentiate between the estimates obtained in the previous section with a goal in mind to pick the one which is without the outlier. Dissimilarity criteria of this kind are used in signal processing. For instance, [15] defines dissimilarity as pairwise Euclidean distance between two signals.

Consider the estimates $\hat{x}^{\text{av}}_{2,j}(t)\ \forall j \in \{1,\dots,n_2\}$ obtained in the previous section, we define their dissimilarity at time t as

$$D_{jk}(t) = \int_0^t e^{-\beta(t-\tau)} \mid \hat{x}^{\text{av}}_{2,j}(\tau) - \hat{x}^{\text{av}}_{2,k}(\tau) \mid \mathrm{d}\tau \quad \text{for } \beta > 0, \tag{14}$$

where $\hat{x}^{\text{av}}_{2,j}(\tau)$ is the average estimate of the unmeasured nodes except the node j at time τ.

This definition seems to require all the average estimates from $\tau = 0$ to the current time $\tau = t$, but this integral can be computed recursively as

$$D_{jk}(t) = e^{-\beta\delta t} D_{jk}(t-\delta t) + \int_{t-\delta t}^{t} e^{-\beta(t-\tau)} \mid \hat{x}^{\text{av}}_{2,j}(\tau) - \hat{x}^{\text{av}}_{2,k}(\tau) \mid \mathrm{d}\tau,$$

which might be simplified with a suitable approximation, e.g.,

$$D_{jk}(t) \simeq e^{-\beta\delta t} D_{jk}(t-\delta t) + \delta t \mid \hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t) \mid .$$

The matrix $D = [D_{jk}]$ so obtained is called the *Dissimilarity matrix*. Note that D is a non-negative, symmetric matrix with zero diagonal elements. Here, the idea is to measure how far are the estimates from each other.

Rather than instantaneous comparisons, we consider an integral, so as to consider only differences that hold over non-trivial windows of time. The forgetting factor β tunes the weight given to past values, and should be chosen avoiding the two extremes: β near to zero gives too much weight to the initial transient, where the estimates might have large errors, while too large β gives vanishing weight to all past values.

The system is assumed to have an outlier at j_o, i.e., there is a significant difference between the average $x^{\text{av}}_{2,j_o}(t)$ excluding j_o and the average $x^{\text{av}}_{2,k}(t)$ excluding any other node k. Moreover, the outlier is unique. Hence, for any j and k different from j_o, $|x^{\text{av}}_{2,j_o}(t) - x^{\text{av}}_{2,k}(t)|$ is large and $|x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t)|$ is small (at least as an integral over time, as in the dissimilarity matrix). Since each estimate $\hat{x}^{\text{av}}_{2,j}(t)$ converges to the corresponding correct average $x^{\text{av}}_{2,j}(t)$, we also have that $|\hat{x}^{\text{av}}_{2,j_o}(t) - \hat{x}^{\text{av}}_{2,k}(t)|$ is large and $|\hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t)|$ is small, except possibly for an initial transient. For this reason, we can say that $\hat{j}_o(t)$ is the detected position of the outlier at time t if $\hat{j}_o(t)$th row sum of the dissimilarity matrix at time t is the largest:

$$\hat{j}_o(t) = \underset{j}{\operatorname{argmax}} \sum_{\substack{k=1 \\ k\neq j}}^{n} D_{jk}(t). \tag{15}$$

The above processes as depicted in Figure 6 can be put altogether as follows:

1. Using a bank of n_2 observers of the form (5), compute $\hat{x}^{\text{av}}_{2,j}(t)$ for all j from 1 to n_2.
2. Compute $D_{jk}(t)$ defined in (14).
3. From the dissimilarity matrix, detect the outlier $\hat{j}_o(t)$ at time t given by (15). Then choose the corresponding average estimate $\hat{x}^{\text{av}}_{2,\hat{j}_o}(t)$ obtained from the bank of observers, which excludes $x_{n_1+\hat{j}_o}(t)$.

Now, we illustrate the method with Example 1. We will see that indeed $\hat{j}_o(t)$ converges to j_o, the actual position of the outlier.

Example 1 (continued)**.** *The position of the outlier in Example 1 is $j_o = 1$ but here we assume that this information*

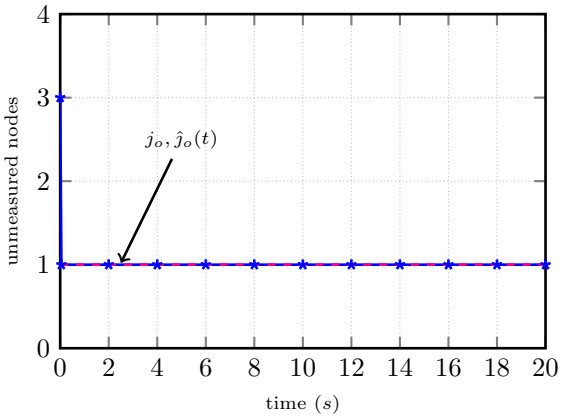

Figure 8: Detected position of the outlier in the set of the unmeasured nodes at time t i.e. $\hat{j}_o(t)$ in Example 1. $j_o = 1$ is the actual position of the outlier. At first, the method identifies unmeasured node 3 as the outlier but it converges very quickly to j_o.

is unknown. We want to detect the outlier position and obtain the average estimate excluding the outlier. Note that the condition (12) is satisfied with $\text{rank}(A_{12}) = 4 = n_2$. *Therefore, we can design a bank of* n_2 *observers of the form (5) each of which estimates the average of all but one unmeasured node at a time. Here, each observer is designed with* $\alpha = 3$, *and the parameters* ℓ_j^T, g_j^T, h_j^T *computed with (7) and initialization* $w_j(0) = -\ell_j^T y(0)$. *The estimates obtained by this bank of observers are depicted in Figure 7. We compute the dissimilarity matrix with* $\beta = 10$ *in (14) and then follow the process described above to detect the outlier using (15). Figure 8 shows that indeed the proposed method detects the outlier position* $j_o = 1$, *i.e., node* $n_1 + j_o = 7$ *in the network depicted in Figure 2. In Figure 8, it can be seen that at first, the method identifies unmeasured node 3 (node 9 in the network) as the outlier but it converges very quickly to the actual position, i.e.,* $j_o = 1$ *(node* $n_1 + j_o = 7$ *in the network). One possible reason for this delay in detection could be the delay in convergence of the observed value to the original value.*

Note that we already have the average estimates $\hat{x}^{\text{av}}_{2,j}(t)$ *for all* $j \in \{1, \ldots, n_2\}$, *obtained from the bank of observers. From them, we obtain* $\hat{x}^{\text{av}}_{2,\hat{j}_o}(t)$, *which is initially equal to* $\hat{x}^{\text{av}}_{2,3}(t)$ *and then equal to* $\hat{x}^{\text{av}}_{2,1}(t)$, *as illustrated by the dash-starred black line in Figure 7. This is consistent with the quantity we aim at reconstructing: the average estimate of the unmeasured nodes excluding the outlier, i.e.,* $\hat{x}^{\text{av}}_{2,1}(t)$.

5. System matrices are partially unknown

In this section, we consider the case when there exists an outlier in the system but the system matrices in (1) are partially unknown. Assuming that this outlier is caused by a fault, we call the initial system which is without the fault, the nominal system. We aim to detect the outlier and to compute the average without the outlier using the detection method proposed in Section 4 even when the system with the fault is unknown, and only the nominal system is known. Recall that, in the proposed detection method, we deploy a bank of observers of the form (5), designed with the faulty system matrices. However, in this section, since the faulty system is unknown, we propose to design the observers with the nominal system.

Let us define the nominal system as

$$\begin{cases} \dot{x}(t) = & \tilde{A}x(t) + \tilde{B}u(t) \\ y(t) = & Cx(t), \end{cases} \tag{16}$$

which describes the same physical system as system (1) but without the fault causing the outlier. The system (1) and the system (16) are related in such a way that

$$A = \tilde{A} + \Delta \text{ and } B = \tilde{B} + \Psi, \tag{17}$$

where $\Delta \in \mathbb{R}^{n\times n}$ and $\Psi \in \mathbb{R}^{n\times m}$ are the matrices which describe the fault responsible for the outlier.

The block structure of the matrices corresponding to the state partition $x(t) = [x_1^T(t), x_2^T(t)]^T$ as in Section 2 is

$$\begin{aligned} \tilde{A} &= \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}, \ \Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}, \\ \tilde{B} &= \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}, \ \Psi = \begin{bmatrix} \Psi_1 \\ \Psi_2 \end{bmatrix}, \ C = \begin{bmatrix} I_{n_1} & \mathbf{0}_{n_1,n_2} \end{bmatrix}. \end{aligned} \tag{18}$$

Let us consider the following fault model:

$$\text{Fault model: } \begin{cases} \Delta = e_{n_1+j_o} r^T + \rho \tilde{A} e_{n_1+j_o} e_{n_1+j_o}^T, \\ \Psi = e_{n_1+j_o} \psi^T. \end{cases} \tag{19}$$

where $e_i \in \mathbb{R}^n$ is the standard ith basis vector, $\rho \in \mathbb{R}$ is a scalar, $r^T \in \mathbb{R}^n$ and $\psi^T \in \mathbb{R}^m$ are row vectors, that can be arbitrary, n_1 is the number of measured nodes and j_o is the true position of the outlier. This means that Δ is a matrix which is non-zero only in its $(n_1 + j_o)$th row and column. Moreover, its $(n_1 + j_o)$th column is proportional to the corresponding column of $\tilde{A}$, while no such restriction is assumed on its $(n_1 + j_o)$th row, which can be arbitrary. The matrix Ψ is a matrix with non-zero entries only in its $(n_1 + j_o)$th row which is equal to the row vector ψ^T.

In terms of the network system, this corresponds to altering only the neighborhood of node $n_1 + j_o$. More precisely, the row vector r^T represents the change in the influence of the in-neighbors of the node $(n_1 + j_o)$ on it. It allows for arbitrary changes in the incoming edges and in their weights. The term $\rho \tilde{A} e_{n_1+j_o} e_{n_1+j_o}^T$ in (19) represents the change in the influence of node $(n_1 + j_o)$ on its out-neighbors. Due to this term, all entries of the $(n_1 + j_o)$th column of A, other than the one on the diagonal, are equal to the corresponding entry of $\tilde{A}$, multiplied by $(1+\rho)$. This means that the outgoing edges from $n_1 + j_o$ are unchanged, and their weights are all multiplied by a same scalar factor $(1 + \rho)$, which describes a change in the strength of the influence of node $n_1 + j_o$ on its out-neighbors. The

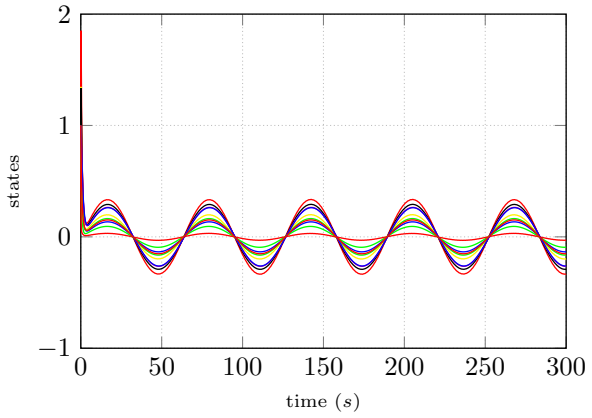

Figure 9: State trajectories of the nominal system described in Example 2.

row vector ψ^T changes the effect of the input, again in a localized way, only on the node $n_1 + j_0$.

Now, we illustrate the nominal system and the fault model with an example.

Example 2. *Consider the network depicted in Figure 2. The input is given by* $u(t) = 0.2 * \sin(0.1t)$ *and* $x(0) = \mathbf{1}_{10}$. *Here, we give a nominal system with no outlier, and a fault as in (19), such that the resulting faulty system is the system with an outlier that was presented in Example 1. The nominal system matrix* $\tilde{A}$ *has the following blocks (according to the partition in (18)):*

$$\tilde{A}_{11} = \begin{bmatrix} -3.25 & 0.98 & 0.84 & 0 & 0 & 0 \\ 0.61 & -5.33 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3.53 & 0.48 & 0 & 0 \\ 0 & 0 & 0.25 & -3.05 & 0 & 0 \\ 0 & 0 & 0 & 0.13 & -1.69 & 0 \\ 0 & 0.54 & 0 & 0 & 0.85 & -2.18 \end{bmatrix},$$

$$\tilde{A}_{12} = \begin{bmatrix} 17.15 & 0.35 & 0.35 & 0 \\ 0 & 0 & 0.76 & 0.76 \\ 0 & 0.22 & 0 & 0 \\ 11.27 & 0 & 0 & 0.23 \\ 7.84 & 0.16 & 0 & 0 \\ 0 & 0.35 & 0 & 0 \end{bmatrix},$$

$$\tilde{A}_{21} = \begin{bmatrix} 0 & 0.89 & 0 & 0 & 0 & 0.16 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\tilde{A}_{22} = \begin{bmatrix} -11.76 & 0 & 0 & 0 \\ 0.98 & -1.4 & 0 & 0 \\ 0 & 0.64 & -1.16 & 0 \\ 0 & 0 & 0.4 & -2.0 \end{bmatrix}.$$

Matrices B *and* C *are the same as in Example 1, i.e.,* $B = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T$ *and* $C = \begin{bmatrix} I_6 & \mathbf{0}_{6,4} \end{bmatrix}$.

It can be seen in Figure 9 that there is no outlier in the nominal system. Then we consider a faulty system matrix $A = \tilde{A} + \Delta$ *and* $B = \tilde{B} + \Psi$, *where* Δ *and* Ψ *are given by (19), with* $j_o = 1$, $r^T = [0, -0.89, 0, 0, 0, -0.14, -11.55, 0, 0, 0]$, $\rho = -\frac{48}{49}$ *and* $\Psi = \mathbf{0}_{n,1}$. *This faulty system is the system considered in Example 1, where the state* $n_1 + j_o = 7$ *is an outlier as shown in Figure 3.*

Now, we proceed towards the problem of detecting the outlier while designing the observers with the known nominal system instead of the faulty system. Recall that, to estimate $x^{\text{av}}_{2,j}(t)$, the average of the unmeasured nodes excluding the element at the jth position, we use observers of the form (5) with observer gains obtained by fixing $\alpha > 0$ and computing:

$$\ell_j^T = q_j^T(\tilde{A}_{22} + \alpha I_{n_2})\tilde{A}_{12}^{\dagger}, \tag{20a}$$

$$h_j^T = -\ell_j^T(\alpha I_{n_1} + \tilde{A}_{11}) + q_j^T\tilde{A}_{21}, \tag{20b}$$

$$g_j^T = q_j^T\tilde{B}_2 - \ell_j^T\tilde{B}_1, \tag{20c}$$

where $\tilde{A}_{12}^{\dagger}$ is the Moore-Penrose pseudo-inverse of $\tilde{A}_{12}$,

Now, we present a lemma to describe an interesting property of the error dynamics when the observer gains are computed as in (20).

Lemma 1. *With the fault model given by (19) and* $\text{rank}(\tilde{A}_{12}) = n_2$, *given that the bank of observers are of the form (5) and the observers gains computed as in (20), then the estimation error* $\xi_j(t) = \hat{x}^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,j}(t)$ *is such that its dynamics is as follows:*

$$\dot{\xi}_j(t) = -\alpha\xi_j(t) + \begin{cases} 0 & j = j_o \\ \frac{1}{n_2 - 1}\left(r_1^T x_1(t) + (r_2^T - \rho\alpha e_{j_o}^T)x_2(t) + \psi^T u(t)\right) & j \neq j_o. \end{cases} \tag{21}$$

Proof. Let us consider the estimation error $\xi_j(t)$.

$$\begin{aligned} \dot{\xi}_j(t) = &-\alpha\xi_j(t) \\ &+(q_j^T(\tilde{A}_{21} + \Delta_{21}) - \alpha\ell_j^T - \ell_j^T(\tilde{A}_{11} + \Delta_{11}) - h_j^T)x_1(t) \\ &+(q_j^T(\tilde{A}_{22} + \Delta_{22}) - \ell_j^T(\tilde{A}_{12} + \Delta_{12}) + \alpha q_j^T)x_2(t) \\ &+(q_j^T(\tilde{B}_2 + \Psi_2) - g_j^T - \ell_j^T(\tilde{B}_1 + \Psi_1))u(t). \end{aligned}$$

Rearranging the equation by separating the terms with $\Delta_{11}, \Delta_{12}, \Delta_{21}, \Delta_{22}, \Psi_1$ and Ψ_2, we have

$$\begin{aligned} \dot{\xi}_j(t) = -\alpha\xi_j(t) &+(q_j^T\tilde{A}_{21} - \alpha\ell_j^T - \ell_j^T\tilde{A}_{11} - h_j^T)x_1(t) \\ &+(q_j^T\tilde{A}_{22} - \ell_j^T\tilde{A}_{12} + \alpha q_j^T)x_2(t) \\ &+(q_j^T\tilde{B}_2 - g_j^T - \ell_j^T\tilde{B}_1)u(t) \\ &+(q_j^T\Delta_{21} - \ell_j^T\Delta_{11})x_1(t) \\ &+(q_j^T\Delta_{22} - \ell_j^T\Delta_{12})x_2(t) \\ &+(q_j^T\Psi_2 - g_j^T - \ell_j^T\Psi_1)u(t). \end{aligned} \tag{22}$$

As shown in Section 3, the rank condition $\text{rank}(\tilde{A}_{12}) = n_2$ ensures that computing ℓ_j, g_j and h_j as in (20), we have

$$-\alpha\ell_j^T + q_j^T\tilde{A}_{21} - \ell_j^T\tilde{A}_{11} - h_j^T = 0, \tag{23a}$$

$$+\alpha q_j^T + q_j^T\tilde{A}_{22} - \ell_j^T\tilde{A}_{12} = 0, \text{and} \tag{23b}$$

$$q_j^T\tilde{B}_2 - g_j^T - \ell_j^T\tilde{B}_1 = 0. \tag{23c}$$

Therefore, from (22) and (23), we have

$$\begin{aligned}\dot{\xi}_j(t) = -\alpha\xi_j(t) &+(q_j^T\Delta_{21} - \ell_j^T\Delta_{11})x_1(t)\\ &+(q_j^T\Delta_{22} - \ell_j^T\Delta_{12})x_2(t)\\ &+(q_j^T\Psi_2 - g_j^T - \ell_j^T\Psi_1)u(t).\end{aligned} \tag{24}$$

From (19), using the notation $r^T = [r_1^T, r_2^T]$, we have

$$\begin{aligned}\Delta_{11} &= \mathbf{0}_{n_1,n_1}, & \Psi_1 &= \mathbf{0}_{n_1,1},\\ \Delta_{12} &= \rho\tilde{A}_{12}e_{j_o}e_{j_o}^T, & &\\ \Delta_{21} &= e_{j_o}r_1^T, & \Psi_2 &= e_{j_o}\psi^T,\\ \Delta_{22} &= e_{j_o}r_2^T + \rho\tilde{A}_{22}e_{j_o}e_{j_o}^T. & &\end{aligned}$$

Substituting these values in (24), we have

$$\begin{aligned}\dot{\xi}_j(t) = -\alpha\xi_j(t) &+(q_j^Te_{j_o}r_1^T)x_1(t) + (q_j^Te_{j_o}r_2^T)x_2(t)\\ &+\rho\left(q_j^T\tilde{A}_{22} - \ell_j^T\tilde{A}_{12}\right)e_{j_o}e_{j_o}^Tx_2(t)\\ &+(q_j^Te_{j_o}\psi^T)u(t).\end{aligned} \tag{25}$$

Now, let us consider the second last term of (25). From (23b), it can be written as

$$\rho(q_j^T\tilde{A}_{22} - \ell_j^T\tilde{A}_{12})e_{j_o}e_{j_o}^T = -\rho\alpha q_j^Te_{j_o}e_{j_o}^T.$$

Notice that, if $j = j_o$, we have $q_j^Te_{j_o} = 0$, while for all $j \neq j_o$ we have $q_j^Te_{j_o} = \frac{1}{n_2-1}$. Therefore, we conclude that the coefficient of $x_1(t)$ in (25) is

$$q_j^Te_{j_o}r_1^T = \begin{cases}\mathbf{0}_{1,n_1} & \text{if } j = j_o\\ \frac{1}{n_2-1}r_1^T & \text{if } j \neq j_o,\end{cases} \tag{26}$$

the coefficient of $x_2(t)$ in (25) is

$$q_j^Te_{j_o}r_2^T - \rho\alpha q_j^Te_{j_o}e_{j_o}^T = \begin{cases}\mathbf{0}_{1,n_2} & \text{if } j = j_o\\ \frac{1}{n_2-1}(r_2^T - \rho\alpha e_{j_o}^T) & \text{if } j \neq j_o.\end{cases} \tag{27}$$

and the coefficient of $u(t)$ in (25) is

$$q_j^Te_{j_o}\psi^T = \begin{cases}\mathbf{0}_{1,m} & \text{if } j = j_o\\ \frac{1}{n_2-1}\psi^T & \text{if } j \neq j_o,\end{cases} \tag{28}$$

which ends the proof of the lemma. □

Now, we comment on the average estimates in the following theorem.

Theorem 3. *With the fault model given by (19) and* $\text{rank}(\tilde{A}_{12}) = n_2$*, given that the bank of observers are of the form (5) and the observer gains are computed as (20), then*

i.) for $j = j_o$, $\hat{x}^{\text{av}}_{2,j_o}(t) \to x^{\text{av}}_{2,j_o}(t)$ *as* $t \to \infty$.

ii.) for all $j, k \neq j_o$, $(\hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t)) \to (x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t))$ *as* $t \to \infty$.

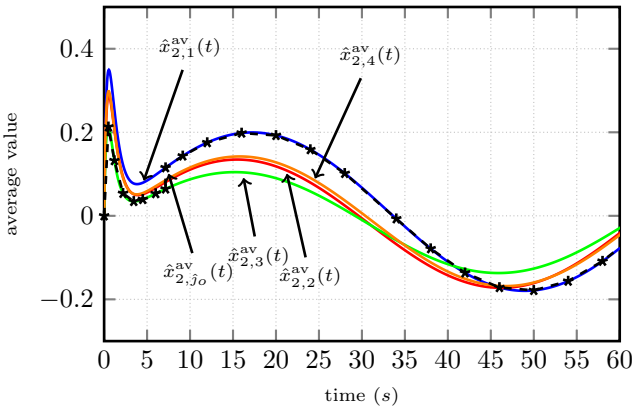

Figure 10: Estimated average trajectories $\hat{x}^{\text{av}}_{2,j}(t)$ $\forall j \in \{1,\dots,n_2\}$, in Example 2. The observer gains were computed using nominal matrices instead of the faulty matrices. The dashed-star line represents $\hat{x}^{\text{av}}_{2,j_o}(t)$.

Proof. i.) It follows immediately from Lemma 1, since $\xi_{j_o}(t) = x^{\text{av}}_{2,j_o}(t) - \hat{x}^{\text{av}}_{2,j_o}(t)$ has a stable dynamics $\dot{\xi}_{j_o}(t) = -\alpha\xi_{j_o}(t)$.

ii.) We study the difference $(x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t)) - (\hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t)) = \xi_j(t) - \xi_k(t)$. From (21), for $j, k \neq j_o$, such difference has stable dynamics

$$\dot{\xi}_j(t) - \dot{\xi}_k(t) = -\alpha(\xi_j(t) - \xi_k(t)).$$

□

In Section 4, the observers were designed using the knowledge of the matrices A and B, and hence it was possible to ensure convergence of the estimates $\hat{x}^{\text{av}}_{2,j}(t)$ to the true averages $x^{\text{av}}_{2,j}(t)$. Here, the observers are designed with the nominal matrices $\tilde{A}$ and $\tilde{B}$, while the true averages are from the faulty system with matrices A and B, so in general convergence of the estimates cannot be ensured. However, Theorem 3 gives two important convergence results.

First, for the estimate of the average without the outlier ($\hat{x}^{\text{av}}_{2,j_o}(t)$), convergence to the true average $x^{\text{av}}_{2,j_o}(t)$ is still ensured, despite the use of matrix $\tilde{A}$ instead of the unknown matrix A. Hence, in case the outlier position j_o is known or can be correctly detected, then the trimmed-average $x^{\text{av}}_{2,j_o}(t)$ can be correctly estimated.

Second, for j and k different from j_o, although the estimates $\hat{x}^{\text{av}}_{2,j}(t)$ and $\hat{x}^{\text{av}}_{2,k}(t)$ can be wrong with respect to the true averages $x^{\text{av}}_{2,j}(t)$ and $x^{\text{av}}_{2,k}(t)$, Theorem 3 ensures that their difference $(\hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t))$ converges to the true difference $(x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t))$.

Moreover, since $u(t)$ and $x(t)$ are assumed to remain bounded, we can see from (21) that the estimates $\hat{x}^{\text{av}}_{2,j}(t)$ and $\hat{x}^{\text{av}}_{2,k}(t)$ remain bounded, which avoids overflow issues in numerical computation.

To see the importance of these two convergence results for our detection method, recall that we use a bank of scalar observers and the dissimilarity matrix defined

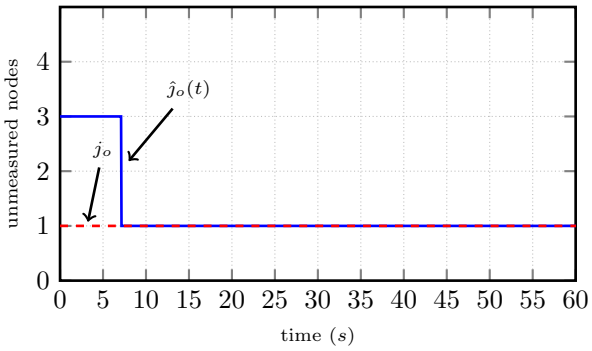

Figure 11: Detected position of the outlier in the set of the unmeasured nodes at time t i.e. $\hat{\jmath}_o(t)$ in Example 2. $j_o = 1$ is the actual position of the outlier. At first, the method identifies node 3 as the outlier but it converges quickly to j_o.

in (14), which involves pairwise differences of estimates $|\hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t)|$. In this paper, we restrict our attention to systems where a localized fault at an unmeasured node results in the appearence of a single outlier state trajectory, corresponding to such node. As discussed in Sect. 4.2.1, this implies that, for any j and k different from j_o, $|x^{\text{av}}_{2,j_o}(t) - x^{\text{av}}_{2,k}(t)|$ is large and $|x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t)|$ is small (at least as an integral over time). Then, our detection method is capable of reconstructing the correct outlier position j_o if these two properties of the true averages are preserved for the estimated averages.

For j and k different from j_o, the second part of Theorem 3 ensures that the difference $(\hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t))$ converges to the true difference $(x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t))$. Hence, the property that the true difference $|x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t)|$ is small implies that also the estimated difference $|\hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t)|$ is small.

The case of the differences $|\hat{x}^{\text{av}}_{2,j_o}(t) - \hat{x}^{\text{av}}_{2,k}(t)|$ involving the outlier is more delicate. The first part of Theorem 3 ensures that $\hat{x}^{\text{av}}_{2,j_o}(t)$ converges to the true $x^{\text{av}}_{2,j_o}(t)$, but gives no information about $\hat{x}^{\text{av}}_{2,k}(t)$. In most cases, as illustrated in our examples, $\hat{x}^{\text{av}}_{2,k}(t)$ is a wrong estimate of $x^{\text{av}}_{2,k}(t)$, but the difference $|\hat{x}^{\text{av}}_{2,j_o}(t) - \hat{x}^{\text{av}}_{2,j}(t)|$ remains large, thus leading to a correct detection of the outlier position. When this happens, also the correct reconstruction of $x^{\text{av}}_{2,j_o}(t)$ is ensured, since $\hat{x}^{\text{av}}_{2,j_o}(t)$ converges to $x^{\text{av}}_{2,j_o}(t)$. In some particular cases, the wrong estimate $\hat{x}^{\text{av}}_{2,k}(t)$ might happen to be very near to $\hat{x}^{\text{av}}_{2,j_o}(t)$. As discussed above, we also know that all estimates $\hat{x}^{\text{av}}_{2,j}(t)$ with $j \neq j_o$ are near to each other, and hence in this case they are all near to $\hat{x}^{\text{av}}_{2,j_o}(t)$. This might lead to a wrong detection of the outlier position j_o but nevertheless ensures that the trimmed average $x^{\text{av}}_{2,j_o}(t)$ is reconstructed with only a very small error. Now we illustrate this with the help of Example 2.

Example 2 (continued)**.** *Consider the nominal and the faulty systems described in Example 2. The position of the outlier is $j_o = 1$ but this information is assumed to be unknown and moreover, the system is also partially known: only the nominal matrix $\tilde{A}$ is known, not the fault. Our goal is to detect the outlier position and obtain the average estimate excluding the outlier. Note that the condition (12) is satisfied as* $\text{rank}(\tilde{A}_{12}) = 4 = n_2$. *Therefore, a bank of n_2 observers of the form (5) can be designed. Here, for all $j \in \{1, \ldots, n_2\}$, we take $\alpha = 3$, $w_j(0) = -\ell_j^T y(0)$ and the parameters ℓ_j^T, g_j^T, h_j^T are computed as in (20). The estimates obtained by this bank of observers are depicted in Figure 10. We compute the dissimilarity matrix with $\beta = 10$ in (14) and then follow the proposed method to detect the outlier using (15). Figure 11 shows that indeed the proposed method detects the outlier position $j_o = 1$. In Figure 11, it can be seen that at first the method identifies unmeasured node 3 as the outlier but it converges quickly to the true position, i.e., $j_o = 1$ (node $n_1 + j_o = 7$ in the network).*

Remark 3: *As a possible variation of our method, the bank of scalar observers for reconstruction of $x^{\text{av}}_{2,j}(t)$ for all j might be replaced by a full-order state observer $\hat{x}(t) = A\hat{x}(t) + Bu(t) + L(y(t) - C\hat{x}(t))$, followed by computation of averages $\hat{x}^{\text{av,full}}_{2,j}(t) := [\mathbf{0}_{1,n_1}, q_j^T]\hat{x}(t)$ for all j. Then, estimates $\hat{x}^{\text{av,full}}_{2,j}(t)$ can be used instead of $\hat{x}^{\text{av}}_{2,j}(t)$ to compute the dissimilarity matrix (14) and to find the outlier position (15) and the trimmed average without the outlier.*

Notice that the rank condition (12) that ensures the existence of the bank of tunable scalar observers implies observability of the system (1). Hence, in the case where matrices A and B are known (as in Section 3), the full-order observer can be designed to have stable error dynamics with any desired rate of convergence.

However, when the observer is designed using nominal matrices, different from the faulty ones of the system (as in the current section), the estimates obtained from the full-order observer do not share the useful convergence properties of the bank of scalar observers. Theorem 3 ensures that $\xi_{j_o}(t) = x^{\text{av}}_{2,j_0}(t) - \hat{x}^{\text{av}}_{2,j_0}(t)$ tends to zero, and the differences $\xi_j(t) - \xi_k(t) = (x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t)) - (\hat{x}^{\text{av}}_{2,j}(t) - \hat{x}^{\text{av}}_{2,k}(t))$ also tend to zero.

The full-order observer designed with nominal matrices is $\dot{\hat{x}} = \tilde{A}\hat{x}(t) + \tilde{B}u(t) + L(y(t) - C\hat{x}(t))$, and has error dynamics $\dot{e}(t) = (\tilde{A} - LC)e(t) + \Delta x(t) + \Psi u(t)$. Because of the terms $\Delta x(t) + \Psi u(t)$, the error $e(t)$ does not tend to zero, and also when considering $[\mathbf{0}_{1,n_1}, q_{j_0}^T]e(t) = x^{\text{av}}_{2,j_0}(t) - \hat{x}^{\text{av,full}}_{2,j_0}(t)$ and $[\mathbf{0}_{1,n_1}, q_j^T]e(t) - [\mathbf{0}_{1,n_1}, q_k^T]e(t) = (x^{\text{av}}_{2,j}(t) - x^{\text{av}}_{2,k}(t)) - (\hat{x}^{\text{av,full}}_{2,j}(t) - \hat{x}^{\text{av,full}}_{2,k}(t))$ there is no guarantee that they tend to zero. Therefore, the outlier position and the trimmed average without the outlier might not be obtained correctly.

6. Outlier detection in a faulty metal plate

In this section, we illustrate the outlier detection method using the nominal system on a faulty thermal diffusion system. This system is inspired by [2, Section IV]. Here, we deal with a network given by spatial discretization of a

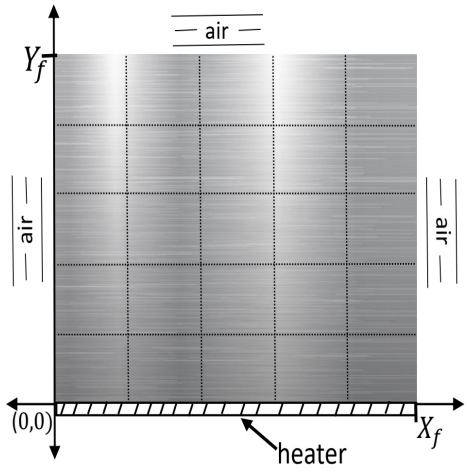

Figure 12: A metal plate with a heater attached on one side and surrounded by air on the other sides.

thermal diffusion system consisting of a rectangular metal plate attached to a heater on one side and surrounded by air on the other sides as shown in Figure 12. Here, we detect the faulty region with only the knowledge of the diffusion properties of the metal plate and the diffusion equations governing the heat diffusion.

6.1. Thermal diffusion system

Consider a rectangular metal plate as shown in Figure 12. Let $(X, Y) \in \mathcal{D} := [0, X_f] \times [0, Y_f]$ correspond to a point on the metal plate. Let $T(X, Y, t)$ be the temperature of the metal plate at the position (X, Y) at time t. The heat transfer in the metal plate is described by the two-dimensional heat conduction equation

$$\frac{\partial T}{\partial t} = \lambda \frac{\partial^2 T}{\partial X^2} + \gamma \frac{\partial^2 T}{\partial Y^2}, \quad (X, Y) \in int(\mathcal{D}). \tag{29}$$

where $int(\mathcal{D})$ denote the interior region of the plate. The constants λ and γ denote the diffusion coefficient along the X-axis and Y-axis respectively. Moreover, there is an exchange of heat between the plate, the heater and the air which is described by the following boundary conditions. The heat exchange with the air is described with the Robin boundary condition [16]

$$\frac{\partial T}{\partial \nu} = -\eta_a (T - T_a), \quad (X, Y) \in \mathcal{S}_a. \tag{30}$$

where S_a is the set of contact points with the air, ν is the outward unit vector normal to $\mathcal{S}_a$, and T_a is the temperature of the air, and η_a is the ratio of coefficient of thermal conductivity and the coefficient of heat transfer between air and the metal plate. For simplicity, we suppose $T_a = 0$ for any t, X and Y. The heat exchange with the heater is described by another Robin boundary condition

$$\frac{\partial T}{\partial Y} = \eta_h (T - u), \quad X \in [0, X_f], Y = 0. \tag{31}$$

where u is the temperature of the heater and η_h is the ratio of coefficient of thermal conductivity and the coefficient of heat transfer between heater and the metal plate. The heater is assumed to have a uniform temperature distribution, that is, u is independent of X and Y.

We discretize (29) and (31) with step size $\delta = \Delta X = \Delta Y$ using central difference quotients. In discretization, let the number of the cells along each axis be n_c so that the total number of cells will be n_c^2. We obtain a network with grid structure with dynamics (16), where $x \in \mathbb{R}^{n_c^2}$ is a vector of spatially discretized temperature T. The numbering of the nodes starts from the bottom left, and follows columns, each from bottom to top as in Figure 14. To describe the entries $\tilde{a}_{ij}$ of $\tilde{A}$, we need to distinguish different cases, depending on the position of the cell corresponding to vertex i. The non-zero entries of matrix $\tilde{A} = [\tilde{a}_{ij}]$ are given as follows. If i is not a boundary node then

$$\tilde{a}_{ij} = \begin{cases} -2(\lambda + \gamma) & \text{if } j = i \\ \gamma & \text{if } j = i + 1, i - 1 \\ \lambda & \text{if } j = i + n_c, i - n_c. \end{cases}$$

If i is a boundary node attached to the heater but not a corner node then, from the discretization of (29) and (31), we have

$$\tilde{a}_{ij} = \begin{cases} -2(\lambda + \gamma + \delta\eta_h\lambda) & \text{if } j = i \\ 2\gamma & \text{if } j = i + 1 \\ \lambda & \text{if } j = i + n_c, i - n_c. \end{cases}$$

For the boundaries with air, the discretization of (29) and (30) is used. If i is on the upper, left or right boundary, except for the corner nodes, we have $\tilde{a}_{ij} = -2(\lambda + \gamma + \delta\eta_a\lambda)$ if $j = i$. If i is on the upper boundary, the non-zero non-diagonal entries are $\tilde{a}_{ij} = 2\gamma$ for $j = i - 1$ and $\tilde{a}_{ij} = \lambda$ for $j = i + n_c, i - n_c$. Similarly, if i is on the left border, we have $\tilde{a}_{ij} = \gamma$ for $j = i - 1, i + 1$ and $\tilde{a}_{ij} = 2\lambda$ for $j = i + n_c$, and then on the right, we have $\tilde{a}_{ij} = \gamma$ for $j = i - 1, i + 1$ and $\tilde{a}_{ij} = 2\lambda$ for $j = i - n_c$. Finally, if i is an upper-left or upper-right corner node, we have $\tilde{a}_{ij} = -2(\lambda + \gamma + \delta\eta_a\lambda + \delta\eta_a\gamma)$ for $j = i$. If i is on the upper-left corner, we have $\tilde{a}_{ij} = 2\gamma$ for $j = i - 1$ and $\tilde{a}_{ij} = 2\lambda$ for $j = i + n_c$. Similarly, if i is on the upper-right corner, we have $\tilde{a}_{ij} = 2\gamma$ for $j = i - 1$ and $\tilde{a}_{ij} = 2\lambda$ for $j = i - n_c$.

For a corner node in the bottom, the discretization of (29), (30) and (31) is used. If i is either a bottom-left or bottom-right corner node, we have $\tilde{a}_{ij} = -2(\lambda+\gamma+\delta\eta_a\lambda+\delta\eta_h\gamma)$ for $j = i$. For non-zero non-diagonal entries, we have $\tilde{a}_{ij} = 2\gamma$ for $j = i + 1$ and $\tilde{a}_{ij} = 2\lambda$ for $j = i + n_c$ if i is on the bottom-left corner. Similarly, if i is on the bottom-right corner, we have $\tilde{a}_{ij} = 2\gamma$ for $j = i + 1$ and $\tilde{a}_{ij} = 2\lambda$ for $j = i - n_c$. The entries of the matrix $B = [b_{ij}] \in \mathbb{R}^{n_c^2 \times 1}$ are

$$b_{i1} = \begin{cases} 2\eta_h\delta\gamma & \text{if } i = 1 \bmod n_c \\ 0 & \text{otherwise.} \end{cases}$$

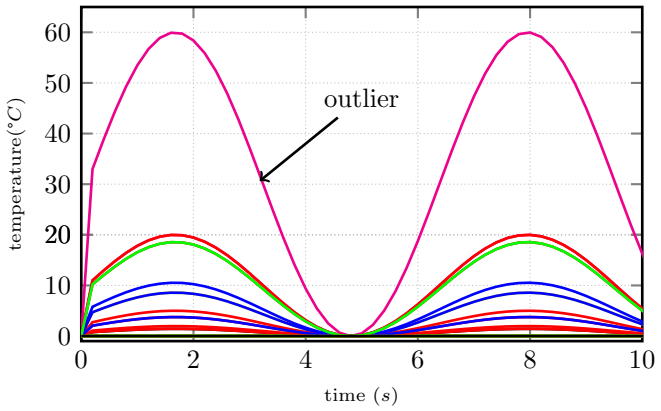

Figure 13: State trajectories of the spatially disctretized equations describing the faulty metal plate.

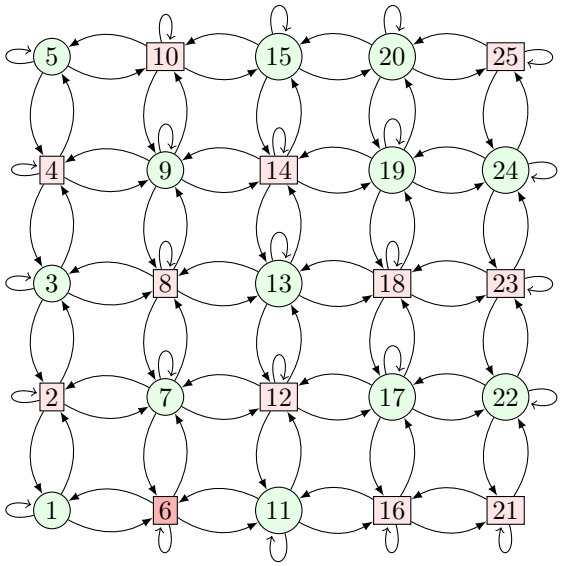

Figure 14: Graphical representation of the discretized metal plate. In this graph, the nodes correspond to the cells obtained after discretization and edges show the heat transfer between them. The circle nodes in green are the measured nodes and the red square nodes are the unmeasured nodes. The nodes $\{1, 6, 11, 16, 21\}$ receive input directly and the node 6 in dark red is the outlier.

and the matrix C depends on the choice of the measured nodes.

6.2. Illustration of the detection method

In this subsection, we show the effectiveness of the proposed detection method for spatially discretized thermal diffusion network with $\delta = 1$, and $n_c = 5, \lambda = \gamma = 29.1, \eta_h = 1.3 \times 10^4, \eta_a = 10^3$ and input $u(t) = 10 + 10 * \sin(t)$. Initially the plate is kept at $0°C$. On discretization, we get a network with grid structure as shown in Figure 14. The nodes $\{1, 6, 11, 16, 21\}$ are attached to the heater.

We assume that there is a defect in a region corresponding to the discretization cell number 6. The metal plate corresponding to this region is defective and has different coefficients of diffusion than the rest part of the plate. This defect can be the result of the formation of brown stains, the formation of oxides on the metal plate, casting defects, welding defects, or rolling defects to name a few. In our example, the fault scales the local coefficients of diffusion λ and γ by one-third. This fault is represented as faulty matrices given by $A = \tilde{A} + \Delta$ and $B = \tilde{B} + \Psi$ with Δ satisfying (19) with $r^T = -\frac{2}{3}(e_6^T \tilde{A} - (e_6^T \tilde{A} e_6) e_6^T)$, $\rho = -\frac{2}{3}$ and $\Psi = \mathbf{0}_{n,1}$. It can be seen in Figure 13 that node 6 is an outlier.

We choose $\{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24\}$ as the set of measured nodes, and hence the set of unmeasured nodes is $\{2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25\}$. Now, we can apply a suitable permutation to rearrange the nodes such that the first $n_1 = 13$ nodes are measured and the rest n_2 are unmeasured so as to obtain a block structure as in (18). Note that, node 6 in the network is the third node in the set of unmeasured nodes, so we have $j_o = 3$. Now, we assume that we neither know the fault nor the position of the outlier. Therefore, for estimation, we use the system matrices obtained by the discretization of the thermal diffusion equation described in the previous subsection. Recall from Theorem 2 that in order to design a bank of scalar observers, the sub-matrix $\tilde{A}_{12}$ needs to be of full column rank. We find that indeed $\text{rank}(\tilde{A}_{12}) = 12 = n_2$. Therefore, a bank of scalar observers of the form (5) can be designed. Here, we take $\alpha = 3$ and $w_j(0) = -\ell_j^T y(0)$ for all $j \in \{1, \dots 12\}$.

Figure 15 shows the different average estimates excluding one node at a time $\hat{x}^{\text{av}}_{2,j}(t)$ and the corresponding true averages $x^{\text{av}}_{2,j}(t)$, depicted by solid and dotted lines respectively. The true averages $x^{\text{av}}_{2,j}(t)$ clearly show that j_o is an outlier: for all j except j_o, the averages $x^{\text{av}}_{2,j}(t)$ are near to each other, while $x^{\text{av}}_{2,j_o}(t)$ is more distant. As predicted by Theorem 3, the estimate $\hat{x}^{\text{av}}_{2,j_o}(t)$ quickly converges to the true trimmed-average $x^{\text{av}}_{2,j_o}(t)$. For $j \neq j_o$, it can be seen that all the estimates $\hat{x}^{\text{av}}_{2,j}(t)$ for $j \neq j_o$, have a significant error ξ_j from the corresponding true averages, but such errors tend to be the same for all j, consistently with Lemma 1. Hence, the fact that the true averages are near to each other results in the estimates also being near to each other, despite their error.

Using $\beta = 10$ in (14) and then computing (15), we detect the position of the outlier as shown in Figure 16.

It can be seen in Figure 16 that the detected position of the outlier is $\hat{\jmath} = j_o$. The process identifies the outlier from the beginning as it is evident in Figure 15 that the difference between the average estimates $\hat{x}^{\text{av}}_{2,j_o}(t)$ and $\hat{x}^{\text{av}}_{2,j}(t)$ is very large since the beginning.

7. Concluding remarks

Average state reconstruction with the help of some sensor measurements can give unexpected results if there are some outliers among the unmeasured states. A method to

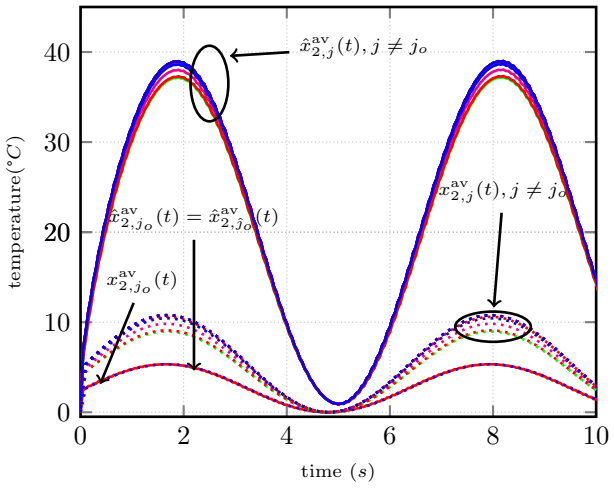

Figure 15: In solid lines, estimated average trajectories of the unmeasured nodes excluding one at a time, $\hat{x}^{\text{av}}_{2,j}$. The red trajectory corresponds to the average estimate without the outlier, i.e., $\hat{x}^{\text{av}}_{2,j_o}(t)$. Moreover, in this case $\hat{x}^{\text{av}}_{2,j_o}(t) = \hat{x}^{\text{av}}_{2,\hat{j}_o}(t)$. The dotted trajectories correspond to the true averages $x^{\text{av}}_{2,j}(t)$.

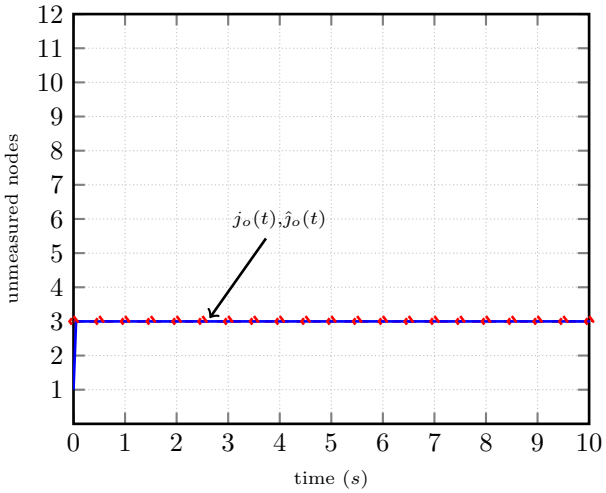

Figure 16: Detected position of the outlier $\hat{j}_o$. It can be seen that the detected position converges to the true position of the outlier j_o very quickly. It is found that the outlier is at the third position in the set of unmeasured nodes i.e. $j_o = 3$, which is at node 6 in the network.

estimate the average excluding the outlier has been proposed. For that, a design of a scalar and tunable observer has been given along with the condition under which a bank of these observers can be designed to estimate the average of the unmeasured nodes while excluding an element at every possible position. Moreover, the problem of detection of the existing outlier when the system matrices might be known or partially unknown has also been addressed by proposing a dissimilarity based matrix inspired from the euclidean distance-based dissimilarity matrix used in signal processing.

Future works will be focused on cases of multiple outliers and sequential methods such as group testing for detection. The sequential methods can be useful in reducing the number of observers required.

Acknowledgement

This work is supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-AdG no. 694209, Scale-FreeBack, website: http://scale-freeback.eu)

References

[1] M. U. B. Niazi, C. Canudas-de-Wit, A. Kibangou, Average state estimation in large-scale clustered network systems, IEEE Transactions on Control of Network Systems 7 (4) (2020) 1736–1745. doi:10.1109/TCNS.2020.2999304.

[2] T. Sadamoto, T. Ishizaki, J. I. Imura, Average state observers for large-scale network systems, IEEE Transactions on Control of Network Systems 4 (4) (2017) 761–769. doi:10.1109/TCNS.2016.2550866.

[3] M. U. B. Niazi, D. Deplano, C. Canudas-de-Wit, A. Y. Kibangou, Scale-free estimation of the average state in large-scale systems, IEEE Control Systems Letters 4 (1) (2020) 211–216.

[4] D. M. Hawkins, Identification of Outliers, Chapman and Hall, 1980. doi:10.1007/978-94-015-3994-4.

[5] P. J. Rousseeuw, A. M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons, Inc., USA, 1987.

[6] C. C. Aggarwal, Outlier Analysis, Vol. 2, Springer International Publishing, 2017. doi:10.1007/978-3-319-47578-3.

[7] A. Arning, A. Rakesh, P. Raghavan, Method for deviation in large databases, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (1996) 164–169.

[8] E. M. Knorr, R. T. Ng, V. Tucakov, Distance-based outliers: Algorithms and applications, The VLDB Journal 3 (2000) 237–253. doi:10.1007/s007780050006.

[9] V. Hautamaki, I. Karkkainen, P. Franti, Outlier detection using k-nearest neighbour graph, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., Vol. 3, 2004, pp. 430–433 Vol.3.

[10] Y. Zhang, N. Meratnia, P. Havinga, Outlier detection techniques for wireless sensor networks: A survey, IEEE Communications Surveys Tutorials 12 (2) (2010) 159–170.

[11] J. Branch, B. Szymanski, C. Giannella, Ran Wolff, H. Kargupta, In-network outlier detection in wireless sensor networks, in: 26th IEEE International Conference on Distributed Computing Systems (ICDCS'06), 2006, pp. 51–51.

[12] S.-J. Yim, C. Yoon-Hwa, Neighbor-based malicious node detection in wireless sensor networks, Wireless Sensor Networks 4 (September) (2012) 361–374.

[13] H. Wang, H. Li, J. Fang, H. Wang, Robust Gaussian Kalman filter with outlier detection, IEEE Signal Processing Letters 25 (8) (2018) 1236–1240. doi:10.1109/LSP.2018.2851156.

[14] U. Pratap, C. Canudas-de-Wit, F. Garin, Average state estimation in presence of outliers, 59th IEEE Conference on Decision and Control (CDC) (2020) 6058–6063.doi:10.1109/CDC42340.2020.9303809.

[15] Y. Xu, S. M. Salapaka, C. L. Beck, Aggregation of graph models and Markov chains by deterministic annealing, IEEE Transactions on Automatic Control 59 (10) (2014) 2807–2812. doi:10.1109/TAC.2014.2319473.

[16] D. Zill, M. Cullen, Differential Equations with Boundary-Value Problems, Brooks/Cole Cengage Learning, Canada, 2009.