On identification of parameterized switched linear systems

Ye Yuan\(^1\), Yorie Nakahira\(^2\) and Claire J. Tomlin\(^2\)

1. Department of of Electrical Engineering and Computer Sciences, University of California, Berkeley, United States
Email: (yy311,tomlinj)@berkeley.edu.

2. Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, United States
Email: ynakahir@caltech.edu.

Abstract: Motivated by applications in fault/attack detection of engineering systems, we formulate in this paper a parameterized switched linear systems identification problem. First, we show that the identifiability of parameterized switched linear systems is equivalent to the global identifiability of a parameterized linear time-invariant system. We propose a new necessary and sufficient condition to check the identifiability of a parameterized linear time-invariant system. Second, using the prior knowledge that the difference between the nominal system and the perturbed system is small (with only a small number of parameters that have been changed), we formulate the fault localization problem as a sparse linear regression problem. We investigate the performance of the proposed algorithm when there is a finite number of samples, and propose a necessary and sufficient condition to determine the number of samples required to achieve a “close-to-true” solution. Potential applications of this work can be: a) development of a sensor placement strategy as well as a real-time monitoring algorithm for the early detection of faults (e.g., material aging, line tripping) or attacks in power grids; b) understanding the functionality of genes, proteins and metabolites in disease dynamics, which could potentially lead to early disease detection and drug design.

Key Words: System Identification, Fault Detection, Finite-sample guarantee.

1 Introduction

Engineering networks, such as power and water distribution networks, are critical infrastructures that possess strict safety and reliability constraints [4]. It is therefore important to design monitoring schemes against unexpected faults and attacks in such networks [23][24]. The monitoring of such networks requires fault detection and fault localization. Fault detection is the process of detecting the occurrence of a fault when it happens [24]. Fault localization is the process of inferring the actual location of a fault. When performing such tasks, it is essential to detect the occurrence with high confidence, accuracy, and efficiency [37, 47].

To tackle this challenge, we model such engineering systems as parameterized switched linear systems. Previous studies have modeled faults as the result of external disturbances [12]. This study instead models faults as the result of changes in internal parameters [1]. One example of our model is a line trip in power systems, that is, an abrupt change in branch admittance [24]. In such a set-up, we separate our overall goal into two technical problems: the first problem is to test whether such a parameterized system is identifiable, that is, whether each set of parameters will always generate a unique set of input-output data. The second problem is to detect and locate a fault using a limited number of samples.

If the identifiability test fails, then there is no way to identify those parameters, regardless of the number of samples [1][3][8][16]. Therefore, if a non-identifiable parameterized linear system switches to a system with a different set of parameters which generates the same input-output data, then it is impossible to detect the parameter change. In this study, we have developed new tools to check the identifiability of parameterized switched linear systems.

Previous studies have used the general hybrid system method for identification [17][22][27][29]. We instead solve the fault localization problem as a sparse linear regression, under the assumption that only a small number of parameters may change. This is justifiable as several system components rarely fail at the same time. Solving this problem as a sparse linear regression requires fewer samples than do traditional methods, enabling online implementation. While most of the literature in system identification concerns the asymptotic behavior of the identification algorithms [2][8][18][40], here we propose a necessary condition and a sufficient condition to guarantee the performance of proposed identification algorithm on a finite and limited number of samples.

Our contribution is summarized below:

**Fundamental limitations in identifiability:** A necessary and sufficient condition (Theorem 4.3) for the identifiability of a switched system has been proposed. It is a generalization and simplification of the sufficient condition presented in [1].

**Algorithms for detection and identification:** Algorithms for detection and identification have been proposed. The identification algorithm takes advantage of the problem structure to reduce the number of required samples.

**Detection and identification accuracy guarantee:** For noisy systems, we have shown that the fault location can be correctly identified under some mild assumptions (Theorem 5.5). For non-stationary systems, the new result provides a theoretical guarantee for parameter estimation accuracy.

Notations

Let \( I \) be the identity matrix and \( e_i^T \) the \( r \)th canonical basis vector. For a real vector \( a \in \mathbb{R}^n \), \( a[i] \) denotes the \( i \)-th parameter.

\[ a[i] = \begin{cases} 1 & \text{if } i = k \in [n], (k \text{ is the index of parameter } a) \\ 0 & \text{otherwise} \end{cases} \]

This work is supported in part by NSF under CPS:ActionWebs (CNS-0931843) and CPS-FORCES (CNS1239166), by NASA under grants NNX12AR18A and UCSCMCA-14-022 (UARC), by ONR under grants N00014-12-1-0609, N000141310341 (Embedded Humans MURI), and MIP-5710002646 (SMARTS MURI), and by AFOSR under grants UPenn-N00014-12-1-0609, N000141310341 (Embedded Humans MURI), and AFOSR under grants UPenn-N000141310341 (Embedded Humans MURI), and by AFOSR under grants UPenn-N000141310341 (Embedded Humans MURI), and by AFOSR under grants UPenn-N000141310341 (Embedded Humans MURI).

Due to page limit, we would like to refer the reader to a complete version which is downloadable from [http://hybrid.eecs.berkeley.edu/~yeyuan/switchedid.pdf](http://hybrid.eecs.berkeley.edu/~yeyuan/switchedid.pdf). The Matlab code is downloadable from [http://yy311.github.io/software.html](http://yy311.github.io/software.html).
element for $1 \leq i \leq n$. Given a matrix $M$, denote $M_i$ to be the $i$-th row of $M$, $M_{ij}$ to be the element in $i$-th row and $j$-th column, $M_{i,R}$ to be its submatrix $M_{i,R} = [M_{ij}], i \in L, j \in R$. Let $vec(M)$ denote the vector formed by stacking the columns of $M$ into one long vector $(M \in \mathcal{R}^{m \times n})$, i.e.,

$$vec(M) = [M_{11} \ M_{21} \ldots \ M_{m1} \ M_{12} \ldots \ M_{mn}]^T.$$ 

Let $\|M\|_\infty = \max |M_{ij}|$ and $\|M\|_1 = \max \sum |M_{ij}|$. Let $\lambda_{max}(M)$ and $\lambda_{min}(M)$ denote the maximum and minimum eigenvalue of a square matrix $M$ respectively. Let $\Re(\cdot)$ be the range space, $\Re(\cdot)$ be the null space and $\otimes$ be the Kronecker product.

### 2 Preliminary – identifiability of LTI systems–

In this section, we review some results on the global identifiability of a parameterized linear time-invariant (LTI) system in [1][8]. Consider a system in the linear state-space form

$$x(t+1) = Ax(t) + Bu(t),$$

$$y(t) = Cx(t).$$

Assume that the input/output properties of this system are available, for example, the Markov parameters $H_k = CA^kB, k = 0, 1, 2, \ldots$. Given an affine parameterization $(A, B, C)(\cdot) : \Omega \rightarrow \mathcal{R}^N$, we are concerned with the condition when $(A, B, C)(\alpha)$ can be uniquely identified from input-output data for any $\alpha \in \Omega$. We denoted that condition to be global identifiability.

**Definition 2.1** [Global Identifiability] A parameterization $(A, B, C)(\cdot) : \Omega \rightarrow \mathcal{R}^N$ is said to be globally identifiable in $\alpha \in \Omega$ if for all $\alpha_i (i = 1, 2)$, the following conditions (i) and (ii) imply $\alpha_1 = \alpha_2$, where

(i) $C(\alpha_1)A^k(\alpha_1)B(\alpha_1) = C(\alpha_2)A^k(\alpha_2)B(\alpha_2), k = 0, 1, 2, \ldots$

(ii) $(A, B, C)(\alpha)$ is minimal [9]

**Proposition 2.2** A parameterization $(A, B, C)(\cdot) : \Omega \rightarrow \mathcal{R}^N$ is said to be globally identifiable in $\alpha \in \Omega$ if for all $\alpha_i (i = 1, 2)$, there exists $S \in \mathcal{R}^{CN}$ such that $\alpha_1 = \alpha_2$ is implied by the following conditions

$$SA(\alpha_1) = A(\alpha_2)S,$$

$$SB(\alpha_1) = B(\alpha_2),$$

$$C(\alpha_1) = C(\alpha_2)S.$$ 

Based on this definition and this proposition, a number of sufficient conditions have been proposed in [8] and [16].

### 3 Problem description

In this section, we shall first introduce the system model and then consider the identifiability of parameterized switched linear systems. In this paper, we consider the following discrete-time parameterized switched linear system

$$x(t+1) = A(\alpha_i)x(t) + B(\alpha_i)u(t),$$

$$y(t) = C(\alpha_i)x(t),$$

where $x(t) \in \mathcal{R}^n, u \in \mathcal{R}^m, A(\alpha_1) \in \mathcal{R}^{n \times n}, B(\alpha_1) \in \mathcal{R}^{n \times m}$ and $C(\alpha_1) \in \mathcal{R}^{p \times n}$. We let $N \triangleq n(n + m + p)$ be the total number of elements. We assume that $(A, B, C)(\alpha_i)$ is minimal for $i = 1, 2$.

The system equations are assumed to be derived from first principles and the elements of the $(A, B, C)$ matrices are therefore either zeros or known functions of some unknown parameters (with certain constraints). Thus if there exists $q$ unknown parameters $\alpha_i \in \Omega_i \subset \mathcal{R}^q$, then the matrices may be written as $A(\alpha_1), B(\alpha_1), C(\alpha_1), \alpha_1 \rightarrow \mathcal{R}^{n \times n}, B(\cdot) : \Omega_1 \rightarrow \mathcal{R}^{n \times m}$ and $C(\cdot) : \Omega_1 \rightarrow \mathcal{R}^{p \times n}$. We illustrate this parameterization through an example.

**Example 1** Consider a parameterization of $(A, B, C)(\alpha_1)$ in which $\alpha_1 = [k_1, k_2, k_3]$

$$A(\alpha_1) = \begin{bmatrix} k_1 & 1 + 2k_2 & k_3 \\ 0 & 1 & 0 \\ k_3 & 0 & 1 \end{bmatrix}, B(\alpha_1) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, C(\alpha_1) = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$ 

We assume that the parameterization for $(A, B, C)(\alpha_1)$ is an affine parameterization (i.e., $f : \mathcal{R}^m \rightarrow \mathcal{R}^n$ is affine if $f(\alpha) = c + l(\alpha)$ where $l$ is linear and $c$ is a constant). For example, the $(A, B, C)(\alpha_1)$ in Example 1 is an affine parameterization. Next, we consider the scenario that the parameters in the nominal system $(A, B, C)(\alpha_1)$ switch to a perturbed “faulty” one at an unknown time $t = T$. For any $t \geq T$, the system has been switched to $(A, B, C)(\alpha_2)$, as shown in Fig. 1.

Assuming $\alpha_1$ is known, the main goal of this paper is to answer the following questions:

**Identifiability:** What is the condition when both the switching time $T$ and a new parameter set $\alpha_2$ can be uniquely identified from input-output data?

**Parameter estimation:** What is the number of samples (after the occurrence of the fault at $T$) that is enough to achieve the desired estimation accuracy for a noisy system?

### 4 Identifiability of parameterized switched linear systems

In section, we shall provide new results on identifiability of parameterized switched linear systems. First of all, we show that the identifiability of parameterized switched linear systems is equivalent to the global identifiability of a parameterized linear time-invariant system. Then a new necessary and sufficient condition is proposed to check the identifiability of a parameterized linear time-invariant system.

**Theorem 4.1** Considering a parameterized switched linear system in eq. (1), both systems $(A, B, C)(\alpha_1)$ and $(A, B, C)(\alpha_2)$ and switching time $T$ can be identified from input-output data if and only if the affine parameterization $(A, B, C)(\alpha)$ is globally identifiable.

---

**Fig. 1:** A diagram for a parameterized switched linear system. The nominal system $(A, B, C)(\alpha_1)$ has been switched to an unknown faulty system $(A, B, C)(\alpha_2)$ at an unknown event time $T$. 
Proof: If the the affine parameterization \((A, B, C)(\alpha)\) is not globally identifiable, then there exist at least two systems (i.e., \(\Sigma_1 \triangleq (A, B, C)(\alpha_1)\) and \(\Sigma_2 \triangleq (A, B, C)(\alpha_2)\)) that give the same input-output, then we can simply design a non-identifiable parameterized switched system \((A, B, C)(\alpha_1)\) and \((A, B, C)(\alpha_2)\) for any switching time \(T\).

On the other hand, there are two scenarios:

1. if there exists a parameterized linear switched system is not identifiable, then there exist two different parameterization (i.e., \(\alpha_1 \neq \alpha_2\)) of \((A, B, C)(\cdot)\) namely \((A, B, C)(\alpha_1)\) and \((A, B, C)(\alpha_2)\) gives the same input-output, which contradicts the definition of the global identifiability of \((A, BC)(\alpha)\).

2. if the switching time \(T\) is not identifiable, then there exist two different switching schemes, i.e., \((A, B, C)(\alpha_2)\) with switching time \(T_2\) and \((A, B, C)(\alpha_3)(\alpha_2 \neq \alpha_3)\) with switching time \(T_3\), that generate the same input-output data. In other words, \((A, B, C)(\alpha_2)\) and \((A, B, C)(\alpha_3)\) generate the same input-output data for anytime after \(T \geq \max(T_2, T_3)\), which contradicts the definition of the global identifiability of \((A, BC)(\alpha)\).

In other words, we can convert to the global identifiability of a switched-time system to the global identifiability of an LTI system. Based on this, we shall focus on the global identifiability of a parameterized LTI system, and propose new results on checking its identifiability. The result is a generalization of a sufficient condition in \([8]\) to a necessary and sufficient condition.

**Theorem 4.2** Let

\[
Z(\alpha_2, \alpha_1) = \begin{bmatrix} A^T(\alpha_2) \otimes I - I \otimes A(\alpha_1) \\ B^T(\alpha_2) \otimes I \\ -I \otimes C(\alpha_1) \\ 0 \end{bmatrix}
\]

\[(2)\]

\[
M(\alpha) = \begin{bmatrix} \text{vec}(A(\alpha)) \\ \text{vec}(B(\alpha)) \\ \text{vec}(C(\alpha)) \\ 0 \end{bmatrix}
\]

\[(3)\]

\[
\hat{Y}(\alpha_1, \alpha_2) = [Z(\alpha_1, \alpha_2) \quad M(\alpha_2 - \alpha_1)]
\]

\[(4)\]

An affine parameterization \((A, B, C)(\alpha) : \Omega_1 \to \mathbb{R}^N\) is globally identifiable if and only if any of the following condition holds:

(i) For any \(\alpha_i \in \Omega_1 (i = 1, 2)\),

\[
\Re(\{Z(\alpha_1, \alpha_2)\}) \cap \Re(M(\alpha_2 - \alpha_1)) = \{0\},
\]

\[(5)\]

(ii) The variable \(G(\alpha_1, \alpha_2) \triangleq \det[\hat{Y}(\alpha_1, \alpha_2)]^T \hat{Y}(\alpha_1, \alpha_2)\) satisfies

\[
G(\alpha_1, \alpha_2) \neq 0 \forall \alpha_1, \alpha_2 \in \Omega_1.
\]

\[(6)\]

**Remark 1** The global identifiability is a property for a given parameterization of a nominal noisless system.

**Remark 2** The condition (ii) is more tractable than the condition (i), which requires us to searching \(\alpha_2\) in eq. \((5)\) in a space of symbolic matrices.

**Remark 3** The condition (ii) is equivalent to showing

\[
G(\alpha_1, \alpha_2) > 0
\]

for any \(\alpha_1\) and \(\alpha_2\) in \(\Omega_1\). Given the knowledge of the nominal system \((A, B, C)(\alpha_1)\), \(G\) becomes a function of \(\alpha_2\), and therefore it is easier to check its positivity. To check the positivity of a multi-variant polynomial \(G\), it is sufficient to check whether it can be decomposed to a sum of squares \([31]\).

We first provide an example to demonstrate Theorem 4.2. The following example \([8]\).

**Example 2** We now illustrate Theorem 4.2 through the following example \([8]\). Recall Example 1,

\[
A(\alpha) = \begin{bmatrix} k_1 & 1 + 2k_2 \\ 0 & k_1 + k_3 \end{bmatrix}, \quad B(\alpha) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C(\alpha) = \begin{bmatrix} 0 & 1 \end{bmatrix}.
\]

The transfer function \(G(z, \alpha) = \frac{1 + 2k_2}{z^2 - (2k_2 + k_3^2)z + k_1(k_1 + k_3)}\) and the parameterization is globally identifiable, by verifying either \((5)\) or \((6)\).

**Proof:** (Theorem 4.2) We first show the sufficient part, the global identifiability is implied that if there exists \(S \in \mathcal{GL}(n)\) such that

\[
A(\alpha_1)S = SA(\alpha_2)
\]

\[
B(\alpha_1) = S B(\alpha_2)
\]

\[
C(\alpha_1)S = C(\alpha_2),
\]

has a unique solution for all \(\alpha_1, \alpha_2 \in \Omega_1\).

Let \(Q_2 = I - S\), and take the vec operator \([2]\) at both sides of these equations, we obtain

\[
(\hat{A}^T(\alpha_2) \otimes I)\text{vec}\{Q_2\} - (I \otimes A(\alpha_1))\text{vec}\{Q_2\} = \text{vec}(A(\alpha_2) - A(\alpha_1))
\]

\[
(\hat{B}^T(\alpha_2) \otimes I)\text{vec}\{Q_2\} = \text{vec}(B(\alpha_2) - B(\alpha_1))
\]

\[
- (I \otimes C(\alpha_1))\text{vec}\{Q_2\} = \text{vec}(C(\alpha_2) - C(\alpha_1)).
\]

Since the parameterization is affine, the above equations can be written as

\[
Z(\alpha_2, \alpha_1)\text{vec}\{Q_2\} = M(\alpha_2 - \alpha_1)
\]

\[(7)\]

The global identifiability is implied there is no solution to the above eq. \((7)\), or in other words

\[
\Re(\{Z(\alpha_2, \alpha_1)\}) \cap \Re(M(\alpha_2 - \alpha_1)) = \{0\}.
\]

Next, we show the necessity part by contradiction. If there exists a nonzero vector in the intersection of the range space, or equivalently, there exists a nonzero vector \(q\) such that

\[
Z(\alpha_2, \alpha_1)q = M(\alpha_2 - \alpha_1).
\]

Then there exists a matrix \(Q \neq I\) such that

\[
A(\alpha_1)Q = QA(\alpha_2)
\]

\[
B(\alpha_1) = QB(\alpha_2)
\]

\[
C(\alpha_1)Q = C(\alpha_2).
\]
If \( Q \in G \mathcal{L}(n) \), then the parameterization is not identifiable by definition. If \( Q \notin G \mathcal{L}(n) \) (i.e., \( Q \) is not invertible), we consider the controllability matrix

\[
W_c(\alpha_1) = [B(\alpha_1), A(\alpha_1)B(\alpha_1), \ldots, A^{n-1}(\alpha_1)B(\alpha_1)] = [QB(\alpha_2), QA(\alpha_2)B(\alpha_2), \ldots, QA^{n-1}(\alpha_2)B(\alpha_2)] = QW_c(\alpha_2),
\]

which leads to \( \Re(W_c(\alpha_1)) \neq \{0\} \). This contradicts to the minimality of the parameterization. □

Alternatively, one can also obtain a negative result using the following condition.

**Proposition 4.3** Let \((A, B, C)() : \Omega_1 \rightarrow \mathbb{R}^N\) be an affine parameterization, assume that \((A, B, C)(\alpha)\) is minimal and define \( \Psi(\alpha) \triangleq \{Z(\alpha, \alpha); \frac{\partial M(\alpha)}{\partial \alpha}\} \), then \((A, B, C)(\alpha)\) is not globally identifiable if (for some constant integer \( r \))

1) \( \text{rank}(\Psi(\alpha)) \neq r \) for some \( \alpha \in \{||\alpha - \alpha_1||_2 < \epsilon\} \), or

2) \( \text{rank}(\Psi(\alpha)) = r \) for all \( \alpha \in \{||\alpha - \alpha_1||_2 < \epsilon\} \), and

\[
\det(\Psi(\alpha)^T\Psi(\alpha)) \neq 0.
\]

The Theorem 4.3 is proved using the following lemma.

**Lemma 4.4** Let \( \Omega \) be an open set in \( \mathbb{R}^n \) and \( f : \Omega \rightarrow \mathbb{R}^m \) be a \( C^k \) map with \( k \geq 1 \). Then if \( \frac{\partial f}{\partial x} \) has constant rank \( r \) in a neighborhood of \( x \), \( f \) is locally injective if and only if \( r = n \).

### 5 Switched systems identification

#### 5.1 Identification

Once \( \hat{T} \) has been detected using the results in the Appendix, there is a need for practical methods for on-line fault detection in a reliable, fast and robust manner to avoid cascading failure. Cascading failures present severe threats to power grid reliability and security, and thus timely detection of cascades is of significant importance. Conventional ways to perform the fault localization is to apply system identification techniques, for example, the prediction error method [2] to identify \( A(\alpha_2) \) using the data after the occurrence of the fault \( (t \geq \hat{T}) \)

\[
x(t + 1) = A(\alpha_2)x(t) + u(t) + w(t).
\]

Once \( A(\alpha_2) \) is estimated, one can compare it with \( A(\alpha_1) \). However, the identification process would require a large number of samples to get a good identification performance as analyzed below.

Define \( X_M = [x(T), \ldots, x(T + M - 1)] \), \( Y_M = [x(T + 1) - u(T), \ldots, x(T + M) - u(T + M - 1)] \), \( W_M = [w(T), \ldots, w(T + M - 1)] \), then the maximum likelihood estimator is

\[
\hat{A}_i = \arg \min_{A_i} \| (Y_M)^T - X_M^TA_i^T \|_2.
\]

The identification error can be computed

\[
\hat{A}_i - A_i(\alpha_2) = (X_MX_M^T)^{-1}X_M(W_M)^T,
\]

which goes to 0 (unbiased) when the number of sample \( M \) goes to \( \infty \). Instead, we would like to locate the fault which is then equivalently to identifying \( A(\alpha_1 - \alpha_2) \). We can convert the above fault-location problem to an optimization problem by stacking successive measurements together. The difference between \( \alpha_1 \) and \( \alpha_2 \) is therefore the location of the fault.

\[
E_M = A(\alpha_1 - \alpha_2)X_M + W_M, \quad \text{in which}
\]

\[
E_M = [e(T) \ e(T + 1) \ \ldots \ e(T + M)].
\]

For any state-variable \( i \)

\[
E_{M_i} = A_i(\alpha_1 - \alpha_2)X_M + W_{M_i}. \quad \text{(10)}
\]

The unknown parameter \( A_i(\alpha_1 - \alpha_2) \) for all \( i \) is sparse since, there are only a very small number of parameters that have been changed. We can solve the \( i \)-th row of \( A_i(\alpha_1 - \alpha_2) \) using the following optimization

\[
a_i^0(M) \triangleq \arg \min_w \|w^T\|_0 \quad \text{subject to: } E_{M_i} = w^TX_M. \quad \text{(11)}
\]

We further can convexify the above optimization \( \text{(11)} \) and obtain a computationally efficient optimization below (known as Lasso). Due to the low computational complexity nature of Lasso, a recursive algorithm can be proposed to localize the fault in Algorithm [1]

\[
a_i^1(M) \triangleq \arg \min_w \|E_{M_i} - w^TX_M\|_2 + \lambda\|w^T\|_1. \quad \text{(12)}
\]

**Algorithm 1** Location for faults using recursive Lasso algorithm

\[
\begin{align*}
\text{for } i = 1, \ldots, n & \text{ do } \\
& \text{ for } M = 1, \ldots, n \text{ do } \\
& \quad \text{ Form } E_{M_i} \text{ and } X_M \text{ from data } \\
& \quad \text{ Solve optimization in eq. (12) and obtain } a_i^1(M) \\
& \quad \text{ end for } \\
& \quad \text{ if } a_i^0(M) \text{ converges (practically, we test if } \|a_i^1(M - 1) - a_i^1(M)\|_2 \leq \epsilon) \text{ then } \\
& \quad \quad \text{ Break; } \\
& \quad \text{ end if } \\
& \quad \text{ We obtain an estimate for the fault } \hat{A}_i(\alpha_1 - \alpha_2) = a_i^1(M). \\
& \text{ end for } \\
& \text{ An estimate for the faults } \\
& \quad \hat{A}^T(\alpha_1 - \alpha_2) = [\hat{A}_1^T(\alpha_1 - \alpha_2) \ \ldots \ \hat{A}_n^T(\alpha_1 - \alpha_2)].
\end{align*}
\]

#### 5.2 Performance guarantee – noiseless case –

In this section, we shall provide theoretical guarantee for the proposed algorithm in the previous section in order to solve Problem 1.2. This is challenging partially due to the basis functions for the optimization is not orthogonal [36]. These sufficient conditions that we are going to provide are somewhat conservative and there are a large number of examples that the algorithm works well in practice even when such conditions fail to be satisfied [4, 41], yet it is important to understand why this algorithm works and how many samples it requires to correctly localize the faults in different settings.
Even when the identifiability condition has been satisfied, there are requirements on input design to allow the distinction between $\alpha_1$ and $\alpha_2$ for the actual data. For instance, if $x(T) = 0$ and $u(t) = 0$ for any $t \geq T$, then $y_M = 0$ for all $M$ in the noiseless case, and therefore, we can not identify $A(\alpha_1 - \alpha_2)$ independent of methods. Such input requirements are typically called “persistence of excitation”\cite{44}.

**Definition 5.1** We say a system is persistently excited by an input $x(T)$ and $u(t)$ for $t \geq T$, if there exists a positive integer $M^o$ such that there do not exist two parameter sets $(A, B, C)(\alpha_1) \neq (A, B, C)(\alpha_2)$ that generate $D_M^o$, in which $D_M = \{y(T), y(T+1), \ldots, y(T+M)\}$.

Given a parameterized switched linear system, if the system is identifiable and the output is persistently excited, then there does not exist $\alpha_3$ such that

$$E_M = A_i(\alpha_1 - \alpha_3)X_M, \ \forall \ M, i.$$\textbf{Remark 4} The identifiability and persistent excitation guarantee that the uniqueness of solution in the optimization when the number of sample is large.

Next, we are now characterizing the smallest $M^o$ so that $D_M$ can differentiate any two parameter sets in the noiseless case, which would then be a lower bound for the noisy case. When the system is persistently excited, from Cayley-Hamilton theorem\cite{9}, we can set $M^o = n$ that suffices to differentiate any two different parameter sets $\alpha_1$ and $\alpha_2$ in the noiseless case.

**Definition 5.2** The spark of a given matrix $X_M^T$, i.e., $\text{Spark}(X_M^T)$, is the smallest number of columns of $X_M^T$ that are linearly dependent.

The spark of our dictionary matrix depends on the dynamics as well as the sample number $M$.

**Proposition 5.3**\cite{33} For any vector $w \in \mathbb{R}^n$, there exists one unique signal $w$, such that $E_M^T = X_M^Tw$ with $\|w\|_0 = s$ if $\text{Spark}(X_M^T) > 2s$.

**Remark 5** This Proposition offers a sufficient condition for checking whether the true solution $A_i(\alpha_1 - \alpha_2)$ has sparsity $s$ and $\text{Spark}(X_M^T(\alpha_1 - \alpha_2)) > 2s$, then there does not exist sparser solution to eq. (11), the true solution is the only solution of the optimization (11). We can set $M^o = \text{max}(M^o(i))$.

Having shown the theoretical guarantee for $l_0$ optimization, i.e., the solution of optimization (11) is the same as the true solution when the sample number is at least $M^o$. Next, we provide a sufficient condition to guarantee the uniqueness of the lasso solution which can link the solution of optimization (11) to that of optimization (12). Then, using the similar deduction, we conclude that when the lasso in eq. (12) has a unique solution, the solution is the true solution.

**Lemma 5.4**\cite{44} For any $E_M$, $X_M$, and $\lambda > 0$, let $S$ be the true support of $A_i(\alpha_1 - \alpha_2)$, if

$$\mathcal{N}(\{X_M^T\})_{S,\text{SUSC},S} = \{0\},\tag{13}$$

then the lasso solution is unique.

**Remark 6** A summary of the results in the subsection: 1) If a parameterization is identifiable and minimal and the system is persistently excited, then $\alpha_2$ is uniquely identifiable.

2) In addition, given $M$ and if eq. (13) and condition in Proposition 5.3 are satisfied, then the solution of the lasso in eq. (12) is equal to the true solution. If eq. (13) and condition in Proposition 5.3 are not satisfied, we can still apply the same algorithm but can hardly guarantee the closeness between solution of the lasso in eq. (12) and the true solution.

**5.3 Performance guarantee – nois y case –**

We first define the following terms.

- Let $x_M = [x(T) - \hat{x}(T), x(T+1) - \hat{x}(T+1), \ldots, x(T+M-1) - \hat{x}(T+M-1)]$, where $T$ is the time at which the system switch is detected.
- Denote the maximum eigenvalue of $\Sigma$ as $\gamma = \lambda_{\text{max}}(\Sigma)$.
- Let $\Sigma = E[x^T(T)x(T)]$. Denote the maximum eigenvalue of $\Sigma$ as $\gamma = \lambda_{\text{max}}(\Sigma)$.
- Let $\delta$ be the maximum entry of $\max\{A(\alpha_1 - \alpha_2)\}$.
- Let $\sigma$ denote the maximum singular value of $A(\alpha_1 - \alpha_2)$. We further assume that $\sigma < 1$.

**Theorem 5.5** Consider the problem of learning the support $S(\|S\|_0 = s)$ of row $A_i$ from the discrete-time trajectory $\{x(t)\}_{T \leq t \leq T+M}$. If the number of samples

$$M \geq \max\left\{144s\varphi \alpha^2 \beta^2 \lambda^2 \log \frac{4n\varphi}{\delta}, 9\lambda^2 \alpha^{-2}\xi \log \frac{4n}{\delta}\right\},$$

then with probability greater than $1 - \delta$, the proposed algorithm recovers the correct signed support of $A_i(\alpha_1 - \alpha_2)$ for any $\lambda \leq (1 + \alpha/3)^{-1}A_{\text{min}}^{\text{SC}}$. Here $\varphi = 8n\left\{(1-\sigma)^{-2} + \gamma + 2(1-\sigma)^{-1/2}\right\}$ and $\xi = 8n\gamma(1-\sigma)^{-1}$.

The proof to this theorem relies on a number of technical lemmas (the proofs to these lemmas are in Appendix) as follows:

**Lemma 5.6**\cite{33} The optimization (12) correctly recovers the support of $A_i$ if all of the following conditions hold:

$$\|\hat{G}\|_\infty \leq \min\left\{\frac{\alpha_0}{3}, A_{\text{min}}^{\text{SC}}, \frac{\lambda}{4\sigma}\right\}.$$\textbf{Lemma 5.7} Recall that $Q \triangleq E[|X_M^T X_M^T|]$, $\hat{Q} \triangleq \frac{1}{M}X_M^T X_M^T$, then the following inequality holds

$$\mathbb{P}(\|\hat{Q}_{S,\text{SC}}S, S\|_1 > \epsilon) \leq 2ns \exp\left\{-\frac{\epsilon^2 M}{\varphi}\right\}.\tag{14}$$
Proof: See Appendix.

Lemma 5.8 Recall that \( \hat{G} \triangleq \frac{1}{M} X_M (E_M)^T \), then its tail probability satisfies

\[
P(\|\hat{G}\|_\infty > \epsilon) \leq 2n \exp \left\{ - \frac{Me^2}{\xi} \right\}.
\]  

(15)

Proof: See Appendix.

Remark 7 Our result has additional terms involving \( \gamma \) in addition to the terms obtained in \( \ref{5.5} \). These terms are to take into account the effect of the first system before it has been switched.

Remark 8 There exists literature that studies the estimation accuracy when \( X_M \) are i.i.d. Gaussian variables \( \ref{5.5} \). This iid condition is relaxed in this paper, but our bound might be loose. An immediate future direction is to improve the tightness of Theorem \ref{5.5}.

Proof: (to Theorem \ref{5.5}) To prove this, we impose that both probability of condition on \( Q \) failing bounded by \( \delta_1 \) and the probability of the condition on \( \hat{G} \) failing bounded by \( \delta_2 \), in here \( \delta_1 + \delta_2 = \delta \).

First, we choose \( \lambda \leq (1 + \alpha / 3)^{-1} \frac{A_{\text{min}} \beta}{4k} \) which would lead to \( \lambda \alpha \leq \frac{A_{\text{min}} \beta}{4k} - \lambda \). Using Lemma \ref{5.8} pick \( \epsilon = \frac{\lambda \alpha}{\xi} \) and set \( \delta_1 \geq 2n \exp \left( - \frac{M \lambda^2 \alpha^2}{144s^2} \right) \), then

\[
\lambda^2 \geq \left( \log \frac{2n}{\delta_1} \right) \left( 8n \gamma + (1 - \sigma)^{-1} \right) 9 \alpha^{-2} M^{-1}.
\]

We obtain that the number of samples must satisfy to make the above inequality hold

\[
M \geq 9 \lambda^{-2} \alpha^{-2} (8n \gamma + (1 - \sigma)^{-1}) \log \frac{2n}{\delta_1}.
\]

Similarly, using Lemma \ref{5.7} we set \( \epsilon = \frac{\alpha \beta}{12 \sqrt{\pi}} \) and \( \delta_2 \geq 2ns \exp \left( - \frac{\alpha^2 \beta M}{144s^2} \right) \), then

\[
M \geq 144s \varphi \alpha^{-2} \beta^{-2} \log \frac{2ns}{\delta_2}.
\]

After some further manipulation, we can obtain the lower bound for \( M \) by letting \( \delta_1 = \delta_2 = \frac{\delta}{2} \).

6 Simulations

Consider a linear system that has 300 variables, \( A(\alpha_1) \) is a random matrix, \( B = C = I \) and let the noise variance \( \sigma^2 = .01 \). At time instant \( t = 0 \), faults occurred in \([1, 1]\) and \([1, 5]\), i.e., \( A(\alpha_2 - \alpha_1)_{[1, 1]} = 1 \) and \( A(\alpha_2 - \alpha_1)_{[1, 5]} = 1 \).

Following the procedure in Algorithm \ref{1} we aim to localize these faults. In Fig. 2 the estimated \( A_{1}(\alpha_2 - \alpha_1) \) are depicted starting at the time point \( k = 5 \) till \( k = 15 \). When \( k = 10 \), the estimated fault is close to the estimated when \( k = 9 \). The support \( \text{supp}(A_{1}(\alpha_2 - \alpha_1)) \) remains unchanged afterwards, which is consistent with the faults location. We apply the same algorithm to a randomly chosen variable 10.

Fig. 2: (above) Plot of \( A_{1}(\alpha_2 - \alpha_1) \) over different sample numbers. Every trajectory represents an estimated element in \( A_{1}(\alpha_2 - \alpha_1) \). (below) A randomly chosen node (e.g., 10) and the plot of \( A_{10}(\alpha_2 - \alpha_1) \) over different sample numbers. Every trajectory represents an estimated element in \( A_{10}(\alpha_2 - \alpha_1) \).

The elements of estimated \( A_{10}(\alpha_2 - \alpha_1) \) are depicted starting at the time point \( k = 5 \) till \( k = 15 \). The estimated elements are fluctuating around 0 caused by measurement noise.

Further experiments are performed and their associated discussions shall be demonstrated in the Appendix.

7 Discussions

We model fault/attack detection problems as parameterized switched linear system identification problems. We propose a necessary and sufficient condition to check the identifiability of a parameterized switched linear system. The concept of identifiability is important since it provides a theoretical guarantee on the meaningfulness of the identified parameters using any identification algorithm. Assuming that only a small number of parameters are changed, we formulate the fault localization problem as a sparse linear regression problem. We propose a sufficient condition that links the number of samples that is required to achieve a “close-to-true” solution. Currently, the identifiability and estimation algorithm only applies to linear systems. Future work may aim to extend our model to nonlinear systems.

These tools have powerful applications in both engineering and biological systems. In engineering systems, they can guide the design of reliable infrastructures for power and water distribution. In biological systems, they can be used to monitor the reaction dynamics of genes, proteins, and metabolites \([10, 11, 33, 46]\), thereby facilitating disease.
8 Acknowledgements

We acknowledge Dr. Jorge Gonçalves, Dr. Yilin Mo, Dr. Keith Glover, Dr. Lennart Ljung, Dr. John C. Doyle, and Mr. Roel Dobbe with their great supports.

References


A Proofs

The proof to the Theorem 5.5 relies on a number of technical lemmas as follows:

Lemma A.1 [Laplace transform method] Let $X_1, \ldots, X_n$ be independent random variables and $S = \sum_i X_i$ and the moment generating function then we have

$$P(S > t) \leq \inf_{\theta > 0} \left\{ e^{-\theta t} \prod_i m_X_i(\theta) \right\}.$$

Lemma A.2 [Frechet inequality]

$$P(A) + P(B) - 1 \leq P(A \cap B) \leq \min(P(A), P(B)).$$

Proof: [Lemma 5.7] Let $z$ be a vector of independent and identically distributed Gaussian random variables with covariance $I_{(M+1)n}$. The $(i,j)$-th entry of $\hat{Q}$ can be decomposed into

$$\hat{Q}_{ij} = \frac{1}{M} \sum_{k=1}^{M} x_i(T + k)x_j(T + k)$$

$$= \frac{1}{M} \left[ x_i(T + 1) \cdots x_i(T + M) \right] \begin{bmatrix} x_j(T + 1) \\ \vdots \\ x_j(T + M) \end{bmatrix}$$

$$= \frac{1}{M} z^T R(i,j) z,$$

where

$$R(i, j) = 0.5 \left( \Psi_i^T \Psi_j + \Psi_j^T \Psi_i \right)$$

and

$$\Psi_j = \begin{bmatrix} 0 & 0 & 0 & \cdots \\ (\Sigma^{1/2})_{ij} & 0 & 0 & \cdots \\ (A\Sigma^{1/2})_{ij} & I_j & 0 & \cdots \\ (A^2\Sigma^{1/2})_{ij} & A_j & I_j & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots \\ (A^{M-1}\Sigma^{1/2})_{ij} & (A^{M-2})_j & (A^{M-3})_j & \cdots & I_j \end{bmatrix}.$$}

Since $R(i, j)$ is a real symmetric matrix, it permits the eigenvalue decomposition $R(i, j) = U^T \Lambda U$, where $\Lambda$ is a diagonal matrix with eigenvalues $\lambda_i (i = 1 : Mn)$ and $U$ is a unitary matrix. The expected value of $\hat{Q}_{ij}$ satisfies

$$Q^*_{ij} \triangleq E[Q_{ij}] = E\left[ \frac{1}{M} z^T U^T \Lambda U z \right]$$

$$= E\left[ \frac{1}{M} \sum_{i=1}^{Mn} v_i z_i^2 \right] = \frac{1}{M} \sum_{i=1}^{Mn} v_i.$$

Last equation holds since $E[z_i^2] = 1$ for any $i$.

Let $m_v$ denote the moment generating function of the random variable $v$, then we bound the distribution of $|\hat{Q}_{ij} - Q^*_{ij}| > \epsilon$
as follows

\[
P(Q_{ij} - Q_{ij}^* > \epsilon) \\
\quad = P \left( \frac{1}{M} \sum_{i=1}^{M} v_i (z_i^2 - 1) > \epsilon \right) \\
\quad = P \left( \frac{1}{M} \sum_{i=1}^{M} v_i z_i^2 > \epsilon + \frac{1}{M} \sum_{i=1}^{M} v_i \right) \\
\quad \leq \inf_{\psi > 0} \left\{ e^{-\psi t} \prod_{i=1}^{M} m_i v_i z_i^2 / M(\psi) \right\}, \quad t = \epsilon + \frac{1}{M} \sum_{i=1}^{M} v_i \\
\quad \leq \inf_{\psi} \left\{ e^{-\psi t} \prod_{i=1}^{M} (1 - 2v_i \psi / M)^{-1/2} \right\} \\
\quad = \inf_{\psi > 0} \exp \left\{ -\psi t + \log \left( \prod_{i=1}^{M} (1 - 2v_i \psi / M)^{-1/2} \right) \right\} \\
\quad = \inf_{\psi > 0} \exp \left\{ -\psi t - \frac{1}{2} \sum_{i=1}^{M} \log (1 - 2v_i \psi / M) \right\} \\
\quad \leq \inf_{\psi > 0} \exp \left\{ -\psi \left( \epsilon + \frac{1}{M} \sum_{i=1}^{M} v_i \right) + \ldots \\
\quad + \frac{1}{2} \sum_{i=1}^{M} \left( 2v_i \psi / M + (2v_i \psi / M)^2 \right) \right\} \\
\quad = \inf_{\psi > 0} \left\{ -\psi \epsilon + 2\psi^2 \sum_{i=1}^{M} \left( v_i / M \right)^2 \right\} \\
\quad = \exp \left\{ -\frac{M^2 \epsilon^2}{\sum_{i=1}^{M} v_i^2} \right\} \\
\quad = \exp \left\{ -\frac{M \epsilon^2}{8n (1 - \sigma)^{-2} + (1 - \gamma)^{-2} + 2(1 - \sigma)^{-1} (1 - \gamma)^{-1}} \right\} \\
\quad = \exp \left\{ -\frac{M \epsilon^2}{n \epsilon^2 \theta} \right\}
\]

where \( \varphi = \{8(1 - \sigma)^{-2} + (1 - \gamma)^{-2} + 2(1 - \sigma)^{-1} (1 - \gamma)^{-1} \}^{-1} \). Using Frechet inequalities, we obtain (14).

In the above derivation, inequality (i) holds due to the Laplace transform method in Lemma A.1 and inequality (ii) used the fact that \( \forall |x| < 0.5, \quad \log(1 - x) > -x - x^2 \) and inequality (iii) holds due to Lemma A.3 and the increasing property of an exponential function.

\[\square\]

**Lemma A.3** Let \( v_i \) denote the \( i \)-th largest eigenvalue of \( R(i, j) \), and assume \( \sigma = \sigma_{\text{max}}(A) < 1 \), then the following inequalities hold

\[
\sum_{i=1}^{M} v_i^2 \leq Mn \left\{ (1 - \sigma)^{-2} + (1 - \gamma)^{-2} + 2(1 - \sigma)^{-1} (1 - \gamma)^{-1} \right\}.
\]
Proof: Let

\[ \Psi_1^j = \begin{bmatrix}
0 & 0 & 0 & \cdots \\
0 & 0 & 0 & \cdots \\
0 & I_j & 0 & \cdots \\
0 & A_j & I_j & 0 & \cdots \\
0 & (A^2)_j & A_j & I_j & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & (A^{M-2})_j & (A^{M-3})_j & \cdots & I_j
\end{bmatrix} \quad \text{(S1)} \]

\[ \Psi_2^j = \begin{bmatrix}
0 & 0 & 0 & \cdots \\
0 & 0 & 0 & \cdots \\
0 & 0 & (\Sigma_{1/2})_j & 0 & \cdots \\
0 & 0 & (A\Sigma_{1/2})_j & 0 & \cdots \\
0 & 0 & (A^2\Sigma_{1/2})_j & 0 & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 0 & \cdots \\
\end{bmatrix} \quad \text{(S2)} \]

Notice that the following condition holds

\[ \sigma_{\text{max}}(\Psi_j) \leq \sigma_{\text{max}}(\Psi_1^j) + \sigma_{\text{max}}(\Psi_2^j) \leq (1 - \sigma_{\text{max}})^{-1} + (1 - \gamma)^{-1} \]

The maximum eigenvalue satisfies

\[ |v_i| \leq \sigma_{\text{max}}(0.5(\Psi_i^T \Psi_j + \Psi_j^T \Psi_i)) \leq \sigma_{\text{max}}(\Psi_i^T \Psi_j) \leq \sigma_{\text{max}}(\Psi_i)\sigma_{\text{max}}(\Psi_j) \leq (1 - \sigma)^{-2} + (1 - \gamma)^{-2} + 2(1 - \sigma)^{-1}(1 - \gamma)^{-1} \]

\[ \square \]

Proof: (Lemma 5.8) We now attempt to bound the value of \( G = \frac{1}{T}X(W_j)^T \). Define

\[ H(j) = 0.5(\Phi_i^T \Psi_j + \Psi_j^T \Phi_i), \quad \text{(S3)} \]

\[ \Phi_i = \begin{bmatrix} 0 & 0 & I_i & 0 & 0 & I_i & \cdots & \cdots & I_i \\
0 & 0 & 0 & I_i & 0 & 0 & \cdots & \cdots & I_i \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\end{bmatrix}, \quad \text{(S4)} \]

\[ G_j = \frac{1}{M}z^T H(j)z. \quad \text{(S5)} \]
Let $q_i$ be the eigenvalue of $H(j)$. By using similar argument with Lemma 5.7,

\[
\mathbb{P}(G_{ij} > \epsilon) = \mathbb{P}\left( \frac{1}{M} \sum_{i=1}^{M_n} q_i z_i^2 > \epsilon \right)
\leq \inf_{\psi > 0} \left\{ e^{-\psi} \prod_{i=1}^{M_n} m_{q_i z_i^2 / M(\psi)} \right\}
\leq \inf_{\psi > 0} \left\{ e^{-\psi} \prod_{i=1}^{M_n} (1 - 2q_i \psi / M)^{-1/2} \right\}
\leq \inf_{\psi > 0} \exp \left\{ -\psi \epsilon - \frac{1}{2} \sum_{i=1}^{M_n} \log (1 - 2q_i \psi / M) \right\}
\leq \inf_{\psi > 0} \exp \left\{ -\psi \epsilon + \frac{\psi^2}{2} \sum_{i=1}^{M_n} q_i^2 \right\}
\leq \exp \left\{ -\frac{\epsilon^2 M^2}{8 \sum_{i=1}^{M_n} q_i^2} \right\}
\leq \exp \left\{ -\frac{\epsilon^2 M}{8n \{(1 - \sigma)^{-1}(1 - \gamma)^{-1}\}} \right\}
\leq \exp \left\{ -\frac{M \epsilon^2 \phi}{n} \right\}
\]

where $\phi = 8(1 - \sigma)(1 - \gamma)$. Using Frechet inequalities, we obtain (15).

**Lemma A.4** [33] Let $q_i$ denote the $i$-th largest eigenvalue of $H(j)$ and assume $\sigma = \sigma_{\text{max}}(A) < 1$, then

\[
\sum_{i=1}^{M_n} q_i = 0,
\sum_{i=1}^{M_n} q_i^2 \leq nM \{(1 - \sigma)^{-1}(1 - \gamma)^{-1}\}.
\]

**Proof:** This proof is similarly to Lemma A.3 From

\[
|q_i| \leq \sigma_{\text{max}}(\Psi_i)\sigma_{\text{max}}(\Psi_j)
\leq (1 - \sigma)^{-1}(1 - \gamma)^{-1},
\]

we obtain

\[
\sum_{i=1}^{M_n} q_i^2 \leq Mn \sigma_{\text{max}}(H(j))
\leq Mn (1 - \sigma)^{-1}(1 - \gamma)^{-1}.
\]

**B Detection**

In this section, under the assumption that the parameterization of $(A,B,C)(\alpha_1)$ is identifiable, we focus on the detection problem, i.e., how to detect the fault occurrence time $T$ and the fault location, i.e., $(A,B,C)(\alpha_1 - \alpha_2)$ using data.

**Example 3** We consider a parameterization of $(A,B,C)(\alpha_1)$ in which $\alpha_1 = [k_1, k_2, k_3]$ as in Example 2. At time $T$, a fault occurs which leads to a change to one of the parameters, e.g., $k_2$. Then the parameters are switched to \( \alpha_2 = [k_1, k_2 + \delta, k_3] \), i.e.,

\[
A(\alpha_2) = \begin{bmatrix} k_1 & 1 + 2k_2 + 2\delta \k_2 \k_3 \end{bmatrix},
\]

\[
B(\alpha_2) = \begin{bmatrix} 0 \1 \end{bmatrix}, C(\alpha_2) = \begin{bmatrix} 1 & 0 \end{bmatrix}.
\]
If we were able to identify \((A,B,C)(\alpha_1 - \alpha_2)\) from data, i.e.,
\[
A(\alpha_1 - \alpha_2) = \begin{bmatrix} 0 & 2\delta_k \\ 0 & 0 \end{bmatrix},
\]
\[
B(\alpha_1 - \alpha_2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C(\alpha_1 - \alpha_2) = \begin{bmatrix} 1 & 0 \end{bmatrix},
\]
then this would tell us, since the parameterization is known a priori, that \(k_2\) has been changed by \(\delta_k\).

Consider a parameterized switched linear system \((A,B,C)(\alpha_1)\) (assuming that it is identifiable using results in the previous section) with process noises
\[
x(t + 1) = A(\alpha_1)x(t) + B(\alpha_1)u(t) + w(t) \\
y(t) = C(\alpha_1)x(t),
\]
(S1)
in which the noise \(w\) is the white Gaussian process \(w(t) \sim \mathcal{N}(0, I)\).

The system switches to \((A,B,C)(\alpha_2)\) at an unknown time \(T\). Let us consider the simplest case that \(B(\alpha_i) = C(\alpha_i) = I\) for \(i = 1, 2\) for now (which will be relaxed in Section C), and \(u(k)\) is known.

In such a setup, we can define one-step ahead prediction
\[
y(t + 1|t) = A(\alpha_1)x(t) + u(t).
\]
(S2)
We can define the error between measurement and prediction, i.e., \(e(t + 1) \triangleq y(t + 1|t) - y(t + 1)\). It is easy to see that \(e(t + 1)\) should be Gaussian when \(t + 1 < T\). There are many methods in the literature to test whether a series can be considered to be white noise, for example, difference sign test, or rank test. Here we introduce the turning point test as one example [25]; it examines a series \(e[i](t)\) to test whether it is purely random.

C Assumption relaxation

Here in the section, we shall relax the assumption we made in the paper on \(B = C = I\). Consider a discrete-time linear system \((A,B,C)(\alpha)\) in eq. (1) and we can compute its parametric transfer function
\[
y(k) = G(z, \theta_1)u(k), \\
\theta_1 = h(\alpha_1).
\]
(S1)

When fault occurs, i.e., \(\alpha_1\) has been switched to \(\alpha_2\), we again compute the difference between the observation and the prediction
\[
y(k|k - 1) = G(z, \theta_1)u(k) \\
y(k) = G(z, \theta_2)u(k) \\
\theta_i = h(\alpha_i) \forall i = \{1, 2\},
\]
(S2)
and obtain
\[
e(k) \triangleq y(k|k - 1) - y(k) = (G(z, \theta_1) - G(z, \theta_2))u(k),
\]
where \(z^{-1}\) is the time delay operator.

Example 4 We shall use example 3.5 in [1], that
\[
A = \begin{bmatrix} \alpha_1 & 1 + 2\alpha_2 \\ 0 & \alpha_1 + \alpha_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix},
\]
\[
G(z, \theta) = \frac{1 + 2\alpha_1 z}{z^2 - (2\alpha_1 + \alpha_2)z + \alpha_1(\alpha_1 + \alpha_2)} \triangleq \frac{\theta_4 + \theta_5 z}{z^2 - \theta_1 z + \theta_2}.
\]
Therefore, according to the definition
\[
\theta = \begin{bmatrix} -(2\alpha_1 + \alpha_2) \\ \alpha_1(\alpha_1 + \alpha_2) \\ 2\alpha_1 \\ 0 \\ 1 \end{bmatrix} \triangleq h(\alpha).
\]
(S3)

Remark 9 From this example, \(h(\cdot)\) can be nonlinear even though \((A,B,C)(\cdot)\) is an affine parameterization. Also, note that one parameter change in \(\alpha\) could lead to changes in a large number of parameters in \(\theta\).
Example 5 (Continued) In this example, for any \( k \geq T + 2 \), we have

\[
e(k) - \theta_1 e(k - 1) + \theta_2 e(k - 2) = \theta_3 u(k - 1) + \theta_4 u(k - 2),
\]

in here, \( \theta_1 - \theta_2 = [\theta_1 \theta_2 \theta_3 \theta_4]^T \).

We can convert it back to a state-space form with an affine parameterization in \( \theta_1 - \theta_2 \)

\[
\begin{bmatrix}
  e(k) \\
  e(k - 1)
\end{bmatrix} =
\begin{bmatrix}
  1 & -1 \\
  1 & 0
\end{bmatrix}
\begin{bmatrix}
  e(k - 1) \\
  e(k - 2)
\end{bmatrix} +
\begin{bmatrix}
  \theta_3 \\
  \theta_4
\end{bmatrix}
\begin{bmatrix}
  u(k - 1) \\
  u(k - 2)
\end{bmatrix}.
\]

We now generalize this idea formally, given a multi-input-multi-output (MIMO) system

\[
y(t) = G(z, \theta)u(t),
\]

(S4)

\[
y(t) = [y_1(t), \ldots, y_p(t)]^T
\]

and \( u(t) = [u_1(t), \ldots, u_m(t)]^T \) are the outputs and inputs at time \( t \), respectively.

Let \( p_i = \sum_{k=1}^{K_{yi}} \theta^{yi}(k)z^{-k} \) be the lowest common multiple of the denominator of \( G_{ij}(z) \) for all \( j \), and let \( \theta_{ij} \triangleq p_iG_{ij} = \sum_{k=1}^{K_{yi}} \theta^{yi}(k)z^{-k} \). Then model of system (S4) can be written as (\( \forall i \))

\[
y_i(t) = [A_{yi}(t), A_{ui}(t)]] \quad W_i \triangleq A(t)W_i, \quad t \geq T + K_{yi}
\]

(S5)

where

\[
A_{yi}(t) = [y_1(t - 1), \ldots, y_i(t - K_{yi})] \ldots
\]

\[
y_q(t - 1), \ldots, y_q(t - K_{yq})],
\]

\[
A_{ui}(t) = [u_1(t - 1), \ldots, u_i(t - K_{ui})] \ldots
\]

\[
u_p(t - 1), \ldots, u_p(t - K_{up})],
\]

and

\[
W_i = \begin{bmatrix}
  \theta_{yi}(1) & \ldots & \theta_{yi}(K_{yi}) & \ldots & \theta_{ui}(K_{ui})
\end{bmatrix}^T.
\]

(S7)

For any \( i \), the model (S5) can be further written in a more compact form

\[

\begin{bmatrix}
  y_i(T + K_{yi}) \\
  y_i(T + K_{yi} + 1) \\
  \vdots \\
  y_i(T + K_{yi} + M)
\end{bmatrix} =
\begin{bmatrix}
  A(T + K_{yi}) \\
  A(T + K_{yi} + 1) \\
  \vdots \\
  A(T + K_{yi} + M)
\end{bmatrix}
\begin{bmatrix}
  W_i, \quad t \geq T + K_{yi}
\end{bmatrix}.
\]

(S8)

We can then apply recursive lasso in Algorithm 1 to identify \( W_i \) for all \( i \) and then obtain a good estimation of \( \theta_1 - \theta_2 \).

Remark 10 Once, we obtain \( \theta_1 - \theta_2 \), we can compute the \( \alpha_2 - \alpha_1 \) since \( h \) is known given a parameterization.

D Further experiments

In this section, we shall perform some further experiments on a number of systems that have 100 * 100 variables \( s \in \{1, 2, 3, 4, 5, 6\} \), in which \( A(\alpha_1) \) is a randomly generated matrix, \( B = C = I \) and the noise variance \( \sigma^2 = .01 \). At time instant \( t = 0 \), faults occurred in \([1, 1]\) and \([1, 5]\), i.e., \( A(\alpha_2 - \alpha_1)[1, 1] = 1 \) and \( A(\alpha_2 - \alpha_1)[1, 5] = 1 \). Following the procedure in Algorithm 1, we aim to localize these faults. In the following Figure, the estimated \( A_1(\alpha_2 - \alpha_1) \) are depicted starting at the time point \( k = 5 \) till \( k = 15 \).

For most of the figures below, we observe that when \( k = 10 \), the estimated fault is close to the estimated when \( k = 9 \). The support \( \text{supp}(A_1(\alpha_2 - \alpha_1)) \) remains unchanged afterwards which are consistent with the faults location. It indicates that the proposed recursive algorithm can correctly localize the faults and the number of samples that requires is somehow independent of the dimension of the system but the number of faults associated with a node. This inconstant experimental results (with respect to Theorem 5.5) motivate further analysis to have a tighter bound on the number of required samples.

The mean square error (MSE) in the following figures is the \( l_2 \) penalty in eq. (12) indicating the fitness error, we aim to have it close to 0. On the other hand, we would like to have the solution is sparse (having a small number of nonzero entries). Observing how the MSE and sparsity of the solution evolve gives a hint on how to intuitively tune \( \lambda \) in eq. (12). More specifically, when we observe that the MSE is very close to 0 but the solution is not very sparse, then we can increase \( \lambda \) and the other way around. Note that, we may use different \( \lambda \) for different state-variable and different sample number in Algorithm 1 though in all the experiments we perform, we use a single \( \lambda = 2 \).
Fig. S3: Identification results for randomly generated systems with 100, 200, 300, 400, 500, 600 state variables. For figure number $2s - 1$, in which $s \in \{1, 2, 3, 4, 5, 6\}$, it shows the mean square error of the identification; for figure number $2s$, it shows the localization results for one of the state-variables. Both figures are generated by a systems with $100 \times s$ state variables.