Inverse Problems for Matrix Exponential in System Identification: System Aliasing

Zuogon Yue, Johan Thunberg* and Jorge Gonçalves

***Abstract*— This note addresses identification of the A-matrix in continuous time linear dynamical systems on state-space form. If this matrix is partially known or known to have a sparse structure, such knowledge can be used to simplify the identification. We begin by introducing some general conditions for solvability of the inverse problems for matrix exponential. Next, we introduce "system aliasing" as an issue in the identification of slow sampled systems. Such aliasing give rise to non-unique matrix logarithms. As we show, by imposing additional conditions on and prior knowledge about the A-matrix, the issue of system aliasing can, at least partially, be overcome. Under conditions on the sparsity and the norm of the A-matrix, it is identifiable up to a finite equivalence class.**

I. Introduction

Time-series models in engineering, economics and biology can often be represented by state-space models. In the example of gene regulatory networks (evolving close to equilibria), the structure of the right-hand side defines pathways, from which conclusions can be drawn about possible diseases. For continuous-time linear dynamics, the state-space form has an A-matrix, which reveals the direct connections between the states. In many dynamical systems, as the aforementioned ones, the structure of this matrix is either partially known or known to be sparse. This paper investigates what such criteria can be used to identify the A-matrix.

Estimating continuous-time systems from discrete-time measured data is an important part of the field of systems identification, see e.g. [1]–[3]. However, with low sampling rates, the identification of continuous-time systems becomes particularly challenging, manifested in the lack of comprehensive studies. In the presence of "system aliasing", the discrete-time signals do not contain certain information about the continuous-time signals. As a result, even though the discrete-time system can be identified, the selection of the continuous-time model may be ambiguous. This note sheds light on how this issue of system aliasing complicates the identification.

Before addressing the issues related to system aliasing, we first recall some results on matrix logarithms and exponentials – the questions of existence and uniqueness are addressed. The results are given as algebraic conditions for obtaining a unique A-matrix within a set. In contrast to these results, the note proceeds by first providing the minimal sampling frequency such that system aliasing is avoided. Then we consider the issue of system aliasing. When know that the A-matrix is sparse, we observe that allowing for "aliased" representations might lead to A-matrices that are more sparse among the aliased solutions, thus exposing more structures in the state generation. This note gives a mathematical definition of "system aliases", and study how to select the sparest one among the aliases of the underlying systems. Refer to [4] for more preliminaries and all proofs.

This work was supported by Fonds National de la Recherche Luxembourg (9247977 & 8864515).

All authors are with Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6, avenue du Swing, L-4367 Belvaux, Luxembourg.

*For correspondence, johan.thunberg@uni.lu

II. Preliminaries

Let

$$A = \bar{A} + E \text{ and } D \in \mathbb{R}^{p\times n^2},$$

where $\bar{A} \in \mathbb{R}^{n\times n}$ and $E \in \mathscr{S} \subseteq \mathbb{R}^{n\times n}$. This note addresses properties that must hold for $(A, D, \mathscr{S})$ or $(\bar{A}, E, D, \mathscr{S})$ in order to guarantee that E can be determined from $D\mathrm{vec}(\exp(A))$. In the following cited definitions and theorems, we adopt the notations in [5].

A. Principal logarithm

Theorem 1 (principal logarithm [5, Thm. 1.31])**.** *Let $P \in \mathbb{C}^{n\times n}$ have no eigenvalues on $\mathbb{R}^-$. There is a unique logarithm X of P all of whose eigenvalues lie in the strip $\{z : -\pi < \mathrm{im}(z) < \pi\}$. We refer to X as the principal logarithm of P and write $X = \mathrm{Log}(P)$. If P is real then its principal logarithm is real.*

Let $\mathcal{G}(h) = \{z \in \mathbb{C} : -\pi/h < \mathrm{im}(z) < \pi/h, h \in \mathbb{R}\}$. We denote the set of real matrices in $\mathbb{R}^{n\times n}$ whose eigenvalues lie in the strip $\mathcal{G}(1)$ by $\mathscr{A}(n)$. By restricting the set for which $\bar{A}$ and E belong to $\mathscr{A}(n)$, it follows that

$$E = \mathrm{Log}(\exp(A)) - \bar{A},$$

is one-to-one. Throughout the text, the notations $\exp(\cdot)$ and $e^{(\cdot)}$ are used interchangeably. We use $\log(\cdot)$ for general *primary* matrix logarithms and $\mathrm{Log}(\cdot)$ for principal logarithms.

Theorem 2 (Gantmacher [5, Thm. 1.27])**.** *Let $P \in \mathbb{C}^{n\times n}$ be nonsingular with the Jordan canonical form*

$$Z^{-1}PZ = J = \mathrm{diag}(J_1, J_2, ..., J_p) \tag{1a}$$

$$J_k = J_k(\lambda_k) = \begin{bmatrix} \lambda_k & 1 & & \\ & \lambda_k & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k \end{bmatrix} \in \mathbb{C}^{m_k\times m_k}. \tag{1b}$$

Then all solutions to $e^A = P$ are given by

$$A = ZU\,\mathrm{diag}(L_1^{j_1}, L_2^{j_2}, ..., L_p^{j_p})U^{-1}Z^{-1}, \tag{2}$$

where

$$L_k^{j_k} = \log(J_k(\lambda_k)) + 2j_k\pi i I_{m_k}; \tag{3}$$

$\log(J_k(\lambda_k))$ *denotes*

$$f(J_k) := \begin{bmatrix} f(\lambda_k) & f'(\lambda_k) & \cdots & \frac{f^{(m_k-1)}(\lambda_k)}{(m_k-1)!} \\ & f(\lambda_k) & \ddots & \vdots \\ & & \ddots & f'(\lambda_k) \\ & & & f(\lambda_k) \end{bmatrix}$$

with f the principal branch of the logarithm, defined by $\mathrm{Im}(\log(z)) \in (-\pi, \pi]$; j_k is an arbitrary integer; and U is an arbitrary nonsingular matrix that commutes with J.

Theorem 3 (classification of logarithms [5, Thm. 1.28])**.** *Let the nonsingular matrix $P \in \mathbb{C}^{n\times n}$ have the Jordan canonical form (1) with p Jordan blocks, and let $s \le p$ be the number of distinct eigenvalues of A. Then $e^A = P$ has a countable infinity of solutions that are primary functions of P, given by*

$$A_j = Z\,\mathrm{diag}(L_1^{j_1}, L_2^{(j_2)}, ..., L_p^{(j_p)})Z^{-1}, \tag{4}$$

where $L_k^{j_k}$ is defined in (3), corresponding to all possible choices of the integers $j_1, ..., j_p$, subject to the constraint that $j_i = j_k$ whenever $\lambda_i = \lambda_k$.

B. Fréchet Derivatives

Definition 1 ([5])**.** The *Fréchet derivative* of the matrix function $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ at a point $X \in \mathbb{C}^{n\times n}$ is a linear map

$$\begin{array}{ccc} \mathbb{C}^{n\times n} & \xrightarrow{L} & \mathbb{C}^{n\times n} \\ E & \longmapsto & L(X, E) \end{array}$$

such that for all $E \in \mathbb{C}^{n\times n}$

$$f(X + E) - f(X) - L(X, E) = o(\|E\|).$$

The Fréchet derivative exists for matrix functions exp and Log (principal logarithm) and it is unique. It holds that [5, p. 238]

$$L_{\exp}(X, E) = \int_0^1 e^{X(1-s)} E e^{Xs} ds,$$

$$L_{\mathrm{Log}}(X, E) = \int_0^1 (t(X - I) + I)^{-1} E (t(X - I) + I)^{-1} dt.$$

C. Fréchet derivatives for the vector representation

The vector representation of the Fréchet derivatives $L_{\exp}(X, E)$ and $L_{\mathrm{Log}}(X, E)$ have the structure of being given as a matrix multiplied by $\mathrm{vec}(E)$.

$$\mathrm{vec}(L_{\exp}(X, E)) = K(X, 0)\mathrm{vec}(E), \tag{5}$$

$$\mathrm{vec}(L_{\mathrm{Log}}(X, E)) = K(X, 0)^{-1}\mathrm{vec}(E), \tag{6}$$

$$K(X, E) = \big(I \otimes \exp(X)\big)\psi\big((X + E)^T \oplus (-X)\big)\big). \tag{7}$$

Here $K(X, E)$ can be seen as an extension of the object K(X) defined in [5, Thm. 10.13]. The vector representation of $e^A = e^{\bar{A}+E}$ can be written as

$$\mathrm{vec}(e^{\bar{A}+E}) = \mathrm{vec}(e^{\bar{A}}) + K(\bar{A}, E)E. \tag{8}$$

Moreover, let us define the maps $f_{\bar{A}} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n^2}$ and $g_{\bar{A}} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n^2}$ by

$$f_{\bar{A}}(E) = K(\bar{A}, E)\mathrm{vec}(E) \quad \text{and} \quad g_{\bar{A}}(E) = K(\bar{A}, 0)\mathrm{vec}(E).$$

III. Problem Formulations

A. General notation

Definition 2.

$$\mathscr{E}(A, D, h, \mathscr{S}) = \Big\{A^* \in \mathbb{R}^{n\times n} : \ D \in \mathbb{R}^{n^2\times n^2}, h \in \mathbb{R},$$
$$A^* = \arg\min_{\tilde{A}\in\mathscr{S}} \|D\mathrm{vec}(\exp(hA)) - D\mathrm{vec}(\exp(h\tilde{A}))\|_2\Big\},$$

where $\mathscr{S} \subseteq \mathbb{R}^{n\times n}$ contains A.

This set will be used to define the important concept of *system aliasing* in continuous-time linear system identification. Considering the special case of the set $\mathscr{S} \subseteq \mathscr{A}(n)$, we introduce the following set.

Definition 3.

$$\mathscr{E}_L(\bar{A}, E, D, \mathscr{S}) = \Big\{E^* : \bar{A} + E^* \in \mathscr{A}(n), \ D \in \mathbb{R}^{n^2\times n^2},$$
$$E^* = \arg\min_{\tilde{E}\in\mathscr{S}} \|D\mathrm{vec}(\exp(\bar{A} + E)) - D\mathrm{vec}(\exp(\bar{A} + \tilde{E}))\|_2\Big\}$$

where $\mathscr{S} \subseteq \mathbb{R}^{n\times n}$ contains E.

If E is "sufficiently" small in norm, there is an approximated problem that we could investigate, where we can use Fréchet derivatives to approximate $\exp(\bar{A} + E)$ at $\bar{A}$ in the direction of E.

Definition 4.

$$\mathscr{E}_S(\bar{A}, E, D, \mathscr{S}) = \Big\{E^* : \bar{A} + E^* \in \mathscr{A}(n), \ D \in \mathbb{R}^{n^2\times n^2},$$
$$E^* = \arg\min_{\tilde{E}\in\mathscr{S}} \|DK(\bar{A}, 0)\big(\mathrm{vec}(E) - \mathrm{vec}(\tilde{E})\big)\|_2\Big\},$$

where $\mathscr{S} \subseteq \mathbb{R}^{n\times n}$ is a linear subspace containing E.

Here we use the first order approximation of the exponential matrix. Then we assume that $\mathscr{S}$ is a linear subspace and we want to find out in what directions in this space the first order approximation is good. To be more precise, the question is for what $(\bar{A}, E, D, \mathscr{S})$ it holds that

$$\frac{\left(\sup_{E'\in\mathcal{E}_S(\bar{A},tE,D,\mathscr{S})} \|E' - tE\|\right)}{|t|} \to 0 \quad \text{as} t \to 0. \tag{9}$$

B. Continuous-time linear system identification

Consider the linear dynamical system

$$\begin{cases} \dot{x}(t) = Ax(t), \\ y(t) = Cx(t), \end{cases} \tag{10}$$

where $A = \bar{A} + E$; $C \in \mathbb{R}^{p\times n}$ has full rank, $x(t) \in \mathbb{R}^n$ and $y(t) \in \mathbb{R}^p$.

With the general notation given in Section III-A, we can give a definition on *system aliasing* only using the A-matrix and the sampling period h, which does not necessarily depend on specific identification methods.

Definition 5 (System aliasing)**.** Given $A \in \mathscr{S}$ and $h \in \mathbb{R}^+$, if there exists $\hat{A} \neq A \in \mathscr{E}(A, I, h, \mathscr{S})$, then $\hat{A}$ is

called a *system alias* of A with respect to $\mathscr{S}$. By default, choose $\mathscr{S}(A) \coloneqq \{\tilde{A} \in \mathbb{R}^{n\times n} : \max\{\mathrm{im}(\mathrm{eig}(\tilde{A}))\} \leq \max\{\mathrm{im}(\mathrm{eig}(A))\}\}$.

We are particularly interested in $\mathscr{E}(A, I, h, \mathscr{S}) = \{A\}$, i.e. there is *system aliasing*. Note that the concept of *system aliasing* does not depend on specific data. It only depends on system dynamics (e.g. the A-matrix in (10)) and sampling frequencies. If the D matrix is specifically constructed by data instead of I, $\mathscr{E}(A, D, h, \mathscr{S}) = \{A\}$, where A denotes the ground truth. This tells us that the underlying system is identifiable from the given data (see Section IV-B). Obviously, if we have system aliasing for the system with a specific sampling frequency, without extra prior information on A (see Section V), the system is not identifiable.

IV. No system aliasing

A. The minimal sampling frequency

Provided with the definition of *system aliasing*, a question is: for what (A, h) does it hold that $\mathscr{E}(A, I, h, \mathscr{S}(A)) = \{A\}$.

To make principal matrix logarithm $\mathrm{Log}(\cdot)$ well-defined, assume that $\exp(hA)$ has no negative real eigenvalues. According to Theorem 1 and 2, it always holds that $\mathrm{Log}(\exp(hA))/h \in \mathscr{E}(A, I, h, \mathscr{S}(A))$. To avoid *system aliasing*, we have to force $\mathrm{Log}(\exp(hA))/h = A$ to be satisfied. It is equivalent to $\mathrm{eig}(hA) \in \mathcal{G}(1)$.

Given no other information on the system, consider the identification problem of A using full-state measurement. The only way to find the unique estimation is to decrease the sampling period h until the ground truth falls into the strip of $\mathcal{G}(h)$, and then use the principal logarithm. Otherwise, we would be bothered by *system aliases* of A and unable to make a decision, unless we know extra prior information on A. For full-state measurement, identifiability is guaranteed by selecting appropriate h such that there is no *system aliases*. For the general case of identification using output measurement, the issue is studied in Section IV-B.

Theorem 4 (Nyquist-Shannon-like sampling theorem)**.** *To uniquely obtain A from A_d by taking the principal matrix logarithm, where A_d is identified from sampled data, the sampling frequency ω* (rad/s) *must satisfy*

$$\omega \geq 2\max\{|\,\mathrm{im}\,(\lambda_i(A))\,|,\ i = 1, \ldots, n\}\,.$$

Equivalently, the sampling period h should satisfy

$$h \leq \min\{\pi/|\,\mathrm{im}\,(\lambda_i(A))\,|,\ i = 1, \ldots, n\}\,.$$

B. Partial information

Suppose all A's in $\mathscr{A}(n)$, which implies there is no system aliasing, i.e. the case in Definition 3. Now consider the identifiability problem of (10) from data with low sampling frequency $2\pi/h$. To be precise, it is to find out for what $(\bar{A}, E, D, \mathscr{S})$, it holds that $\mathscr{E}_L(\bar{A}, E, D, \mathscr{S}) = \{E\}$.

Lemma 5. *For $\bar{A}$, E, D, $\mathscr{S} \ni E$, if there is linear subspace $\mathscr{L}$ such that $f_{\bar{A}}(\mathscr{S}) \subseteq \mathscr{L}$, then $\mathscr{E}_L(\bar{A}, E, D, \mathscr{S}) = \{E\}$ if*

$$\mathscr{L} \cap \ker(D) = \{0\}.$$

Lemma 6. *For $\bar{A}$, E, D and $\mathscr{S} \ni E$, (9) holds if and only if*

$$g_{\bar{A}}^{-1}(\ker(D)) \cap \mathscr{S} = \{0\}.$$

Proposition 7. *If*

$$\mathscr{S} \subseteq \{\tilde{E} : \mathrm{im}(\tilde{E}) \cup \mathrm{im}(\bar{A}\tilde{E}) \cup \mathrm{im}(\bar{A}^2\tilde{E}) \cup \ldots \cup \mathrm{im}(\bar{A}^{n-1}\tilde{E}) \subseteq \mathrm{im}(C^T)\}.$$

and

$$D = (X^T \otimes C),$$

where $X \in \mathbb{R}^{n\times n}$ is a non-singular matrix, then $\mathscr{E}_L(\bar{A}, E, D, \mathscr{S}) = \{E\}$ and

$$E = \log\left(\begin{bmatrix} C \\ Z \end{bmatrix}^{-1} \begin{bmatrix} Ce^{A}X \\ Ze^{\bar{A}}X \end{bmatrix} X^{-1}\right) - \bar{A},$$

where Z is any matrix in $\mathbb{R}^{(n-p)\times n}$ such that $\mathrm{im}(Z^T) = \ker(C)$.

Proposition 8. *Suppose the system (10) is initialized at k different initial points $x_{0,1}, x_{0,2}, \ldots, x_{0,k} \in \mathbb{R}^n$, $C \in \mathbb{R}^{p\times n}$ is a full rank matrix and $\mathscr{S}$ is an l-dimensional linear subspace. At the time $t = 1$, for initial point i, $y_i(1) = Ce^{A}x_{0,i}$, where $i \in \{1, 2, \ldots, k\}$.*

For almost all linearly independent vectors $x_{0,1}, x_{0,2}, \ldots x_{0,k}$ in $\mathbb{R}^n$ and almost all matrices $\bar{A}$ in $\mathbb{R}^{n\times n}$, (9) holds where

$$D = \begin{bmatrix} x_{0,1} & x_{0,2} & \ldots & x_{0,k} \end{bmatrix}^T \otimes C,$$

if and only if

$$l \leq kp.$$

Remark 1. In Proposition 8, instead of having k different initial points one can have one intial point and sample $y(t)$ at the times $t = 1$, $t = 2$ etc. for almost all $\bar{A}$, E and x_0 such that $[x_0, e^{A}x_0, e^{2A}x_0, \ldots, e^{(k-1)A}]$ has full rank (9) holds when

$$D = (\begin{bmatrix} x_0^T & (e^{A}x_0)^T & \ldots & (e^{(k-1)A}x_0)^T \end{bmatrix} \otimes C,$$

$l \leq kp$ and $\mathscr{S}$ is an l-dimensional linear subset.

V. System aliasing and bounded constraints

In the previous section we hinted that the conditions for no *system aliasing* follow as a consequence of bounded eigenvalues. In this section we follow this path and explicitly formulate an optimization problem to deal with identification in the presence of *system aliases*.

Consider the case of full-state measurements, i.e. $C = I$ in (10), and h is NOT chosen small enough such that $\mathscr{E}(A, I, h, \mathscr{S}(A)) = \{A\}$. In order to find out A among the aliases we need extra information, for instance, properties of A known *a priori*. Here we assume that the ground truth A is the sparest solution in $\mathscr{E}(A, I, h, \mathscr{S}(\kappa))$ and $\kappa \in \mathbb{R}$ as an upper bound that can be roughly estimated. The set $\mathscr{S}(\kappa)$ will be defined after giving Definition 6. A is chosen by solving the following optimization problem

$$\underset{\hat{A}\in\mathscr{E}(A,I,h,\mathscr{S}(\kappa))}{\text{minimize}} \ \|\hat{A}\|_0. \tag{11}$$

We need to calculate $\mathscr{E}(A, I, h, \mathscr{S}(\kappa))$ from data. Given the measurement $X_1 = [x(h), x(2h), \ldots, x(Nh)]$, $X_2 = [x(0), x(h), \ldots, x\big((N-1)h\big)]$, let $\hat{A}_d$ be an estimation of the A-matrix in the corresponding discrete-time state space representation. In the deterministic case[1] as (10), $\hat{A}_d = X_1 X_2^T (X_2 X_2^T)^{-1}$ and $\hat{A}_d = \exp(hA)$. Hence,

$$\mathscr{E}(A, I, h, \mathscr{S}(\kappa)) = \left\{ \tilde{A} \in \mathscr{S}(\kappa) : \exp(h\tilde{A}) = \hat{A}_d \right\}, \quad (12)$$

and define

$$\mathcal{S} \coloneqq \left\{ \tilde{A} \in \mathbb{R}^{n\times n} : \exp(h\tilde{A}) = \hat{A}_d \right\}. \quad (13)$$

To formulate $\mathscr{S}(\kappa)$, we introduce a special norm of A, which is equivalent to the Frobenius norm up to a change of coordinates.

Definition 6 (Z-weighted norm)**.** Let $h_Z(A) = Z^{-1}AZ$, where Z is the matrix defined in Theorem 3. Then the norm is defined as $\|h_Z(\cdot)\|_F = \|\cdot\|_F \circ h_Z$.

Since we assume that $\hat{A}_d$ is fixed, i.e., the data X is not used in the optimization problems defined here, the matrix Z is constant. One can observe that

$$\|h_Z(\hat{A})\|_F = \mathrm{vec}(\hat{A})^T (Z^T \otimes Z^{-1})^T (Z^T \otimes Z^{-1}) \mathrm{vec}(\hat{A})$$

is a proper $(Z^T \otimes Z^{-1})^T (Z^T \otimes Z^{-1})$-weighted vector norm in terms of $\mathrm{vec}(\hat{A})$. Using $\|h_Z(\cdot)\|_F$ is on the one hand simplifying the analysis we conduct throughout this section, and on the other explicitly penalizes the imaginary part of the eigenvalues without "distorting" them through the transformation by Z.

Now we define $\mathscr{S}(\kappa)$ using the norm $\|h_Z(\cdot)\|_F$. The basic idea is that one should exclude such A's whose imaginary parts of eigenvalues are too large, which implies their system response will show wild fluctuation. To make our assumption and the problem (11) practically meaningful, instead of $\mathbb{R}^{n\times n}$, we restrict $\mathscr{S}$ to be a norm bounded subset

$$\mathscr{S}(\kappa) = \left\{ \tilde{A} \in \mathbb{R}^{n\times n} : \|h_Z(\tilde{A})\|_F \le \kappa \right\}. \quad (14)$$

In the following we will show that the feasible set of (11) has only finite elements, which implies it can be solved at least by brute force methods. Recall that the set $\mathcal{S}$ in (13) is countable according to Theorem 3.

Let $M \coloneqq \mathrm{diag}(m_1, m_2, \ldots, m_p)$, $\mathbf{j} \coloneqq [j_1, j_2, \ldots, j_p]$ and $\beta \coloneqq [\beta_1, \beta_2, \ldots, \beta_p]$, where $\log(\lambda_k) \triangleq \alpha_k + i\pi\beta_k$, $k = 1, \ldots, p$, and j_k, λ_k are defined in Theorem 2. A function $\mathscr{I}$ is defined as

$$\mathscr{I}(\mathbf{j}, \delta) \coloneqq \delta^T M \delta + (2\mathbf{j} + \beta)^T M \delta, \quad (15)$$

where $\mathbf{j}, \delta \in \mathbb{Z}^p$. Moreover, it satisfies $\mathscr{I}(\mathbf{j}, \delta) = \mathscr{I}(0, \mathbf{j} + \delta) - \mathscr{I}(0, \mathbf{j})$, which follows by noticing

$$\begin{aligned}\mathscr{I}(\mathbf{j}, \delta) = {} & (\delta + \mathbf{j} + \beta/2)^T M (\delta + \mathbf{j} + \beta/2) \\ & - (\mathbf{j} + \beta/2)^T M (\mathbf{j} + \beta/2). \end{aligned} \quad (16)$$

[1]For stochastic cases, $\hat{A}_d$ is consistently estimated by *Prediction Error Minimization* or *Maximum Likelihood* methods [1], and $\lim_{N\to\infty} \mathbb{E}(\hat{A}_d(N)) = \exp(hA)$. If only finite samples are available, we cannot obtain the exactly equivalent $\mathscr{E}(A, I, h, \mathscr{S}(\kappa))$ from data.

Moreover, let A_0 denote a special matrix logarithm for which all j_k $(k = 1, \ldots, p)$ in (3) are equal to 0.

Definition 7 (equivalence relations)**.** An *equivalence* relation "$\sim$" is defined on $\mathcal{S}$ as a binary relation: for any $A_1, A_2 \in \mathcal{S}$, $\mathbf{j}^{(1)}$ and $\mathbf{j}^{(2)}$ are defined for A_1, A_2, respectively, we say $A_1 \sim A_2$ if $\mathscr{I}(\mathbf{j}^{(1)}, \mathbf{j}^{(2)} - \mathbf{j}^{(1)}) = 0$.

Lemma 9. *Let $\mathcal{S}$ be the set defined in (12) and parametrized by (4) in Theorem 3. For any $A_1, A_2 \in \mathcal{S}$, $\|h_Z(A_1)\|_F = \|h_Z(A_2)\|_F$ if and only if $A_1 \sim A_2$.*

Lemma 10. *Given any $\bar{A} \in \mathcal{S}$, there exist finite $A_i \in \mathcal{S}$ that satisfies $A_i \sim \bar{A}$.*

Lemma 11. *There exists finite $A_i \in \mathcal{S}$ such that $\|h_Z(A_i)\|_F \le \kappa$.*

Proposition 12 (lower boundness of logarithms)**.** *Let $\mathcal{S}$ be the set defined in (12). Given any $\bar{A} \in \mathcal{S}$, there exists $M(\bar{A}) > 0$, such that for any $A \in \{A \in \mathcal{S} : A \nsim \bar{A}\}$, it holds that*

$$\big| \|h_Z(A)\|_F - \|h_Z(\bar{A})\|_F \big| \ge M.$$

Proposition 13. *Let $\mathcal{S}$ be the set defined in (12). For any $\bar{A} \in \mathcal{S}$, there exist $\kappa_l, \kappa_u \in \mathbb{R}$ in $\mathscr{S}(\kappa_l, \kappa_u) = \{\tilde{A} \in \mathbb{R}^{n\times n} : \kappa_l \le \|h_Z(\tilde{A})\|_F \le \kappa_u\}$ such that (11) has a unique optimal point in the sense of the equivalence relation in Definition 7.*

VI. Conclusions

This paper addresses identification of continuous-time dynamical systems with sparse topologies. The key assumption is that the sampling frequency is low. Under this assumption a realization/identification problem comes to surface, which has largely been overlooked in the community. First we propose the minimal sampling frequency that guarantees no system aliasing. Allowing system aliasing, one needs to search over a collection of matrix logarithms to find the sparsest one. We provide theoretical results for when a unique solution exists up to a finite equivalence class.

References

[1] L. Ljung, *System Identification: Theory for the User*, ser. Prentice-Hall information and system sciences series. Prentice Hall PTR, 1999.

[2] H. Garnier and L. Wang, *Identification of Continuous-time Models from Sampled Data*. Springer London, 2008.

[3] N. K. Sinha and G. P. Rao, *Identification of continuous-time systems: Methodology and computer implementation*. Springer Science & Business Media, 2012, vol. 7.

[4] Y. Zuogong, T. Johan, and G. Jorge, "Inverse Problems for Matrix Exponential in System Identification: System Aliasing," *arXiv preprint*, 2016.

[5] N. Higham, *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008.