The Cramer–Rao Lower Bound for Bilinear Systems

Qiyue Zou, *Student Member, IEEE*, Zhiping Lin, *Senior Member, IEEE*, and Raimund J. Ober, *Senior Member, IEEE*

***Abstract*—Estimation of the unknown parameters that characterize a bilinear system is of primary importance in many applications. The Cramer–Rao lower bound (CRLB) provides a lower bound on the covariance matrix of any unbiased estimator of unknown parameters. It is widely applied to investigate the limit of the accuracy with which parameters can be estimated from noisy data. Here it is shown that the CRLB for a data set generated by a bilinear system with additive Gaussian measurement noise can be expressed explicitly in terms of the outputs of its derivative system which is also bilinear. A connection between the nonsingularity of the Fisher information matrix and the local identifiability of the unknown parameters is exploited to derive local identifiability conditions of bilinear systems using the concept of the derivative system. It is shown that for bilinear systems with piecewise constant inputs, the CRLB for uniformly sampled data can be efficiently computed through solving a Lyapunov equation. In addition, a novel method is proposed to derive the asymptotic CRLB when the number of acquired data samples approaches infinity. These theoretical results are illustrated through the simulation of surface plasmon resonance experiments for the determination of the kinetic parameters of protein–protein interactions.**

***Index Terms*—Bilinear systems, Cramer–Rao lower bound (CRLB), Fisher information matrix, local identifiability, parameter estimation, surface plasmon resonance experiments, system identification.**

I. Introduction

BILINEAR systems are an important class of nonlinear systems because of their wide range of applications in a number of different fields, including engineering, biomedical science, economics, etc. A fundamental problem in these applications is to estimate/identify the unknown parameters of a bilinear system from its output observations [1]–[4]. The question therefore naturally arises concerning the accuracy of the estimation that can be achieved based on the assumed bilinear system model and observed noisy outputs. The Cramer–Rao lower bound (CRLB) gives a lower bound on the covariance matrix of any unbiased estimator of unknown parameters [5], [6]. It is commonly used to evaluate the performance of an estimation/identification algorithm and provide guidance to improve the experimental design. The purpose of this paper is to derive an explicit expression of the CRLB for noisy data sets generated by a bilinear system, from the perspective of system theory.

The CRLB for estimating unknown parameters of stationary time series has received considerable attention in the literature [7]–[9]. Recently, the CRLB or Fisher information matrix for one-dimensional (1-D) dynamic nonstationary systems with deterministic input and Gaussian measurement noise has been investigated in [10]. The calculation of the Fisher information matrix for the 1-D data is performed in terms of the derivative system with respect to the unknown parameters and by using the solution to a Lyapunov equation. The above approach has been extended to multidimensional (n-D) data sets generated by n-D linear separable-denominator systems and applied to the analysis of n-D nuclear magnetic resonance (NMR) spectroscopy data sets [11].

Here we generalize the results in [10] to bilinear systems and continue to explore some system theoretical insights of the approach. It is shown that the Fisher information matrix for the output data samples of a multiple-input-multiple-output (MIMO) bilinear system can be expressed in terms of the outputs of its derivative system which is also an MIMO bilinear system. The use of the notion of the derivative system brings two main benefits. First, we can study properties of the Fisher information matrix and the CRLB from a system theoretic point of view, e.g., the local identifiability conditions discussed in Section II. Second, for uniformly sampled data sets generated by bilinear systems with piecewise constant inputs, the CRLB can be efficiently computed using algorithms based on the solution to a Lyapunov equation. It is important to note that different notions of bilinear systems exist in the literature. In [12] and [13], the CRLB was calculated for specific equations that do not immediately reduce to the class of systems considered here, although our general approach may also be applicable to the type of systems considered in [12] and [13].

The organization of the paper is as follows. In Section II, we apply the concept of the derivative system to obtain an explicit expression of the Fisher information matrix for noise corrupted data sets generated by an MIMO time-invariant bilinear system. Provided some weak regularity conditions hold the nonsingularity of the Fisher information matrix is equivalent to the local identifiability of the system. We consider the question of local identifiability in two contexts. First, we answer the question under which conditions a finite number of inputs exist that lead to an identifiable data set. Second, we address the question under which conditions for a given input the resulting data set leads to local identifiability of the parameters. For the uniformly

Manuscript received May 7, 2004; revised May 5, 2005. The research was supported in part by grants by the National Institutes of Health (RO1 GM58538). For reprints and correspondence, readers should contact coauthor R. J. Ober. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yuri I. Abramovich.

Q. Zou was with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798. He is now with the Electrical Engineering Department, University of California, Los Angeles, CA 90095 USA (e-mail: zouqiyue@pmail.ntu.edu.sg).

Z. Lin is with the Centre for Signal Processing, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ezplin@ntu.edu.sg).

R. J. Ober is with the Eric Jonsson School of Electrical Engineering and Computer Science, University of Texas at Dallas, Richardson, TX 75083-0688 USA, and also with the Center for Immunology, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75235-8576 USA (e-mail: ober@utdallas.edu).

Digital Object Identifier 10.1109/TSP.2005.863006

sampled data sets generated by a bilinear system with piecewise constant inputs, it is shown in Section III that the CRLB can be efficiently calculated through solving a Lyapunov equation and that the asymptotic CRLB can be derived without explicitly computing the Fisher information matrix. In Section IV, the theoretical results presented in the paper are illustrated by the simulation of surface plasmon resonance experiments aimed at estimating kinetic constants of protein–protein interactions. Conclusions are presented in Section V. Proofs are given in the Appendix.

II. Cramer–Rao Lower Bound for MIMO Time-Invariant Bilinear Systems

A. General Approach

Consider the state–space model of a general MIMO time-invariant bilinear system given by (see [14])

$$\dot{\mathbf{x}}_{\boldsymbol{\theta}}(t) = \mathbf{A}_{\boldsymbol{\theta}}\mathbf{x}_{\boldsymbol{\theta}}(t) + \sum_{m=1}^{M}\mathbf{F}_{\boldsymbol{\theta},m}u_m(t)\mathbf{x}_{\boldsymbol{\theta}}(t) + \mathbf{B}_{\boldsymbol{\theta}}\mathbf{u}(t) \quad (1)$$

$$\mathbf{y}_{\boldsymbol{\theta}}(t) = \mathbf{C}_{\boldsymbol{\theta}}\mathbf{x}_{\boldsymbol{\theta}}(t), \quad t \geq t^{[0]} \quad (2)$$

where $\mathbf{x}_{\boldsymbol{\theta}}(t) \in \mathbb{R}^{N\times 1}$ is the state vector, $\mathbf{y}_{\boldsymbol{\theta}}(t) \in \mathbb{R}^{R\times 1}$ is the system output vector, $\mathbf{A}_{\boldsymbol{\theta}} \in \mathbb{R}^{N\times N}, \mathbf{B}_{\boldsymbol{\theta}} \in \mathbb{R}^{N\times M}, \mathbf{C}_{\boldsymbol{\theta}} \in \mathbb{R}^{R\times N}, \mathbf{F}_{\boldsymbol{\theta},m} \in \mathbb{R}^{N\times N}, m = 1,\ldots,M$ are the system matrices depending on the unknown parameter vector $\boldsymbol{\theta} := [\theta_1 \ \cdots \ \theta_K]^T$, $\mathbf{x}_{\boldsymbol{\theta}}(t^{[0]}) = \mathbf{x}_{\boldsymbol{\theta},0}$ is the initial state vector, which can also depend on the parameter vector $\boldsymbol{\theta}$, and $\mathbf{u}(t) \in \mathbb{R}^{M\times 1}$ is the input vector with components $u_1(t),\ldots,u_M(t)$, which are independent of $\boldsymbol{\theta}$. For convenience of exposition, we use the notation $\Phi := \{\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1},\ldots,\mathbf{F}_{\boldsymbol{\theta},M}\}$ to represent the bilinear system with state vector $\mathbf{x}_{\boldsymbol{\theta}}(t)$, input $\mathbf{u}(t)$, output $\mathbf{y}_{\boldsymbol{\theta}}(t)$, system matrices $\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1},\ldots,\mathbf{F}_{\boldsymbol{\theta},M}$, and initial state $\mathbf{x}_{\boldsymbol{\theta},0}$, as defined in (1) and (2). The rth element of $\mathbf{y}_{\boldsymbol{\theta}}(t)$ is represented by $y_{\boldsymbol{\theta},r}(t), r = 1,\ldots,R$, i.e., $\mathbf{y}_{\boldsymbol{\theta}}(t) := [y_{\boldsymbol{\theta},1}(t) \ \cdots \ y_{\boldsymbol{\theta},R}(t)]^T$. Similarly, the rth row of $\mathbf{C}_{\boldsymbol{\theta}}$ is denoted by $\mathbf{c}_{\boldsymbol{\theta},r}^T, r = 1,\ldots,R$, i.e., $\mathbf{C}_{\boldsymbol{\theta}} = [\mathbf{c}_{\boldsymbol{\theta},1} \ \cdots \ \mathbf{c}_{\boldsymbol{\theta},R}]^T$, etc.

In this section, we will derive some general results and properties of the CRLB for bilinear systems for the class of *admissible* inputs $\mathbb{V}$, which are assumed to be piecewise continuous, have a finite number of discontinuities and are defined on finite or semi-finite left-closed intervals whose left boundary point is $t^{[0]}$. The main reason for considering the class of admissible inputs is that there exists a unique solution when the input to a bilinear system is from this class. Specifically, when the input is piecewise continuous with a finite number of discontinuities, the state and output vectors will be continuous.

Assume that we have acquired noise corrupted samples $\mathbf{s}(t_j), t^{[0]} = t_0 < t_1 < \cdots < t_{J-1}$, of the measured output of the bilinear system defined by (1) and (2), i.e.,

$$\mathbf{s}(t_j) = \mathbf{y}_{\boldsymbol{\theta}}(t_j) + \mathbf{n}(t_j) \quad (3)$$

where $\mathbf{y}_{\boldsymbol{\theta}}(t_j)$ is the noise free output and $\mathbf{n}(t_j)$ is the measurement noise at the sampling point $t^{[0]} = t_0 < t_1 < \cdots < t_{J-1}$. We assume that the measurement noise components are zero-mean, Gaussian distributed and temporally uncorrelated. The probability density function $p_{\boldsymbol{\theta}}(\mathbf{s})$ for the acquired data set $\mathbf{s} := \{\mathbf{s}(t_j),\ j = 0,\ldots,J-1\}$ is given by

$$p_{\boldsymbol{\theta}}(\mathbf{s}) = \prod_{j=0}^{J-1} p_{\boldsymbol{\theta}}(\mathbf{s}(t_j)) = \prod_{j=0}^{J-1} \frac{1}{(2\pi)^{\frac{R}{2}}|\boldsymbol{\Lambda}_{\mathbf{n}}(t_j)|^{\frac{1}{2}}} \times \exp\left(-\frac{1}{2}[\mathbf{s}(t_j) - \mathbf{y}_{\boldsymbol{\theta}}(t_j)]^T\boldsymbol{\Lambda}_{\mathbf{n}}^{-1}(t_j)[\mathbf{s}(t_j) - \mathbf{y}_{\boldsymbol{\theta}}(t_j)]\right)$$

where $\boldsymbol{\Lambda}_{\mathbf{n}}(t_j) := E[\mathbf{n}(t_j)\mathbf{n}(t_j)^T]$ is the covariance matrix of $\mathbf{n}(t_j), j = 0,\ldots,J-1$, which is assumed to be positive definite throughout the paper. Assume that $p_{\boldsymbol{\theta}}(\mathbf{s})$ satisfies the standard regularity conditions (see e.g., [15], [16]). The Fisher information matrix $\mathbf{I}_{\boldsymbol{\theta}}$ is then defined as (see [5], [17])

$$\mathbf{I}_{\boldsymbol{\theta}} = E\left\{\frac{\partial \ln p_{\boldsymbol{\theta}}(\mathbf{s})}{\partial\boldsymbol{\theta}}\left(\frac{\partial \ln p_{\boldsymbol{\theta}}(\mathbf{s})}{\partial\boldsymbol{\theta}}\right)^T\right\}$$

where $(\partial/(\partial\boldsymbol{\theta})) = [(\partial/(\partial\theta_1)) \ \cdots \ (\partial/(\partial\theta_K))]^T$. If $\mathbf{I}_{\boldsymbol{\theta}}$ is positive definite for all $\boldsymbol{\theta} \in \Theta$, where Θ is assumed to be an open subset of the Euclidean space $\mathbb{R}^{K\times 1}$, by the CRLB any unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ has a variance such that

$$\mathrm{var}(\hat{\boldsymbol{\theta}}) \geq \mathbf{I}_{\boldsymbol{\theta}}^{-1},$$

where $\mathrm{var}(\hat{\boldsymbol{\theta}}) \geq \mathbf{I}_{\boldsymbol{\theta}}^{-1}$ is interpreted as meaning that the matrix $(\mathrm{var}(\hat{\boldsymbol{\theta}}) - \mathbf{I}_{\boldsymbol{\theta}}^{-1})$ is positive semidefinite.

In the following theorem, we first show that the derivative system (with respect to the given parameter vector $\boldsymbol{\theta}$) of a general MIMO time-invariant bilinear system is also an MIMO time-invariant bilinear system. The Fisher information matrix for the sampled output data of the bilinear system with Gaussian measurement noise is then expressed using the output samples of its derivative system.

Theorem 2.1 (Appendix A): Consider the bilinear system represented by $\Phi := \{\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1},\ldots,\mathbf{F}_{\boldsymbol{\theta},M}\}$. Assume that the partial derivatives of $\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1},\ldots,\mathbf{F}_{\boldsymbol{\theta},M}$ and $\mathbf{x}_{\boldsymbol{\theta},0}$ with respect to the elements of $\boldsymbol{\theta}$ exist for $\boldsymbol{\theta} \in \Theta$. Let

$$\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t) := \begin{bmatrix}\tilde{\mathbf{y}}_{\boldsymbol{\theta},1}(t)\\ \vdots \\ \tilde{\mathbf{y}}_{\boldsymbol{\theta},R}(t)\end{bmatrix}, \quad \text{with } \tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t) := \begin{bmatrix}\frac{\partial y_{\boldsymbol{\theta},r}(t)}{\partial\theta_1}\\ \vdots \\ \frac{\partial y_{\boldsymbol{\theta},r}(t)}{\partial\theta_K}\end{bmatrix}, \quad r = 1,\ldots,R, \quad t \geq t^{[0]}.$$

Then, we have the following results.

1) $\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t), t \geq t^{[0]}$, is the output of the derivative system $\tilde{\Phi} := \{\tilde{\mathbf{A}}_{\boldsymbol{\theta}}, \tilde{\mathbf{B}}_{\boldsymbol{\theta}}, \tilde{\mathbf{C}}_{\boldsymbol{\theta}}, \tilde{\mathbf{F}}_{\boldsymbol{\theta},1},\ldots,\tilde{\mathbf{F}}_{\boldsymbol{\theta},M}\}$ given by

$$\dot{\tilde{\mathbf{x}}}_{\boldsymbol{\theta}}(t) = \tilde{\mathbf{A}}_{\boldsymbol{\theta}}\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t) + \sum_{m=1}^{M}\tilde{\mathbf{F}}_{\boldsymbol{\theta},m}u_m(t)\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t) + \tilde{\mathbf{B}}_{\boldsymbol{\theta}}\mathbf{u}(t) \quad (4)$$

$$\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t) = \tilde{\mathbf{C}}_{\boldsymbol{\theta}}\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t), \quad t \geq t^{[0]} \quad (5)$$

which is an MIMO time-invariant bilinear system with state vector $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t), t \geq t^{[0]}$, and has the same input $\mathbf{u}$ as Φ. The state vector $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}$, initial state $\tilde{\mathbf{x}}_{\boldsymbol{\theta},0}$, and system matrices $\tilde{\mathbf{A}}_{\boldsymbol\theta}, \tilde{\mathbf{B}}_{\boldsymbol\theta}, \tilde{\mathbf{C}}_{\boldsymbol\theta}, \tilde{\mathbf{F}}_{\boldsymbol\theta,1}, \ldots, \tilde{\mathbf{F}}_{\boldsymbol\theta,M}$ are given as follows, which will be adopted throughout the paper:

$$\tilde{\mathbf{x}}_{\boldsymbol\theta}(t) := \begin{bmatrix} \partial_1 \mathbf{x}_{\boldsymbol\theta}(t) \\ \vdots \\ \partial_K \mathbf{x}_{\boldsymbol\theta}(t) \end{bmatrix}, \quad t \geq t^{[0]}, \quad \tilde{\mathbf{x}}_{\boldsymbol\theta,0} := \begin{bmatrix} \partial_1 \mathbf{x}_{\boldsymbol\theta}(t^{[0]}) \\ \vdots \\ \partial_K \mathbf{x}_{\boldsymbol\theta}(t^{[0]}) \end{bmatrix}$$
$$\tilde{\mathbf{A}}_{\boldsymbol\theta} := \mathrm{diag}\{\partial_1 \mathbf{A}_{\boldsymbol\theta}, \ldots, \partial_K \mathbf{A}_{\boldsymbol\theta}\}$$
$$\tilde{\mathbf{B}}_{\boldsymbol\theta} := \begin{bmatrix} \partial_1 \mathbf{B}_{\boldsymbol\theta} \\ \vdots \\ \partial_K \mathbf{B}_{\boldsymbol\theta} \end{bmatrix}, \quad \tilde{\mathbf{C}}_{\boldsymbol\theta} := \begin{bmatrix} \tilde{\mathbf{C}}_{\boldsymbol\theta,1} \\ \vdots \\ \tilde{\mathbf{C}}_{\boldsymbol\theta,R} \end{bmatrix} \quad \text{with}$$
$$\tilde{\mathbf{C}}_{\boldsymbol\theta,r} := \mathrm{diag}\left\{\partial_1 \mathbf{c}_{\boldsymbol\theta,r}^T, \ldots, \partial_K \mathbf{c}_{\boldsymbol\theta,r}^T\right\}, \quad r = 1, \ldots, R$$
$$\tilde{\mathbf{F}}_{\boldsymbol\theta,m} := \mathrm{diag}\{\partial_1 \mathbf{F}_{\boldsymbol\theta,m}, \ldots, \partial_K \mathbf{F}_{\boldsymbol\theta,m}\}, \quad m = 1, \ldots, M \quad (6)$$

where for $k = 1, \ldots, K$

$$\partial_k \mathbf{x}_{\boldsymbol\theta}(t) := \begin{bmatrix} \mathbf{x}_{\boldsymbol\theta}(t) \\ \frac{\partial \mathbf{x}_{\boldsymbol\theta}(t)}{\partial \theta_k} \end{bmatrix}, \quad t \geq t^{[0]}$$
$$\partial_k \mathbf{x}_{\boldsymbol\theta}(t^{[0]}) := \begin{bmatrix} \mathbf{x}_{\boldsymbol\theta,0} \\ \frac{\partial \mathbf{x}_{\boldsymbol\theta,0}}{\partial \theta_k} \end{bmatrix}, \quad \partial_k \mathbf{A}_{\boldsymbol\theta} := \begin{bmatrix} \mathbf{A}_{\boldsymbol\theta} & 0 \\ \frac{\partial \mathbf{A}_{\boldsymbol\theta}}{\partial \theta_k} & \mathbf{A}_{\boldsymbol\theta} \end{bmatrix}$$
$$\partial_k \mathbf{B}_{\boldsymbol\theta} := \begin{bmatrix} \mathbf{B}_{\boldsymbol\theta} \\ \frac{\partial \mathbf{B}_{\boldsymbol\theta}}{\partial \theta_k} \end{bmatrix}, \quad \partial_k \mathbf{c}_{\boldsymbol\theta,r}^T := \begin{bmatrix} \frac{\partial \mathbf{c}_{\boldsymbol\theta,r}^T}{\partial \theta_k} & \mathbf{c}_{\boldsymbol\theta,r}^T \end{bmatrix}$$
$$\partial_k \mathbf{F}_{\boldsymbol\theta,m} := \begin{bmatrix} \mathbf{F}_{\boldsymbol\theta,m} & 0 \\ \frac{\partial \mathbf{F}_{\boldsymbol\theta,m}}{\partial \theta_k} & \mathbf{F}_{\boldsymbol\theta,m} \end{bmatrix}, \quad m = 1, \ldots, M. \quad (7)$$

2) For the data points $\mathbf{s}(t_j) = \mathbf{y}_{\boldsymbol\theta}(t_j) + \mathbf{n}(t_j)$, where $\mathbf{y}_{\boldsymbol\theta}(t_j)$ is the sampled output of the bilinear system Φ, and $\mathbf{n}(t_j)$ is temporally independent Gaussian noise with zero mean and variance matrix $\boldsymbol\Lambda_{\mathbf{n}}(t_j)$, the Fisher information matrix is given by

$$\mathbf{I}_{\boldsymbol\theta} = \sum_{j=0}^{J-1} \tilde{\mathbf{Y}}_{\boldsymbol\theta}(t_j) \boldsymbol\Lambda_{\mathbf{n}}^{-1}(t_j) \tilde{\mathbf{Y}}_{\boldsymbol\theta}^T(t_j). \quad (8)$$

Here $\tilde{\mathbf{Y}}_{\boldsymbol\theta}(t_j)$ is defined as $\tilde{\mathbf{Y}}_{\boldsymbol\theta}(t_j) := [\tilde{\mathbf{y}}_{\boldsymbol\theta,1}(t_j) \ \cdots \ \tilde{\mathbf{y}}_{\boldsymbol\theta,R}(t_j)]$, for $j = 0, \ldots, J-1$.

Note that in the above theorem, $\tilde{\mathbf{Y}}_{\boldsymbol\theta}(t_j)$ is a $K \times R$ matrix, while $\tilde{\mathbf{y}}_{\boldsymbol\theta}(t)$ is an $RK \times 1$ vector.

B. Local Identifiability

The parameter vector $\boldsymbol\theta$ is said to be locally identifiable if there exists an open neighborhood of $\boldsymbol\theta$ containing no other parameter vector that is observably equivalent to $\boldsymbol\theta$ [18]. The following Theorem 2.2 quoted from [19] (see also [20]) states that under some weak regularity conditions the local identifiability of an unknown parameter vector is equivalent to the nonsingularity of its associated Fisher information matrix. This connection between local identifiability and the nonsingularity of the Fisher information matrix is of importance in itself. It is also relevant for the calculation of the CRLB which is typically expressed in terms of the inverse of the Fisher information matrix $\mathbf{I}_{\boldsymbol\theta}$. For the remaining part of the paper we need to impose the standard weak regularity conditions (see, e.g., [19]). In our context, this implies in particular that we will assume the partial derivatives of $\mathbf{A}_{\boldsymbol\theta}, \mathbf{B}_{\boldsymbol\theta}, \mathbf{C}_{\boldsymbol\theta}, \mathbf{F}_{\boldsymbol\theta,1}, \ldots, \mathbf{F}_{\boldsymbol\theta,M}$ and $\mathbf{x}_{\boldsymbol\theta,0}$ with respect to $\boldsymbol\theta$ are continuous for all $\boldsymbol\theta \in \Theta$.

Theorem 2.2 [19]: Let $\boldsymbol\theta \in \Theta$. Then $\boldsymbol\theta$ is locally identifiable if and only if the Fisher information matrix $\mathbf{I}_{\boldsymbol\theta}$ is nonsingular.

For a bilinear system with a specific input, we can determine the local identifiability based on the nonsingularity of the Fisher information matrix $\mathbf{I}_{\boldsymbol\theta}$ by Theorems 2.1 and 2.2. However, a disadvantage is that we need to calculate the output samples of the derivative system first. In the study of local identifiability, we are sometimes interested in determining whether there exist some admissible inputs such that the unknown parameter vector is locally identifiable based on the system matrices and the initial state vector of a given system without specifying a specific input and computing its output. In such a study, we may also allow a finite number of admissible inputs to be applied to the same system one after another. In terms of practical experiments, this amounts to conducting a finite number of independent experiments on the same system sequentially, each with a different input. The system theoretic notion of reachability will play an important role in our study of identifiability. We study the question of local identifiability in two contexts. First, we ask the question: Given a bilinear system, under which conditions do there exist a finite number of inputs and output sampling points such that the parameter vector is locally identifiable? Second, we consider the problem of assessing local identifiability for a given input, which will be presented in the next section.

Before proceeding, we first review some important notions from mathematical system theory [21].

Definition 2.1: Consider a system with a set of inputs denoted by $\mathbb{S}$. We define the following:

1) a state $\mathbf{x}$ is said to be *reachable* from initial state $\mathbf{x}_0$ via inputs $\mathbb{S}$ if there exists an input in $\mathbb{S}$ such that the path of its corresponding states starts at $\mathbf{x}_0$ and passes through $\mathbf{x}$;
2) an output $\mathbf{y}$ is said to be *reachable* from initial state $\mathbf{x}_0$ via inputs $\mathbb{S}$ if there exists a state reachable from $\mathbf{x}_0$ via inputs $\mathbb{S}$ such that its corresponding output is $\mathbf{y}$.

The following lemma characterizes the span of reachable states and outputs of a bilinear system via admissible inputs $\mathbb{V}$.

Lemma 2.1 (Appendix B): Consider the MIMO bilinear system represented by $\Phi := \{\mathbf{A}_{\boldsymbol\theta}, \mathbf{B}_{\boldsymbol\theta}, \mathbf{C}_{\boldsymbol\theta}, \mathbf{F}_{\boldsymbol\theta,1}, \ldots, \mathbf{F}_{\boldsymbol\theta,M}\}$, and denote $\Omega_{\mathbf{x}_{\boldsymbol\theta}}$ the span of reachable states and $\Omega_{\mathbf{y}_{\boldsymbol\theta}}$ the span of reachable outputs from a given $\mathbf{x}_{\boldsymbol\theta,0}$ via admissible inputs $\mathbb{V}$. Define matrices $\mathbf{O}^{(0)}, \ldots, \mathbf{O}^{(N-1)}$ as

$$\mathbf{O}^{(0)} := \mathbf{O}_0, \mathbf{O}^{(1)} := [\mathbf{O}_0 \quad \mathbf{O}_1], \ldots,$$
$$\mathbf{O}^{(N-1)} := [\mathbf{O}_0 \quad \cdots \quad \mathbf{O}_{N-1}]. \quad (9)$$

Here $\mathbf{O}_0, \ldots, \mathbf{O}_{N-1}$ are given by

$$\mathbf{O}_0 := [\mathbf{A}_{\boldsymbol\theta}\mathbf{x}_{\boldsymbol\theta,0} \quad \mathbf{B}_{\boldsymbol\theta} + \mathbf{B}'_{\boldsymbol\theta}],$$
$$\mathbf{O}_n := [\mathbf{A}_{\boldsymbol\theta}\mathbf{O}_{n-1} \quad \mathbf{F}_{\boldsymbol\theta,1}\mathbf{O}_{n-1} \quad \cdots \quad \mathbf{F}_{\boldsymbol\theta,m}\mathbf{O}_{n-1}],$$
$$n = 1, \ldots, N-1 \quad (10)$$

where $\mathbf{B}'_{\boldsymbol\theta} = [\mathbf{F}_{\boldsymbol\theta,1}\mathbf{x}_{\boldsymbol\theta,0} \ \cdots \ \mathbf{F}_{\boldsymbol\theta,m}\mathbf{x}_{\boldsymbol\theta,0}]$. Then there exists an integer p with $0 \leq p \leq N-1$ such that

$$\mathrm{range}\left\{\mathbf{O}^{(0)}\right\} \subset \mathrm{range}\left\{\mathbf{O}^{(1)}\right\} \subset \cdots \subset \mathrm{range}\left\{\mathbf{O}^{(p)}\right\}$$
$$= \mathrm{range}\left\{\mathbf{O}^{(p+1)}\right\} = \cdots = \mathrm{range}\left\{\mathbf{O}^{(N-1)}\right\}$$

and

$$\Omega_{\mathbf{y}_{\boldsymbol\theta}} = \mathrm{range}\left\{\left[\mathbf{C}_{\boldsymbol\theta}\mathbf{x}_{\boldsymbol\theta,0} \quad \mathbf{C}_{\boldsymbol\theta}\mathbf{O}^{(p)}\right]\right\}.$$

Recall that in Theorem 2.1, for the output data set generated by a single input $\mathbf{u} \in \mathbb{V}$, we have

$$\mathbf{I}_{\boldsymbol{\theta}} = \sum_{j=0}^{J-1} \tilde{\mathbf{Y}}_{\boldsymbol{\theta}}(t_j)\boldsymbol{\Lambda}_{\mathbf{n}}^{-1}(t_j)\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}^T(t_j).$$

The above expression can be generalized to multiple or even infinitely many inputs. Consider an arbitrary set of admissible inputs denoted by Γ. For each input $\mathbf{u} \in \Gamma$, let $\Upsilon_{\mathbf{u}}$ denote the set of output sample points, $\tilde{\mathbf{x}}_{\boldsymbol{\theta},\mathbf{u}}(t), t \geq t_0$, and $\tilde{\mathbf{y}}_{\boldsymbol{\theta},\mathbf{u}}(t), t \geq t_0$, the state and output vectors of the derivative system $\tilde{\Phi}$, respectively. For each (vector) sample $\mathbf{v} \in \Upsilon_{\mathbf{u}}$, let $t_{\mathbf{v},\mathbf{u}}$ denote the corresponding sampling instant of $\mathbf{v}$, and $\boldsymbol{\Lambda}_{\mathbf{n}}(t_{\mathbf{v},\mathbf{u}})$ the noise covariance matrix at $t_{\mathbf{v},\mathbf{u}}$. Then we have

$$\mathbf{I}_{\boldsymbol{\theta}} = \sum_{\mathbf{u}\in\Gamma}\sum_{\mathbf{v}\in\Upsilon_{\mathbf{u}}} \tilde{\mathbf{Y}}_{\boldsymbol{\theta}}(t_{\mathbf{v},\mathbf{u}})\boldsymbol{\Lambda}_{\mathbf{n}}^{-1}(t_{\mathbf{v},\mathbf{u}})\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}^T(t_{\mathbf{v},\mathbf{u}}) \tag{11}$$

where $\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}(t_{\mathbf{v},\mathbf{u}}) := [\tilde{\mathbf{y}}_{\boldsymbol{\theta},\mathbf{u},1}(t_{\mathbf{v},\mathbf{u}}) \ \ldots \ \tilde{\mathbf{y}}_{\boldsymbol{\theta},\mathbf{u},R}(t_{\mathbf{v},\mathbf{u}})]$, for $j = 0,\ldots,J-1$. Using the Fisher information matrix $\mathbf{I}_{\boldsymbol{\theta}}$ defined in (11) for the set of inputs Γ and the set of output sampling points $\Upsilon_{\mathbf{u}}$, and making use of Lemma 2.1, we can now obtain necessary and sufficient conditions for the local identifiability with respect to a finite number of inputs.

Theorem 2.3 (Appendix C): Consider the bilinear system represented by $\Phi := \{\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1}, \ldots, \mathbf{F}_{\boldsymbol{\theta},M}\}$. The derivative system of Φ is represented by $\tilde{\Phi} := \{\tilde{\mathbf{A}}_{\boldsymbol{\theta}}, \tilde{\mathbf{B}}_{\boldsymbol{\theta}}, \tilde{\mathbf{C}}_{\boldsymbol{\theta}}, \tilde{\mathbf{F}}_{\boldsymbol{\theta},1}, \ldots, \tilde{\mathbf{F}}_{\boldsymbol{\theta},M}\}$, where the dimension of $\tilde{\mathbf{A}}_{\boldsymbol{\theta}}$ is $\tilde{N}\times\tilde{N}$ with $\tilde{N} = 2KN$. Define matrices $\tilde{\mathbf{O}}^{(0)},\ldots,\tilde{\mathbf{O}}^{(\tilde{N}-1)}$ as

$$\tilde{\mathbf{O}}^{(0)} := \tilde{\mathbf{O}}_0, \quad \tilde{\mathbf{O}}^{(1)} := [\tilde{\mathbf{O}}_0 \quad \tilde{\mathbf{O}}_1], \ldots, \\ \tilde{\mathbf{O}}^{(\tilde{N}-1)} := [\tilde{\mathbf{O}}_0 \quad \cdots \quad \tilde{\mathbf{O}}_{\tilde{N}-1}]. \tag{12}$$

Here, $\tilde{\mathbf{O}}_0,\ldots,\tilde{\mathbf{O}}_{\tilde{N}-1}$ are given by

$$\tilde{\mathbf{O}}_0 := [\tilde{\mathbf{A}}_{\boldsymbol{\theta}}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \tilde{\mathbf{B}}_{\boldsymbol{\theta}} + \tilde{\mathbf{B}}'_{\boldsymbol{\theta}}], \\ \tilde{\mathbf{O}}_{\tilde{n}} := [\tilde{\mathbf{A}}_{\boldsymbol{\theta}}\tilde{\mathbf{O}}_{\tilde{n}-1} \quad \tilde{\mathbf{F}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}_{\tilde{n}-1} \quad \ldots \quad \tilde{\mathbf{F}}_{\boldsymbol{\theta},M}\tilde{\mathbf{O}}_{\tilde{n}-1}], \\ \tilde{n} = 1,\ldots,\tilde{N}-1 \tag{13}$$

where $\tilde{\mathbf{B}}'_{\boldsymbol{\theta}} = [\tilde{\mathbf{F}}_{\boldsymbol{\theta},1}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \ \cdots \ \tilde{\mathbf{F}}_{\boldsymbol{\theta},M}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0}]$. Then there exists a finite number of admissible inputs and output sampling points such that the parameter vector $\boldsymbol{\theta}$ is locally identifiable with respect to a finite number of inputs, i.e., the associated Fisher information matrix $\mathbf{I}_{\boldsymbol{\theta}}$ is nonsingular if and only if

$$\text{rank}\left\{\left[\mathbf{C}_{\boldsymbol{\theta},1}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \mathbf{C}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}^{(\tilde{p})} \quad \mathbf{C}_{\boldsymbol{\theta},2}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \mathbf{C}_{\boldsymbol{\theta},2}\tilde{\mathbf{O}}^{(\tilde{p})} \\ \cdots \quad \mathbf{C}_{\boldsymbol{\theta},R}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \mathbf{C}_{\boldsymbol{\theta},R}\tilde{\mathbf{O}}^{(\tilde{p})}\right]\right\} = K$$

where $\tilde{p}$ is the integer such that

$$\text{range}\left\{\tilde{\mathbf{O}}^{(0)}\right\} \subset \text{range}\left\{\tilde{\mathbf{O}}^{(1)}\right\} \subset \cdots \subset \text{range}\left\{\tilde{\mathbf{O}}^{(\tilde{p})}\right\} \\ = \text{range}\left\{\tilde{\mathbf{O}}^{(\tilde{p}+1)}\right\} = \cdots = \text{range}\left\{\tilde{\mathbf{O}}^{(\tilde{N}-1)}\right\}. \tag{14}$$

This theorem provides a criterion in terms of the system matrices and the initial state vector of the derivative system for the existence of a finite number of inputs that will lead to the local identifiability of the parameter vector. An important aspect of this criterion is that it is given in terms of the system matrices and initial conditions without any reference to inputs. However, the result does not provide a direct means of constructing the inputs that lead to local identifiability. After assessing whether or not inputs exist that lead to local identifiability specific inputs need to be found that achieve this. This requires criteria for the local identifiability given a specific input. The derivation of such criteria is one of the topics of the next section.

III. Piecewise Constant Inputs

In this section, we apply the general results derived in Section II for admissible inputs to a special class of *piecewise constant* inputs $\mathbb{U}$, which are vector-valued piecewise constant functions with a finite number of steps. A piecewise constant input $\mathbf{u} \in \mathbb{U}$ can be represented by

$$\mathbf{u}(t) = \sum_{l=0}^{L-1} \mathbf{u}^{[l]}\beta^{[l]}(t), \quad t^{[0]} \leq t < t^{[L]} \tag{15}$$

where $\mathbf{u}^{[l]} := [u_1^{[l]} \ \cdots \ u_M^{[l]}]^T, l = 0,\ldots,L-1$, are constant vectors, and $\beta^{[l]}(t), l = 0,\ldots,L-1$, are the indicator functions defined by

$$\beta^{[l]}(t) = \begin{cases} 1, & \text{for } t \in \left[t^{[l]}, t^{[l+1]}\right) \\ 0, & \text{for } t \notin \left[t^{[l]}, t^{[l+1]}\right). \end{cases}$$

Here, $t^{[0]},\ldots,t^{[L]}$ denote the starting and ending points of the time intervals with $t^{[0]} < \cdots < t^{[L]}$, where $t^{[L]}$ can be either finite or infinite. Note that $\mathbf{u}^{[l]}$ could be a zero vector, and that for a piecewise constant input $\mathbf{u} \in \mathbb{U}$ as defined in (15) we are only interested in the output $\mathbf{y}_{\boldsymbol{\theta}}(t)$ for $t^{[0]} \leq t < t^{[L]}$.

The restriction of our study to the special class of piecewise constant inputs $\mathbb{U}$ is of great interest for several reasons. First and most important, the output of a bilinear system can be expressed in closed form for an input $\mathbf{u} \in \mathbb{U}$ (see, e.g., [22]), while the output of a bilinear system is typically expressed by the infinite Volterra series for an admissible input $\mathbf{u} \in \mathbb{V}$ (see [21]). The closed form expression of the output greatly facilitates the derivation of several new results on the CRLB and local identifiability with respect to a specific input to be presented in this section. Second, the two classes $\mathbb{U}$ and $\mathbb{V}$ are closely related because $\mathbb{U} \subset \mathbb{V}$ and $\mathbb{U}$ is dense in $\mathbb{V}$ [23], [24]. For further details on the approximation of a system with general admissible inputs to the case when the inputs are restricted to piecewise constant inputs (see [23] and [24]).

Lemma 3.1 [22]: Consider a bilinear system represented by $\Phi := \{\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1}, \ldots, \mathbf{F}_{\boldsymbol{\theta},M}\}$, with a specific input $\mathbf{u} \in \mathbb{U}$. The derivative system of Φ is represented by $\tilde{\Phi} := \{\tilde{\mathbf{A}}_{\boldsymbol{\theta}}, \tilde{\mathbf{B}}_{\boldsymbol{\theta}}, \tilde{\mathbf{C}}_{\boldsymbol{\theta}}, \tilde{\mathbf{F}}_{\boldsymbol{\theta},1}, \ldots, \tilde{\mathbf{F}}_{\boldsymbol{\theta},M}\}$, with an initial state vector $\tilde{\mathbf{x}}_{\boldsymbol{\theta},0}$. Assume that all the eigenvalues of $\mathbf{A}_{\boldsymbol{\theta}} + \mathbf{F}_{\boldsymbol{\theta}}^{[l]}$ are in the open left-half plane, where $\mathbf{F}_{\boldsymbol{\theta}}^{[l]} := \sum_{m=1}^{M}\mathbf{F}_{\boldsymbol{\theta},m}u_m^{[l]}, l = 0,\ldots,L-1$.

Let $\tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]} := \mathrm{diag}\{\partial_1 \mathbf{F}_{\boldsymbol{\theta}}^{[l]}, \ldots, \partial_K \mathbf{F}_{\boldsymbol{\theta}}^{[l]}\}, l = 0, \ldots, L-1$, where for $k = 1, \ldots, K$

$$\partial_k \mathbf{F}_{\boldsymbol{\theta}}^{[l]} := \begin{bmatrix} \mathbf{F}_{\boldsymbol{\theta}}^{[l]} & 0 \\ \frac{\partial \mathbf{F}_{\boldsymbol{\theta}}^{[l]}}{\partial \theta_k} & \mathbf{F}_{\boldsymbol{\theta}}^{[l]} \end{bmatrix} = \sum_{m=1}^{M} \begin{bmatrix} \mathbf{F}_{\boldsymbol{\theta},m} & 0 \\ \frac{\partial \mathbf{F}_{\boldsymbol{\theta},m}}{\partial \theta_k} & \mathbf{F}_{\boldsymbol{\theta},m} \end{bmatrix} u_m^{[l]}.$$

Then, the state and output of the derivative system $\tilde{\Phi}$ are given by

$$\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t) = \sum_{l=0}^{L-1} \left[\tilde{\mathbf{Q}}_{\boldsymbol{\theta}}^{[l]}(t) \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) \right) - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \right] \beta^{[l]}(t) \tag{16}$$

$$\begin{aligned} \tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t) &= \tilde{\mathbf{C}}_{\boldsymbol{\theta}} \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t) \\ &= \sum_{l=0}^{L-1} \left[\tilde{\mathbf{C}}_{\boldsymbol{\theta}} \tilde{\mathbf{Q}}_{\boldsymbol{\theta}}^{[l]}(t) \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) \right) - \tilde{\mathbf{C}}_{\boldsymbol{\theta}} \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \right] \beta^{[l]}(t), \\ & t^{[0]} \le t < t^{[L]} \end{aligned} \tag{17}$$

where $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[0]}) = \tilde{\mathbf{x}}_{\boldsymbol{\theta},0}$, $\tilde{\mathbf{Q}}_{\boldsymbol{\theta}}^{[l]}(t) := e^{(\tilde{\mathbf{A}}_{\boldsymbol{\theta}} + \tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]})(t - t^{[l]})}$, for $l = 0, \ldots, L-1$, and $\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}$ is defined as

$$\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} := \begin{cases} -\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[0]}), & \text{for } l = -1 \\ \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta}} + \tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]} \right)^{-1} \tilde{\mathbf{B}}_{\boldsymbol{\theta}} \mathbf{u}^{[l]}, & \text{for } l = 0, \ldots, L-1. \end{cases} \tag{18}$$

Remark 3.1: The assumption that all the eigenvalues of $\mathbf{A}_{\boldsymbol{\theta}} + \mathbf{F}_{\boldsymbol{\theta}}^{[l]}, l = 0, \ldots, L-1$, are in the open left-half plane implies that all the eigenvalues of $\tilde{\mathbf{A}}_{\boldsymbol{\theta}} + \tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]}, l = 0, \ldots, L-1$, are also in the open left-half plane, since $\tilde{\mathbf{A}}_{\boldsymbol{\theta}} + \tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]}, l = 0, \ldots, L-1$, is a block lower triangular matrix with $\mathbf{A}_{\boldsymbol{\theta}} + \mathbf{F}_{\boldsymbol{\theta}}^{[l]}$ as its diagonal block submatrices. Consequently, $\tilde{\mathbf{A}}_{\boldsymbol{\theta}} + \tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]}, l = 0, \ldots, L-1$, is invertible, and $\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}$ is well defined. Note also that the definition of $\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[-1]} := -\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[0]})$ in (18) is not needed in the above theorem. It is defined for the convenience of several theorems to follow. For a specific input $\mathbf{u} \in \mathbb{U}$ as defined in (15), the state and output vectors are continuous in the interval $[t^{[0]}, t^{[L]})$. In particular, $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l]}) = \lim_{t \to t^{[l]}-} \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t), l = 1, \ldots, L-1$. Hence, we can express $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l]})$ in the recursion, shown at the bottom of the page, where $\tilde{\mathbf{w}}_{\boldsymbol{\theta}}(t^{[l]})$ is given in (18).

With Theorem 2.1 and Lemma 3.1, we can readily derive a closed form expression of the Fisher information matrix $\mathbf{I}_{\boldsymbol{\theta}}$ for a bilinear system with a specific input $\mathbf{u} \in \mathbb{U}$ when the output of the bilinear system is sampled (uniformly or nonuniformly) at $t^{[0]} = t_0 < t_1 < \ldots < t_{J-1} < t^{[L]}$. The local identifiability with respect to a specific input from $\mathbb{U}$, or just local identifiability in short in this section, can then be determined by checking the nonsingularity of $\mathbf{I}_{\boldsymbol{\theta}}$. However, checking the nonsingularity of $\mathbf{I}_{\boldsymbol{\theta}}$ directly is computationally rather inefficient, particularly for a large number of data samples. When the output of a bilinear system is sampled uniformly, it is possible to develop a simplified method for checking the nonsingularity of $\mathbf{I}_{\boldsymbol{\theta}}$, as stated in the following theorem.

Theorem 3.1 (Appendix D): Consider the bilinear system represented by $\Phi := \{\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1}, \ldots, \mathbf{F}_{\boldsymbol{\theta},M}\}$, whose derivative system of Φ is represented by $\tilde{\Phi} := \{\tilde{\mathbf{A}}_{\boldsymbol{\theta}}, \tilde{\mathbf{B}}_{\boldsymbol{\theta}}, \tilde{\mathbf{C}}_{\boldsymbol{\theta}}, \tilde{\mathbf{F}}_{\boldsymbol{\theta},1}, \ldots, \tilde{\mathbf{F}}_{\boldsymbol{\theta},M}\}$, with a specific input $\mathbf{u} \in \mathbb{U}$ as defined in (15). Assume the following.

1) The output signal is uniformly sampled with sampling period $T^{[l]}$ in the lth interval, i.e., at
$$t^{[l],[j]} = t^{[l]} + jT^{[l]}, \quad l = 0, \ldots, L-1; \quad j = 0, \ldots, J^{[l]} - 1, \text{ with } t^{[l],[J^{[l]}-1]} < t^{[l+1]}$$
where $t^{[l],[j]}$ denotes the jth sampling instant in the lth interval, and $J^{[l]}$, with $J^{[l]} \ge \tilde{N}+1$, is the total number of samples acquired in the lth interval.
2) All the eigenvalues of $\mathbf{A}_{\boldsymbol{\theta}} + \mathbf{F}_{\boldsymbol{\theta}}^{[l]}, l = 0, \ldots, L-1$, are in the open left-half plane.
3) The noise covariance matrix $\boldsymbol{\Lambda}_{\mathbf{n}}(t^{[l],[j]}) = \sigma^2 \mathbf{I}$, for $l = 0, \ldots, L-1; j = 0, \ldots, J^{[l]} - 1$.

Then, we have the following results.

i) The Fisher information matrix $\mathbf{I}_{\boldsymbol{\theta}}$ for the given data set is given by

$$\begin{aligned} \mathbf{I}_{\boldsymbol{\theta}} = & \frac{1}{\sigma^2} \sum_{r=1}^{R} \tilde{\mathbf{C}}_{\boldsymbol{\theta},r} \left\{ \sum_{l=0}^{L-1} \left[\sum_{j=0}^{J^{[l]}-1} \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]} \right)^j \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) \right) \right. \right. \\ & \times \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) \right)^T \left(\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]} \right)^j \right)^T \\ & - \sum_{j=0}^{J^{[l]}-1} \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]} \right)^j \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \right)^T + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \right)^T \right) \\ & - \sum_{j=0}^{J^{[l]}-1} \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \right)^T + \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \tilde{\mathbf{x}}_{\boldsymbol{\theta}}^T\left(t^{[l]}\right) \right) \left(\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]} \right)^j \right)^T \\ & \left. \left. + \sum_{j=0}^{J^{[l]}-1} \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \right)^T \right] \right\} \tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T \end{aligned} \tag{19}$$

where $\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]} := e^{(\tilde{\mathbf{A}}_{\boldsymbol{\theta}} + \tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]})T^{[l]}}$.

ii) The parameter vector $\boldsymbol{\theta}$ is locally identifiable if and only if $\mathrm{rank}\{[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}' \ \ldots \ \tilde{\mathbf{C}}_{\boldsymbol{\theta},R}\tilde{\mathbf{O}}']\} = K$, where $\tilde{\mathbf{O}}' \in \mathbb{R}^{\tilde{N} \times L(\tilde{N}+1)}$ is defined as $\tilde{\mathbf{O}}' := [\tilde{\mathbf{O}}'_0 \ \ldots \ \tilde{\mathbf{O}}'_{L-1}]$, and $\tilde{\mathbf{O}}'_l, l = 0, \ldots, L-1$, is given by

$$\begin{aligned} \tilde{\mathbf{O}}'_l := \Big[\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \quad \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) \right) \quad \tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]} \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) \right) \\ \ldots \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]} \right)^{\tilde{N}-1} \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) \right) \Big]. \end{aligned} \tag{20}$$

iii) When $J^{[l]} \to \infty, l = 0, \ldots, L-1$, the local identifiability condition is the same as that in part ii) except

$$\tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right) = \begin{cases} \tilde{\mathbf{x}}_{\boldsymbol{\theta},0}, & \text{for } l = 0, \\ \tilde{\mathbf{Q}}_{\boldsymbol{\theta}}^{[l-1]}\left(t^{[l]}\right) \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l-1]}) \right) - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}, & \text{for } l = 1, \ldots, L-1 \end{cases}$$

that the expression for $\tilde{\mathbf{O}}'_l$, $l = 0, \ldots, L-1$, given in (20) now reduces to

$$\tilde{\mathbf{O}}'_l = \Big[\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} \quad \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\right) \quad \tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\right) \quad \ldots \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{\tilde{N}-1}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\right)\Big]. \quad (21)$$

Although the Fisher information matrix could be calculated from part i) of Theorem 3.1, it is computationally rather inefficient to directly compute the sum in (19), particularly when the number of samples is large. If we are only interested in the local identifiability, the simplified nonsingularity condition presented in part ii) of Theorem 3.1 can be used. In particular, when the number of data samples approaches infinity, we are unable to determine the local identifiability through calculating the sum in (19) directly. Nevertheless, the local identifiability in such a case can be checked easily based on $\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}, l = -1, \ldots, L-1$, which depends only on the system matrices, the initial state vector of the derivative system and the specific piecewise constant input.

From the above discussion, we see that there is still a need to calculate the Fisher information matrix efficiently if we are interested not just in the local identifiability, but also in the CRLB. We now propose an alternative method for calculating the Fisher information matrix efficiently through the solution to a Lyapunov equation. Standard results on Lyapunov equations can be found in [25, pp. 177–180], [26].

Theorem 3.2 (Appendix E): Assume that the data model and all the assumptions are the same as in Theorem 3.1. Then the Fisher information matrix for the given data set is

$$\mathbf{I}_{\boldsymbol{\theta}} = \frac{1}{\sigma^2}\sum_{r=1}^{R}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\left\{\sum_{l=0}^{L-1}\left[\tilde{\mathbf{P}}_1^{[l]} - \tilde{\mathbf{P}}_2^{[l]} - \left(\tilde{\mathbf{P}}_2^{[l]}\right)^T + J^{[l]}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right]\right\}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T \quad (22)$$

where $\tilde{\mathbf{P}}_1^{[l]}$ and $\tilde{\mathbf{P}}_2^{[l]}$ are obtained as follows. $\tilde{\mathbf{P}}_1^{[l]}, l = 0, \ldots, L-1$, is the unique solution to the following Lyapunov equation:

$$\begin{aligned}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\tilde{\mathbf{P}}_1^{[l]}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^T - \tilde{\mathbf{P}}_1^{[l]} &= -\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)^T \\ &\quad + \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J^{[l]}}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right) \\ &\quad \times \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)^T\left(\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J^{[l]}}\right)^T.\end{aligned}$$

$\tilde{\mathbf{P}}_2^{[l]}, l = 0, \ldots, L-1$, is given by

$$\tilde{\mathbf{P}}_2^{[l]} = \left(\mathbf{I} - \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J^{[l]}}\right)\left(\mathbf{I} - \tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{-1} \times \left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right).$$

For finite data samples, we can calculate the CRLB by directly inverting $\mathbf{I}_{\boldsymbol{\theta}}$ when it is nonsingular. This is because the inverse of a nonsingular $\mathbf{I}_{\boldsymbol{\theta}}$ is well defined for finite data samples. However, when the number of data samples approaches infinity, i.e., $J^{[l]} \to \infty, l = 0, \ldots, L-1$, the asymptotic CRLB cannot be computed either by part i) of Theorem 3.1 or Theorem 3.2 in general, since the term $\sum_{j=0}^{J^{[l]}-1}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]})^T\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T$ in (19) and the term $J^{[l]}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]})^T\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T$ in (22) can tend to infinity, leading $\mathbf{I}_{\boldsymbol{\theta}}$ to infinity as well. To overcome this difficulty, we propose a novel method for calculating the asymptotic CRLB without computing $\mathbf{I}_{\boldsymbol{\theta}}$ directly.

Given that the nonsingularity condition in part iii) of Theorem 3.1 holds for the asymptotic case, the following theorem proves the existence of the asymptotic CRLB and gives an explicit expression. The asymptotic CRLB for the limiting case of infinite data samples is defined as the limit of the CRLB for the corresponding finite data sample situations. This limit exists since the Fisher information matrices form a monotonically increasing sequence of positive semidefinite matrices.

The CRLB is often used to provide guidance for experimental design. In this context the question often arises how many data samples should be acquired. It is therefore important to know what the CRLB provides in the limiting case where an infinite number of data samples are available.

Theorem 3.3 (Appendix F): Assume that the data model is the same as in Theorem 3.1, except that the number of equidistant samples in $[t^{[l]}, t^{[l+1]})$ tends to infinity, i.e., $J^{[l]} = J', l = 0, \ldots, L-1$, and $J' \to \infty$. Assume the nonsingularity condition in part iii) of Theorem 3.1 holds. Then the asymptotic CRLB is given by

$$\begin{aligned}\mathrm{var}(\hat{\boldsymbol{\theta}}) &\geq \lim_{J' \to \infty}\mathbf{I}_{\boldsymbol{\theta}}^{-1}(J') \\ &= \begin{cases}\sigma^2\tilde{\mathbf{U}}^{\perp}((\tilde{\mathbf{U}}^{\perp})^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}^{\perp})^{-1}(\tilde{\mathbf{U}}^{\perp})^T, & \text{if } \mathrm{rank}(\tilde{\mathbf{U}}) < K \\ 0, & \text{if } \mathrm{rank}(\tilde{\mathbf{U}}) = K\end{cases}\end{aligned}$$

where $\tilde{\mathbf{P}}, \tilde{\mathbf{U}}$ and $\tilde{\mathbf{U}}^{\perp}$ are defined as follows.

i) Construction of $\tilde{\mathbf{P}}$:

$$\tilde{\mathbf{P}} := \sum_{r=1}^{R}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\left\{\sum_{l=0}^{L-1}\left[\tilde{\mathbf{P}}_1^{[l]} - \tilde{\mathbf{P}}_2^{[l]} - \left(\tilde{\mathbf{P}}_2^{[l]}\right)^T\right]\right\}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T$$

where $\tilde{\mathbf{P}}_1^{[l]}, l = 0, \ldots, L-1$, is the unique solution to the following Lyapunov equation:

$$\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\tilde{\mathbf{P}}_1^{[l]}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^T - \tilde{\mathbf{P}}_1^{[l]} = -\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\right)^T$$

and $\tilde{\mathbf{P}}_2^{[l]}, l = 0, \ldots, L-1$, is given by

$$\tilde{\mathbf{P}}_2^{[l]} = \left(\mathbf{I} - \tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{-1}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right).$$

ii) Construction of $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{U}}^{\perp}$: Represent the span of all $\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}, r = 1, \ldots, R, l = 0, \ldots, L-1$, by Ψ, and let G denote the rank of Ψ. Then $\tilde{\mathbf{U}} \in \mathbb{R}^{K \times G}$ is defined as a full rank matrix such that the column space of $\tilde{\mathbf{U}}$ is equal to Ψ, i.e., $\mathrm{range}\{\tilde{\mathbf{U}}\} = \Psi$. For $G < K, \tilde{\mathbf{U}}^{\perp} \in \mathbb{R}^{K \times (K-G)}$ is defined as a full rank matrix such that

$$\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}^{\perp} = 0 \quad \text{and} \quad \mathrm{rank}\{[\tilde{\mathbf{U}} \quad \tilde{\mathbf{U}}^{\perp}]\} = K.$$

In the next section, we illustrate the theoretical results presented in this paper using a simulation example of surface plasmon resonance experiments for the determination of the kinetic parameters of protein–protein interactions.

IV. Example

Surface plasmon resonance (SPR) (see, e.g., [27] and [28]) occurs under certain conditions from a conducting film at the interface between two media of different refractive index. Biosensors such as instruments by the BIAcore company offer a technique for monitoring protein–protein interactions in real time using an optical detection principle based on SPR. In the experiments one of the proteins (ligand) is coupled to a sensor chip and the second protein (analyte) is flowed across the surface coupled ligand using a micro-fluidic device. The SPR response reflects a change in mass concentration at the detector surface as molecules bind or dissociate from the sensor chip. The measured response data can be used to estimate the kinetic constants of protein–protein interactions.

In this simulation example we use the theoretical results presented in the previous sections to analyze the SPR experiments for one-to-one protein–protein interactions that can be modeled by the differential equation

$$\dot{R}(t) = k_a(R_{\max} - R(t))C_0(t) - k_d R(t), \quad t \geq t^{[0]} \tag{23}$$

where $R(t)$ is the measured SPR response in resonance units (RU), $R(t^{[0]}) = 0$, k_a and k_d are the kinetic association and dissociation constants of the interaction respectively, $R_{\max}$ is the maximum analyte binding capacity in RU, $C_0(t)$ is the concentration value of the analyte in the flow cell which can be controlled in the experiments, and the initial SPR response is assumed to be zero.

Let $\mathbf{x}_{\boldsymbol{\theta}}(t) := R(t), \mathbf{u}(t) := C_0(t), \mathbf{y}_{\boldsymbol{\theta}}(t) := R(t), t \geq t^{[0]}$, and $\mathbf{x}_{\boldsymbol{\theta},0} := R(t^{[0]}) = 0$, (23) becomes the following bilinear system $\Phi = \{\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1}\}$:

$$\dot{\mathbf{x}}_{\boldsymbol{\theta}}(t) = \mathbf{A}_{\boldsymbol{\theta}}\mathbf{x}_{\boldsymbol{\theta}}(t) + \mathbf{F}_{\boldsymbol{\theta},1}u_1(t)\mathbf{x}_{\boldsymbol{\theta}}(t) + \mathbf{B}_{\boldsymbol{\theta}}\mathbf{u}(t) \tag{24}$$

$$\mathbf{y}_{\boldsymbol{\theta}}(t) = \mathbf{C}_{\boldsymbol{\theta}}\mathbf{x}_{\boldsymbol{\theta}}(t), \quad t \geq t^{[0]} \tag{25}$$

where $\mathbf{x}_{\boldsymbol{\theta}}(t^{[0]}) = \mathbf{x}_{\boldsymbol{\theta},0} = R(t^{[0]}) = 0, \mathbf{A}_{\boldsymbol{\theta}} = -k_d, \mathbf{B}_{\boldsymbol{\theta}} = k_a R_{\max}, \mathbf{C}_{\boldsymbol{\theta}} = 1, \mathbf{F}_{\boldsymbol{\theta},1} = -k_a$. The unknown parameter vector to be estimated in the experiments is $\boldsymbol{\theta} = [k_a \ k_d \ R_{\max}]^T$.

A practical SPR experiment may consist of an association phase ($t^{[0]} \leq t < t^{[1]}$) and a dissociation phase ($t^{[1]} \leq t < t^{[2]}$), or one of these two phases. During the association phase analyte is flowed across the ligand on the sensor chip with constant concentration C_0 up to time $t^{[1]}$, i.e., $C_0(t) = C_0, t^{[0]} \leq t < t^{[1]}$. The dissociation phase immediately follows the association phase and is characterized by analyte free buffer being flowed across the sensor chip, i.e., $C_0(t) = 0, t^{[1]} \leq t < t^{[2]}$. Hence, a two-phase SPR experiment can be modeled by the bilinear system $\Phi = \{\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1}\}$ with a two-phase piecewise constant input

$$\mathbf{u}(t) = \mathbf{u}^{[0]}\beta^{[0]}(t) + \mathbf{u}^{[1]}\beta^{[1]}(t), \quad t^{[0]} \leq t < t^{[2]} \tag{26}$$

where

$$\mathbf{u}^{[0]} = C_0, \quad \beta^{[0]}(t) = \begin{cases} 1, & \text{for } t \in [t^{[0]}, t^{[1]}) \\ 0, & \text{for } t \notin [t^{[0]}, t^{[1]}) \end{cases}$$

$$\mathbf{u}^{[1]} = 0, \quad \beta^{[1]}(t) = \begin{cases} 1, & \text{for } t \in [t^{[1]}, t^{[2]}) \\ 0, & \text{for } t \notin [t^{[1]}, t^{[2]}). \end{cases}$$

Note that in the two-phase SPR experiment the output samples are obtained from $\mathbf{y}_{\boldsymbol{\theta}}(t)$ for $t^{[0]} \leq t < t^{[2]}$.

A. Derivative System

The first step is the calculation of the derivative system by Theorem 2.1. We represent the derivative system of $\Phi = \{\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1}\}$ by $\tilde{\Phi} := \{\tilde{\mathbf{A}}_{\boldsymbol{\theta}}, \tilde{\mathbf{B}}_{\boldsymbol{\theta}}, \tilde{\mathbf{C}}_{\boldsymbol{\theta}}, \tilde{\mathbf{F}}_{\boldsymbol{\theta},1}\}$, where $\tilde{\mathbf{A}}_{\boldsymbol{\theta}}, \tilde{\mathbf{B}}_{\boldsymbol{\theta}}, \tilde{\mathbf{C}}_{\boldsymbol{\theta}}, \tilde{\mathbf{F}}_{\boldsymbol{\theta},1}$ are given as follows.

$$\tilde{\mathbf{A}}_{\boldsymbol{\theta}} := \operatorname{diag}\{\partial_1\mathbf{A}_{\boldsymbol{\theta}}, \partial_2\mathbf{A}_{\boldsymbol{\theta}}, \partial_3\mathbf{A}_{\boldsymbol{\theta}}\} \quad \text{where}$$

$$\partial_1\mathbf{A}_{\boldsymbol{\theta}} = \begin{bmatrix} -k_d & 0 \\ 0 & -k_d \end{bmatrix}, \quad \partial_2\mathbf{A}_{\boldsymbol{\theta}} = \begin{bmatrix} -k_d & 0 \\ -1 & -k_d \end{bmatrix},$$

$$\partial_3\mathbf{A}_{\boldsymbol{\theta}} = \begin{bmatrix} -k_d & 0 \\ 0 & -k_d \end{bmatrix}.$$

$$\tilde{\mathbf{B}}_{\boldsymbol{\theta}} := \begin{bmatrix} \partial_1\mathbf{B}_{\boldsymbol{\theta}} \\ \partial_2\mathbf{B}_{\boldsymbol{\theta}} \\ \partial_3\mathbf{B}_{\boldsymbol{\theta}} \end{bmatrix} \quad \text{where}$$

$$\partial_1\mathbf{B}_{\boldsymbol{\theta}} = \begin{bmatrix} k_a R_{\max} \\ R_{\max} \end{bmatrix}, \quad \partial_2\mathbf{B}_{\boldsymbol{\theta}} = \begin{bmatrix} k_a R_{\max} \\ 0 \end{bmatrix},$$

$$\partial_3\mathbf{B}_{\boldsymbol{\theta}} = \begin{bmatrix} k_a R_{\max} \\ k_a \end{bmatrix}.$$

$$\tilde{\mathbf{C}}_{\boldsymbol{\theta}} := \operatorname{diag}\{\partial_1\mathbf{C}_{\boldsymbol{\theta},1}, \partial_2\mathbf{C}_{\boldsymbol{\theta},1}, \partial_3\mathbf{C}_{\boldsymbol{\theta},1}\} \quad \text{where}$$

$$\partial_1\mathbf{C}_{\boldsymbol{\theta},1} = [0 \quad 1], \quad \partial_2\mathbf{C}_{\boldsymbol{\theta},1} = [0 \quad 1],$$

$$\partial_3\mathbf{C}_{\boldsymbol{\theta},1} = [0 \quad 1].$$

$$\tilde{\mathbf{F}}_{\boldsymbol{\theta},1} := \operatorname{diag}\{\partial_1\mathbf{F}_{\boldsymbol{\theta},1}, \partial_2\mathbf{F}_{\boldsymbol{\theta},1}, \partial_3\mathbf{F}_{\boldsymbol{\theta},1}\} \quad \text{where}$$

$$\partial_1\mathbf{F}_{\boldsymbol{\theta},1} = \begin{bmatrix} -k_a & 0 \\ -1 & -k_a \end{bmatrix}, \quad \partial_2\mathbf{F}_{\boldsymbol{\theta},1} = \begin{bmatrix} -k_a & 0 \\ 0 & -k_a \end{bmatrix},$$

$$\partial_3\mathbf{F}_{\boldsymbol{\theta},1} = \begin{bmatrix} -k_a & 0 \\ 0 & -k_a \end{bmatrix}.$$

Since the initial state $\mathbf{x}_{\boldsymbol{\theta},0}$ of Φ is equal to zero, the initial state vector $\tilde{\mathbf{x}}_{\boldsymbol{\theta},0}$ of $\tilde{\Phi}$ is also equal to zero, i.e., $\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} = [0\ 0\ 0\ 0\ 0\ 0]^T$. Obviously, the partial derivatives of $\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1}$, and $\mathbf{x}_{\boldsymbol{\theta},0}$ with respect to the elements of $\boldsymbol{\theta}$ are continuously differentiable functions of $\boldsymbol{\theta}$ for all $\boldsymbol{\theta}$.

B. Local Identifiability

Before computing the CRLB we first check whether there exist a finite number of inputs such that $\boldsymbol{\theta}$ will be locally identifiable. We then examine the local identifiability for a specific two-phase input.

1) Case 1: Existence of Inputs That Lead to Local Identifiability: In this case we assume that we can freely select an input to the bilinear system model Φ from the set of admissible inputs $\mathbb{V}$, repeat the experiment for another input from the same set, and measure the corresponding output samples for each experiment conducted. Simple calculations give

$$\begin{aligned}&\text{rank}\,\{[\,\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0}\quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}^{(0)}\,]\}\\ &\quad=\text{rank}\left\{\begin{bmatrix}0 & R_{\max} & -k_dR_{\max} & -2k_aR_{\max}\\ 0 & 0 & -k_aR_{\max} & 0\\ 0 & k_a & -k_ak_d & -k_a^2\end{bmatrix}\right\}=3\end{aligned}$$

for positive k_a, k_d and $R_{\max}$. The positivity assumption is restriction since the constants are naturally positive. Since the size of $\boldsymbol{\theta}$ is 3, by Theorem 2.3 $\boldsymbol{\theta}$ is locally identifiable with respect to a finite number of inputs from the set of admissible inputs.

2) Case 2: Specific Two-Phase Piecewise Constant Input With Uniform Sampling: We next exploit Theorem 3.1 to find out whether the same parameter vector $\boldsymbol{\theta}$ is still locally identifiable with respect to the specific two-phase piecewise constant input defined in (26), assuming that the output is sampled uniformly. Denote the sampling intervals of the association and dissociation phases as $T^{[0]}$ and $T^{[1]}$, respectively. By simple calculations, we obtain the equation shown at the bottom of the page, where $a_0 := e^{-(C_0k_a+k_d)T^{[0]}}, a_1 := e^{-k_dT^{[1]}}$, and $a_0' := e^{-(C_0k_a+k_d)(t^{[1]}-t^{[0]})}$. It is easy to verify that

$$\begin{aligned}&\text{rank}\{\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}'\}\\ &\quad=\text{rank}\left\{\left[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}\quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}+\tilde{\mathbf{x}}_{\boldsymbol{\theta},0})\right.\right.\\ &\qquad\left.\left.\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[1]}(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[1]}+\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[1]}))\right]\right\}=3.\end{aligned}$$

Thus, by Theorem 3.1, $\boldsymbol{\theta}$ is locally identifiable with respect to the specific two-phase piecewise constant input defined in (26) and with uniformly sampled output. Although it happens that for this particular example the parameter vector is locally identifiable with respect to a finite number of inputs from the set of admissible inputs, and also with respect to the specific two-phase piecewise constant input defined in (26), this is not necessarily so in general, as we will see later in this section.

We next consider the nonsingularity criterion in part iii) of Theorem 3.1 for the asymptotic situation where an infinite number of data set points are available. Of the necessary expressions all have already been established with the exception of

$$\begin{aligned}&\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[1]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[1]}-\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}\right)\\ &\quad=\begin{bmatrix}\frac{a_1C_0k_dR_{\max}}{(C_0k_a+k_d)^2}\\ \frac{a_1C_0k_aR_{\max}\left(T^{[1]}C_0k_a+T^{[1]}k_d+1\right)}{(C_0k_a+k_d)^2}\\ \frac{a_1C_0k_a}{C_0k_a+k_d}\end{bmatrix}.\end{aligned}$$

Hence (note that $\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[-1]}=-\tilde{\mathbf{x}}_{\boldsymbol{\theta},0}$), we have

$$\begin{aligned}&\text{rank}\{\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}''\}\\ &\quad=\text{rank}\left\{\left[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}\quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}-\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[-1]})\right.\right.\\ &\qquad\left.\left.\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[1]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[1]}-\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}\right)\right]\right\}=3.\end{aligned}$$

It then follows from Theorem 3.1 that $\boldsymbol{\theta}$ is also locally identifiable with respect to the specific two-phase piecewise constant input defined in (26) for sufficiently large $J^{[0]}$ and $J^{[1]}$. This is of course an obvious result since local identifiability in the finite data case implies local identifiability in the infinite data case.

C. CRLB and Asymptotic CRLB

Since $\boldsymbol{\theta}$ is locally identifiable with respect to a two-phase piecewise constant input in the SPR experiment with uniformly sampled finite data, the next step is to apply Theorems 3.2 and 3.3 to numerically calculate the CRLB and asymptotic CRLB. Here we use simulated data so that we could conveniently select various experimental settings. For comparison, typical numerical values from [29] are assigned to the unknown parameters, i.e.,

$$k_a=1478\text{M}^{-1}\text{s}^{-1}, k_d=4.5\times10^{-3}\text{s}^{-1},\quad R_{\max}=7.75\ \text{RU}.$$

The sampling intervals are chosen as $T^{[0]}=T^{[1]}=1$ s, and the noise variance is assumed to be $\sigma^2=1$. Fig. 1 plots the CRLB in terms of the standard deviations of k_a, k_d and $R_{\max}$ as functions of C_0 and the number of data samples. Obviously, it shows that increasing the number of samples improves the accuracy of estimation. As can be seen from the figure, when the number of samples is sufficiently large, e.g., $J^{[0]}=J^{[1]}=1000$, the CRLB approaches the asymptotic CRLB, which is the lowest possible CRLB, given fixed sampling intervals. The plot also reveals that the concentration value C_0 has an influence on the accuracy of parameter estimation. From Fig. 1(a), the optimal values of C_0 corresponding to the lowest variances of k_a for different number of data samples lie between 1.0×10^{-5} M and 2.0×10^{-5} M, and for C_0 greater than the optimal values the variance increases slowly with C_0. On the other hand, the variances of k_d and $R_{\max}$ decrease with the increase of C_0, but remain almost constant when C_0 is greater than 2.0×10^{-5} M. Therefore, a good

$$\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}=\begin{bmatrix}-\frac{C_0k_dR_{\max}}{(C_0k_a+k_d)^2}\\ \frac{C_0k_aR_{\max}}{(C_0k_a+k_d)^2}\\ -\frac{C_0k_a}{C_0k_a+k_d}\end{bmatrix},\quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}+\tilde{\mathbf{x}}_{\boldsymbol{\theta},0}\right)=\begin{bmatrix}\frac{a_0C_0R_{\max}(T^{[0]}C_0^2k_a^2+T^{[0]}C_0k_ak_d-k_d)}{(C_0k_a+k_d)^2}\\ \frac{a_0C_0k_aR_{\max}(T^{[0]}C_0k_a+T^{[0]}k_d+1)}{(C_0k_a+k_d)^2}\\ -\frac{a_0C_0k_a}{C_0k_a+k_d}\end{bmatrix},$$

$$\begin{aligned}&\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[1]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[1]}+\tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[1]}\right)\right)\\ &\quad=\begin{bmatrix}\frac{a_1C_0R_{\max}\left[\left(t^{[1]}-t^{[0]}\right)a_0'C_0^2k_a^2+\left(t^{[1]}-t^{[0]}\right)a_0'C_0k_ak_d-a_0'k_d+k_d\right]}{(C_0k_a+k_d)^2}\\ \frac{a_1C_0k_aR_{\max}\left[\left(t^{[1]}-t^{[0]}\right)a_0'C_0k_a+\left(t^{[1]}-t^{[0]}\right)a_0'k_d+T^{[1]}a_0'C_0k_a+T^{[1]}a_0'k_d-T^{[1]}C_0k_a-T^{[1]}k_d+a_0'-1\right]}{(C_0k_a+k_d)^2}\\ \frac{a_1C_0k_a(1-a_0')}{C_0k_a+k_d}\end{bmatrix}.\end{aligned}$$

choice of C_0 for practical two-phase SPR experiments would be around the value of 2.0×10^{-5} M.

D. Analytical Solution of Asymptotic CRLB

For the two-phase SPR experimental model with identical uniform sampling interval T for both the association and the dissociation phases, i.e., $T^{[0]} = T^{[1]} = T$, it is in fact possible to give an explicit expression for the asymptotic CRLB for the unknown parameters k_a, k_d and $R_{\max}$. The following results are obtained by applying Theorem 3.3 with some algebraic manipulations. The detailed derivations are omitted here but can be found in [30]. (See the equation at the bottom of the next page.)

E. One-Phase SPR Experiment With Uniform Sampling

Finally, we show that the same parameter vector $\boldsymbol{\theta}$ is not locally identifiable with respect to the one-phase piecewise constant input $\mathbf{u}(t) = \mathbf{u}^{[0]}\beta^{[0]}(t),\ t^{[0]} \le t < t^{[1]}$, for the same bilinear system in this example. Note that the corresponding experiment consists of only an association phase, and the output samples are obtained from $\mathbf{y}_{\boldsymbol{\theta}}(t)$ for $t^{[0]} \le t < t^{[1]}$.

Based on the results in Case 2 of Subsection IV-B, it is easy to check that ($\tilde{N} = 2KN = 6$ here)

$$\begin{aligned}\text{rank}\{\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}'\} &= \text{rank}\left\{\left[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta},0}\right)\right.\right.\\ &\qquad \left.\left.\cdots \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}\right)^5\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta},0}\right)\right]\right\} = 2 < 3.\end{aligned}$$

By Theorem 3.1, $\boldsymbol{\theta}$ is not locally identifiable with respect to the above given one-phase piecewise constant input. Similarly, it is easy to show that $\boldsymbol{\theta}$ is not locally identifiable with respect to the same input in the asymptotic case either.

V. Conclusion

In this paper an explicit expression of the Cramer–Rao lower bound (CRLB) has been derived for the problem of estimating the unknown parameters of a bilinear system from noise corrupted output samples. The concept of derivative system allows us to express the Fisher information matrix in terms of the output samples of its associated derivative system. The establishment of the relationship between local identifiability and reachability has led to the derivation of a necessary and sufficient condition for local identifiability with respect to a finite number of admissible inputs. This criterion is based only on the system matrices and the initial state vector of the derivative system. For the special but important class of piecewise constant inputs with uniformly sampled output data sets, we have derived a criterion for local identifiability with respect to a specific input. Moreover, an alternative method was introduced to calculate the Fisher information matrix and the CRLB through the solution of a Lyapunov equation. This approach was exploited to derive an expression for the limit of the CRLB as the number of data points approaches infinity. A simulation example of surface plasmon resonance experiments has illustrated the applicability of the theoretical results and computational methods presented in this paper.

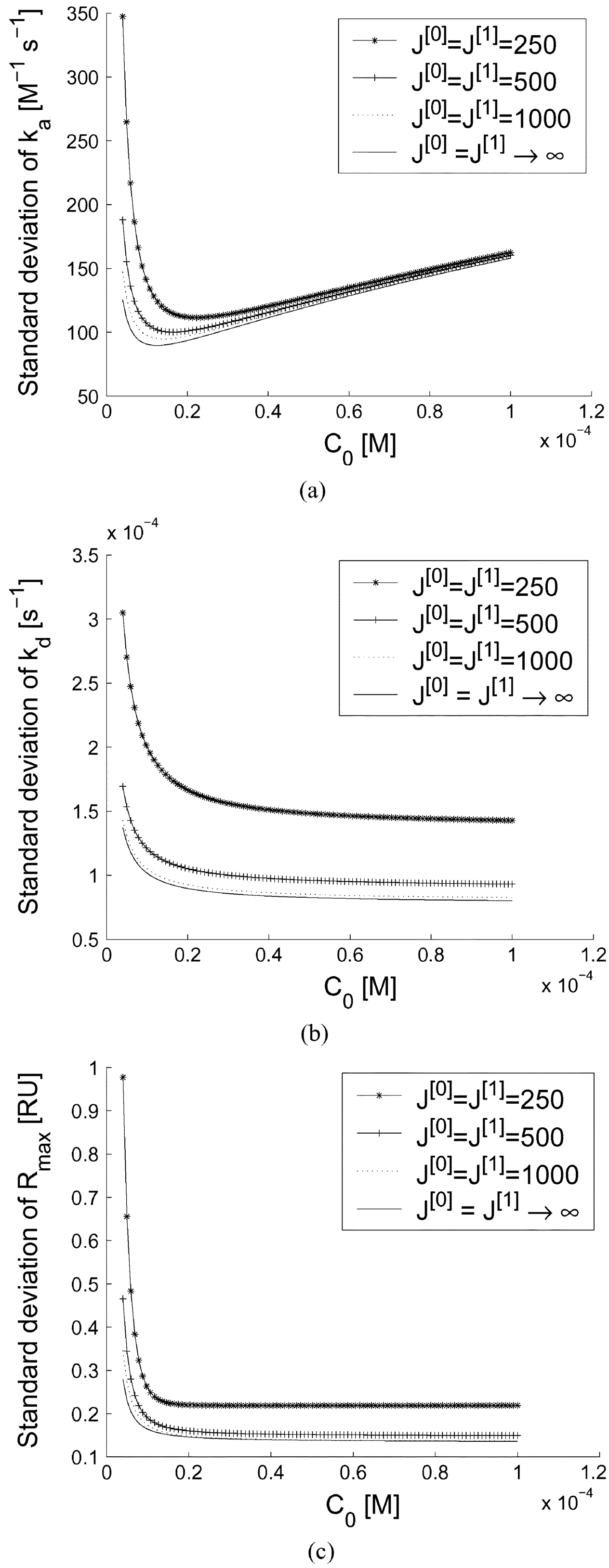


Fig. 1. CRLB for simulated two-phase one-to-one SPR experimental data with $T^{[0]} = T^{[1]} = 1$ s and $\sigma^2 = 1$. (a), (b), and (c) plot the standard deviations of the estimates of k_a, k_d and $R_{\max}$, respectively, for different concentration values and different numbers of samples acquired in the association and dissociation phases.

APPENDIX

A. Proof of Theorem 2.1

1) By assumption the partial derivatives of $\mathbf{A}_{\boldsymbol{\theta}}, \mathbf{B}_{\boldsymbol{\theta}}, \mathbf{C}_{\boldsymbol{\theta}}, \mathbf{F}_{\boldsymbol{\theta},1}, \ldots, \mathbf{F}_{\boldsymbol{\theta},M}$ and $\mathbf{x}_{\boldsymbol{\theta},0}$ with respect to θ_k ($k = 1, \ldots, K$ throughout the proof) exist for $\boldsymbol{\theta} \in \Theta$. Hence, the partial derivatives of $\mathbf{x}_{\boldsymbol{\theta}}(t)$ and $\mathbf{y}_{\boldsymbol{\theta},r}(t), r = 1, \ldots, R$, with respect to θ_k also exist for $\boldsymbol{\theta} \in \Theta$ and $t \geq t^{[0]}$. Since the input $\mathbf{u}(t), t \geq t^{[0]}$, is piecewise continuous, it follows that $\mathbf{x}_{\boldsymbol{\theta}}(t)$ and $(\partial \mathbf{x}_{\boldsymbol{\theta}}(t)/\partial \theta_k)$ are partially differentiable with respect to t on $t \geq t^{[0]}$ with the possible exception of the discontinuities of $\mathbf{u}$. Also, the partial derivative of $(\partial \mathbf{x}_{\boldsymbol{\theta}}(t)/\partial t)$ with respect to θ_k exists for $\boldsymbol{\theta} \in \Theta$ and $t \geq t^{[0]}$ with the possible exception of the discontinuities of $\mathbf{u}$. With the exception of the discrete discontinuities of $\mathbf{u}$, we have $(\partial \dot{\mathbf{x}}_{\boldsymbol{\theta}}(t)/\partial \theta_k) = (\partial^2 \mathbf{x}_{\boldsymbol{\theta}}(t)/\partial \theta_k \partial t) = (\partial^2 \mathbf{x}_{\boldsymbol{\theta}}(t)/\partial t \partial \theta_k), t \geq t^{[0]}$ (see [31,p. 359]).

Taking the partial derivative of (1) with respect to θ_k and using the product formula (see, e.g., [11, Lemma 2.3]) give

$$\frac{\partial \dot{\mathbf{x}}_{\boldsymbol{\theta}}(t)}{\partial \theta_k} = \begin{bmatrix} \frac{\partial \mathbf{A}_{\boldsymbol{\theta}}}{\partial \theta_k} & \mathbf{A}_{\boldsymbol{\theta}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\boldsymbol{\theta}}(t) \\ \frac{\partial \mathbf{x}_{\boldsymbol{\theta}}(t)}{\partial \theta_k} \end{bmatrix} + \sum_{m=1}^{M} \begin{bmatrix} \frac{\partial \mathbf{F}_{\boldsymbol{\theta},m}}{\partial \theta_k} & \mathbf{F}_{\boldsymbol{\theta},m} \end{bmatrix} u_m(t) \begin{bmatrix} \mathbf{x}_{\boldsymbol{\theta}}(t) \\ \frac{\partial \mathbf{x}_{\boldsymbol{\theta}}(t)}{\partial \theta_k} \end{bmatrix} + \frac{\partial \mathbf{B}_{\boldsymbol{\theta}}}{\partial \theta_k} \mathbf{u}(t), \quad t \geq t^{[0]}. \tag{27}$$

With the exception of the discontinuities of $\mathbf{u}$, combining (1) and (27) yields

$$\begin{aligned}\frac{\partial}{\partial t} \partial_k \mathbf{x}_{\boldsymbol{\theta}}(t) =& \frac{\partial}{\partial t} \begin{bmatrix} \mathbf{x}_{\boldsymbol{\theta}}(t) \\ \frac{\partial \mathbf{x}_{\boldsymbol{\theta}}(t)}{\partial \theta_k} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathbf{x}_{\boldsymbol{\theta}}(t)}{\partial t} \\ \frac{\partial^2 \mathbf{x}_{\boldsymbol{\theta}}(t)}{\partial t \partial \theta_k} \end{bmatrix} = \begin{bmatrix} \dot{\mathbf{x}}_{\boldsymbol{\theta}}(t) \\ \frac{\partial \dot{\mathbf{x}}_{\boldsymbol{\theta}}(t)}{\partial \theta_k} \end{bmatrix} \\ =& \begin{bmatrix} \mathbf{A}_{\boldsymbol{\theta}} & 0 \\ \frac{\partial \mathbf{A}_{\boldsymbol{\theta}}}{\partial \theta_k} & \mathbf{A}_{\boldsymbol{\theta}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\boldsymbol{\theta}}(t) \\ \frac{\partial \mathbf{x}_{\boldsymbol{\theta}}(t)}{\partial \theta_k} \end{bmatrix} + \sum_{m=1}^{M} \begin{bmatrix} \mathbf{F}_{\boldsymbol{\theta},m} & 0 \\ \frac{\partial \mathbf{F}_{\boldsymbol{\theta},m}}{\partial \theta_k} & \mathbf{F}_{\boldsymbol{\theta},m} \end{bmatrix} u_m(t) \begin{bmatrix} \mathbf{x}_{\boldsymbol{\theta}}(t) \\ \frac{\partial \mathbf{x}_{\boldsymbol{\theta}}(t)}{\partial \theta_k} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_{\boldsymbol{\theta}} \\ \frac{\partial \mathbf{B}_{\boldsymbol{\theta}}}{\partial \theta_k} \end{bmatrix} \mathbf{u}(t) \\ =& \partial_k \mathbf{A}_{\boldsymbol{\theta}} \partial_k \mathbf{x}_{\boldsymbol{\theta}}(t) + \sum_{m=1}^{M} \partial_k \mathbf{F}_{\boldsymbol{\theta},m} u_m(t) \partial_k \mathbf{x}_{\boldsymbol{\theta}}(t) + \partial_k \mathbf{B}_{\boldsymbol{\theta}} \mathbf{u}(t), \quad t \geq t_0. \end{aligned} \tag{28}$$

Also

$$\partial_k \mathbf{x}_{\boldsymbol{\theta}}(t^{[0]}) = \begin{bmatrix} \mathbf{x}_{\boldsymbol{\theta}}(t^{[0]}) \\ \frac{\partial \mathbf{x}_{\boldsymbol{\theta}}(t^{[0]})}{\partial \theta_k} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{\boldsymbol{\theta},0} \\ \frac{\partial \mathbf{x}_{\boldsymbol{\theta},0}}{\partial \theta_k} \end{bmatrix}. \tag{29}$$

Since $\mathbf{y}_{\boldsymbol{\theta},r}(t) = \mathbf{C}_{\boldsymbol{\theta},r} \mathbf{x}_{\boldsymbol{\theta}}(t), r = 1, \ldots, R, t \geq t^{[0]}$, taking the partial derivative of $\mathbf{y}_{\boldsymbol{\theta},r}(t)$ with respect to θ_k gives

$$\frac{\partial \mathbf{y}_{\boldsymbol{\theta},r}(t)}{\partial \theta_k} = \partial_k \mathbf{C}_{\boldsymbol{\theta},r} \partial_k \mathbf{x}_{\boldsymbol{\theta}}(t), \quad t \geq t^{[0]}. \tag{30}$$

For $r = 1, \ldots, R$, since

$$\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t) = \begin{bmatrix} \frac{\partial \mathbf{y}_{\boldsymbol{\theta},r}(t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \mathbf{y}_{\boldsymbol{\theta},r}(t)}{\partial \theta_K} \end{bmatrix}, \quad \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t) = \begin{bmatrix} \partial_1 \mathbf{x}_{\boldsymbol{\theta}}(t) \\ \vdots \\ \partial_K \mathbf{x}_{\boldsymbol{\theta}}(t) \end{bmatrix}, \quad t \geq t^{[0]}$$

stacking the corresponding equations from (28) and (30) gives

$$\dot{\tilde{\mathbf{x}}}_{\boldsymbol{\theta}}(t) = \tilde{\mathbf{A}}_{\boldsymbol{\theta}} \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t) + \sum_{m=1}^{M} \tilde{\mathbf{F}}_{\boldsymbol{\theta},m} u_m(t) \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t) + \tilde{\mathbf{B}}_{\boldsymbol{\theta}} \mathbf{u}(t) \tag{31}$$

$$\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t) = \tilde{\mathbf{C}}_{\boldsymbol{\theta},r} \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t), \quad t \geq t^{[0]}. \tag{32}$$

The desired derivative system $\tilde{\Phi}$ is then obtained by stacking the corresponding equations from (32) as

$$\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t) = \tilde{\mathbf{C}}_{\boldsymbol{\theta}} \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t), \quad t \geq t^{[0]}. \tag{33}$$

The initial condition of $\tilde{\Phi}$ is given by stacking the corresponding equations from (29) as $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[0]}) = \tilde{\mathbf{x}}_{\boldsymbol{\theta},0}$. Clearly, $\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t)$, $t \geq t^{[0]}$, is the output of the derivative system $\tilde{\Phi} := \{\tilde{\mathbf{A}}_{\boldsymbol{\theta}}, \tilde{\mathbf{B}}_{\boldsymbol{\theta}}, \tilde{\mathbf{C}}_{\boldsymbol{\theta}}, \tilde{\mathbf{F}}_{\boldsymbol{\theta},1}, \ldots, \tilde{\mathbf{F}}_{\boldsymbol{\theta},m}\}$. Note that each element of $\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t)$ is a continuous function of t for $t \geq t^{[0]}$.

$$\begin{aligned} \mathrm{var}(\hat{k}_a) \geq& \left[\lim_{J' \to \infty} \mathbf{I}_{\boldsymbol{\theta}}^{-1}(J')\right]_{11} \\ =& \frac{\sigma^2 \left(-a_0^6 a_1^4 - a_0^4 a_1^6 + 6a_0^4 a_1^4 - a_0^6 a_1^2 - a_0^2 a_1^6 - 6a_0^2 a_1^2 + a_0^4 + a_1^4 + a_0^2 + a_1^2\right) (C_0 k_a + k_d)^2}{T^2 a_0^2 \left(a_0^2 + 1\right) a_1^2 \left(a_1^2 + 1\right) C_0^4 k_a^2 R_{\max}^2}. \\ \mathrm{var}(\hat{k}_d) \geq& \left[\lim_{J' \to \infty} \mathbf{I}_{\boldsymbol{\theta}}^{-1}(J')\right]_{22} = \frac{\sigma^2 (C_0 k_a + k_d)^2 (1 - a_1)^3 (a_1 + 1)^3}{T^2 a_1^2 \left(a_1^2 + 1\right) C_0^2 k_a^2 R_{\max}^2}. \\ \mathrm{var}(\hat{R}_{\max}) \geq& \left[\lim_{J' \to \infty} \mathbf{I}_{\boldsymbol{\theta}}^{-1}(J')\right]_{33} \\ =& \frac{\sigma^2}{T^2 a_0^2 \left(a_0^2 + 1\right) a_1^2 \left(a_1^2 + 1\right) C_0^4 k_a^4} \left[\left(-a_0^4 a_1^6 + 3a_0^4 a_1^4 - a_0^2 a_1^6 - 3a_0^4 a_1^2 + 3a_0^2 a_1^4 - 3a_0^2 a_1^2 + a_0^4 \right.\right. \\ & \left. + a_0^2\right) C_0^2 k_a^2 + \left(-2a_0^4 a_1^6 + 6a_0^4 a_1^4 - 2a_0^2 a_1^6 - 6a_0^4 a_1^2 + 6a_0^2 a_1^4 - 6a_0^2 a_1^2 + 2a_0^4 + 2a_0^2\right) C_0 k_a k_d \\ & \left. + \left(-a_0^6 a_1^4 - a_0^4 a_1^6 + 6a_0^4 a_1^4 - a_0^6 a_1^2 - a_0^2 a_1^6 - 6a_0^2 a_1^2 + a_0^4 + a_0^2 + a_1^4 + a_1^2\right) k_d^2\right]. \end{aligned}$$

2) From a classic result on the Fisher information matrix (see, e.g., [5] and [17])

$$\mathbf{I}_{\boldsymbol{\theta}} = E\left\{\frac{\partial \ln p_{\boldsymbol{\theta}}(\mathbf{s})}{\partial \boldsymbol{\theta}}\left(\frac{\partial \ln p_{\boldsymbol{\theta}}(\mathbf{s})}{\partial \boldsymbol{\theta}}\right)^T\right\} = \sum_{j=0}^{J-1}\left(\frac{\partial \mathbf{y}_{\boldsymbol{\theta}}(t_j)}{\partial \boldsymbol{\theta}^T}\right)^T \boldsymbol{\Lambda}_{\mathbf{n}}^{-1}(t_j)\left(\frac{\partial \mathbf{y}_{\boldsymbol{\theta}}(t_j)}{\partial \boldsymbol{\theta}^T}\right)$$

where $(\partial \mathbf{y}_{\boldsymbol{\theta}}(t_j)/\partial \boldsymbol{\theta}^T) := [(\partial \mathbf{y}_{\boldsymbol{\theta}}(t_j)/\partial \theta_1) \ \dots \ (\partial \mathbf{y}_{\boldsymbol{\theta}}(t_j)/\partial \theta_K)]$. It then follows directly that the Fisher information matrix is given by

$$\mathbf{I}_{\boldsymbol{\theta}} = \sum_{j=0}^{J-1} \tilde{\mathbf{Y}}_{\boldsymbol{\theta}}(t_j)\boldsymbol{\Lambda}_{\mathbf{n}}^{-1}(t_j)\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}^T(t_j),$$

where $\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}(t_j) := [\tilde{\mathbf{y}}_{\boldsymbol{\theta},1}(t_j) \ \dots \ \tilde{\mathbf{y}}_{\boldsymbol{\theta},R}(t_j)], j = 0, \dots, J-1$. □

B. Proof of Lemma 2.1

Let $\mathbf{x}'_{\boldsymbol{\theta}}(t) := \mathbf{x}_{\boldsymbol{\theta}}(t) - \mathbf{x}_{\boldsymbol{\theta},0}$ and $\mathbf{y}'_{\boldsymbol{\theta}}(t) := \mathbf{y}_{\boldsymbol{\theta}}(t) - \mathbf{C}_{\boldsymbol{\theta}}\mathbf{x}_{\boldsymbol{\theta},0}, t \geq t^{[0]}$. By substitution, (1) and (2) become

$$\dot{\mathbf{x}}'_{\boldsymbol{\theta}}(t) = \mathbf{A}_{\boldsymbol{\theta}}\mathbf{x}'_{\boldsymbol{\theta}}(t) + \sum_{m=1}^{M} \mathbf{F}_{\boldsymbol{\theta},m}u_m(t)\mathbf{x}'_{\boldsymbol{\theta}}(t) + (\mathbf{B}_{\boldsymbol{\theta}} + \mathbf{B}'_{\boldsymbol{\theta}})\mathbf{u}(t) + \mathbf{A}_{\boldsymbol{\theta}}\mathbf{x}_{\boldsymbol{\theta},0} \quad (34)$$

$$\mathbf{y}'_{\boldsymbol{\theta}}(t) = \mathbf{C}_{\boldsymbol{\theta}}\mathbf{x}'_{\boldsymbol{\theta}}(t), \quad t \geq t^{[0]} \quad (35)$$

where $\mathbf{x}'_{\boldsymbol{\theta}}(t^{[0]}) = 0$, $\mathbf{B}'_{\boldsymbol{\theta}} := [\mathbf{F}_{\boldsymbol{\theta},1}\mathbf{x}_{\boldsymbol{\theta},0} \ \dots \ \mathbf{F}_{\boldsymbol{\theta},M}\mathbf{x}_{\boldsymbol{\theta},0}]$. Let $\Omega_{\mathbf{x}'_{\boldsymbol{\theta}}}$ denote the span of reachable states and $\Omega_{\mathbf{y}'_{\boldsymbol{\theta}}}$ the span of reachable outputs of the new system (34)–(35) from $\mathbf{x}'_{\boldsymbol{\theta}}(t^{[0]}) = 0$ via admissible inputs $\mathbb{V}$.

The next part of the proof is similar to that for Lemmas 4.1 in [21] with some straightforward generalizations from SIMO bilinear systems to their MIMO counterparts. We only sketch the major steps here. We can first prove that $\Omega_{\mathbf{x}'_{\boldsymbol{\theta}}}$ is included in the column space of $\mathbf{O}^{(\infty)} := [\mathbf{O}_0 \ \dots \ \mathbf{O}_{N-1} \ \mathbf{O}_N \ \dots]$, where $\mathbf{O}_0, \dots, \mathbf{O}_{N-1}$ are defined in (10), $\mathbf{O}_N$ is defined as $\mathbf{O}_N := [\mathbf{A}_{\boldsymbol{\theta}}\mathbf{O}_{N-1} \ \mathbf{F}_{\boldsymbol{\theta},1}\mathbf{O}_{N-1} \ \dots \ \mathbf{F}_{\boldsymbol{\theta},m}\mathbf{O}_{N-1}]$, and similarly for $\mathbf{O}_{N+1}, \mathbf{O}_{N+2}$ and so on. Using the input-output Volterra series expansion in [21] and expressing $e^{\mathbf{A}_{\boldsymbol{\theta}}t}$ in terms of the power series of $\mathbf{A}_{\boldsymbol{\theta}}$ by Cayley–Hamilton theorem, we obtain that $\Omega_{\mathbf{x}'_{\boldsymbol{\theta}}} \subseteq \text{range}\{\mathbf{O}^{(\infty)}\}$. Next, we can show that $\Omega_{\mathbf{x}'_{\boldsymbol{\theta}}} \supseteq \text{range}\{\mathbf{O}^{(\infty)}\}$ using the same argument in the Proof of Lemmas 4.1 in [21]. Therefore, $\Omega_{\mathbf{x}'_{\boldsymbol{\theta}}} = \text{range}\{\mathbf{O}^{(\infty)}\}$. Combining this result with the known fact that $\text{range}\{\mathbf{O}^{(N-1)}\} = \text{range}\{\mathbf{O}^{(\infty)}\}$ (see [21, Lemma 4.2]), it follows immediately that $\Omega_{\mathbf{x}'_{\boldsymbol{\theta}}} = \text{range}\{\mathbf{O}^{(N-1)}\}$, and furthermore, there exists an integer p with $0 \leq p \leq N-1$ such that

$$\text{range}\{\mathbf{O}^{(0)}\} \subset \text{range}\{\mathbf{O}^{(1)}\} \subset \dots \subset \text{range}\left\{\mathbf{O}^{(p)}\right\} = \text{range}\left\{\mathbf{O}^{(p+1)}\right\} = \dots = \text{range}\left\{\mathbf{O}^{(N-1)}\right\}$$

and

$$\Omega_{\mathbf{x}'_{\boldsymbol{\theta}}} = \text{range}\{\mathbf{O}^{(p)}\} \quad \text{and} \quad \Omega_{\mathbf{y}'_{\boldsymbol{\theta}}} = \text{range}\{\mathbf{C}_{\boldsymbol{\theta}}\mathbf{O}^{(p)}\}.$$

Since $\mathbf{x}_{\boldsymbol{\theta}}(t) = \mathbf{x}'_{\boldsymbol{\theta}}(t) + \mathbf{x}_{\boldsymbol{\theta},0}$ and $\mathbf{y}_{\boldsymbol{\theta}}(t) = \mathbf{y}'_{\boldsymbol{\theta}}(t) + \mathbf{C}_{\boldsymbol{\theta}}\mathbf{x}_{\boldsymbol{\theta},0}, t \geq t^{[0]}$, it is obvious that

$$\Omega_{\mathbf{x}_{\boldsymbol{\theta}}} = \text{range}\left\{\left[\mathbf{x}_{\boldsymbol{\theta},0} \quad \mathbf{O}^{(p)}\right]\right\} \quad \text{and}$$
$$\Omega_{\mathbf{y}_{\boldsymbol{\theta}}} = \text{range}\left\{\left[\mathbf{C}_{\boldsymbol{\theta}}\mathbf{x}_{\boldsymbol{\theta},0} \quad \mathbf{C}_{\boldsymbol{\theta}}\mathbf{O}^{(p)}\right]\right\}. \quad \square$$

C. Proof of Theorem 2.3

Consider an arbitrary set of admissible inputs denoted by Γ. For each input $\mathbf{u} \in \Gamma$, let $\Upsilon_{\mathbf{u}}$ denote the set of output sampling points, $\tilde{\mathbf{x}}_{\boldsymbol{\theta},\mathbf{u}}(t), t \geq t_0$, and $\tilde{\mathbf{y}}_{\boldsymbol{\theta},\mathbf{u}}(t), t \geq t_0$, the state and output vectors of the derivative system $\tilde{\Phi}$, respectively. For each (vector) sample $\mathbf{v} \in \Upsilon_{\mathbf{u}}$, let $t_{\mathbf{v},\mathbf{u}}$ denote the corresponding sampling instant of $\mathbf{v}$, and $\boldsymbol{\Lambda}_{\mathbf{n}}(t_{\mathbf{v},\mathbf{u}})$ the noise covariance matrix at $t_{\mathbf{v},\mathbf{u}}$. The Fisher information matrix $\mathbf{I}_{\boldsymbol{\theta}}$ is then given in (11), which is repeated here for convenience.

$$\mathbf{I}_{\boldsymbol{\theta}} = \sum_{\mathbf{u}\in\Gamma}\sum_{\mathbf{v}\in\Upsilon_{\mathbf{u}}} \tilde{\mathbf{Y}}_{\boldsymbol{\theta}}(t_{\mathbf{v},\mathbf{u}})\boldsymbol{\Lambda}_{\mathbf{n}}^{-1}(t_{\mathbf{v},\mathbf{u}})\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}^T(t_{\mathbf{v},\mathbf{u}})$$

where $\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}(t_{\mathbf{v},\mathbf{u}}) := [\tilde{\mathbf{y}}_{\boldsymbol{\theta},\mathbf{u},1}(t_{\mathbf{v},\mathbf{u}}) \ \dots \ \tilde{\mathbf{y}}_{\boldsymbol{\theta},\mathbf{u},R}(t_{\mathbf{v},\mathbf{u}})]$, for $j = 0, \dots, J-1$. By Theorem 2.2, to show local identifiability it is equivalent to showing that $\mathbf{I}_{\boldsymbol{\theta}}$ is nonsingular. Since the size of $\mathbf{I}_{\boldsymbol{\theta}}$ is $K \times K$ and all the covariance matrices $\boldsymbol{\Lambda}_{\mathbf{n}}(t_{\mathbf{v},\mathbf{u}})$ are positive definite, nonsingularity of $\mathbf{I}_{\boldsymbol{\theta}}$ is equivalent to that the span of all the column vectors of $\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}(t_{\mathbf{v},\mathbf{u}}), \mathbf{v} \in \Upsilon_{\mathbf{u}}$ for all $\mathbf{u} \in \Gamma$, is of dimension K, i.e., the span of all the vectors $\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t_{\mathbf{v},\mathbf{u}}), r = 1, \dots, R$, has a dimension of K.

Let $\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta}}}$ denote the span of reachable outputs of $\tilde{\Phi}$, and $\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}}$ the span of the vectors $\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t)$ for all the reachable outputs $\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t)$ of $\tilde{\Phi}, r = 1, \dots, R, t \geq t^{[0]}$. From bilinear system theory [21], the dimension of the span of any set of outputs $\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t)$ of $\tilde{\Phi}$ is no bigger than that of $\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta}}}$, and there exists a finite set of admissible inputs such that the span of the corresponding outputs is of the same dimension as that of the span of all the reachable outputs. Similarly, the dimension of the span of any set of the vectors $\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t), t \geq t^{[0]}$, is no bigger than that of $\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}}$, and there exists a finite set of admissible inputs such that the span of the corresponding $\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t), t \geq t^{[0]}$, is of the same dimension as that of $\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}}, r = 1, \dots, R$. It is then clear that the existence of a nonsingular $\mathbf{I}_{\boldsymbol{\theta}}$ is equivalent to that the span of the vectors $\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t), r = 1, \dots, R, t \geq t^{[0]}$, for all the reachable outputs $\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t), t \geq t^{[0]}$, is of dimension K.

By Lemma 2.1, $\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta}}} = \text{range}\{[\tilde{\mathbf{C}}_{\boldsymbol{\theta}}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \ \ \tilde{\mathbf{C}}_{\boldsymbol{\theta}}\tilde{\mathbf{O}}^{(\tilde{p})}]\}$, for some $\tilde{p}$ with $0 \leq \tilde{p} \leq \tilde{N} - 1$. Hence, $\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}} = \text{range}\{[\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \ \tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{O}}^{(\tilde{p})}]\}, r = 1, \dots, R$. Then the span of the vectors $\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t), r = 1, \dots, R, t \geq t^{[0]}$, for all the reachable outputs $\tilde{\mathbf{y}}_{\boldsymbol{\theta}}(t), t \geq t^{[0]}$, is given by

$$\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},1}} \oplus \Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},2}} \oplus \dots \oplus \Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},R}} = \Big[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}^{(\tilde{p})} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},2}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},2}\tilde{\mathbf{O}}^{(\tilde{p})} \dots \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},R}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},R}\tilde{\mathbf{O}}^{(\tilde{p})}\Big].$$

Therefore, there exist a finite number of admissible inputs and output sampling points such that the associated Fisher information matrix such that $\mathbf{I}_{\boldsymbol{\theta}}$ is nonsingular if and only if $\{\Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},1}} \oplus \Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},2}} \oplus \dots \oplus \Omega_{\tilde{\mathbf{y}}_{\boldsymbol{\theta},R}}\}$ is of dimension K, i.e.,

$$\text{rank}\left\{\Big[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}^{(\tilde{p})} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},2}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},2}\tilde{\mathbf{O}}^{(\tilde{p})} \quad \dots \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},R}\tilde{\mathbf{x}}_{\boldsymbol{\theta},0} \quad \tilde{\mathbf{C}}_{\boldsymbol{\theta},R}\tilde{\mathbf{O}}^{(\tilde{p})}\Big]\right\} = K$$

for some $\tilde{p}$ with $0 \leq \tilde{p} \leq \tilde{N} - 1$. □

D. Proof of Theorem 3.1

i) From the definition of $\tilde{\mathbf{Q}}_{\boldsymbol{\theta}}^{[l]}(t)$ in Lemma 3.1, we have $\tilde{\mathbf{Q}}_{\boldsymbol{\theta}}^{[l]}(t^{[l],[j]}) = e^{(\tilde{\mathbf{A}}_{\boldsymbol{\theta}}+\tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]})(t^{[l],[j]}-t^{[l]})} = e^{(\tilde{\mathbf{A}}_{\boldsymbol{\theta}}+\tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]})(jT^{[l]})} = (\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]})^j$. Substituting $\tilde{\mathbf{Q}}_{\boldsymbol{\theta}}^{[l]}(t^{[l],[j]}) = (\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]})^j$ into $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l],[j]})$ in (16) gives

$$\tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l],[j]}\right) = \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^j\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right) - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]},$$
$$l = 0,\ldots,L-1; j = 0,\ldots,J^{[l]}-1. \quad (36)$$

By Theorem 2.1 and recalling (36), we have

$$\begin{aligned}\mathbf{I}_{\boldsymbol{\theta}} &= \sum_{l=0}^{L-1}\sum_{j=0}^{J^{[l]}-1}\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}\left(t^{[l],[j]}\right)\boldsymbol{\Lambda}_{\mathbf{n}}^{-1}\left(t^{[l],[j]}\right)\tilde{\mathbf{Y}}_{\boldsymbol{\theta}}^T\left(t^{[l],[j]}\right)\\
&= \frac{1}{\sigma^2}\sum_{r=1}^{R}\sum_{l=0}^{L-1}\sum_{j=0}^{J^{[l]}-1}\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}\left(t^{[l],[j]}\right)\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}^T\left(t^{[l],[j]}\right)\\
&= \frac{1}{\sigma^2}\sum_{r=1}^{R}\sum_{l=0}^{L-1}\sum_{j=0}^{J^{[l]}-1}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l],[j]}\right)\tilde{\mathbf{x}}_{\boldsymbol{\theta}}^T\left(t^{[l],[j]}\right)\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T\\
&= \frac{1}{\sigma^2}\sum_{r=1}^{R}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\left\{\sum_{l=0}^{L-1}\left[\sum_{j=0}^{J^{[l]}-1}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^j\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)\right.\right.\\
&\quad\times\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)^T\left(\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^j\right)^T\\
&\quad-\sum_{j=0}^{J^{[l]}-1}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^j\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right)\\
&\quad-\sum_{j=0}^{J^{[l]}-1}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T + \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\tilde{\mathbf{x}}_{\boldsymbol{\theta}}^T\left(t^{[l]}\right)\right)\left(\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^j\right)^T\\
&\quad\left.\left.+\sum_{j=0}^{J^{[l]}-1}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right]\right\}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T. \end{aligned} \quad (37)$$

ii) By Theorem 2.2, to show local identifiability it suffices to show that $\mathbf{I}_{\boldsymbol{\theta}}$ is nonsingular. From Remark 3.1, all the eigenvalues of $\tilde{\mathbf{A}}_{\boldsymbol{\theta}} + \tilde{\mathbf{F}}_{\boldsymbol{\theta}}^{[l]}$ are also in the open left-half plane. Hence, all the eigenvalues of $\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}$ are in the open unit disc, and none of them is equal to one. From the proof in part i), the Fisher information matrix $\mathbf{I}_{\boldsymbol{\theta}}$ is given by

$$\begin{aligned}\mathbf{I}_{\boldsymbol{\theta}} &= \frac{1}{\sigma^2}\sum_{r=1}^{R}\sum_{l=0}^{L-1}\sum_{j=0}^{J^{[l]}-1}\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}\left(t^{[l],[j]}\right)\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}^T\left(t^{[l],[j]}\right)\\
&= \frac{1}{\sigma^2}\sum_{r=1}^{R}\sum_{l=0}^{L-1}\sum_{j=0}^{J^{[l]}-1}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l],[j]}\right)\tilde{\mathbf{x}}_{\boldsymbol{\theta}}^T\left(t^{[l],[j]}\right)\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T.\end{aligned}$$

Let $\tilde{\mathbf{O}}'' := [\tilde{\mathbf{O}}_0'' \ \ldots \ \tilde{\mathbf{O}}_{L-1}'']$, where $\tilde{\mathbf{O}}_l'' := [\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l],[0]}) \ \ldots \ \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l],[J^{[l]}-1]})], l = 0,\ldots,L-1$, with $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l],[j]}), j = 0,\ldots,J^{[l]}-1$, being given in (36). Thus, $\mathbf{I}_{\boldsymbol{\theta}}$ being nonsingular is equivalent to the space spanned by all the vectors $\tilde{\mathbf{y}}_{\boldsymbol{\theta},r}(t^{[l],[j]}), r = 1,\ldots,R, j = 0,\ldots,J^{[l]}-1$ for $l = 0,\ldots,L-1$ is of full rank, or, $\text{rank}\{[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}'' \ \ldots \ \tilde{\mathbf{C}}_{\boldsymbol{\theta},R}\tilde{\mathbf{O}}'']\} = K$. It remains to show that $\text{rank}\{[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}'' \ \ldots \ \tilde{\mathbf{C}}_{\boldsymbol{\theta},R}\tilde{\mathbf{O}}'']\} = K$ is equivalent to $\text{rank}\{[\tilde{\mathbf{C}}_{\boldsymbol{\theta},1}\tilde{\mathbf{O}}' \ \ldots \ \tilde{\mathbf{C}}_{\boldsymbol{\theta},R}\tilde{\mathbf{O}}']\} = K$, for which it is sufficient to show $\text{range}\{\tilde{\mathbf{O}}''\} = \text{range}\{\tilde{\mathbf{O}}'\}$.

Consider $\tilde{\mathbf{O}}_0''$ and $\tilde{\mathbf{O}}_0'$. $\tilde{\mathbf{O}}_0'$ is of size $(\tilde{N}+1)\times(\tilde{N}+1)$ while $\tilde{\mathbf{O}}_0''$ is of size $(\tilde{N}+1)\times J^{[0]}$, with $J^{[0]} \geq \tilde{N}+1$. We first show that all the columns of $\tilde{\mathbf{O}}_0''$ can be expressed by linear combinations of the columns of $\tilde{\mathbf{O}}_0'$. Expand $\det(\lambda\mathbf{I}-\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]})$ as $\det(\lambda\mathbf{I}-\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}) = \lambda^{\tilde{N}} + \alpha_1\lambda^{\tilde{N}-1} + \ldots + \alpha_{\tilde{N}-1}\lambda + \alpha_{\tilde{N}}$. By Cayley–Hamilton theorem, $(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]})^{\tilde{N}} + \alpha_1(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]})^{\tilde{N}-1} + \ldots + \alpha_{\tilde{N}-1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]} + \alpha_{\tilde{N}}\mathbf{I} = 0$. We then have

$$\begin{aligned}\tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[0],[\tilde{N}]}\right) &= \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}\right)^{\tilde{N}}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[0]}\right)\right) - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]}\\
&= -\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]} - \alpha_{\tilde{N}}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[0]}\right)\right)\\
&\quad - \alpha_{\tilde{N}-1}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[0]}\right)\right) - \ldots\\
&\quad - \alpha_1\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}\right)^{\tilde{N}-1}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[0]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[0]}\right)\right).\end{aligned}$$

Hence, $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[0],[\tilde{N}]})$ is a linear combination of the $(\tilde{N}+1)$ columns of $\tilde{\mathbf{O}}_0'$ with coefficients of $-1, -\alpha_{\tilde{N}}, -\alpha_{\tilde{N}-1}, \ldots, -\alpha_2$, and $-\alpha_1$. Using Cayley–Hamilton theorem repeatedly, it can be similarly argued that $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[0],[j]}), j = \tilde{N}+1,\ldots,J^{[0]}-1$, can be expressed as linear combinations of the columns of $\tilde{\mathbf{O}}_0'$. Hence, $\tilde{\mathbf{O}}_0''$ and $\tilde{\mathbf{O}}_0'$ are related by

$$\tilde{\mathbf{O}}_0'' = \tilde{\mathbf{O}}_0'\left[\begin{array}{c|c}\underbrace{\begin{matrix}-1 & -1 & \ldots & -1 & -1 & -1\\ 1 & 0 & \ldots & 0 & 0 & -\alpha_{\tilde{N}}\\ 0 & 1 & \ldots & 0 & 0 & -\alpha_{\tilde{N}-1}\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & 0 & \ldots & 1 & 0 & -\alpha_2\\ 0 & 0 & \ldots & 0 & 1 & -\alpha_1\end{matrix}}_{\mathbf{H}} & \text{X}\end{array}\right]$$

where X is an $(\tilde{N}+1)\times(J^{[0]}-\tilde{N}-1)$ constant matrix whose value is of no interest here. Since none of the eigenvalues of $\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]}$ are equal to one, we then have

$$\begin{aligned}\det(\mathbf{H}) &= (-1)^{\tilde{N}+1}(1+\alpha_1+\ldots+\alpha_{\tilde{N}})\\
&= (-1)^{\tilde{N}+1}\det(\lambda\mathbf{I}-\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[0]})\Big|_{\lambda=1} \neq 0.\end{aligned}$$

Thus, $\mathbf{H}$ is of full rank, which implies that all the columns of $\tilde{\mathbf{O}}_0'$ can also be expressed as linear combinations of the columns of $\tilde{\mathbf{O}}_0''$. Therefore, $\text{range}\{\tilde{\mathbf{O}}_0''\} = \text{range}\{\tilde{\mathbf{O}}_0'\}$. It can be similarly shown that $\text{range}\{\tilde{\mathbf{O}}_l''\} = \text{range}\{\tilde{\mathbf{O}}_l'\}, l = 1,\ldots,L-1$. Therefore, $\text{range}\{\tilde{\mathbf{O}}''\} = \text{range}\{\tilde{\mathbf{O}}'\}$.

iii) From the proof in part ii), it is clear that the nonsingularity condition (20) holds for any $J^{[l]}, l = 0,\ldots,L-1$, with $J^{[l]} \geq \tilde{N}+1$, and hence holds for $J^{[l]} \to \infty$. To show expression (20) reduces to (21) for $J^{[l]} \to \infty, l = 0,\ldots,L-1$, it is equivalent to show that $\lim_{J^{[l-1]}\to\infty}\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l]}) = -\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}$. Since

$\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[0]}) = -\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[-1]}$ from (18), we only need to show that $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l]}) = -\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}$ for $l = 1, \ldots, L-1$. From the proof in part i), $\tilde{\mathbf{Q}}^{[l-1]}(t^{[l-1],[J^{[l-1]}]}) = (\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l-1]})^{J^{[l-1]}}$. Since all the eigenvalues of $\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l-1]}$ are in the open unit disc, $\lim_{J^{[l-1]}\to\infty} \tilde{\mathbf{Q}}^{[l-1]}(t^{[l-1],[J^{[l-1]}]}) = \lim_{J^{[l-1]}\to\infty}(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l-1]})^{J^{[l-1]}} = 0$, for $l = 1, \ldots, L$. Therefore

$$\begin{aligned}\lim_{J^{[l-1]}\to\infty} &\tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\\ &= \lim_{J^{[l-1]}\to\infty} \tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l-1],[J^{[l-1]}]})\\ &= \lim_{J^{[l-1]}\to\infty} \tilde{\mathbf{Q}}^{[l-1]}(t^{[l-1],[J^{[l-1]}]})\\ &\quad\times\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l-1]}\right)\right) - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\\ &= -\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}, \quad l = 1, \ldots, L-1. \qquad \square\end{aligned}$$

E. Proof of Theorem 3.2

For $l = 0, \ldots, L-1$, let

$$\begin{aligned}\tilde{\mathbf{P}}_1^{[l]} :=& \sum_{j=0}^{J^{[l]}-1}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^j\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)\\ &\times\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)^T\left(\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^j\right)^T\\ \tilde{\mathbf{P}}_2^{[l]} :=& \sum_{j=0}^{J^{[l]}-1}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^j\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right).\end{aligned}$$

Substituting $\tilde{\mathbf{P}}_1^{[l]}$ and $\tilde{\mathbf{P}}_2^{[l]}$ into (19), we have

$$\begin{aligned}\mathbf{I}_{\boldsymbol{\theta}} =& \frac{1}{\sigma^2}\sum_{r=1}^{R}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\left\{\sum_{l=0}^{L-1}\left[\tilde{\mathbf{P}}_1^{[l]} - \tilde{\mathbf{P}}_2^{[l]} - \left(\tilde{\mathbf{P}}_2^{[l]}\right)^T\right.\right.\\ &\left.\left. + J^{[l]}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right]\right\}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T.\end{aligned}$$

Since all the eigenvalues of $\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}, l = 0, \ldots, L-1$, are in the open unit disc, $\tilde{\mathbf{P}}_1^{[l]}, l = 0, \ldots, L-1$, is the unique solution to the following Lyapunov equation:

$$\begin{aligned}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}&\tilde{\mathbf{P}}_1^{[l]}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^T - \tilde{\mathbf{P}}_1^{[l]}\\ &= -\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)^T\\ &\quad + \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J^{[l]}}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)\\ &\quad\times\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)^T\left(\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J^{[l]}}\right)^T.\end{aligned}$$

As $\sum_{j=0}^{J^{[l]}-1}(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]})^j = (\mathbf{I} - (\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]})^{J^{[l]}})(\mathbf{I} - \tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]})^{-1}, \tilde{\mathbf{P}}_2^{[l]}, \quad l = 0, \ldots, L-1$ is given by

$$\begin{aligned}\tilde{\mathbf{P}}_2^{[l]} =& \left(\mathbf{I} - \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J^{[l]}}\right)\left(\mathbf{I} - \tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{-1}\\ &\times\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right). \qquad \square\end{aligned}$$

F. Proof of Theorem 3.3

With $J^{[l]} = J'$ for $l = 0, \ldots, L-1$, the Fisher information matrix in Theorem 3.2 can be rewritten in terms of J' as

$$\begin{aligned}\mathbf{I}_{\boldsymbol{\theta}}(J') =& \frac{1}{\sigma^2}\sum_{r=1}^{R}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\left\{\sum_{l=0}^{L-1}\left[\tilde{\mathbf{P}}_{1,J'}^{[l]} - \tilde{\mathbf{P}}_{2,J'}^{[l]}\right.\right.\\ &\left.\left. - \left(\tilde{\mathbf{P}}_{2,J'}^{[l]}\right)^T + J'^{[l]}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right]\right\}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T\end{aligned}$$

where $\tilde{\mathbf{P}}_{1,J'}^{[l]}, l = 0, \ldots, L-1$, is the unique solution to the following Lyapunov equation:

$$\begin{aligned}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}&\tilde{\mathbf{P}}_{1,J'}^{[l]}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^T - \tilde{\mathbf{P}}_{1,J'}^{[l]}\\ &= -\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)^T\\ &\quad + \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J'}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)\\ &\quad\times\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\right)^T\left(\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J'}\right)^T \qquad (38)\end{aligned}$$

and $\tilde{\mathbf{P}}_{2,J'}^{[l]}, l = 0, \ldots, L-1$, is given by

$$\begin{aligned}\tilde{\mathbf{P}}_{2,J'}^{[l]} =& \left(\mathbf{I} - \left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{J'}\right)\left(\mathbf{I} - \tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{-1}\\ &\times\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T + \tilde{\mathbf{x}}_{\boldsymbol{\theta}}\left(t^{[l]}\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right). \qquad (39)\end{aligned}$$

Let

$$\tilde{\mathbf{P}}_{J'} := \sum_{r=1}^{R}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\left\{\sum_{l=0}^{L-1}\left[\tilde{\mathbf{P}}_{1,J'}^{[l]} - \tilde{\mathbf{P}}_{2,J'}^{[l]} - \left(\tilde{\mathbf{P}}_{2,J'}^{[l]}\right)^T\right]\right\}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T. \qquad (40)$$

Then

$$\mathbf{I}_{\boldsymbol{\theta}}(J') = \frac{1}{\sigma^2}\left(\tilde{\mathbf{P}}_{J'} + J'\sum_{r=1}^{R}\sum_{l=0}^{L-1}\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T\right).$$

Since the nonsingularity condition in part iii) of Theorem 3.1 holds for the asymptotic case, there exists a sufficiently large integer $J_0 \geq \tilde{N} + 1$ such that $\mathbf{I}_{\boldsymbol{\theta}}(J')$ is nonsingular for all $J' > J_0$, which in turn implies that $\mathbf{I}_{\boldsymbol{\theta}}(J')$ is positive definite as the Fisher information matrix is always positive semidefinite. In the remainder of the proof, we assume $J' > J_0$. From the proof in part iii) of Theorem 3.1, we have $\lim_{J'\to\infty}\tilde{\mathbf{x}}_{\boldsymbol{\theta}}(t^{[l]}) = -\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}$ and $\lim_{J'\to\infty}(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]})^{J'} = 0$, for $l = 0, \ldots, L-1$. Hence, taking the limit of $J' \to \infty$ on both sides of (38) gives

$$\begin{aligned}\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}&\tilde{\mathbf{P}}_1^{[l]}\left(\tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^T - \tilde{\mathbf{P}}_1^{[l]}\\ &= -\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\right)\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]} - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\right)^T\end{aligned}$$

where $\tilde{\mathbf{P}}_1^{[l]} := \lim_{J'\to\infty}\tilde{\mathbf{P}}_{1,J'}^{[l]}, l = 0, \ldots, L-1$. Similarly, taking the limit of $J' \to \infty$ on both sides of (39) gives

$$\begin{aligned}\tilde{\mathbf{P}}_2^{[l]} = \left(\mathbf{I} - \tilde{\mathbf{A}}_{\boldsymbol{\theta},d}^{[l]}\right)^{-1}&\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T - \tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l-1]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\right),\\ &l = 0, \ldots, L-1\end{aligned}$$

where $\tilde{\mathbf{P}}_2^{[l]} := \lim_{J'\to\infty} \tilde{\mathbf{P}}_{2,J'}^{[l]}$. Now taking the limit of $J' \to \infty$ on both sides of (40) gives

$$\tilde{\mathbf{P}} := \lim_{J'\to\infty} \tilde{\mathbf{P}}_{J'} = \sum_{r=1}^{R} \tilde{\mathbf{C}}_{\boldsymbol{\theta},r} \left\{ \sum_{l=0}^{L-1} \left[\tilde{\mathbf{P}}_1^{[l]} - \tilde{\mathbf{P}}_2^{[l]} - \left(\tilde{\mathbf{P}}_2^{[l]}\right)^T \right] \right\} \tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T.$$

For $J' \to \infty$, although $\mathbf{I}_{\boldsymbol{\theta}}(J')$ tends to infinity, the inverse of $\mathbf{I}_{\boldsymbol{\theta}}(J')$ still converges, as will be shown in the following. Using a singular value decomposition, $\sum_{r=1}^{R}\sum_{l=0}^{L-1} \tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]})^T\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T$ can be expressed as

$$\sum_{r=1}^{R}\sum_{l=0}^{L-1} \tilde{\mathbf{C}}_{\boldsymbol{\theta},r}\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\left(\tilde{\mathbf{w}}_{\boldsymbol{\theta}}^{[l]}\right)^T\tilde{\mathbf{C}}_{\boldsymbol{\theta},r}^T = [\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}] \begin{bmatrix} \boldsymbol{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} [\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}]^T \tag{41}$$

where $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{G\times G}$ is diagonal with positive diagonal entries, $\tilde{\mathbf{U}}_s \in \mathbb{R}^{K\times G}, \tilde{\mathbf{U}}_s^{\perp} \in \mathbb{R}^{K\times(K-G)}$, and $[\tilde{\mathbf{U}}_s \ \tilde{\mathbf{U}}_s^{\perp}]$ is orthogonal. By substitution

$$\begin{aligned}
&\mathbf{I}_{\boldsymbol{\theta}}^{-1}(J') \\
&= \sigma^2\left(\tilde{\mathbf{P}}_{J'} + J'[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}]\begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & 0 \\ 0 & 0 \end{bmatrix}[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}]^T\right)^{-1} \\
&= \sigma^2[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}] \\
&\quad\times\left([\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}]^T\tilde{\mathbf{P}}_{J'}[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}] + J'\begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & 0 \\ 0 & 0 \end{bmatrix}\right)^{-1} \\
&\quad\times[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}]^T \\
&= \sigma^2[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}]\begin{bmatrix} \tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s + J'\tilde{\boldsymbol{\Sigma}} & \tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp} \\ (\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s & (\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp} \end{bmatrix}^{-1} \\
&\quad\times[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}]^T.
\end{aligned}$$

Consider first the case when $G = K$, i.e., $\tilde{\mathbf{U}}_s$ has a rank of K and $\tilde{\mathbf{U}}_s^{\perp}$ diminishes in (41). In this case, the asymptotic CRLB is given by

$$\begin{aligned}
\mathrm{var}(\theta) &\geq \lim_{J'\to\infty}\mathbf{I}_{\boldsymbol{\theta}}^{-1}(J') \\
&= \sigma^2\tilde{\mathbf{U}}_s\left[\lim_{J'\to\infty}\left(\tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s + J'\tilde{\boldsymbol{\Sigma}}\right)^{-1}\right]\tilde{\mathbf{U}}_s^T \\
&= \sigma^2\tilde{\mathbf{U}}_s\left[\lim_{J'\to\infty}\frac{1}{J'}\right]\left[\lim_{J'\to\infty}\left(\frac{1}{J'}\tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s + \tilde{\boldsymbol{\Sigma}}\right)^{-1}\right]\tilde{\mathbf{U}}_s^T \\
&= 0
\end{aligned}$$

since $\lim_{J'\to\infty}((1/J')\tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s) = 0$ and $\tilde{\boldsymbol{\Sigma}}$ is of full rank.

Next, consider the case when $N < K$. Let $\tilde{\mathbf{Z}}_1 := \tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s + J'\tilde{\boldsymbol{\Sigma}}, \tilde{\mathbf{Z}}_2 := \tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp}, \tilde{\mathbf{Z}}_3 := (\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp}$, and $\tilde{\boldsymbol{\Delta}} := \tilde{\mathbf{Z}}_3 - \tilde{\mathbf{Z}}_2^T\tilde{\mathbf{Z}}_1^{-1}\tilde{\mathbf{Z}}_2$ ($\tilde{\boldsymbol{\Delta}}$ is called the Schur complement of $\tilde{\mathbf{Z}}_1$). That $\mathbf{I}_{\boldsymbol{\theta}}(J')$ is positive definite implies $\begin{bmatrix} \tilde{\mathbf{Z}}_1 & \tilde{\mathbf{Z}}_2 \\ \tilde{\mathbf{Z}}_2^T & \tilde{\mathbf{Z}}_3 \end{bmatrix}$ is also positive definite. Then $\tilde{\mathbf{Z}}_1$ and $\tilde{\boldsymbol{\Delta}}$ are positive definite (see [32. Th. 7.7.6]). Using the formula of the inverse of block matrices [25]

$$\begin{bmatrix} \tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s + J'\tilde{\boldsymbol{\Sigma}} & \tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp} \\ \left(\tilde{\mathbf{U}}_s^{\perp}\right)^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s & \left(\tilde{\mathbf{U}}_s^{\perp}\right)^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp} \end{bmatrix}^{-1} = \begin{bmatrix} \tilde{\mathbf{Z}}_1^{-1} + \tilde{\mathbf{Z}}_1^{-1}\tilde{\mathbf{Z}}_2\tilde{\boldsymbol{\Delta}}^{-1}\tilde{\mathbf{Z}}_2^T\tilde{\mathbf{Z}}_1^{-1} & -\tilde{\mathbf{Z}}_1^{-1}\tilde{\mathbf{Z}}_2\tilde{\boldsymbol{\Delta}}^{-1} \\ -\tilde{\boldsymbol{\Delta}}^{-1}\tilde{\mathbf{Z}}_2^T\tilde{\mathbf{Z}}_1^{-1} & \tilde{\boldsymbol{\Delta}}^{-1} \end{bmatrix}.$$

For $J' \to \infty, \lim_{J'\to\infty}\tilde{\mathbf{Z}}_1^{-1} = \lim_{J'\to\infty}(\tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s + J'\tilde{\boldsymbol{\Sigma}})^{-1} = \lim_{J'\to\infty}(\tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}_s + J'\tilde{\boldsymbol{\Sigma}})^{-1} = 0$. Since $\mathbf{I}_{\boldsymbol{\theta}}(J')$ is positive definite, for any nonzero vector $\mathbf{b} \in \mathbb{R}^{(K-G)\times 1}$

$$\mathbf{b}^T\left(\tilde{\mathbf{U}}_s^{\perp}\right)^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp}\mathbf{b} = \sigma^2\mathbf{b}^T\left(\tilde{\mathbf{U}}_s^{\perp}\right)^T\mathbf{I}_{\boldsymbol{\theta}}(J')\tilde{\mathbf{U}}_s^{\perp}\mathbf{b} > 0$$

which shows that $\lim_{J'\to\infty}\tilde{\boldsymbol{\Delta}} = \lim_{J'\to\infty}(\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp} = (\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}_s^{\perp}$ is positive definite. Therefore

$$\lim_{J'\to\infty}\begin{bmatrix} \tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s + J'\tilde{\boldsymbol{\Sigma}} & \tilde{\mathbf{U}}_s^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp} \\ (\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s & (\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}_{J'}\tilde{\mathbf{U}}_s^{\perp} \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & ((\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}_s^{\perp})^{-1} \end{bmatrix}.$$

The asymptotic CRLB is then given in terms of $\tilde{\mathbf{U}}_s$ as

$$\begin{aligned}
\mathrm{var}(\hat{\boldsymbol{\theta}}) &\geq \lim_{J'\to\infty}\mathbf{I}_{\boldsymbol{\theta}}^{-1}(J') \\
&= \sigma^2\left[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}\right]\begin{bmatrix} 0 & 0 \\ 0 & \left(\left(\tilde{\mathbf{U}}_s^{\perp}\right)^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}_s^{\perp}\right)^{-1} \end{bmatrix}[\tilde{\mathbf{U}}_s \quad \tilde{\mathbf{U}}_s^{\perp}]^T \\
&= \sigma^2\tilde{\mathbf{U}}_s^{\perp}\left(\left(\tilde{\mathbf{U}}_s^{\perp}\right)^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}_s^{\perp}\right)^{-1}\left(\tilde{\mathbf{U}}_s^{\perp}\right)^T.
\end{aligned}$$

Finally, we show that $\tilde{\mathbf{U}}_s^{\perp}((\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}_s^{\perp})^{-1}(\tilde{\mathbf{U}}_s^{\perp})^T = \tilde{\mathbf{U}}^{\perp}((\tilde{\mathbf{U}}^{\perp})^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}^{\perp})^{-1}(\tilde{\mathbf{U}}^{\perp})^T$. As $\mathrm{range}\{\tilde{\mathbf{U}}^{\perp}\} = \mathrm{range}\{\tilde{\mathbf{U}}_s^{\perp}\}$, there exists a nonsingular matrix $\tilde{\mathbf{V}} \in \mathbb{R}^{(K-G)\times(K-G)}$ such that $\tilde{\mathbf{U}}_s^{\perp} = \tilde{\mathbf{U}}^{\perp}\tilde{\mathbf{V}}$. It follows that

$$\begin{aligned}
&\tilde{\mathbf{U}}_s^{\perp}((\tilde{\mathbf{U}}_s^{\perp})^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}_s^{\perp})^{-1}(\tilde{\mathbf{U}}_s^{\perp})^T \\
&= \tilde{\mathbf{U}}^{\perp}\tilde{\mathbf{V}}((\tilde{\mathbf{U}}^{\perp}\tilde{\mathbf{V}})^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}^{\perp}\tilde{\mathbf{V}})^{-1}(\tilde{\mathbf{U}}^{\perp}\tilde{\mathbf{V}})^T \\
&= \tilde{\mathbf{U}}^{\perp}\big((\tilde{\mathbf{U}}^{\perp})^T\tilde{\mathbf{P}}\tilde{\mathbf{U}}^{\perp}\big)^{-1}(\tilde{\mathbf{U}}^{\perp})^T. \qquad \square
\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers, whose valuable comments and suggestions helped to improve the quality of a previous version of the paper.

REFERENCES

[1] R. S. Baheti, R. R. Mohler, and H. A. Spang III, "Second-order correlation method for bilinear system identification," *IEEE Trans. Autom. Control*, vol. AC-25, no. 6, pp. 1141–1146, Dec. 1980.
[2] W. Favoreel, B. D. Moor, and P. V. Overschee, "Subspace identification of bilinear systems subject to white inputs," *IEEE Trans. Autom. Control*, vol. 44, no. 6, pp. 1157–1165, Jun. 1999.
[3] Z. Zhu and H. Leung, "Adaptive identification of nonlinear systems with application to chaotic communications," *IEEE Trans. Circuits Syst., I Reg. Papers*, vol. 47, no. 7, pp. 1072–1080, Jul. 2000.
[4] N. Kalouptsidis, P. Koukoulas, and V. J. Mathews, "Blind identification of bilinear systems," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 484–499, Feb. 2003.
[5] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993, vol. I.
[6] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, vol. I.
[7] B. Friedlander, "On the computation of the Cramer–Rao bound for ARMA parameter estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 4, pp. 721–727, Aug. 1984.
[8] B. Hanzon, *Identifiability, Recursive Identification and Spaces of Linear Dynamical Systems: Part 1, CWI Tract*. Amsterdam, The Netherlands: Centrum voor Wiskunde and Informatica, 1989, vol. 63.
[9] P. Stoica and B. C. Ng, "Performance bounds for blind channel estimation," in *Signal Processing Advances in Wireless and Mobile Communications*. ser. Trends in Channel Estimation and Equalization, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2001, vol. I, ch. 2.

[10] R. J. Ober, "The Fisher information matrix for linear systems," *Syst. Control Lett.*, vol. 47, pp. 221–226, 2002.

[11] R. J. Ober, Q. Zou, and Z. Lin, "Calculation of the Fisher information matrix for multidimensional data sets," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2679–2691, Oct. 2003.

[12] G. S. J. Young and R. Chellappa, "3-D motion estimation using a sequence of noisy stereo images: Models, estimation, and uniqueness results," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 8, pp. 735–759, Aug. 1990.

[13] A. K. R. Chowdhury and R. Chellappa, "Stochastic approximation and rate-distortion analysis for robust structure and motion estimation," *Int. J. Comput. Vision*, vol. 55, pp. 27–53, Oct. 2003.

[14] R. R. Mohler, *Nonlinear Systems, Vol. II, Applications to Bilinear Control*. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[15] S. Zacks, *The Theory of Statistical Inference*. New York: Wiley, 1971.

[16] P. E. Caines, *Linear Stochastic Systems*. New York: Wiley, 1988.

[17] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, no. 5, pp. 720–741, May 1989.

[18] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[19] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, vol. 39, pp. 577–591, May 1971.

[20] P. Stoica and T. Soderstrom, "On nonsingular information matrices and local identifiability," *Int. J. Control*, vol. 36, pp. 323–329, 1982.

[21] W. J. Rugh, *Nonlinear System Theory: The Volterra/Wiener Approach*. Baltimore, MD: The Johns Hopkins Univ. Press, 1981.

[22] R. J. Ober and E. S. Ward, "A system theoretic formulation of NMR experiments," *J. Math. Chem.*, vol. 20, pp. 47–65, 1996.

[23] K. A. Grasse, "Reachability of interior states by piecewise constant controls," *Forum Math.*, vol. 7, pp. 607–628, 1995.

[24] R. J. Ober and E. S. Ward, "On the role of reachability and observability in NMR experimentation," *J. Math. Chem.*, vol. 26, pp. 15–26, 1999.

[25] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[26] F. R. Gantmacher, *The Theory of Matrices*. New York: Chelsea, 1959, vol. I and II.

[27] S. Lofas and B. Johnsson, "A novel hydrogel matrix on gold surfaces in surface plasmon resonance sensors for fast and efficient covalent immobilization of ligands," *J. Chem. Soc. Chem. Commun.*, vol. 21, pp. 1526–1528, 1990.

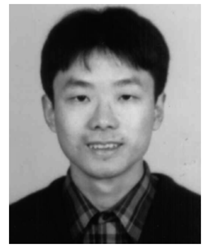
[28] P. Schuck, "Reliable determination of binding affinity and kinetics using surface plasmon resonance biosensors," *Curr. Opin. Biotechnol.*, vol. 8, pp. 498–502, 1997.

[29] R. J. Ober and E. S. Ward, "The influence of signal noise on the accuracy of kinetic constants measured by surface plasmon resonance experiments," *Anal. Biochem.*, vol. 273, pp. 49–59, 1999.

[30] Q. Zou, "System Theory in Biomedical Applications," M. Eng. Thesis, School of Electrical and Electronic Engineering, Nanyang Technological Univ., Singapore, 2003.

[31] T. M. Apostol, *Mathematical Analysis*, 2nd ed. Reading, MA: Addison-Wesley, 1974.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.



Qiyue Zou (S'06) received the B.Eng. and M.Eng. degrees in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2001 and 2004, respectively. He is currently working toward the Ph.D. degree at the Electrical Engineering Department at University of California, Los Angeles.

His current research interests include statistical signal processing, system theory, information theory, and wireless communications.

Zhiping Lin (SM'00) received the Ph.D. degree in information engineering from the University of Cambridge, Cambridge, U.K., in 1987.

He was with the University of Calgary, AB, Canada, from 1987 to 1988, with Shantou University, China, from 1988 to 1993, and with DSO National Laboratories, Singapore, from 1993 to 1999. Since February 1999, he has been an Associate Professor at Nanyang Technological University (NTU), Singapore. He is also the Program Director of the Bio-Signal Processing, Center for Signal Processing, NTU. He is a Co-Editor of *Multidimensional Systems and Signal Processing* and an Associate Editor of *Circuits, Systems and Signal Processing*. His research interests include multidimensional systems and signal processing, array signal processing, and biomedical signal processing.

Raimund J. Ober (S'87–M'87–SM'95) received the Ph.D. degree in electrical engineering from Cambridge University, Cambridge, U.K., in 1987.

From 1987 to 1990, he was a Research Fellow at Girton College and the Engineering Department, Cambridge University. In 1990, he joined the University of Texas at Dallas, Richardson, where he is currently a Professor with the Department of Electrical Engineering. He is also Adjunct Professor at the University of Texas Southwestern Medical Center, Dallas. He is an Associate Editor of *Systems and Control Letters* and *Mathematics of Control, Signals, and Systems*. His research interests include signal and imaging processing and their applications to biological problems, fluorescence microscopy and biosensors.