The Cramer Rao Lower Bound for Bilinear Systems

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

EDICS: 2-ESTM, 2-PERF

The research was supported in part by grants by the National Institutes of Health (RO1 GM58538).

* Corresponding author

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

R. J. Ober is with the Center for Systems, Communications and Signal Processing EC33, 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, NB9.106, University of Texas Southwestern Medical Center at Dallas, 6000 Harry Hines Boulevard, Dallas, TX 75235-8576, USA (e-mail: ober@utdallas.edu).

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 certain Lyapunov equations. 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 an example from surface plasmon resonance experiments for the determination of the kinetic parameters of protein-protein interactions.

Keywords: Bilinear systems; Cramer Rao lower bound; 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 can 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 (1D) dynamic non-stationary systems with deterministic input and Gaussian measurement noise has been investigated in [10].

The calculation of the Fisher information matrix for the 1D data is performed in terms of the derivative system with respect to the system parameters and by using the solution to a Lyapunov equation. The above approach has been extended to multidimensional (nD) data sets generated by nD linear separable-denominator systems and applied to the analysis of nD 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 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 III. Second, for uniformly sampled data sets generated by bilinear systems with piecewise constant inputs, the CRLB can be efficiently computed using algorithms based on solutions to certain Lyapunov equations. Although there have been some papers on the calculation of the CRLB for some specific bilinear models in the literature, such as [12], to our best knowledge, however, an explicit expression of the CRLB for a general bilinear system model has not been available so far.

The organization of the paper is as follows. In Section II we apply the concept of 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. In Section III the nonsingularity conditions of the Fisher information matrix are derived. Provided some weak regularity conditions hold the nonsingularity of the Fisher information matrix is equivalent to the local identifiability of the system. For the uniformly sampled data sets generated by a bilinear system with piecewise constant inputs, it is shown in Section IV that the CRLB can be efficiently calculated through solving certain Lyapunov equations, and that the asymptotic CRLB can be derived without explicitly computing the Fisher information matrix. In Section V the theoretical results presented in the paper are illustrated by an example from surface plasmon resonance experiments aimed at estimating kinetic constants of protein-protein interactions.

Notation

A^T	transpose of matrix A
$\oplus$	sum of vector spaces
$\mathbb{R}^{n\times m}$	space of $n \times m$ real matrices
$\theta = \left[\begin{array}{ccc} \theta_1 & \dots & \theta_K \end{array}\right]^T$	parameter vector
$I(\theta)$	Fisher information matrix for the parameter vector θ

I	identity matrix of appropriate dimension
σ^2	variance of Gaussian noise
$E\{x\}$	expected value of random variable x
$\mathrm{var}\,(x)$	variance of random variable x
$\mathrm{rank}\,\{A\}$	rank of matrix A
$\mathrm{range}\,\{A\}$	span of the columns of matrix A
$\det\,(A)$	determinant of matrix A
$\mathrm{diag}\,\{M_1,\ \ldots,\ M_r\}$	block diagonal matrix whose diagonal block entries are M_1, …, M_r
$\Phi = \{A,\ B,\ C,\ F_1,\ \ldots,\ F_m\}$	bilinear system with system matrices A, B, C, F_1, …, F_m
$\Phi' = \{\mathcal{A},\ \mathcal{B},\ \mathcal{C},\ \mathcal{F}_1,\ \ldots,\ \mathcal{F}_m\}$	derivative system of Φ with system matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{F}_1$, …, $\mathcal{F}_m$
$\mathbb{V}$	set of admissible inputs
$\mathbb{U}$	set of piecewise constant inputs
$u^{[l]}$	l^{th} constant input vector of a piecewise constant input
β_l	indicator function for the l^{th} interval

II. Cramer Rao Lower Bound

Consider the state-space model of a general MIMO time-invariant bilinear system given by (see [13])

$$\dot{x}_\theta(t) = Ax_\theta(t) + \sum_{q=1}^{m} F_q u_q(t) x_\theta(t) + Bu(t), \qquad x_\theta(t^{[0]}) = x_0, \tag{1}$$

$$y_\theta(t) = Cx_\theta(t), \quad t \geq t^{[0]}, \tag{2}$$

where $x_\theta(t) \in \mathbb{R}^{n\times 1}$ is the state vector, $u(t) \in \mathbb{R}^{m\times 1}$ is the input vector with components $u_1(t)$, …, $u_m(t)$, $y_\theta(t) \in \mathbb{R}^{p\times 1}$ is the system output vector, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $F_q \in \mathbb{R}^{n\times n}$, $q = 1, \ldots, m$, are the system matrices depending on the unknown parameter vector $\theta := \begin{bmatrix} \theta_1 & \ldots & \theta_K \end{bmatrix}^T$, and x_0 is the initial state vector, which can also depend on the parameter vector θ. For convenience of exposition, we use the notation $\Phi := \{A,\ B,\ C,\ F_1,\ \ldots,\ F_m\}$ to represent the bilinear system with state vector $x_\theta(t)$, input $u(t)$, output $y_\theta(t)$, system matrices A, B, C, F_1, …, F_m, and initial state x_0, as defined in (1)-(2). The i^{th} element of $y_\theta(t)$ is represented by $y_{\theta,i}(t)$, $i = 1, \ldots, p$, i.e., $y_\theta(t) := \begin{bmatrix} y_{\theta,1}(t) & \ldots & y_{\theta,p}(t) \end{bmatrix}^T$. Similarly, the i^{th} row of C is denoted by C_i, $i = 1, \ldots, p$, i.e., $C = \begin{bmatrix} C_1^T & \ldots & C_p^T \end{bmatrix}^T$.

In this paper two classes of input functions are considered for the bilinear system defined in (1)-(2) [14], [15]:

(i) The set of admissible inputs $\mathbb{V}$: these inputs 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 $t^{[0]}$.

(ii) The set of piecewise constant inputs $\mathbb{U}$: they are vector-valued piecewise constant functions with a finite number of steps. A piecewise constant input $u \in \mathbb{U}$ can be represented by

$$u(t) = \sum_{l=0}^{L-1} u^{[l]} \beta_l(t), \quad t^{[0]} \leq t < t^{[L]}, \tag{3}$$

where $u^{[l]} := \left[\begin{array}{ccc} u_1^{[l]} & \dots & u_m^{[l]} \end{array} \right]^T$, $l = 0, \dots, L-1$, are constant vectors, and $\beta_l(t)$, $l = 0, \dots, L-1$, 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, $t^{[0]}, \ \dots, \ t^{[L]}$ denote the starting and ending points of the time intervals with $t^{[0]} < \dots < t^{[L]}$, where $t^{[L]}$ can be either finite or infinite. Note that $u^{[l]}$ could be a zero vector, and that for a piecewise constant input $u \in \mathbb{U}$ as defined in (3) we are only interested in the output $y_\theta(t)$ for $t^{[0]} \leq t < t^{[L]}$.

Obviously, $\mathbb{U} \subset \mathbb{V}$. Unlike a linear system, it is difficult to express the output of a bilinear system in terms of its input and system matrices in a simple closed form. For an admissible input $u \in \mathbb{V}$, the input-output map can be represented by the infinite Volterra series (see [16]). For a piecewise constant input $u \in \mathbb{U}$, the input-output description can be further simplified, as shown in the following lemma whose proof is standard (see also [17]).

Lemma 2.1: Assume that the input of the bilinear system $\Phi = \{A, \ B, \ C, \ F_1, \ \dots, \ F_m\}$ is piecewise constant as defined by (3). Let $F^{[l]} := \sum_{q=1}^{m} F_q u_q^{[l]}$ and assume that $A + F^{[l]}$ is invertible, $l = 0, \dots, L-1$. Then the output of the system is given by

$$y_\theta(t) = \sum_{l=0}^{L-1} \left[C Q_l(t) \left(W^{[l]} + x_\theta(t^{[l]}) \right) - C W^{[l]} \right] \beta_l(t), \quad t^{[0]} \leq t < t^{[L]}, \tag{4}$$

where $Q_l(t) := e^{(A+F^{[l]})(t-t^{[l]})}$ and $W^{[l]} := (A + F^{[l]})^{-1} B u^{[l]}$, $l = 0, \dots, L-1$, and $x_\theta(t^{[l]})$ is given by

$$x_\theta(t^{[l]}) = \begin{cases} x_0, & \text{for } l = 0, \\ Q_{l-1}(t^{[l]}) \left(W^{[l-1]} + x_\theta(t^{[l-1]}) \right) - W^{[l-1]}, & \text{for } l = 1, \dots, L-1. \end{cases}$$

Assume that we have acquired noise corrupted samples $s_{\theta,i}(j)$, $i = 1, \dots, p$, $j = 0, \dots, J-1$, of the measured output of the bilinear system defined by (1)-(2), i.e.,

$$s_{\theta,i}(j) = y_{\theta,i}(t_j) + w_i(j), \tag{5}$$

where $y_{\theta,i}(t_j)$ is the i^{th} noise free output element at the sampling point t_j and $w_i(j)$ is the measurement noise component, $i = 1, \ldots, p$, $j = 0, \ldots, J-1$, $t^{[0]} \leq t_0 < t_1 < \ldots < t_{J-1}$. We assume that the measurement noise components have independent Gaussian distributions with zero mean and variance $\sigma_{i,j}^2$, $i = 1, \ldots, p$, $j = 0, \ldots, J-1$. Hence the probability density function $p(S;\theta)$ for the acquired data set $S := \{s_{\theta,i}(j),\ i = 1, \ldots, p,\ j = 0, \ldots, J-1\}$ is given by

$$p(S;\theta) = \prod_{i=1}^{p} \prod_{j=0}^{J-1} \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \exp\left(-\frac{1}{2\sigma_{i,j}^2} \left[s_{\theta,i}(j) - y_{\theta,i}(t_j) \right]^2 \right).$$

The parameter space Θ, i.e. the set of all possible values for the parameter vector θ, is assumed to be an open subset of the Euclidean space $\mathbb{R}^{K\times 1}$. Also, $p(S;\theta)$ is assumed to satisfy the standard regularity conditions (see e.g. [18], [19]). The Fisher information matrix $I(\theta)$ is then defined as

$$[I(\theta)]_{sr} = E\left\{ \left(\frac{\partial \ln p(S;\theta)}{\partial \theta_s} \right) \left(\frac{\partial \ln p(S;\theta)}{\partial \theta_r} \right) \right\}, \quad 1 \leq s, r \leq K,$$

where $E\{\cdot\}$ is the expected value with respect to the underlying probability measure. If $I(\theta)$ is positive definite for all $\theta \in \Theta$, by the CRLB any unbiased estimator $\hat{\theta}$ of θ has a variance such that

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

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

In the following theorem we first show that the derivative system (with respect to the given parameter vector θ) 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 for Gaussian measurement noise is then expressed using the output samples of its derivative system.

Theorem 2.1: Consider the bilinear system represented by $\Phi = \{A,\ B,\ C,\ F_1,\ \ldots,\ F_m\}$. Assume that the partial derivatives of A, B, C, F_1, …, F_m and x_0 with respect to the elements of θ exist for all $\theta \in \Theta$, and that the input $u(t)$ is independent of the parameter vector θ. Let

$$\mathcal{Y}_\theta(t) := \begin{bmatrix} \mathcal{Y}_{\theta,1}(t) \\ \vdots \\ \mathcal{Y}_{\theta,p}(t) \end{bmatrix}, \quad \textit{with} \ \ \mathcal{Y}_{\theta,i}(t) := \begin{bmatrix} \frac{\partial y_{\theta,i}(t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial y_{\theta,i}(t)}{\partial \theta_K} \end{bmatrix} \ (i = 1, \ldots, p), \quad t \geq t^{[0]}.$$

Then

1.) $\mathcal{Y}_\theta(t)$, $t \geq t^{[0]}$, is the output of the derivative system $\Phi' := \{\mathcal{A},\ \mathcal{B},\ \mathcal{C},\ \mathcal{F}_1,\ \ldots,\ \mathcal{F}_m\}$, which is an MIMO time-invariant bilinear system with state vector $\mathcal{X}_\theta(t)$, $t \geq t^{[0]}$, and has the same input u as Φ. The state vector $\mathcal{X}_\theta$, initial state $\mathcal{X}_0$, and system matrices $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{F}_1$, $\ldots$, $\mathcal{F}_m$ are given as follows, which will be adopted throughout the paper.

$$\mathcal{X}_\theta(t) := \begin{bmatrix} \partial_1 x_\theta(t) \\ \vdots \\ \partial_K x_\theta(t) \end{bmatrix},\ t \geq t^{[0]},\ \mathcal{X}_0 := \begin{bmatrix} \partial_1 x_\theta(t^{[0]}) \\ \vdots \\ \partial_K x_\theta(t^{[0]}) \end{bmatrix},\ \mathcal{A} := \operatorname{diag}\{\partial_1 A,\ \ldots,\ \partial_K A\},$$

$$\mathcal{B} := \begin{bmatrix} \partial_1 B \\ \vdots \\ \partial_K B \end{bmatrix},\ \mathcal{C} := \begin{bmatrix} \mathcal{C}_1 \\ \vdots \\ \mathcal{C}_p \end{bmatrix} \text{ with } \mathcal{C}_i := \operatorname{diag}\{\partial_1 C_i,\ \ldots,\ \partial_K C_i\},\ i = 1, \ldots, p, \tag{6}$$

$$\mathcal{F}_q := \operatorname{diag}\{\partial_1 F_q,\ \ldots,\ \partial_K F_q\},\ q = 1, \ldots, m,$$

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

$$\partial_s x_\theta(t) := \begin{bmatrix} x_\theta(t) \\ \frac{\partial x_\theta(t)}{\partial \theta_s} \end{bmatrix},\ t \geq t^{[0]},\ \partial_s x_\theta(t^{[0]}) := \begin{bmatrix} x_0 \\ \frac{\partial x_0}{\partial \theta_s} \end{bmatrix},\ \partial_s A := \begin{bmatrix} A & 0 \\ \frac{\partial A}{\partial \theta_s} & A \end{bmatrix}, \tag{7}$$

$$\partial_s B := \begin{bmatrix} B \\ \frac{\partial B}{\partial \theta_s} \end{bmatrix}, \partial_s C_i := \begin{bmatrix} \frac{\partial C_i}{\partial \theta_s} & C_i \end{bmatrix},\ \partial_s F_q := \begin{bmatrix} F_q & 0 \\ \frac{\partial F_q}{\partial \theta_s} & F_q \end{bmatrix},\ q = 1, \ldots, m;$$

2.) for the data points $s_{\theta,i}(j) := y_{\theta,i}(t_j) + w_i(j)$, where $y_{\theta,i}(t_j)$ is the sampled output of the bilinear system Φ and $w_i(j)$ is independent Gaussian noise with zero mean and variance $\sigma^2_{i,j}$, $i = 1, \ldots, p$, $j = 0, \ldots, J-1$, $t^{[0]} \leq t_0 < t_1 < \cdots < t_{J-1}$, the Fisher information matrix is given by

$$I(\theta) = \sum_{i=1}^{p} \sum_{j=0}^{J-1} \frac{1}{\sigma^2_{i,j}} P_i \mathcal{Y}_\theta(t_j) \mathcal{Y}_\theta^T(t_j) P_i^T. \tag{8}$$

Here $P_i \in \mathbb{R}^{K \times pK}$, $i = 1, \ldots, p$, is defined as

$$P_i = [\ \underbrace{\mathbf{0}\ \ \ldots\ \ \mathbf{0}}_{(i-1)\ \mathbf{0}s}\ \ I_K\ \ \underbrace{\mathbf{0}\ \ \ldots\ \ \mathbf{0}}_{(p-i)\ \mathbf{0}s}\ ]\ , \tag{9}$$

where $\mathbf{0}$ denotes the $K \times K$ zero matrix and I_K the $K \times K$ identity matrix.

Proof: 1.) By assumption the partial derivatives of A, B, C, F_1, …, F_m and x_0 with respect to θ_s ($s = 1, \ldots, K$ throughout the proof) exist for all $\theta \in \Theta$. Hence, the partial derivatives of $x_\theta(t)$ and $y_{\theta,i}(t)$, $i = 1, \ldots, p$, with respect to θ_s also exist for all $\theta \in \Theta$ and $t \geq t^{[0]}$. Since the input $u(t)$, $t \geq t^{[0]}$, is piecewise continuous, it follows that $x_\theta(t)$ and $\frac{\partial x_\theta(t)}{\partial \theta_s}$ are partially differentiable with respect to t on $t \geq t^{[0]}$ with the possible exception of the discontinuities of u. Also, the partial derivative of $\frac{\partial x_\theta(t)}{\partial t}$ with respect to θ_s exists for all $\theta \in \Theta$ and $t \geq t^{[0]}$ with the possible exception of the discontinuities of u. With the exception of the discrete discontinuities of u, we have $\frac{\partial \dot{x}_\theta(t)}{\partial \theta_s} = \frac{\partial^2 x_\theta(t)}{\partial \theta_s \partial t} = \frac{\partial^2 x_\theta(t)}{\partial t \partial \theta_s}$, $t \geq t^{[0]}$ (see page 359 in [20]).

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

$$\frac{\partial \dot{x}_\theta(t)}{\partial \theta_s} = \left[\begin{array}{cc} \frac{\partial A}{\partial \theta_s} & A \end{array} \right] \left[\begin{array}{c} x_\theta(t) \\ \frac{\partial x_\theta(t)}{\partial \theta_s} \end{array} \right] + \sum_{q=1}^{m} \left[\begin{array}{cc} \frac{\partial F_q}{\partial \theta_s} & F_q \end{array} \right] u_q(t) \left[\begin{array}{c} x_\theta(t) \\ \frac{\partial x_\theta(t)}{\partial \theta_s} \end{array} \right] + \frac{\partial B}{\partial \theta_s} u(t), \quad t \geq t^{[0]}. \tag{10}$$

With the exception of the discontinuities of u, combining (1) and (10) yields

$$\begin{aligned} \frac{\partial}{\partial t} \partial_s x_\theta(t) = &\ \frac{\partial}{\partial t} \left[\begin{array}{c} x_\theta(t) \\ \frac{\partial x_\theta(t)}{\partial \theta_s} \end{array} \right] = \left[\begin{array}{c} \frac{\partial x_\theta(t)}{\partial t} \\ \frac{\partial^2 x_\theta(t)}{\partial t \partial \theta_s} \end{array} \right] = \left[\begin{array}{c} \dot{x}_\theta(t) \\ \frac{\partial \dot{x}_\theta(t)}{\partial \theta_s} \end{array} \right] \\ = &\ \left[\begin{array}{cc} A & 0 \\ \frac{\partial A}{\partial \theta_s} & A \end{array} \right] \left[\begin{array}{c} x_\theta(t) \\ \frac{\partial x_\theta(t)}{\partial \theta_s} \end{array} \right] + \sum_{q=1}^{m} \left[\begin{array}{cc} F_q & 0 \\ \frac{\partial F_q}{\partial \theta_s} & F_q \end{array} \right] u_q(t) \left[\begin{array}{c} x_\theta(t) \\ \frac{\partial x_\theta(t)}{\partial \theta_s} \end{array} \right] + \left[\begin{array}{c} B \\ \frac{\partial B}{\partial \theta_s} \end{array} \right] u(t) \\ = &\ \partial_s A \partial_s x_\theta(t) + \sum_{q=1}^{m} \partial_s F_q u_q(t) \partial_s x_\theta(t) + \partial_s B u(t), \quad t \geq t_0. \end{aligned} \tag{11}$$

Also,

$$\partial_s x_\theta(t^{[0]}) = \left[\begin{array}{c} x_\theta(t^{[0]}) \\ \frac{\partial x_\theta(t^{[0]})}{\partial \theta_s} \end{array} \right] = \left[\begin{array}{c} x_0 \\ \frac{\partial x_0}{\partial \theta_s} \end{array} \right]. \tag{12}$$

Since $y_{\theta,i}(t) = C_i x_\theta(t)$, $i = 1, \ldots, p$, $t \geq t^{[0]}$, taking the partial derivative of $y_{\theta,i}(t)$ with respect to θ_s gives

$$\frac{\partial y_{\theta,i}(t)}{\partial \theta_s} = \partial_s C_i \partial_s x_\theta(t), \quad t \geq t^{[0]}. \tag{13}$$

For $i = 1, \ldots, p$, since

$$\mathcal{Y}_{\theta,i}(t) = \left[\begin{array}{c} \frac{\partial y_{\theta,i}(t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial y_{\theta,i}(t)}{\partial \theta_K} \end{array} \right], \ \mathcal{X}_\theta(t) = \left[\begin{array}{c} \partial_1 x_\theta(t) \\ \vdots \\ \partial_K x_\theta(t) \end{array} \right], \quad t \geq t^{[0]},$$

stacking the corresponding equations from (11) and (13) gives

$$\dot{\mathcal{X}}_\theta(t) = \mathcal{A} \mathcal{X}_\theta(t) + \sum_{q=1}^{m} \mathcal{F}_q u_q(t) \mathcal{X}_\theta(t) + \mathcal{B} u(t), \tag{14}$$

$$\mathcal{Y}_{\theta,i}(t) = \mathcal{C}_i \mathcal{X}_\theta(t), \quad t \geq t^{[0]}. \tag{15}$$

The desired derivative system Φ' is then obtained by stacking the corresponding equations from (15) as

$$\mathcal{Y}_\theta(t) = \mathcal{C} \mathcal{X}_\theta(t), \quad t \geq t^{[0]}. \tag{16}$$

The initial condition of Φ' is given by stacking the corresponding equations from (12) as $\mathcal{X}_\theta(t^{[0]}) = \mathcal{X}_0$. Clearly, $\mathcal{Y}_\theta(t)$, $t \geq t^{[0]}$, is the output of the derivative system $\Phi' := \{\mathcal{A},\ \mathcal{B},\ \mathcal{C},\ \mathcal{F}_1,\ \ldots,\ \mathcal{F}_m\}$. Note that each element of $\mathcal{Y}_\theta(t)$ is a continuous function of t for $t \geq t^{[0]}$.

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

$$[I(\theta)]_{sr} = E\left\{ \left(\frac{\partial \ln p(S;\theta)}{\partial \theta_s}\right) \left(\frac{\partial \ln p(S;\theta)}{\partial \theta_r}\right) \right\} = \sum_{i=1}^{p} \sum_{j=0}^{J-1} \frac{1}{\sigma_{i,j}^2} \frac{\partial y_{\theta,i}(t_j)}{\partial \theta_s} \frac{\partial y_{\theta,i}(t_j)}{\partial \theta_r}, \quad \text{for } 1 \leq s, r \leq K.$$

It then follows directly that the Fisher information matrix is given by

$$I(\theta) = \sum_{i=1}^{p} \sum_{j=0}^{J-1} \frac{1}{\sigma_{i,j}^2} \mathcal{Y}_{\theta,i}(t_j) \mathcal{Y}_{\theta,i}^T(t_j) = \sum_{i=1}^{p} \sum_{j=0}^{J-1} \frac{1}{\sigma_{i,j}^2} P_i \mathcal{Y}_\theta(t_j) \mathcal{Y}_\theta^T(t_j) P_i^T,$$

where $\mathcal{Y}_\theta(t_j)$ is the sampled output of the derivative system Φ' at t_j, $j = 0, \ldots, J-1$, and P_i, $i = 1, \ldots, p$, are defined in (9). $\square$

For the data set generated by a bilinear system with a piecewise constant input $u \in \mathbb{U}$, the following corollary derives an explicit expression of its associated Fisher information matrix.

Corollary 2.1: Assume the bilinear system model and assumptions are the same as in Theorem 2.1, except that the input u is piecewise constant and the data points are sampled from $y_{\theta,i}(t)$, $i = 1, \ldots, p$, at $t^{[0]} \leq t_0 < t_1 < \ldots < t_{J-1} < t^{[L]}$. Assume that $A + F^{[l]}$ is invertible, where $F^{[l]} := \sum_{q=1}^{m} F_q u_q^{[l]}$, $l = 0, \ldots, L-1$. Let $\mathcal{F}^{[l]} := \operatorname{diag}\{\partial_1 F^{[l]}, \ldots, \partial_K F^{[l]}\}$, $l = 0, \ldots, L-1$, where for $s = 1, \ldots, K$

$$\partial_s F^{[l]} := \begin{bmatrix} F^{[l]} & 0 \\ \frac{\partial F^{[l]}}{\partial \theta_s} & F^{[l]} \end{bmatrix} = \sum_{q=1}^{m} \begin{bmatrix} F_q & 0 \\ \frac{\partial F_q}{\partial \theta_s} & F_q \end{bmatrix} u_q^{[l]}.$$

Then,

1.) the output of the derivative system Φ' is given by

$$\mathcal{Y}_\theta(t) = \sum_{l=0}^{L-1} \left[\mathcal{C} \mathcal{Q}_l(t) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]}) \right) - \mathcal{C} \mathcal{W}^{[l]} \right] \beta_l(t), \quad t^{[0]} \leq t < t^{[L]}, \tag{17}$$

where $\mathcal{Q}_l(t) := e^{(\mathcal{A}+\mathcal{F}^{[l]})(t-t^{[l]})}$, $\mathcal{W}^{[l]} := \left(\mathcal{A} + \mathcal{F}^{[l]}\right)^{-1} \mathcal{B} u^{[l]}$, $l = 0, \ldots, L-1$, and

$$\mathcal{X}_\theta(t^{[l]}) = \begin{cases} \mathcal{X}_0, & \text{for } l = 0, \\ \mathcal{Q}_{l-1}(t^{[l]}) \left(\mathcal{W}^{[l-1]} + \mathcal{X}_\theta(t^{[l-1]}) \right) - \mathcal{W}^{[l-1]}, & \text{for } l = 1, \ldots, L-1; \end{cases}$$

2.) the Fisher information matrix is given by

$$\begin{aligned} I(\theta) =& \sum_{i=1}^{p} \sum_{j=0}^{J-1} \frac{1}{\sigma_{i,j}^2} P_i \mathcal{Y}_\theta(t_j) \mathcal{Y}_\theta^T(t_j) P_i^T \\ =& \sum_{i=1}^{p} \sum_{j=0}^{J-1} \sum_{l=0}^{L-1} \frac{1}{\sigma_{i,j}^2} P_i \mathcal{C} \mathcal{Q}_l(t_j) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]}) \right) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]}) \right)^T \mathcal{Q}_l^T(t_j) \mathcal{C}^T P_i^T \beta_l(t_j) \\ & - \sum_{i=1}^{p} \sum_{j=0}^{J-1} \sum_{l=0}^{L-1} \frac{1}{\sigma_{i,j}^2} P_i \mathcal{C} \mathcal{Q}_l(t_j) \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T + \mathcal{X}_\theta(t^{[l]}) \left(\mathcal{W}^{[l]} \right)^T \right) \mathcal{C}^T P_i^T \beta_l(t_j) \end{aligned}$$

$$
- \sum_{i=1}^{p} \sum_{j=0}^{J-1} \sum_{l=0}^{L-1} \frac{1}{\sigma_{i,j}^2} P_i \mathcal{C} \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T + \mathcal{W}^{[l]} \mathcal{X}_\theta^T (t^{[l]}) \right) \mathcal{Q}_l^T (t_j) \mathcal{C}^T P_i^T \beta_l(t_j)
$$

$$
+ \sum_{i=1}^{p} \sum_{j=0}^{J-1} \sum_{l=0}^{L-1} \frac{1}{\sigma_{i,j}^2} P_i \mathcal{C} \mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T \mathcal{C}^T P_i^T \beta_l(t_j). \tag{18}
$$

Proof: 1.) By Theorem 2.1, $\mathcal{Y}_\theta(t)$, $t^{[0]} \leq t < t^{[L]}$, is the output of the derivative system $\Phi' := \{\mathcal{A},\ \mathcal{B},\ \mathcal{C},\ \mathcal{F}_1,\ \ldots,\ \mathcal{F}_m\}$, which is bilinear and has the same piecewise constant input as Φ. Since $\mathcal{A} + \mathcal{F}^{[l]}$ is a block lower-triangular matrix with $A + F^{[l]}$ as its diagonal block submatrices and $A + F^{[l]}$ is invertible, it follows that $\mathcal{A} + \mathcal{F}^{[l]}$ is also invertible, $l = 0, \ldots, L-1$. Using Lemma 2.1 (see also (4)), we can easily derive the explicit expression for $\mathcal{Y}_\theta(t)$, $t^{[0]} \leq t < t^{[L]}$, given in (17).

2.) The Fisher information matrix $I(\theta)$ given in (18) can be readily obtained by substituting $\mathcal{Y}_\theta(t)$ in (17) into expression (8). $\square$

III. Local Identifiability

The parameter vector θ is said to be locally identifiable if there exists an open neighborhood of θ containing no other parameter vector that is observably equivalent to θ [21]. The following Theorem 3.1 quoted from [22] 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 invertibility of the Fisher information matrix is of importance in and of 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 $I(\theta)$.

Theorem 3.1: [22] Let θ be a regular point of the Fisher information matrix $I(\theta)$. Then θ is locally identifiable if and only if $I(\theta)$ is nonsingular.

Specifically, Theorem 3.1 assumes that $\frac{\partial p(S;\theta)}{\partial \theta_s}$, $\frac{\partial \ln p(S;\theta)}{\partial \theta_s}$ and the elements of $I(\theta)$ are continuous functions of θ for each $s = 1, \ldots, K$ and all $\theta \in \Theta$. For the data model given by (5), the explicit expressions of $\frac{\partial p(S;\theta)}{\partial \theta_s}$ and $\frac{\partial \ln p(S;\theta)}{\partial \theta_s}$ are

$$
\frac{\partial p(S;\theta)}{\partial \theta_s} = \left(\prod_{i=1}^{p} \prod_{j=0}^{J-1} \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \right) \left[\sum_{i=1}^{p} \sum_{j=0}^{J-1} \frac{1}{\sigma_{i,j}^2} \left[s_{\theta,i}(j) - y_{\theta,i}(t_j) \right] \right. \tag{19}
$$

$$
\left. \cdot \exp \left(\sum_{i=1}^{p} \sum_{j=0}^{J-1} - \frac{1}{2\sigma_{i,j}^2} \left[s_{\theta,i}(j) - y_{\theta,i}(t_j) \right]^2 \right) \frac{\partial y_{\theta,i}(t_j)}{\partial \theta_s} \right],
$$

and

$$
\frac{\partial \ln p(S;\theta)}{\partial \theta_s} = \sum_{i=1}^{p} \sum_{j=0}^{J-1} \frac{1}{\sigma_{i,j}^2} \left[s_{\theta,i}(j) - y_{\theta,i}(t_j) \right] \frac{\partial y_{\theta,i}(t_j)}{\partial \theta_s}. \tag{20}
$$

From (19)-(20) and Theorem 2.1, it is easy to see that $\frac{\partial p(S;\theta)}{\partial \theta_s}$, $\frac{\partial \ln p(S;\theta)}{\partial \theta_s}$ and the elements of $I(\theta)$ are continuous functions of θ if the partial derivatives of A, B, C, F_1, …, F_m and x_0 with respect to the elements of θ are continuous functions of θ for all $\theta \in \Theta$.

It will be shown shortly that the local identifiability of a bilinear system is closely related to the span of reachable outputs of its associated derivative system with respect to the unknown parameter vector. Before proceeding, we first review some important notions from mathematical system theory [16].

Definition 3.1: Consider a system with a set of inputs denoted by $\mathbb{S}$. Then

1.) a state x is said to be *reachable* from initial state x_0 via inputs $\mathbb{S}$ if there exists an input in $\mathbb{S}$ such that the path of its corresponding states starts at x_0 and passes through x;

2.) an output y is said to be *reachable* from initial state x_0 via inputs $\mathbb{S}$ if there exists a state reachable from x_0 via inputs $\mathbb{S}$ such that its corresponding output is y.

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

Lemma 3.1: Consider the MIMO bilinear system represented by $\Phi = \{A,\ B,\ C,\ F_1,\ \ldots,\ F_m\}$. Let Ω_{x_θ} denote the span of reachable states and Ω_{y_θ} the span of reachable outputs from a given x_0 via admissible inputs $\mathbb{V}$ (or piecewise constant inputs $\mathbb{U}$). Define matrices $O^{[0]}$, …, $O^{[n-1]}$ (n is the size of the state vector) as

$$O^{[0]} := O_0, \quad O^{[1]} := \begin{bmatrix} O_0 & O_1 \end{bmatrix}, \quad \ldots, \quad O^{[n-1]} := \begin{bmatrix} O_0 & \ldots & O_{n-1} \end{bmatrix}. \tag{21}$$

Here O_0, …, O_{n-1} are given by

$$O_0 := \begin{bmatrix} Ax_0 & B + B' \end{bmatrix}, \quad O_v := \begin{bmatrix} AO_{v-1} & F_1 O_{v-1} & \ldots & F_m O_{v-1} \end{bmatrix}, \quad v = 1, \ldots, n-1,$$

where $B' = \begin{bmatrix} F_1 x_0 & \ldots & F_m x_0 \end{bmatrix}$. Then there exists an integer $0 \le r \le n-1$ such that

$$\text{range}\left\{O^{[0]}\right\} \subset \text{range}\left\{O^{[1]}\right\} \subset \ldots \subset \text{range}\left\{O^{[r]}\right\} = \text{range}\left\{O^{[r+1]}\right\} = \ldots = \text{range}\left\{O^{[n-1]}\right\},$$

and

$$\Omega_{x_\theta} = \text{range}\left\{\begin{bmatrix} x_0 & O^{[r]} \end{bmatrix}\right\}, \quad \Omega_{y_\theta} = \text{range}\left\{\begin{bmatrix} Cx_0 & CO^{[r]} \end{bmatrix}\right\}.$$

Proof: Let $x'_\theta(t) := x_\theta(t) - x_0$ and $y'_\theta(t) := y_\theta(t) - Cx_0$, $t \ge t^{[0]}$. By substitution, (1)-(2) become

$$\dot{x}'_\theta(t) = Ax'_\theta(t) + \sum_{q=1}^{m} F_q u_q(t) x'_\theta(t) + (B + B')u(t) + Ax_0, \qquad x'_\theta(t^{[0]}) = 0, \tag{22}$$

$$y'_\theta(t) = Cx'_\theta(t), \quad t \ge t^{[0]}, \tag{23}$$

where $B' := \begin{bmatrix} F_1 x_0 & \ldots & F_m x_0 \end{bmatrix}$. Let $\Omega_{x'_\theta}$ denote the span of reachable states and $\Omega_{y'_\theta}$ the span of reachable outputs of the new system (22)-(23) from $x'_\theta(t^{[0]}) = 0$ via admissible inputs $\mathbb{V}$ (or piecewise constant inputs

$\mathbb{U}$). By Lemmas 4.1 and 4.2 in [16] with some straightforward generalizations from SIMO bilinear systems to their MIMO counterparts, there exists an integer $0 \le r \le n-1$ such that

$$\text{range}\left\{O^{[0]}\right\} \subset \text{range}\left\{O^{[1]}\right\} \subset \ldots \subset \text{range}\left\{O^{[r]}\right\} = \text{range}\left\{O^{[r+1]}\right\} = \ldots = \text{range}\left\{O^{[n-1]}\right\},$$

and

$$\Omega_{x'_\theta} = \text{range}\{O^{[r]}\} \quad \text{and} \quad \Omega_{y'_\theta} = \text{range}\{CO^{[r]}\},$$

where $O^{[0]}$, $\ldots$, $O^{[n-1]}$ are defined in (21). Since $x_\theta(t) = x'_\theta(t) + x_0$ and $y_\theta(t) = y'_\theta(t) + Cx_0$, $t \ge t^{[0]}$, it is obvious that

$$\Omega_{x_\theta} = \text{range}\left\{\left[\begin{array}{cc} x_0 & O^{[r]} \end{array}\right]\right\} \quad \text{and} \quad \Omega_{y_\theta} = \text{range}\left\{\left[\begin{array}{cc} Cx_0 & CO^{[r]} \end{array}\right]\right\}. \qquad \square$$

Using the above lemma, we can now obtain necessary and sufficient conditions for the existence of an output data set generated by a bilinear system with either admissible inputs or piecewise constant inputs such that its associated Fisher information matrix $I(\theta)$ is nonsingular. Criteria for the nonsingularity of the Fisher information matrix when the system is excited by a specific input will be given in the next section.

Theorem 3.2: Consider the bilinear system represented by $\Phi = \{A,\ B,\ C,\ F_1,\ \ldots,\ F_m\}$. Assume that A, B, C, F_1, $\ldots$, F_m and x_0 depend on the unknown parameter vector θ of size K, and their partial derivatives with respect to the elements of θ exist for all $\theta \in \Theta$ and are continuous functions of θ. The derivative system of Φ is represented by $\Phi' = \{\mathcal{A},\ \mathcal{B},\ \mathcal{C},\ \mathcal{F}_1,\ \ldots,\ \mathcal{F}_m\}$, where the dimension of $\mathcal{A}$ is $M \times M$ with $M = 2Kn$. Define matrices $\mathcal{O}^{[0]}$, $\ldots$, $\mathcal{O}^{[M-1]}$ as

$$\mathcal{O}^{[0]} := \mathcal{O}_0, \quad \mathcal{O}^{[1]} := \left[\begin{array}{cc} \mathcal{O}_0 & \mathcal{O}_1 \end{array}\right], \quad \ldots, \quad \mathcal{O}^{[M-1]} := \left[\begin{array}{ccc} \mathcal{O}_0 & \ldots & \mathcal{O}_{M-1} \end{array}\right]. \tag{24}$$

Here $\mathcal{O}_0$, $\ldots$, $\mathcal{O}_{M-1}$ are given by

$$\mathcal{O}_0 := \left[\begin{array}{cc} \mathcal{A}\mathcal{X}_0 & \mathcal{B} + \mathcal{B}' \end{array}\right], \quad \mathcal{O}_v := \left[\begin{array}{cccc} \mathcal{A}\mathcal{O}_{v-1} & \mathcal{F}_1\mathcal{O}_{v-1} & \ldots & \mathcal{F}_m\mathcal{O}_{v-1} \end{array}\right], \quad v = 1, \ldots, M-1, \tag{25}$$

where $\mathcal{B}' = \left[\begin{array}{ccc} \mathcal{F}_1\mathcal{X}_0 & \ldots & \mathcal{F}_m\mathcal{X}_0 \end{array}\right]$. Assume that the measurement noise components have independent Gaussian distributions with zero mean.

Then there exists a finite set of admissible inputs (or a finite set of piecewise constant inputs) and output samples such that the associated Fisher information matrix $I(\theta) \in \mathbb{R}^{K \times K}$ is nonsingular if and only if

$$\text{rank}\left\{\left[\begin{array}{ccccccc} P_1\mathcal{C}\mathcal{X}_0 & P_1\mathcal{C}\mathcal{O}^{[r]} & P_2\mathcal{C}\mathcal{X}_0 & P_2\mathcal{C}\mathcal{O}^{[r]} & \ldots & P_p\mathcal{C}\mathcal{X}_0 & P_p\mathcal{C}\mathcal{O}^{[r]} \end{array}\right]\right\} = K,$$

where P_i, $i = 1, \ldots, p$, are defined in (9), and r is the integer such that

$$\text{range}\left\{\mathcal{O}^{[0]}\right\} \subset \text{range}\left\{\mathcal{O}^{[1]}\right\} \subset \ldots \subset \text{range}\left\{\mathcal{O}^{[r]}\right\} = \text{range}\left\{\mathcal{O}^{[r+1]}\right\} = \ldots = \text{range}\left\{\mathcal{O}^{[M-1]}\right\}. \tag{26}$$

Proof: We prove it for the case of admissible inputs. It can be similarly proven for the case of piecewise constant inputs.

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

$$I_u(\theta) = \sum_{i=1}^{p} \sum_{j=0}^{J-1} \frac{1}{\sigma_{i,j}^2} P_i \mathcal{Y}_\theta(t_j) \mathcal{Y}_\theta^T(t_j) P_i^T.$$

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 $u \in \Gamma$, let Υ_u denote the set of output sample points, $\mathcal{X}_{\theta,u}(t)$ and $\mathcal{Y}_{\theta,u}(t)$ the state and output vectors of the derivative system Φ'. For each sample $v \in \Upsilon_u$, let $t_{v,u}$ denote the corresponding sampling instant of v, and $\sigma_{i,v,u}^2$ the noise variance at $t_{v,u}$, $i = 1, \ldots, p$. Then we have

$$I(\theta) = \sum_{u \in \Gamma} \sum_{i=1}^{p} \sum_{v \in \Upsilon_u} \frac{1}{\sigma_{i,v,u}^2} P_i \mathcal{Y}_{\theta,u}(t_{v,u}) \mathcal{Y}_{\theta,u}^T(t_{v,u}) P_i^T.$$

Since the size of $I(\theta)$ is $K \times K$, nonsingularity of $I(\theta)$ is equivalent to that the span of the vectors $P_i \mathcal{Y}_{\theta,u}(t_{v,u}),\ i = 1, \ldots, p,\ v \in \Upsilon_u$, for all $u \in \Gamma$ is of dimension K.

Let $\Omega_{\mathcal{Y}_\theta}$ denote the span of reachable outputs of Φ', and $\Omega_{P_i \mathcal{Y}_\theta}$ the span of the vectors $P_i \mathcal{Y}_\theta(t)$ for all the reachable outputs $\mathcal{Y}_\theta(t)$ of Φ', $i = 1, \ldots, p$, $t \geq t^{[0]}$. From bilinear system theory [16], the dimension of the span of any set of outputs of Φ' is no bigger than that of $\Omega_{\mathcal{Y}_\theta}$, 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 $P_i \mathcal{Y}_\theta(t)$, $t \geq t^{[0]}$, is no bigger than that of $\Omega_{P_i \mathcal{Y}_\theta}$, and there exists a finite set of admissible inputs such that the span of the corresponding $P_i \mathcal{Y}_\theta(t)$, $t \geq t^{[0]}$, is of the same dimension as that of $\Omega_{P_i \mathcal{Y}_\theta}$, $i = 1, \ldots, p$. It is then clear that the existence of a nonsingular $I(\theta)$ is equivalent to that the span of the vectors $P_i \mathcal{Y}_\theta(t)$, $i = 1, \ldots, p$, $t \geq t^{[0]}$, for all the reachable outputs $\mathcal{Y}_\theta(t)$, $t \geq t^{[0]}$, is of dimension K.

By Lemma 3.1, $\Omega_{\mathcal{Y}_\theta} = \mathrm{range}\left\{\begin{bmatrix} \mathcal{C}\mathcal{X}_0 & \mathcal{C}\mathcal{O}^{[r]} \end{bmatrix}\right\}$. Hence, $\Omega_{P_i \mathcal{Y}_\theta} = \mathrm{range}\left\{\begin{bmatrix} P_i\mathcal{C}\mathcal{X}_0 & P_i\mathcal{C}\mathcal{O}^{[r]} \end{bmatrix}\right\}$, $i = 1, \ldots, p$. Then the span of the vectors $P_i \mathcal{Y}_\theta(t)$, $i = 1, \ldots, p$, $t \geq t^{[0]}$, for all the reachable outputs $\mathcal{Y}_\theta(t)$, $t \geq t^{[0]}$, is given by

$$\Omega_{P_1 \mathcal{Y}_\theta} \oplus \Omega_{P_2 \mathcal{Y}_\theta} \oplus \ldots \oplus \Omega_{P_p \mathcal{Y}_\theta} = \begin{bmatrix} P_1\mathcal{C}\mathcal{X}_0 & P_1\mathcal{C}\mathcal{O}^{[r]} & P_2\mathcal{C}\mathcal{X}_0 & P_2\mathcal{C}\mathcal{O}^{[r]} & \ldots & P_p\mathcal{C}\mathcal{X}_0 & P_p\mathcal{C}\mathcal{O}^{[r]} \end{bmatrix}.$$

Therefore, there exists a finite set of admissible inputs and output samples such that the associated Fisher information matrix $I(\theta) \in \mathbb{R}^{K \times K}$ is nonsingular if and only if $\dim\{\Omega_{P_1 \mathcal{Y}_\theta} \oplus \Omega_{P_2 \mathcal{Y}_\theta} \oplus \ldots \oplus \Omega_{P_p \mathcal{Y}_\theta}\} = K$, i.e.,

$$\mathrm{rank}\left\{\begin{bmatrix} P_1\mathcal{C}\mathcal{X}_0 & P_1\mathcal{C}\mathcal{O}^{[r]} & P_2\mathcal{C}\mathcal{X}_0 & P_2\mathcal{C}\mathcal{O}^{[r]} & \ldots & P_p\mathcal{C}\mathcal{X}_0 & P_p\mathcal{C}\mathcal{O}^{[r]} \end{bmatrix}\right\} = K,$$

for some integer $0 \leq r \leq M - 1$. □

Theorem 3.2 provides a convenient method to check whether the unknown parameter vector can be locally identified from the output samples generated by a bilinear system model with either admissible inputs or piecewise constant inputs.

IV. CRLB FOR UNIFORMLY SAMPLED OUTPUT DATA

Although Theorem 3.2 gives a condition for the existence of some admissible inputs or piecewise constant inputs such that the unknown parameter vector can be locally identified from the corresponding output samples of a bilinear system, it does not tell us what the inputs are, nor how many inputs should be used. In practice, we are often restricted to measuring output samples of a bilinear system with just one input. Therefore, it is important, both theoretically and practically, to know whether the unknown parameter vector can be locally identified from the corresponding output samples of a bilinear system with a specific input. When the output of a bilinear system with an admissible or a piecewise constant input is measured using a nonuniform sampling scheme, the local identifiability can be tested based on the Fisher information matrix obtained by Theorem 2.1 or Corollary 2.1, respectively. However, checking the nonsingularity of the Fisher information matrix directly is computationally rather inefficient, particularly for a large number of data samples. When the output of a bilinear system with a piecewise constant input is sampled uniformly, it is possible to develop a simplified method for checking the nonsingularity of the Fisher information matrix.

Another advantage of uniformly sampling the output of a bilinear system with a piecewise constant input is that the associated Fisher information matrix and the CRLB can be computed efficiently through solving certain Lyapunov equations. Moreover, it is also possible to derive the asymptotic CRLB for infinite uniformly sampled data points.

Throughout the section we assume that all the eigenvalues of $A + F^{[l]}$, $l = 0, \ldots, L-1$, are in the open left-half plane. Since $\mathcal{A} + \mathcal{F}^{[l]}$, $l = 0, \ldots, L-1$, is a block lower triangular matrix with $A + F^{[l]}$ as its diagonal block submatrices, its eigenvalues are also in the open left-half plane, and hence the eigenvalues of $\mathcal{A}_d^{[l]} := e^{(\mathcal{A}+\mathcal{F}^{[l]})T_l}$ ($\mathcal{A}_d^{[l]}$ is of dimension $M \times M$ with $M = 2Kn$) are in the open unit disc, where T_l is the sampling period for the l^{th} interval ($t^{[l]} \leq t < t^{[l+1]}$) of the piecewise constant input, $l = 0, \ldots, L-1$.

Theorem 4.1: Consider the bilinear system represented by $\Phi = \{A,\ B,\ C,\ F_1,\ \ldots,\ F_m\}$. Assume that

1.) A, B, C, F_1, …, F_m and x_0 depend on the unknown parameter vector θ of size K, and their partial derivatives with respect to the elements of θ exist for all $\theta \in \Theta$ and are continuous functions of θ;

2.) the input u is piecewise constant as defined by (3);

3.) the output signal is uniformly sampled with the sampling period T_l in the l^{th} interval of the piecewise constant input, i.e., at

$$t_{j^{[l]}}^{[l]} = t^{[l]} + t^{[l,0]} + j^{[l]}T_l, \quad j^{[l]} = 0, \ldots, J^{[l]} - 1, \quad t^{[l]} \leq t_{j^{[l]}}^{[l]} < t^{[l+1]},$$

where $t_{j^{[l]}}^{[l]}$ denotes the ${j^{[l]}}^{th}$ sampling instant in the l^{th} interval, $t^{[l,0]}$ is the starting time relative to $t^{[l]}$ for sampling in the l^{th} interval , and $J^{[l]}$, with $J^{[l]} \geq M+1$, is the total number of samples acquired in the l^{th} interval, $l = 0, \ldots, L-1$;

4.) the noise components have independent Gaussian distributions with zero mean and variance ${\sigma_{i,j}^{[l]}}^2 = \sigma^2$, $i = 1, \ldots, p$, $j = 0, \ldots, J^{[l]} - 1$, for $l = 0, \ldots, L-1$;

5.) the derivative system of Φ is represented by $\Phi' = \{\mathcal{A},\ \mathcal{B},\ \mathcal{C},\ \mathcal{F}_1,\ \ldots,\ \mathcal{F}_m\}$;

6.) the eigenvalues of $A + F^{[l]}$, $l = 0, \ldots, L-1$, are in the open left-half plane.

Then $I(\theta)$ is nonsingular if and only if $\mathrm{rank}\left\{\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}' & \ldots & P_p\mathcal{C}\mathcal{O}' \end{array}\right]\right\} = K$, where $\mathcal{O}' \in \mathbb{R}^{M \times L(M+1)}$ is defined as $\mathcal{O}' := \left[\begin{array}{ccc} \mathcal{O}'_0 & \ldots & \mathcal{O}'_{L-1} \end{array}\right]$, and $\mathcal{O}'_0$, …, $\mathcal{O}'_{L-1}$ is given by

$$\mathcal{O}'_l := \left[\begin{array}{llll} \mathcal{W}^{[l]} & \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} (\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})) & \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{A}_d^{[l]} (\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})) & \ldots \\ & \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}_d^{[l]}\right)^{M-1} (\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})) & & \end{array}\right], \quad l = 0, \ldots, L-1.$$

Proof: The assumption that all the eigenvalues of $A + F^{[l]}$, $l = 0, \ldots, L-1$, are in the open left-half plane implies that none of the eigenvalues of $\mathcal{A}_d^{[l]}$ is equal to one. By Corollary 2.1, the Fisher information matrix $I(\theta)$ is given by

$$I(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^{p} \sum_{l=0}^{L-1} \sum_{j^{[l]}=0}^{J^{[l]}-1} P_i \mathcal{Y}_\theta(t_{j^{[l]}}^{[l]}) \mathcal{Y}_\theta^T(t_{j^{[l]}}^{[l]}) P_i^T.$$

$I(\theta)$ being nonsingular is equivalent to that the subspace spanned by all the vectors $P_i\mathcal{Y}_\theta(t_{j^{[l]}}^{[l]})$, $i = 1, \ldots, p$, $j^{[l]} = 0, \ldots, J^{[l]} - 1$, for $l = 0, \ldots, L-1$ is of full rank. We also have $\mathcal{Y}_\theta(t_{j^{[l]}}^{[l]}) = \mathcal{C}\mathcal{X}_\theta(t_{j^{[l]}}^{[l]})$, $j^{[l]} = 0, \ldots, J^{[l]} - 1$, for $l = 0, \ldots, L-1$. Let $\bar{\mathcal{O}} := \left[\begin{array}{ccc} \bar{\mathcal{O}}_0 & \ldots & \bar{\mathcal{O}}_{L-1} \end{array}\right]$, where for $l = 0, \ldots, L-1$, $\bar{\mathcal{O}}_l := \left[\begin{array}{ccc} \mathcal{X}_\theta(t_0^{[l]}) & \ldots & \mathcal{X}_\theta(t_{J^{[l]}-1}^{[l]}) \end{array}\right]$ and by Lemma 2.1

$$\mathcal{X}_\theta(t_{j^{[l]}}^{[l]}) = \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}_d^{[l]}\right)^{j^{[l]}} \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right) - \mathcal{W}^{[l]}, \quad j^{[l]} = 0, \ldots, J^{[l]} - 1.$$

Thus, $I(\theta)$ is nonsingular if and only if matrix $\left[\begin{array}{ccc} P_1\mathcal{C}\bar{\mathcal{O}} & \ldots & P_p\mathcal{C}\bar{\mathcal{O}} \end{array}\right]$ is of full rank, i.e., $\mathrm{rank}\left\{\left[\begin{array}{ccc} P_1\mathcal{C}\bar{\mathcal{O}} & \ldots & P_p\mathcal{C}\bar{\mathcal{O}} \end{array}\right]\right\} = K$.

It remains to show that $\mathrm{rank}\left\{\left[\begin{array}{ccc} P_1\mathcal{C}\bar{\mathcal{O}} & \ldots & P_p\mathcal{C}\bar{\mathcal{O}} \end{array}\right]\right\} = K$ is equivalent to $\mathrm{rank}\left\{\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}' & \ldots & P_p\mathcal{C}\mathcal{O}' \end{array}\right]\right\} = K$, for which it is sufficient to show $\mathrm{range}\left\{\bar{\mathcal{O}}\right\} = \mathrm{range}\left\{\mathcal{O}'\right\}$.

Considering the submatrix $\bar{\mathcal{O}}_l$ of $\bar{\mathcal{O}}$ and the submatrix $\mathcal{O}'_l$ ($l = 0, \ldots, L-1$ throughout the remainder of the proof) of $\mathcal{O}'$, we have $\bar{\mathcal{O}}_l = \mathcal{O}'_l V_l$, where $V_l \in \mathbb{R}^{(M+1)\times J^{[l]}}$ is given by

$$V_l = \begin{bmatrix} -1 & -1 & \ldots & -1 & -1 & -1 & -1 & -1 & \ldots & -1 \\ 1 & 0 & \ldots & 0 & 0 & -\alpha_{l,M} & \beta^{[1]}_{l,M} & \beta^{[2]}_{l,M} & \ldots & \beta^{[J^{[l]}-M-1]}_{l,M} \\ 0 & 1 & \ldots & 0 & 0 & -\alpha_{l,M-1} & \beta^{[1]}_{l,M-1} & \beta^{[2]}_{l,M-1} & \ldots & \beta^{[J^{[l]}-M-1]}_{l,M-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & -\alpha_{l,2} & \beta^{[1]}_{l,2} & \beta^{[2]}_{l,2} & \ldots & \beta^{[J^{[l]}-M-1]}_{l,2} \\ 0 & 0 & \ldots & 0 & 1 & -\alpha_{l,1} & \beta^{[1]}_{l,1} & \beta^{[2]}_{l,1} & \ldots & \beta^{[J^{[l]}-M-1]}_{l,1} \end{bmatrix}.$$

Here, $\alpha_{l,s}$, $s = 1, \ldots, M$, is given by

$$\det(\lambda I - \mathcal{A}^{[l]}_d) = \lambda^M + \alpha_{l,1}\lambda^{M-1} + \ldots + \alpha_{l,M-1}\lambda + \alpha_{l,M}.$$

By the Cayley-Hamilton theorem, $(\mathcal{A}^{[l]}_d)^M + \alpha_{l,1}(\mathcal{A}^{[l]}_d)^{M-1} + \ldots + \alpha_{l,M-1}\mathcal{A}^{[l]}_d + \alpha_{l,M} = 0$. We then have

$$\begin{aligned} \mathcal{X}_\theta(t^{[l]}_M) = &\left(\mathcal{A}^{[l]}_d\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}^{[l]}_d\right)^M \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right) - \mathcal{W}^{[l]} \\ = &- \mathcal{W}^{[l]} - \alpha_{l,M}\left(\mathcal{A}^{[l]}_d\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right) - \alpha_{l,M-1}\left(\mathcal{A}^{[l]}_d\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{A}^{[l]}_d \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right) - \ldots \\ &- \alpha_{l,1}\left(\mathcal{A}^{[l]}_d\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}^{[l]}_d\right)^{M-1} \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right). \end{aligned}$$

By induction, we can always find $\beta^{[j''-M]}_{l,s}$, $s = 1, \ldots, M$, $j'' = M+1, \ \ldots, J^{[l]}-1$, such that

$$\begin{aligned} \mathcal{X}_\theta(t^{[l]}_{j''}) = &- \mathcal{W}^{[l]} - \beta^{[j''-M]}_{l,M}\left(\mathcal{A}^{[l]}_d\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{W}^{[l]} - \beta^{[j''-M]}_{l,M-1}\left(\mathcal{A}^{[l]}_d\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{A}^{[l]}_d \mathcal{W}^{[l]} - \ldots \\ &- \beta^{[j''-M]}_{l,1}\left(\mathcal{A}^{[l]}_d\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}^{[l]}_d\right)^{M-1} \mathcal{W}^{[l]}. \end{aligned}$$

Consider now a submatrix of V_l given by

$$V'_l = \begin{bmatrix} -1 & -1 & \ldots & -1 & -1 & -1 \\ 1 & 0 & \ldots & 0 & 0 & -\alpha_{l,M} \\ 0 & 1 & \ldots & 0 & 0 & -\alpha_{l,M-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & -\alpha_{l,2} \\ 0 & 0 & \ldots & 0 & 1 & -\alpha_{l,1} \end{bmatrix}.$$

Its determinant is

$$\det(V'_l) = (-1)^{M+1}(1 + \alpha_{l,1} + \ldots + \alpha_{l,M}) = (-1)^{M+1} \det(\lambda I - \mathcal{A}^{[l]}_d)\Big|_{\lambda=1} \neq 0,$$

since none of the eigenvalues of $\mathcal{A}_d^{[l]}$ is equal to one. Therefore, V_l is of full rank, and $\mathrm{range}\left\{\bar{\mathcal{O}}_l\right\} = \mathrm{range}\left\{\mathcal{O}'_l\right\}$, $l = 0, \ldots, L-1$. Hence, $\mathrm{range}\left\{\bar{\mathcal{O}}\right\} = \mathrm{range}\left\{\mathcal{O}'\right\}$. □

When the bilinear system in Theorem 4.1 is locally identifiable, the next step is to calculate its associated Fisher information matrix and the CRLB. Although the Fisher information matrix could be calculated using Corollary 2.1 in Section II, it is computationally rather inefficient to directly compute the summations in (18), particularly when the number of samples is large. We now propose an alternative method for computing the Fisher information matrix efficiently through solutions to certain Lyapunov equations in the following theorem. Standard results on Lyapunov equations can be found in [23], [24].

Theorem 4.2: Assume that the data model and all the assumptions are the same as in Theorem 4.1. Then the Fisher information matrix for the given data set is

$$\begin{aligned} I(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^{p} P_i \mathcal{C} \Bigg\{ \sum_{l=0}^{L-1} \Bigg[& \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{P}_1^{[l]} \left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \right)^T - \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{P}_2^{[l]} \\ & - \left(\mathcal{P}_2^{[l]}\right)^T \left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \right)^T + J^{[l]} \mathcal{W}^{[l]} \left(\mathcal{W}^{[l]}\right)^T \Bigg] \Bigg\} \mathcal{C}^T P_i^T, \end{aligned} \tag{27}$$

where $\mathcal{P}_1^{[l]}$ and $\mathcal{P}_2^{[l]}$ are obtained as follows.

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

$$\begin{aligned} \mathcal{A}_d^{[l]} \mathcal{P}_1^{[l]} \left(\mathcal{A}_d^{[l]}\right)^T - \mathcal{P}_1^{[l]} = & - \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right)^T \\ & + \left(\mathcal{A}_d^{[l]}\right)^{J^{[l]}} \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right)^T \left(\left(\mathcal{A}_d^{[l]}\right)^{J^{[l]}} \right)^T. \end{aligned}$$

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

$$\mathcal{P}_2^{[l]} = \left(I - \left(\mathcal{A}_d^{[l]}\right)^{J^{[l]}} \right) \left(I - \mathcal{A}_d^{[l]} \right)^{-1} \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]}\right)^T + \mathcal{X}_\theta(t^{[l]}) \left(\mathcal{W}^{[l]}\right)^T \right).$$

Proof: By the uniform sampling assumption and with $\mathcal{A}_d^{[l]} = e^{(\mathcal{A}+\mathcal{F}^{[l]})T_l}$, $l = 0, \ldots, L-1$, the Fisher information matrix in Corollary 2.1 can be rewritten as

$$\begin{aligned} I(\theta) = & \frac{1}{\sigma^2} \sum_{i=1}^{p} P_i \mathcal{C} \Bigg\{ \sum_{l=0}^{L-1} \Bigg[\sum_{j^{[l]}=0}^{J^{[l]}-1} \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}_d^{[l]}\right)^{j^{[l]}} \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right)^T \\ & \cdot \left(\left(\mathcal{A}_d^{[l]}\right)^{j^{[l]}} \right)^T \left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \right)^T \\ & - \sum_{j^{[l]}=0}^{J^{[l]}-1} \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}_d^{[l]}\right)^{j^{[l]}} \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]}\right)^T + \mathcal{X}_\theta(t^{[l]}) \left(\mathcal{W}^{[l]}\right)^T \right) \end{aligned}$$

$$
- \sum_{j^{[l]}=0}^{J^{[l]}-1} \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T + \mathcal{W}^{[l]} \mathcal{X}_\theta^T (t^{[l]}) \right) \left(\left(\mathcal{A}_d^{[l]} \right)^{j^{[l]}} \right)^T \left(\left(\mathcal{A}_d^{[l]} \right)^{\frac{t^{[l,0]}}{T_l}} \right)^T
$$

$$
+ \sum_{j^{[l]}=0}^{J^{[l]}-1} \mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T \Bigg] \Bigg\} \mathcal{C}^T P_i^T .
$$

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

$$
\mathcal{P}_1^{[l]} := \sum_{j^{[l]}=0}^{J^{[l]}-1} \left(\mathcal{A}_d^{[l]} \right)^{j^{[l]}} \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta (t^{[l]}) \right) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta (t^{[l]}) \right)^T \left(\left(\mathcal{A}_d^{[l]} \right)^{j^{[l]}} \right)^T ,
$$

$$
\mathcal{P}_2^{[l]} := \sum_{j^{[l]}=0}^{J^{[l]}-1} \left(\mathcal{A}_d^{[l]} \right)^{j^{[l]}} \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T + \mathcal{X}_\theta (t^{[l]}) \left(\mathcal{W}^{[l]} \right)^T \right) .
$$

We then have

$$
I(\theta) = \frac{1}{\sigma^2} \sum_{i=1}^{p} P_i \mathcal{C} \left\{ \sum_{l=0}^{L-1} \left[\left(\mathcal{A}_d^{[l]} \right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{P}_1^{[l]} \left(\left(\mathcal{A}_d^{[l]} \right)^{\frac{t^{[l,0]}}{T_l}} \right)^T - \left(\mathcal{A}_d^{[l]} \right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{P}_2^{[l]} \right. \right.
$$

$$
\left. \left. - \left(\mathcal{P}_2^{[l]} \right)^T \left(\left(\mathcal{A}_d^{[l]} \right)^{\frac{t^{[l,0]}}{T_l}} \right)^T + J^{[l]} \mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T \right] \right\} \mathcal{C}^T P_i^T .
$$

Since all the eigenvalues of $\mathcal{A}_d^{[l]}$, $l = 0, \ldots, L-1$, are in the open unit disc, $\mathcal{P}_1^{[l]}$, $l = 0, \ldots, L-1$, is the unique solution to the following Lyapunov equation

$$
\begin{aligned}
\mathcal{A}_d^{[l]} \mathcal{P}_1^{[l]} \left(\mathcal{A}_d^{[l]} \right)^T - \mathcal{P}_1^{[l]} = & - \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta (t^{[l]}) \right) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta (t^{[l]}) \right)^T \\
& + \left(\mathcal{A}_d^{[l]} \right)^{J^{[l]}} \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta (t^{[l]}) \right) \left(\mathcal{W}^{[l]} + \mathcal{X}_\theta (t^{[l]}) \right)^T \left(\left(\mathcal{A}_d^{[l]} \right)^{J^{[l]}} \right)^T .
\end{aligned}
$$

As $\sum_{j^{[l]}=0}^{J^{[l]}-1} \left(\mathcal{A}_d^{[l]} \right)^{j^{[l]}} = \left(I - \left(\mathcal{A}_d^{[l]} \right)^{J^{[l]}} \right) \left(I - \mathcal{A}_d^{[l]} \right)^{-1}$, $\mathcal{P}_2^{[l]}$ is given by

$$
\mathcal{P}_2^{[l]} = \left(I - \left(\mathcal{A}_d^{[l]} \right)^{J^{[l]}} \right) \left(I - \mathcal{A}_d^{[l]} \right)^{-1} \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T + \mathcal{X}_\theta (t^{[l]}) \left(\mathcal{W}^{[l]} \right)^T \right), \quad l = 0, \ldots, L-1. \qquad \square
$$

If data samples are finite and the nonsingularity conditions in Theorem 4.1 hold, the CRLB can be calculated easily by inverting $I(\theta)$ in Theorem 4.2. However, when the number of data samples approaches infinity, i.e., $J^{[l]} \to \infty$, in general, $I(\theta)$ cannot be computed in this way since the term $J^{[l]} P_i \mathcal{C} \mathcal{W}^{[l]} \left(\mathcal{W}^{[l]} \right)^T \mathcal{C}^T P_i^T$ in (27) can tend to infinity, $l = 0, \ldots, L-1$. In Theorem 4.4 we will propose a novel method for calculating the asymptotic CRLB without computing $I(\theta)$ directly. Before proceeding, we extend Theorem 4.1 to the case of infinite data samples.

Theorem 4.3: Assume that the data model is the same as in Theorem 4.1 and the number of equidistant samples in each $[t^{[l]},\ t^{[l+1]})$ is equal, i.e., $J^{[l]} = J'$, $l = 0, \ldots, L-1$. The dimension of $\mathcal{A}_d^{[l]}$, $l = 0, \ldots, L-1$, is $M \times M$ with $M = 2Kn$. Then there exists a positive integer J_0 such that for any $J' > J_0$ the associated Fisher information matrix $I_{J'}(\theta)$ is nonsingular if

$$\operatorname{rank}\left\{\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'' & \dots & P_p\mathcal{C}\mathcal{O}'' \end{array}\right]\right\} = K,$$

where $\mathcal{O}'' \in \mathbb{R}^{M\times L(M+1)}$ is defined as

$$\mathcal{O}'' := \left[\begin{array}{ccc} \mathcal{O}''_0 & \dots & \mathcal{O}''_{L-1} \end{array}\right],$$

and $\mathcal{O}''_0$, ..., $\mathcal{O}''_{L-1}$ are given by

$$\begin{aligned} \mathcal{O}''_0 := & \left[\mathcal{W}^{[0]} \quad \left(\mathcal{A}_d^{[0]}\right)^{\frac{t^{[0,0]}}{T_0}} (\mathcal{W}^{[0]} + \mathcal{X}_0) \quad \left(\mathcal{A}_d^{[0]}\right)^{\frac{t^{[0,0]}}{T_0}} \mathcal{A}_d^{[0]} (\mathcal{W}^{[0]} + \mathcal{X}_0) \quad \dots \right. \\ & \left. \left(\mathcal{A}_d^{[0]}\right)^{\frac{t^{[0,0]}}{T_0}} \left(\mathcal{A}_d^{[0]}\right)^{M-1} (\mathcal{W}^{[0]} + \mathcal{X}_0) \right], \\ \mathcal{O}''_l := & \left[\mathcal{W}^{[l]} \quad \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} (\mathcal{W}^{[l]} - \mathcal{W}^{[l-1]}) \quad \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{A}_d^{[l]} (\mathcal{W}^{[l]} - \mathcal{W}^{[l-1]}) \quad \dots \right. \\ & \left. \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}_d^{[l]}\right)^{M-1} (\mathcal{W}^{[l]} - \mathcal{W}^{[l-1]}) \right], \ l = 1, \dots, L-1. \end{aligned}$$

Proof: For any given $J' \geq M+1$, by Theorem 4.1 the associated Fisher information matrix $I_{J'}(\theta)$ is nonsingular if and only if $\operatorname{rank}\left\{\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'_{J'} & \dots & P_p\mathcal{C}\mathcal{O}'_{J'} \end{array}\right]\right\} = K$, where $\mathcal{O}'_{J'} := \left[\begin{array}{ccc} \mathcal{O}'_{0,J'} & \dots & \mathcal{O}'_{L-1,J'} \end{array}\right]$ and

$$\begin{aligned} \mathcal{O}'_{l,J'} := & \left[\mathcal{W}^{[l]} \quad \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} (\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})) \quad \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{A}_d^{[l]} (\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})) \quad \dots \right. \\ & \left. \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \left(\mathcal{A}_d^{[l]}\right)^{M-1} (\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})) \right], \quad l = 0, \dots, L-1. \end{aligned}$$

Note that we use subscript J' in $\mathcal{O}'_{l,J'}$ as $\mathcal{X}_\theta(t^{[l]})$ may depend on J'. It is easy to see that $\mathcal{O}'_{0,J'} = \mathcal{O}''_0$ for all J'. When $J' \to \infty$, we have $t^{[l]} \to \infty$ for $l = 1, \dots, L$. By Corollary 2.1 with $\lim_{t^{[l]}\to\infty} \mathcal{Q}_{l-1}(t^{[l]}) = 0$, $\lim_{t^{[l]}\to\infty} \mathcal{X}_\theta(t^{[l]}) = -\mathcal{W}^{[l-1]}$ for $l = 1, \dots, L-1$. Thus, $\lim_{J'\to\infty} \mathcal{O}'_{l,J'} = \mathcal{O}''_l$ for $l = 1, \dots, L-1$. Hence,

$$\lim_{J'\to\infty} \left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'_{J'} & \dots & P_p\mathcal{C}\mathcal{O}'_{J'} \end{array}\right] = \left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'' & \dots & P_p\mathcal{C}\mathcal{O}'' \end{array}\right],$$

which implies that for any $\epsilon > 0$ there exists a J_ϵ such that

$$\left\|\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'_{J'} & \dots & P_p\mathcal{C}\mathcal{O}'_{J'} \end{array}\right] - \left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'' & \dots & P_p\mathcal{C}\mathcal{O}'' \end{array}\right]\right\| < \epsilon$$

for all $J' > J_\epsilon$, where $\|\cdot\|$ denotes the norm of a matrix. Therefore, if $\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'' & \dots & P_p\mathcal{C}\mathcal{O}'' \end{array}\right]$ is of full rank, i.e., $\operatorname{rank}\left\{\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'' & \dots & P_p\mathcal{C}\mathcal{O}'' \end{array}\right]\right\} = K$, there exists a positive integer J_0 such that $\operatorname{rank}\{\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{O}'_{J'} & \dots & P_p\mathcal{C}\mathcal{O}'_{J'} \end{array}\right]\} = K$ for all $J' > J_0$, which implies $I_{J'}(\theta)$ is nonsingular for all $J' > J_0$.

$\square$

Given the nonsingularity condition 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.

Theorem 4.4: Assume that the data model is the same as in Theorem 4.1, 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 Theorem 4.3 holds. Then the asymptotic CRLB is given by

$$\mathrm{var}(\hat{\theta}) \geq \lim_{J' \to \infty} I_{J'}^{-1}(\theta) = \begin{cases} \sigma^2 U^{\perp} \left(\left(U^{\perp}\right)^T \mathcal{P} U^{\perp} \right)^{-1} (U^{\perp})^T, & \text{if } \mathrm{rank}(U) < K, \\ 0, & \text{if } \mathrm{rank}(U) = K, \end{cases}$$

where $\mathcal{P}$, U and $U^{\perp}$ are defined as follows:

i.) Construction of $\mathcal{P}$:

$$\mathcal{P} := \sum_{i=1}^{p} P_i \mathcal{C} \left\{ \sum_{l=0}^{L-1} \left[\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{P}_1^{[l]} \left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \right)^T - \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \mathcal{P}_2^{[l]} - \left(\mathcal{P}_2^{[l]}\right)^T \left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}} \right)^T \right] \right\} \cdot \mathcal{C}^T P_i^T,$$

where $\mathcal{P}_1^{[0]}$ is the unique solution to the following Lyapunov equation

$$\mathcal{A}_d^{[0]} \mathcal{P}_1^{[0]} \left(\mathcal{A}_d^{[0]}\right)^T - \mathcal{P}_1^{[0]} = -\left(\mathcal{W}^{[0]} + \mathcal{X}_0\right) \left(\mathcal{W}^{[0]} + \mathcal{X}_0\right)^T,$$

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

$$\mathcal{A}_d^{[l]} \mathcal{P}_1^{[l]} \left(\mathcal{A}_d^{[l]}\right)^T - \mathcal{P}_1^{[l]} = -\left(\mathcal{W}^{[l]} - \mathcal{W}^{[l-1]}\right) \left(\mathcal{W}^{[l]} - \mathcal{W}^{[l-1]}\right)^T,$$

$\mathcal{P}_2^{[0]}$ is given by

$$\mathcal{P}_2^{[0]} = \left(I - \mathcal{A}_d^{[0]}\right)^{-1} \left(\mathcal{W}^{[0]} \left(\mathcal{W}^{[0]}\right)^T + \mathcal{X}_0 \left(\mathcal{W}^{[0]}\right)^T \right),$$

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

$$\mathcal{P}_2^{[l]} = \left(I - \mathcal{A}_d^{[l]}\right)^{-1} \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]}\right)^T - \mathcal{W}^{[l-1]} \left(\mathcal{W}^{[l]}\right)^T \right).$$

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

$$U^T U^{\perp} = 0 \quad \text{and} \quad \mathrm{rank}\left\{ \begin{bmatrix} U & U^{\perp} \end{bmatrix} \right\} = K.$$

Proof: With $J^{[l]} = J'$ for $l = 0, \ldots, L-1$, the Fisher information matrix in Theorem 4.2 can be rewritten in terms of J' as

$$I_{J'}(\theta) = \frac{1}{\sigma^2}\left\{\sum_{i=1}^{p} P_i\mathcal{C}\left\{\sum_{l=0}^{L-1}\left[\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\mathcal{P}_{1,J'}^{[l]}\left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\right)^T - \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\mathcal{P}_{2,J'}^{[l]}\right.\right.\right.$$
$$\left.\left.\left. - \left(\mathcal{P}_{2,J'}^{[l]}\right)^T\left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\right)^T\right]\right\}\mathcal{C}^T P_i^T + J'\sum_{i=1}^{p}\sum_{l=0}^{L-1} P_i\mathcal{C}\mathcal{W}^{[l]}\left(\mathcal{W}^{[l]}\right)^T\mathcal{C}^T P_i^T\right\},$$

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

$$\begin{aligned}\mathcal{A}_d^{[l]}\mathcal{P}_{1,J'}^{[l]}\left(\mathcal{A}_d^{[l]}\right)^T - \mathcal{P}_{1,J'}^{[l]} = &-\left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right)\left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right)^T \\ &+\left(\mathcal{A}_d^{[l]}\right)^{J'}\left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right)\left(\mathcal{W}^{[l]} + \mathcal{X}_\theta(t^{[l]})\right)^T\left(\left(\mathcal{A}_d^{[l]}\right)^{J'}\right)^T,\end{aligned} \tag{28}$$

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

$$\mathcal{P}_{2,J'}^{[l]} = \left(I - \left(\mathcal{A}_d^{[l]}\right)^{J'}\right)\left(I - \mathcal{A}_d^{[l]}\right)^{-1}\left(\mathcal{W}^{[l]}\left(\mathcal{W}^{[l]}\right)^T + \mathcal{X}_\theta(t^{[l]})\left(\mathcal{W}^{[l]}\right)^T\right).$$

Let

$$\begin{aligned}\mathcal{P}_{J'} := &\sum_{i=1}^{p} P_i\mathcal{C}\left\{\sum_{l=0}^{L-1}\left[\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\mathcal{P}_{1,J'}^{[l]}\left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\right)^T - \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\mathcal{P}_{2,J'}^{[l]} - \left(\mathcal{P}_{2,J'}^{[l]}\right)^T\right.\right. \\ &\left.\left.\cdot\left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\right)^T\right]\right\}\mathcal{C}^T P_i^T.\end{aligned}$$

Then

$$I_{J'}(\theta) = \frac{1}{\sigma^2}\left(\mathcal{P}_{J'} + J'\sum_{i=1}^{p}\sum_{l=0}^{L-1} P_i\mathcal{C}\mathcal{W}^{[l]}\left(\mathcal{W}^{[l]}\right)^T\mathcal{C}^T P_i^T\right).$$

By Theorem 4.3, $I_{J'}(\theta)$ is nonsingular for all $J' > J_0$, which in turn implies that $I_{J'}(\theta)$ is positive definite as the Fisher information matrix is always positive semidefinite. In the remainder of the proof, we assume $J' > J_0$. When $J' \to \infty$, the limit of $\mathcal{P}_{J'}$ exists and is given by

$$\begin{aligned}\mathcal{P} := &\lim_{J'\to\infty}\mathcal{P}_{J'} \\ = &\sum_{i=1}^{p} P_i\mathcal{C}\left\{\sum_{l=0}^{L-1}\left[\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\mathcal{P}_1^{[l]}\left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\right)^T - \left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\mathcal{P}_2^{[l]} - \left(\mathcal{P}_2^{[l]}\right)^T\left(\left(\mathcal{A}_d^{[l]}\right)^{\frac{t^{[l,0]}}{T_l}}\right)^T\right]\right\} \\ &\cdot\mathcal{C}^T P_i^T,\end{aligned}$$

where $\mathcal{P}_1^{[l]} := \lim_{J'\to\infty} \mathcal{P}_{1,J'}^{[l]}$ and $\mathcal{P}_2^{[l]} := \lim_{J'\to\infty} \mathcal{P}_{2,J'}^{[l]}$, $l = 0, \dots, L-1$. By Corollary 2.1 with $\lim_{t^{[l]}\to\infty} \mathcal{Q}_{l-1}(t^{[l]}) = 0$, $\lim_{t^{[l]}\to\infty} \mathcal{X}_\theta(t^{[l]}) = -\mathcal{W}^{[l-1]}$ for $l = 1, \dots, L-1$. Using $\lim_{J'\to\infty} \left(\mathcal{A}_d^{[l]}\right)^{J'} = 0$, $l = 0, \dots, L-1$, with (28), $\mathcal{P}_1^{[0]}$ is the unique solution to the following Lyapunov equation

$$\mathcal{A}_d^{[0]} \mathcal{P}_1^{[0]} \left(\mathcal{A}_d^{[0]}\right)^T - \mathcal{P}_1^{[0]} = -\left(\mathcal{W}^{[0]} + \mathcal{X}_0\right)\left(\mathcal{W}^{[0]} + \mathcal{X}_0\right)^T,$$

and $\mathcal{P}_1^{[l]}$, $l = 1, \dots, L-1$, is the unique solution to the following Lyapunov equation

$$\mathcal{A}_d^{[l]} \mathcal{P}_1^{[l]} \left(\mathcal{A}_d^{[l]}\right)^T - \mathcal{P}_1^{[l]} = -\left(\mathcal{W}^{[l]} - \mathcal{W}^{[l-1]}\right)\left(\mathcal{W}^{[l]} - \mathcal{W}^{[l-1]}\right)^T.$$

Similarly,

$$\mathcal{P}_2^{[0]} = \left(I - \mathcal{A}_d^{[0]}\right)^{-1} \left(\mathcal{W}^{[0]} \left(\mathcal{W}^{[0]}\right)^T + \mathcal{X}_0 \left(\mathcal{W}^{[0]}\right)^T\right),$$

and

$$\mathcal{P}_2^{[l]} = \left(I - \mathcal{A}_d^{[l]}\right)^{-1} \left(\mathcal{W}^{[l]} \left(\mathcal{W}^{[l]}\right)^T - \mathcal{W}^{[l-1]} \left(\mathcal{W}^{[l]}\right)^T\right), \quad l = 1, \dots, L-1.$$

For $J' \to \infty$, although $I_{J'}(\theta)$ tends to infinity, the inverse of $I_{J'}(\theta)$ still converges, as will be shown in the following. Using a singular value decomposition, $\sum_{i=1}^p \sum_{l=0}^{L-1} P_i \mathcal{C} \mathcal{W}^{[l]} \left(\mathcal{W}^{[l]}\right)^T \mathcal{C}^T P_i^T$ can be expressed as

$$\sum_{i=1}^p \sum_{l=0}^{L-1} P_i \mathcal{C} \mathcal{W}^{[l]} \left(\mathcal{W}^{[l]}\right)^T \mathcal{C}^T P_i^T = \begin{bmatrix} U_s & U_s^\perp \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_s & U_s^\perp \end{bmatrix}^T, \tag{29}$$

where $\Sigma \in \mathbb{R}^{N\times N}$ is diagonal with positive diagonal entries, $U_s \in \mathbb{R}^{K\times N}$, $U_s^\perp \in \mathbb{R}^{K\times(K-N)}$, and $\begin{bmatrix} U_s & U_s^\perp \end{bmatrix}$ is orthogonal. By substitution,

$$\begin{aligned} I_{J'}^{-1}(\theta) &= \sigma^2 \left(\mathcal{P}_{J'} + J' \begin{bmatrix} U_s & U_s^\perp \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_s & U_s^\perp \end{bmatrix}^T\right)^{-1} \\ &= \sigma^2 \begin{bmatrix} U_s & U_s^\perp \end{bmatrix} \left(\begin{bmatrix} U_s & U_s^\perp \end{bmatrix}^T \mathcal{P}_{J'} \begin{bmatrix} U_s & U_s^\perp \end{bmatrix} + J' \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}\right)^{-1} \begin{bmatrix} U_s & U_s^\perp \end{bmatrix}^T \\ &= \sigma^2 \begin{bmatrix} U_s & U_s^\perp \end{bmatrix} \begin{bmatrix} U_s^T \mathcal{P}_{J'} U_s + J'\Sigma & U_s^T \mathcal{P}_{J'} U_s^\perp \\ \left(U_s^\perp\right)^T \mathcal{P}_{J'} U_s & \left(U_s^\perp\right)^T \mathcal{P}_{J'} U_s^\perp \end{bmatrix}^{-1} \begin{bmatrix} U_s & U_s^\perp \end{bmatrix}^T. \end{aligned}$$

Consider first the case when $N = K$, i.e., U_s has a rank of K and $U_s^\perp$ diminishes in (29). In this case, the asymptotic CRLB is given by

$$\begin{aligned} \mathrm{var}(\hat{\theta}) \geq \lim_{J'\to\infty} I_{J'}^{-1}(\theta) &= \sigma^2 U_s \left[\lim_{J'\to\infty} \left(U_s^T \mathcal{P}_{J'} U_s + J'\Sigma\right)^{-1}\right] U_s^T \\ &= \sigma^2 U_s \left[\lim_{J'\to\infty} \frac{1}{J'}\right] \left[\lim_{J'\to\infty} \left(\frac{1}{J'} U_s^T \mathcal{P}_{J'} U_s + \Sigma\right)^{-1}\right] U_s^T \\ &= 0, \end{aligned}$$

since $\lim_{J'\to\infty}\left(\frac{1}{J'}U_s^T\mathcal{P}_{J'}U_s\right)=0$ and Σ is of full rank.

Next, consider the case when $N<K$. Let $Z_1 := U_s^T\mathcal{P}_{J'}U_s + J'\Sigma,\ Z_2 := U_s^T\mathcal{P}_{J'}U_s^\perp,\ Z_3 := \left(U_s^\perp\right)^T\mathcal{P}_{J'}U_s^\perp$, and $\Delta := Z_3 - Z_2^T Z_1^{-1} Z_2$ (Δ is called the Schur complement of Z_1). That $I_{J'}(\theta)$ is positive definite implies $\begin{bmatrix} Z_1 & Z_2 \\ Z_2^T & Z_3 \end{bmatrix}$ is also positive definite. Then Z_1 and Δ are positive definite (see Theorem 7.7.6 in [25]). Using the formula of the inverse of block matrices [23],

$$\begin{bmatrix} U_s^T\mathcal{P}_{J'}U_s + J'\Sigma & U_s^T\mathcal{P}_{J'}U_s^\perp \\ \left(U_s^\perp\right)^T\mathcal{P}_{J'}U_s & \left(U_s^\perp\right)^T\mathcal{P}_{J'}U_s^\perp \end{bmatrix}^{-1} = \begin{bmatrix} Z_1^{-1} + Z_1^{-1}Z_2\Delta^{-1}Z_2^TZ_1^{-1} & -Z_1^{-1}Z_2\Delta^{-1} \\ -\Delta^{-1}Z_2^TZ_1^{-1} & \Delta^{-1} \end{bmatrix}.$$

For $J'\to\infty$, $\lim_{J'\to\infty} Z_1^{-1} = \lim_{J'\to\infty}\left(U_s^T\mathcal{P}_{J'}U_s + J'\Sigma\right)^{-1} = \lim_{J'\to\infty}\left(U_s^T\mathcal{P}U_s + J'\Sigma\right)^{-1} = 0$. Since $I_{J'}(\theta)$ is positive definite, for any nonzero vector $b\in\mathbb{R}^{(K-N)\times 1}$

$$b^T\left(U_s^\perp\right)^T\mathcal{P}_{J'}U_s^\perp b = \sigma^2 b^T\left(U_s^\perp\right)^T I_{J'}(\theta)U_s^\perp b > 0,$$

which shows that $\lim_{J'\to\infty}\Delta = \lim_{J'\to\infty}\left(U_s^\perp\right)^T\mathcal{P}_{J'}U_s^\perp = \left(U_s^\perp\right)^T\mathcal{P}U_s^\perp$ is positive definite. Therefore,

$$\lim_{J'\to\infty}\begin{bmatrix} U_s^T\mathcal{P}_{J'}U_s + J'\Sigma & U_s^T\mathcal{P}_{J'}U_s^\perp \\ \left(U_s^\perp\right)^T\mathcal{P}_{J'}U_s & \left(U_s^\perp\right)^T\mathcal{P}_{J'}U_s^\perp \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \left(\left(U_s^\perp\right)^T\mathcal{P}U_s^\perp\right)^{-1} \end{bmatrix}.$$

The asymptotic CRLB is then given in terms of U_s as

$$\begin{aligned}\operatorname{var}(\hat{\theta}) \geq \lim_{J'\to\infty} I_{J'}^{-1}(\theta) &= \sigma^2\begin{bmatrix} U_s & U_s^\perp \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & \left(\left(U_s^\perp\right)^T\mathcal{P}U_s^\perp\right)^{-1} \end{bmatrix}\begin{bmatrix} U_s & U_s^\perp \end{bmatrix}^T \\ &= \sigma^2 U_s^\perp\left(\left(U_s^\perp\right)^T\mathcal{P}U_s^\perp\right)^{-1}\left(U_s^\perp\right)^T.\end{aligned}$$

Finally, we show that $U_s^\perp\left(\left(U_s^\perp\right)^T\mathcal{P}U_s^\perp\right)^{-1}\left(U_s^\perp\right)^T = U^\perp\left(\left(U^\perp\right)^T\mathcal{P}U^\perp\right)^{-1}\left(U^\perp\right)^T$. As $\operatorname{range}\left\{U^\perp\right\} = \operatorname{range}\left\{U_s^\perp\right\}$, there exists a nonsingular matrix $V\in\mathbb{R}^{(K-N)\times(K-N)}$ such that $U_s^\perp = U^\perp V$. It follows that

$$\begin{aligned}U_s^\perp\left(\left(U_s^\perp\right)^T\mathcal{P}U_s^\perp\right)^{-1}\left(U_s^\perp\right)^T &= U^\perp V\left(\left(U^\perp V\right)^T\mathcal{P}U^\perp V\right)^{-1}\left(U^\perp V\right)^T \\ &= U^\perp\left(\left(U^\perp\right)^T\mathcal{P}U^\perp\right)^{-1}\left(U^\perp\right)^T. \qquad \square\end{aligned}$$

In the next section we illustrate the theoretical results presented in the previous sections using an example from surface plasmon resonance experiments for the determination of the kinetic parameters of protein-protein interactions.

V. EXAMPLE

Surface plasmon resonance (SPR) (see, e.g. [26], [27]) 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. SPR response reflects a change in mass concentration at the detector surface as molecules bind or dissociate from the sensor chip. It can be used to estimate the kinetic constants of protein-protein interactions.

In this 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

$$\dot{R}(t) = k_a \left(R_{max} - R(t)\right) C_0(t) - k_d R(t), \quad t \geq t^{[0]}, \quad R(t^{[0]}) = 0, \tag{30}$$

where $R(t)$ is the measured SPR response in resonance units (RU), 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 $x_\theta(t) := R(t)$, $u(t) := C_0(t)$, $y_\theta(t) := R(t),\ t \geq t^{[0]}$, and $x_0 := R(t^{[0]}) = 0$, (30) becomes the following bilinear system $\Phi = \{A,\ B,\ C,\ F_1\}$

$$\dot{x}_\theta(t) = Ax_\theta(t) + F_1 u(t) x_\theta(t) + Bu(t), \qquad x_\theta(t^{[0]}) = x_0, \tag{31}$$

$$y_\theta(t) = Cx_\theta(t), \quad t \geq t^{[0]}, \tag{32}$$

where $A = -k_d, \quad B = k_a R_{max}, \quad C = 1, \quad F_1 = -k_a$. The unknown parameter vector to be estimated in the experiments is $\theta = \left[\begin{array}{ccc} k_a & k_d & R_{max} \end{array}\right]^T$.

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 $\Phi = \{A,\ B,\ C,\ F_1\}$ with a two-phase piecewise constant input

$$u(t) = u^{[0]}\beta_0(t) + u^{[1]}\beta_1(t), \quad t^{[0]} \leq t < t^{[2]},$$

where

$$u^{[0]} = C_0, \quad \beta_0(t) = \begin{cases} 1, & \text{for } t \in \left[t^{[0]},\ t^{[1]}\right), \\ 0, & \text{for } t \notin \left[t^{[0]}, t^{[1]}\right), \end{cases} \qquad u^{[1]} = 0, \quad \beta_1(t) = \begin{cases} 1, & \text{for } t \in \left[t^{[1]},\ t^{[2]}\right), \\ 0, & \text{for } t \notin \left[t^{[1]}, t^{[2]}\right). \end{cases}$$

Note that in the two-phase SPR experiment the output samples are obtained from $y_\theta(t)$ 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 $\Phi = \{A,\ B,\ C,\ F_1\}$ by $\Phi' = \{\mathcal{A},\ \mathcal{B},\ \mathcal{C},\ \mathcal{F}_1\}$ where $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{F}_1$ are given as follows.

$\mathcal{A} :=\ \mathrm{diag}\,\{\partial_1 A,\ \partial_2 A,\ \partial_3 A\}$ where

$$\partial_1 A = \begin{bmatrix} -k_d & 0 \\ 0 & -k_d \end{bmatrix}, \quad \partial_2 A = \begin{bmatrix} -k_d & 0 \\ -1 & -k_d \end{bmatrix}, \quad \partial_3 A = \begin{bmatrix} -k_d & 0 \\ 0 & -k_d \end{bmatrix}.$$

$\mathcal{B} := \begin{bmatrix} \partial_1 B \\ \partial_2 B \\ \partial_3 B \end{bmatrix}$ where

$$\partial_1 B = \begin{bmatrix} k_a R_{max} \\ R_{max} \end{bmatrix}, \quad \partial_2 B = \begin{bmatrix} k_a R_{max} \\ 0 \end{bmatrix}, \quad \partial_3 B = \begin{bmatrix} k_a R_{max} \\ k_a \end{bmatrix}.$$

$\mathcal{C} :=\ \mathrm{diag}\,\{\partial_1 C_1,\ \partial_2 C_1,\ \partial_3 C_1\}$ where

$$\partial_1 C_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad \partial_2 C_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad \partial_3 C_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

$\mathcal{F}_1 :=\ \mathrm{diag}\,\{\partial_1 F,\ \partial_2 F,\ \partial_3 F\}$ where

$$\partial_1 F_1 = \begin{bmatrix} -k_a & 0 \\ -1 & -k_a \end{bmatrix}, \quad \partial_2 F_1 = \begin{bmatrix} -k_a & 0 \\ 0 & -k_a \end{bmatrix}, \quad \partial_3 F_1 = \begin{bmatrix} -k_a & 0 \\ 0 & -k_a \end{bmatrix}.$$

Since the initial state x_0 of Φ is equal to zero, the initial state vector $\mathcal{X}_0$ of Φ' is also equal to zero, i.e. $\mathcal{X}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$.

B. Local Identifiability

Before computing the CRLB, we first verify whether θ is locally identifiable by the set of admissible inputs or the set of piecewise constant inputs, and then discuss its local identifiability for a specific two-phase piecewise constant input.

Case 1: The set of admissible or piecewise constant inputs: In this case we make the practically impossible assumption that an input of the bilinear system model Φ can be freely selected from the set of admissible inputs $\mathbb{V}$ (or piecewise constant inputs $\mathbb{U}$) and we can repeat the experiment for another input from the same set and measure the corresponding output samples for each experiment conducted. Simple calculations give

$$\operatorname{rank}\left\{\begin{bmatrix} P_1\mathcal{C}\mathcal{X}_0 & P_1\mathcal{C}\mathcal{O}^{[1]} \end{bmatrix}\right\} = \operatorname{rank}\left\{\begin{bmatrix} 0 & R_{max} & -k_dR_{max} & -2k_aR_{max} \\ 0 & 0 & -k_aR_{max} & 0 \\ 0 & k_a & -k_ak_d & -k_a^2 \end{bmatrix}\right\} = 3$$

for positive k_a, k_d and R_{max}, which holds for practical SPR experiments. Note that $P_1 = I_3$ as $p = 1$ and $K = 3$. Since the size of θ is 3, by Theorem 3.2 there exists a finite output data set generated by the bilinear system model with some admissible inputs (or piecewise constant inputs) such that θ is locally identifiable.

Case 2: A specific two-phase piecewise constant input with uniform sampling: We next exploit Theorem 4.1 to find out whether the same parameter vector θ is locally identifiable for a specific two-phase piecewise constant input with uniform sampling. In the remaining part of the example, we assume that the output samples in the association and dissociation phases are acquired with the sampling intervals T_0 and T_1 respectively, and $t^{[0,0]} = t^{[1,0]} = 0$. By simple calculations, we obtain

$$P_1\mathcal{C}\mathcal{W}^{[0]} = \begin{bmatrix} -\frac{C_0k_dR_{max}}{(C_0k_a+k_d)^2} \\ \frac{C_0k_aR_{max}}{(C_0k_a+k_d)^2} \\ -\frac{C_0k_a}{C_0k_a+k_d} \end{bmatrix}, \quad P_1\mathcal{C}\mathcal{A}_d^{[0]}(\mathcal{W}^{[0]} + \mathcal{X}_0) = \begin{bmatrix} \frac{a_0C_0R_{max}(T_0C_0^2k_a^2+T_0C_0k_ak_d-k_d)}{(C_0k_a+k_d)^2} \\ \frac{a_0C_0k_aR_{max}(T_0C_0k_a+T_0k_d+1)}{(C_0k_a+k_d)^2} \\ -\frac{a_0C_0k_a}{C_0k_a+k_d} \end{bmatrix},$$

$$P_1\mathcal{C}\mathcal{A}_d^{[1]}(\mathcal{W}^{[1]} + \mathcal{X}_\theta(t^{[1]})) =$$

$$\begin{bmatrix} \frac{a_1C_0R_{max}[(t^{[1]}-t^{[0]})a_0'C_0^2k_a^2+(t^{[1]}-t^{[0]})a_0'C_0k_ak_d-a_0'k_d+k_d]}{(C_0k_a+k_d)^2} \\ \frac{a_1C_0k_aR_{max}[(t^{[1]}-t^{[0]})a_0'C_0k_a+(t^{[1]}-t^{[0]})a_0'k_d+T_1a_0'C_0k_a+T_1a_0'k_d-T_1C_0k_a-T_1k_d+a_0'-1]}{(C_0k_a+k_d)^2} \\ \frac{a_1C_0k_a(1-a_0')}{C_0k_a+k_d} \end{bmatrix},$$

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

$$\operatorname{rank}\{P_1\mathcal{C}\mathcal{O}'\} = \operatorname{rank}\left\{\begin{bmatrix} P_1\mathcal{C}\mathcal{W}^{[0]} & P_1\mathcal{C}\mathcal{A}_d^{[0]}(\mathcal{W}^{[0]} + \mathcal{X}_0) & P_1\mathcal{C}\mathcal{A}_d^{[1]}(\mathcal{W}^{[1]} + \mathcal{X}_\theta(t^{[1]})) \end{bmatrix}\right\} = 3.$$

Thus, by Theorem 4.1, θ is locally identifiable by a two-phase experiment with uniform sampling.

We next consider the nonsingularity criterion in Theorem 4.3 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

$$P_1\mathcal{C}\mathcal{A}_d^{[1]}(\mathcal{W}^{[1]} - \mathcal{W}^{[0]}) = \begin{bmatrix} \frac{a_1 C_0 k_d R_{max}}{(C_0 k_a + k_d)^2} \\ -\frac{a_1 C_0 k_a R_{max}(T_1 C_0 k_a + T_1 k_d + 1)}{(C_0 k_a + k_d)^2} \\ \frac{a_1 C_0 k_a}{C_0 k_a + k_d} \end{bmatrix}.$$

Hence,

$$\text{rank}\left\{P_1\mathcal{C}\mathcal{O}''\right\} = \text{rank}\left\{\left[\begin{array}{ccc} P_1\mathcal{C}\mathcal{W}^{[0]} & P_1\mathcal{C}\mathcal{A}_d^{[0]}(\mathcal{W}^{[0]} + \mathcal{X}_0) & P_1\mathcal{C}\mathcal{A}_d^{[1]}(\mathcal{W}^{[1]} - \mathcal{W}^{[0]}) \end{array}\right]\right\} = 3.$$

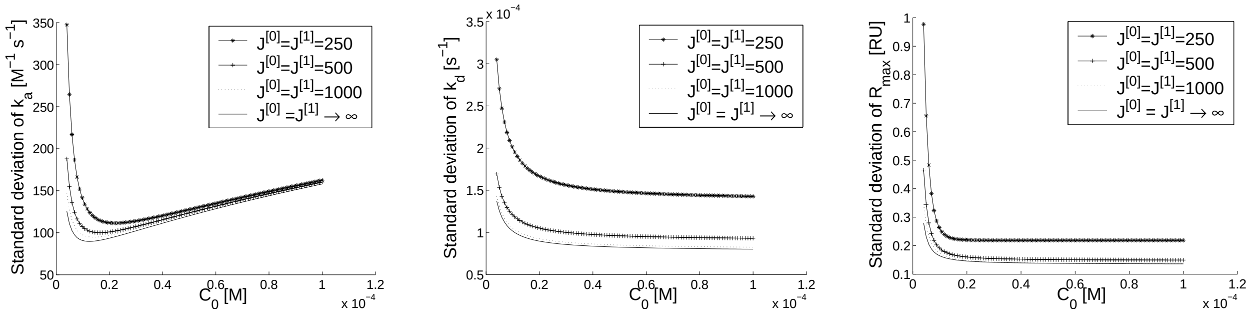
It then follows from Theorem 4.3 that θ is also locally identifiable for sufficiently large $J^{[0]}$ and $J^{[1]}$. This is of course an obvious result since local identifiability in the finite data case implies local identifiability in the infinite data case.

C. CRLB and Asymptotic CRLB

Since θ is locally identifiable by the finite data set uniformly sampled from the output of a two-phase SPR experimental model, the next step is to apply Theorem 4.2 and Theorem 4.4 to numerically calculate the associated CRLB and asymptotic CRLB. Here we use simulated data so that we could conveniently select various experimental settings. For comparison, typical numerical values from [28] are assigned to the unknown parameters, i.e.,

$$k_a = 1478 \text{ M}^{-1}\text{s}^{-1},\; k_d = 4.5 \times 10^{-3} \text{ s}^{-1},\; R_{max} = 7.75 \text{ RU}.$$

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



(a) (b) (c)

Fig. 1. The 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.

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 4.4 with some algebraic manipulations. The detailed derivations are omitted here but can be found in [29].

$$
\begin{aligned}
\mathrm{var}(\hat{k}_a) \geq & \left[\lim_{J'\to\infty} I_{J'}^{-1}(\theta)\right]_{11} \\
= & \frac{\sigma^2\left(-a_0^6a_1^4 - a_0^4a_1^6 + 6a_0^4a_1^4 - a_0^6a_1^2 - a_0^2a_1^6 - 6a_0^2a_1^2 + a_0^4 + a_1^4 + a_0^2 + a_1^2\right)(C_0k_a + k_d)^2}{T^2a_0^2(a_0^2+1)a_1^2(a_1^2+1)C_0^4k_a^2R_{max}^2}.
\end{aligned}
$$

$$
\mathrm{var}(\hat{k}_d) \geq \left[\lim_{J'\to\infty} I_{J'}^{-1}(\theta)\right]_{22} = \frac{\sigma^2\left(C_0k_a + k_d\right)^2(1-a_1)^3(a_1+1)^3}{T^2a_1^2(a_1^2+1)C_0^2k_a^2R_{max}^2}.
$$

$$
\begin{aligned}
\mathrm{var}(\hat{R}_{max}) \geq & \left[\lim_{J'\to\infty} I_{J'}^{-1}(\theta)\right]_{33} \\
= & \frac{\sigma^2}{T^2a_0^2(a_0^2+1)a_1^2(a_1^2+1)C_0^4k_a^4}\left[\left(-a_0^4a_1^6 + 3a_0^4a_1^4 - a_0^2a_1^6 - 3a_0^4a_1^2 + 3a_0^2a_1^4 - 3a_0^2a_1^2 + a_0^4\right.\right. \\
& \left. + a_0^2\right)C_0^2k_a^2 + \left(-2a_0^4a_1^6 + 6a_0^4a_1^4 - 2a_0^2a_1^6 - 6a_0^4a_1^2 + 6a_0^2a_1^4 - 6a_0^2a_1^2 + 2a_0^4 + 2a_0^2\right)C_0k_ak_d \\
& \left. + \left(-a_0^6a_1^4 - a_0^4a_1^6 + 6a_0^4a_1^4 - a_0^6a_1^2 - a_0^2a_1^6 - 6a_0^2a_1^2 + a_0^4 + a_0^2 + a_1^4 + a_1^2\right)k_d^2\right].
\end{aligned}
$$

E. One-phase SPR Experiment with Uniform Sampling

Finally, we show that the same parameter vector θ is not locally identifiable if only a one-phase piecewise constant input $u(t) = u^{[0]}\beta_0(t),\ t^{[0]} \leq t < t^{[1]}$, is applied to the bilinear system in the above example, i.e. the experiment consists of only an association phase. Note that in this one-phase SPR experiment the output samples are obtained from $y_\theta(t)$ for $t^{[0]} \leq t < t^{[1]}$.

Based on the results in Case 2 of Subsection V-B, it is easy to check that

$$\begin{aligned}\operatorname{rank}\{P_1\mathcal{C}\mathcal{O}'\} &= \operatorname{rank}\left\{\left[\begin{array}{cccc} P_1\mathcal{C}\mathcal{W}^{[0]} & P_1\mathcal{C}\mathcal{A}_d^{[0]}(\mathcal{W}^{[0]}+\mathcal{X}_0) & \ldots & P_1\mathcal{C}\left(\mathcal{A}_d^{[0]}\right)^5(\mathcal{W}^{[0]}+\mathcal{X}_0)\end{array}\right]\right\}\\ &= 2 < 3.\end{aligned}$$

By Theorem 4.1, θ is not locally identifiable in this case. Similarly, it is easy to show that θ is not locally identifiable in the asymptotic case either.

References

[1] R. S. Baheti, R. R. Mohler, and H. A. Spang III, "Second-order correlation method for bilinear system identification," *IEEE Trans. Automat. Contr.*, vol. AC-25, 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. Automat. Contr.*, vol. 44, 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*, vol. 47, pp. 1072–1080, Jul. 2000.

[4] N. Kalouptsidis, P. Koukoulas, and V. J. Mathews, "Blind identification of bilinear systems," *IEEE Trans. Signal Processing*, vol. 51, pp. 484–499, Feb. 2003.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing : Estimation Theory*. New Jersey: 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 Processing*, vol. ASSP-32, pp. 721–727, Aug. 1984.

[8] B. Hanzon, *Identifiability, Recursive Identification and Spaces of Linear Dynamical Systems: Part 1, CWI Tract, vol. 63*. Amsterdam: Centrum voor Wiskunde and Informatica, 1989.

[9] P. Stoica and B. C. Ng, "Performance bounds for blind channel estimation," in *Signal Processing Advances in Wireless and Mobile Communications*, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. New Jersey: Prentice-Hall, 2001, vol. I: Trends in Channel Estimation and Equalization, ch. 2.

[10] R. J. Ober, "The Fisher information matrix for linear systems," *Systems and Control Letters*, 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 Processing*, vol. 51, 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. Machine Intell.*, vol. 12, pp. 735–759, Aug. 1990.

[13] R. R. Mohler, *Nonlinear systems, vol. II, Applications to Bilinear Control*. New Jersey: Prentice-Hall, 1991.

[14] J. K. Hale, *Ordinary Differential Equations*, 2nd ed. Florida: Krieger, 1980.

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

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

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

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

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

[20] T. M. Apostol, *Mathematical Analysis*, 2nd ed. Massachusetts: Addison-Wesley, 1974.

[21] L. Ljung, *System Identification: Theory for the User*. New Jersey: Prentice-Hall, 1999.

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

[23] T. Kailath, *Linear Systems*. New Jersey: Prentice-Hall, 1980.

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

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.

[26] 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.

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

[28] 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.

[29] Q. Zou, “System theory in biomedical applications,” M. Eng. thesis, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 2003.