Cramer Rao Lower Bound for Parameter Estimation in Nonlinear Systems

Zhiping Lin, Qiyue Zou, E. Sally Ward, Raimund J. Ober

EDICS: SAS-STAT

Abstract

The Cramer Rao lower bound (CRLB) gives a lower bound on the achievable accuracy of parameter estimates resulting from an unbiased estimation procedure based on a given set of observed noisy data. Calculation of the CRLB, which is the inverse of the Fisher information matrix, for output data sets of a general nonlinear system is a challenging problem and is considered in this paper. It is shown that the Fisher information matrix for a data set generated by a nonlinear system with additive Gaussian measurement noise can be expressed in terms of the outputs of its derivative system that is also a nonlinear system. An example is considered arising from surface plasmon resonance experiments to determine the dynamic parameters of molecular interactions.

Keywords: Cramer Rao lower bound; Fisher information matrix; Parameter estimation; Nonlinear systems; Surface plasmon resonance experiments; Molecular interactions.

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

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

E. S. Ward and R. J. Ober are with the Center for Immunology, University of Texas Southwestern Medical Center at Dallas, 6000 Harry Hines Boulevard, Dallas, TX 75235-8576. R. J. Ober is also with the School of Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX 75083-0688. (e-mail: ober@utdallas.edu).

I. Introduction

A typical approach to studying complicated biomedical phenomenon is to investigate mathematical models that describe the underlying phenomena. Unknown parameters of an underlying model are estimated from experimental data. For example, surface plasmon resonance biosensors are used to estimate the kinetic constants of protein-protein interactions [2], [3]. The accuracy with which the parameters can be determined depends on a variety of factors, measurement noise level, sampling rate, number of repeat experiments, etc. The production of the reagents for such experiments can often be very costly. Therefore an efficient setup of the experimental conditions is of great importance to avoid unnecessary costs in executing the experiments. A classical tool for experiment design is the Fisher information matrix (see e.g. [4]). The inverse of the Fisher information matrix gives a lower bound (the Cramer Rao lower bound (CRLB)) on the covariance matrix of any unbiased estimator of the parameters [5], [6]. It is widely used as a benchmark to evaluate the performance of an estimation algorithm and can provide guidance to improve the experimental design.

The acquired data in biophysical experiments can often be modeled as the output of a linear or nonlinear system, with the sampled output being corrupted by white noise (see e.g. [7] for nuclear magnetic resonance data, [8] for biosensor data). A crucial aspect is that the data arising in biophysical experiments is typically non-stationary, which means that existing results on the CRLB for nonlinear filtering that deal with stationary time series (see, e.g., [9]) cannot be applied to the problem at hand. To date no effective approach appears to be available to compute the CRLB for parameter estimation for the case of nonstationary deterministic nonlinear systems corrupted by measurement noise.

Recently, a systematic approach has been proposed for calculating the CRLB for output data sets of one-dimensional non-stationary linear dynamic systems with deterministic input and Gaussian measurement noise [10]. The approach has been extended to data sets generated by multidimensional linear separable-denominator systems [11]. The above approach is based on the concept of a derivative system associated with the original dynamic system. Here we will show that the concept of derivative system can be further generalized to a nonlinear dynamic system and that the derivative system is again a nonlinear system of the same structure as the original system. The Fisher information matrix can then be calculated by determining the outputs of the derivative system at the time points at which the experimental data is obtained.

II. Fisher Information Matrix and CRLB

We consider a nonlinear system Φ with m inputs and p outputs, described in the state-space form as (see [12])

$$\dot{x}_\theta(t) = f_\theta(x_\theta) + \sum_{k=1}^{m} g_{\theta,k}(x_\theta)u_k(t), \quad x_\theta(t_0) = x_0, \tag{1}$$

$$y_\theta(t) = h_\theta(x_\theta), \quad t \geq t_0, \tag{2}$$

where $\theta := \begin{bmatrix} \theta_1 & \dots & \theta_K \end{bmatrix}^T$ is the unknown parameter vector to be estimated, the state x_θ is assumed to be an element of $\mathbb{R}^n$, the inputs u_1, ..., u_m are real-valued functions of t that are not dependent on θ, $y_\theta \in \mathbb{R}^p$ is the system output, the mappings f_θ, $g_{\theta,1}$, ..., $g_{\theta,m}$ are $\mathbb{R}^n$-valued functions of x_θ and θ, and the output function h_θ is an $\mathbb{R}^p$-valued function of x_θ and θ. Throughout the paper we use $y_{\theta,i}(t)$ and $h_{\theta,i}(x_\theta)$ to represent the i^{th} element of $y_\theta(t)$ and $h_\theta(x_\theta)$, $i = 1, \dots, p$, respectively, i.e. $y_\theta(t) = \begin{bmatrix} y_{\theta,1}(t) & \dots & y_{\theta,p}(t) \end{bmatrix}^T$ and $h_\theta(x_\theta) = \begin{bmatrix} h_{\theta,1}(x_\theta) & \dots & h_{\theta,p}(x_\theta) \end{bmatrix}^T$. Similarly, we use $x_{\theta,l}(t)$, $f_{\theta,l}(x_\theta)$ and $g_{\theta,k,l}(x_\theta)$ ($k = 1, \dots, m$) to represent the l^{th} element of $x_\theta(t)$, $f_\theta(x_\theta)$ and $g_{\theta,k}(x_\theta)$, $l = 1, \dots, n$, respectively. Throughout the paper, we assume:

Assumption 2.1: 1.) The nonlinear system Φ is represented by (1)-(2) where f_θ, $g_{\theta,1}$, ..., $g_{\theta,m}$ and h_θ are *smooth* functions of x_θ and θ, i.e. all entries of f_θ, $g_{\theta,1}$, ..., $g_{\theta,m}$ and h_θ are real-valued functions of x_θ and θ with continuous partial derivatives of any order. The input functions u_1, ..., u_m are assumed to be *piecewise continuous* in t and independent of the parameter vector θ;

2.) The acquired noise corrupted data samples $s_{\theta,i}(j)$, are the measured output of the nonlinear system Φ, i.e., $s_{\theta,i}(j) = y_{\theta,i}(t_j) + w_i(t_j)$, where $y_{\theta,i}(t_j)$ is the i^{th} noise free output element at the sampling instant t_j and $w_i(t_j)$ is the measurement noise component, $t_0 \leq t_1 < t_2 < \ldots < t_J$; the measurement noise components have independent Gaussian distributions with zero mean and variance $\sigma^2_{i,j}$, $i = 1, \dots, p$, $j = 1, \dots, J$.

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 $\mathbb{R}^K$. The probability density function $p(S;\theta)$ for the acquired data set $S := \{s_{\theta,i}(j),\ i = 1, \dots, p,\ j = 1, \dots, J\}$ is assumed to satisfy the standard regularity conditions (see e.g. [13]). The Fisher information matrix $I(\theta)$ is then given by $[I(\theta)]_{sr} = E\left\{\left(\frac{\partial \ln p(S;\theta)}{\partial \theta_s}\right)\left(\frac{\partial \ln p(S;\theta)}{\partial \theta_r}\right)\right\}$, $1 \leq s, r \leq K$, 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 $\mathrm{var}(\hat{\theta}) \geq I^{-1}(\theta)$.

In the following theorem we first show that the derivative system (with respect to the given parameter vector θ) of a general nonlinear dynamic system given by (1)-(2) can also be expressed as a nonlinear system of the same form. In the second part of the theorem this fact is used to show that the Fisher information matrix for the sampled output data of a nonlinear system with white Gaussian measurement noise can be expressed using the output samples of its derivative system.

Theorem 2.1: Consider the nonlinear system Φ represented by (1)-(2) and assume that Assumption 2.1 is satisfied. Let $\mathcal{Y}_\theta(t) := \left[\mathcal{Y}_{\theta,1}^T(t) \dots \mathcal{Y}_{\theta,p}^T(t)\right]^T$, with $\mathcal{Y}_{\theta,i}(t) := \left[\frac{\partial y_{\theta,i}(t)}{\partial \theta_1} \dots \frac{\partial y_{\theta,i}(t)}{\partial \theta_K}\right]^T,\ i = 1, \dots, p,\ t \geq t_0$. Then,

1.) $\mathcal{Y}_\theta(t)$, $t \geq t_0$, is equal to the output of the derivative system Φ' represented by

$$\dot{\mathcal{X}}_\theta(t) = \mathcal{F}_\theta(\mathcal{X}_\theta) + \sum_{k=1}^{m} \mathcal{G}_{\theta,k}(\mathcal{X}_\theta)\mathcal{U}_k(t), \quad \mathcal{X}_\theta(t_0) = \mathcal{X}_0, \tag{3}$$

$$\mathcal{Y}_\theta(t) = \mathcal{H}_\theta(\mathcal{X}_\theta), \quad t \geq t_0, \tag{4}$$

where $\mathcal{X}_0 := \left[x_0^T \ \ \frac{\partial x_0^T}{\partial \theta_1} \dots \frac{\partial x_0^T}{\partial \theta_K}\right]^T$ and for $t \geq t_0$

$$\mathcal{X}_\theta(t) := \left[x_\theta^T(t) \ \ \frac{\partial x_\theta^T(t)}{\partial \theta_1} \dots \frac{\partial x_\theta^T(t)}{\partial \theta_K}\right]^T;$$

$$\mathcal{U}_k(t) := u_k(t),\ k = 1, \dots, m;$$

$$\mathcal{F}_\theta(\mathcal{X}_\theta) := \begin{bmatrix} f_\theta(x_\theta) \\ \frac{\partial f_\theta(x_\theta)}{\partial \theta_1} + \mathbf{J}_{x_\theta}\left[f_\theta(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial f_\theta(x_\theta)}{\partial \theta_K} + \mathbf{J}_{x_\theta}\left[f_\theta(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_K} \end{bmatrix}; \tag{5}$$

$$\mathcal{G}_{\theta,k}(\mathcal{X}_\theta) := \begin{bmatrix} g_{\theta,k}(x_\theta) \\ \frac{\partial g_{\theta,k}(x_\theta)}{\partial \theta_1} + \mathbf{J}_{x_\theta}\left[g_{\theta,k}(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial g_{\theta,k}(x_\theta)}{\partial \theta_K} + \mathbf{J}_{x_\theta}\left[g_{\theta,k}(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_K} \end{bmatrix},$$

$$k = 1, \dots, m;$$

$$\mathcal{H}_\theta(\mathcal{X}_\theta) := \left[\mathcal{H}_{\theta,1}^T(\mathcal{X}_\theta) \dots \mathcal{H}_{\theta,p}^T(\mathcal{X}_\theta)\right]^T, \tag{6}$$

with $\mathcal{H}_{\theta,i}(\mathcal{X}_\theta) := \begin{bmatrix} \frac{\partial h_{\theta,i}(x_\theta)}{\partial \theta_1} + \mathbf{J}_{x_\theta}\left[h_{\theta,i}(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_1} \\ \vdots \\ \frac{\partial h_{\theta,i}(x_\theta)}{\partial \theta_K} + \mathbf{J}_{x_\theta}\left[h_{\theta,i}(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_K} \end{bmatrix}$, $i = 1, \ldots, p$. Here $\mathbf{J}_{x_\theta}\left[f_\theta(x_\theta)\right]$, $\mathbf{J}_{x_\theta}\left[g_{\theta,k}(x_\theta)\right]$ and $\mathbf{J}_{x_\theta}\left[h_{\theta,i}(x_\theta)\right]$ are the Jacobian matrices of $f_\theta(x_\theta)$, $g_{\theta,k}(x_\theta)$ and $h_{\theta,i}(x_\theta)$ with respect to x_θ, respectively.

2.) The Fisher information matrix is given by

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

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

$$P_i := [\ \underbrace{\mathbf{0} \quad \ldots \quad \mathbf{0}}_{(i-1)\ \mathbf{0}s} \quad I_K \quad \underbrace{\mathbf{0} \quad \ldots \quad \mathbf{0}}_{(p-i)\ \mathbf{0}s}\ ], \tag{7}$$

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

Proof: 1.) It follows from Assumption 2.1 that $x_\theta(t)$ is continuous with respect to t and θ, piecewise partially differentiable with respect to t, and partially differentiable with respect to θ_s for all $\theta \in \Theta$ and $t \geq t_0$ (Assume $s = 1, \ldots, K$ and $t \geq t_0$ throughout the proof). Then with the possible exception of the discrete discontinuities of u_k, $k = 1, \ldots, m$, we have (see page 359 in [15]) $\frac{\partial^2 x_\theta(t)}{\partial t \partial \theta_s} = \frac{\partial^2 x_\theta(t)}{\partial \theta_s \partial t} = \frac{\partial \dot{x}_\theta(t)}{\partial \theta_s}$. Taking the partial derivative of (1) with respect to θ_s gives $\frac{\partial \dot{x}_\theta(t)}{\partial \theta_s} = \frac{\partial f_\theta(x_\theta)}{\partial \theta_s} + \mathbf{J}_{x_\theta}\left[f_\theta(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_s} + \sum_{k=1}^{m} \mathbf{J}_{x_\theta}\left[g_{\theta,k}(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_s} u_k(t)$. The partial derivative of $y_{\theta,i}(t)$, $i = 1, \ldots, p$, with respect to θ_s is given by $\frac{\partial y_{\theta,i}(t)}{\partial \theta_s} = \frac{\partial h_{\theta,i}(x_\theta)}{\partial \theta_s} + \mathbf{J}_{x_\theta}\left[h_{\theta,i}(x_\theta)\right] \frac{\partial x_\theta(t)}{\partial \theta_s}$. Since $\mathcal{X}_\theta(t) = \left[x_\theta^T(t) \quad \frac{\partial x_\theta^T(t)}{\partial \theta_1} \ldots \frac{\partial x_\theta^T(t)}{\partial \theta_K}\right]^T$ and $\mathcal{Y}_{\theta,i}(t) = \left[\frac{\partial y_{\theta,i}(t)}{\partial \theta_1} \ldots \frac{\partial y_{\theta,i}(t)}{\partial \theta_K}\right]^T$, $i = 1, \ldots, p$, stacking the corresponding equations produces

$$\begin{aligned} \dot{\mathcal{X}}_\theta(t) &= \left[\left(\frac{\partial x_\theta(t)}{\partial t}\right)^T \quad \left(\frac{\partial^2 x_\theta(t)}{\partial t \partial \theta_1}\right)^T \ldots \left(\frac{\partial^2 x_\theta(t)}{\partial t \partial \theta_K}\right)^T\right]^T \\ &= \left[\dot{x}_\theta^T(t) \quad \frac{\partial \dot{x}_\theta^T(t)}{\partial \theta_1} \ldots \frac{\partial \dot{x}_\theta^T(t)}{\partial \theta_K}\right]^T \\ &= \mathcal{F}_\theta(\mathcal{X}_\theta) + \sum_{k=1}^{m} \mathcal{G}_{\theta,k}(\mathcal{X}_\theta) \mathcal{U}_k(t) \end{aligned}$$

and $\mathcal{Y}_{\theta,i}(t) = \left[\frac{\partial y_{\theta,i}(t)}{\partial \theta_1} \ldots \frac{\partial y_{\theta,i}(t)}{\partial \theta_K}\right]^T = \mathcal{H}_{\theta,i}(\mathcal{X}_\theta)$, $i = 1, \ldots, p$. Since $\mathcal{Y}_\theta(t) = \left[\mathcal{Y}_{\theta,1}^T(t) \ldots \mathcal{Y}_{\theta,p}^T(t)\right]^T$, the derivative system Φ' is then obtained by stacking the corresponding equations as $\mathcal{Y}_\theta(t) = \mathcal{H}_\theta(\mathcal{X}_\theta)$, $t \geq t_0$. The initial condition of Φ' is given by $\mathcal{X}_\theta(t)|_{t=t_0} = \mathcal{X}_0$.

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

$$I(\theta) = \left[\sum_{i=1}^{p} \sum_{j=1}^{J} \frac{1}{\sigma_{i,j}^2} \frac{\partial y_{\theta,i}(t_j)}{\partial \theta_s} \frac{\partial y_{\theta,i}(t_j)}{\partial \theta_r} \right]_{1 \le s,r \le K} \tag{8}$$

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

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

The above theorem shows that the Fisher information matrix arising from the estimation of unknown parameters using the output data set of a nonlinear system represented by (1)-(2) can be conveniently expressed in terms of its associated derivative system which has a similar representation. Although in general there does not exist an analytical solution for a general nonlinear system, there are a number of good numerical methods for solving nonlinear systems numerically. It follows that the Fisher information matrix and the CRLB can be computed using one of these numerical methods.

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 [16]. As a corollary we can obtain a criterion for the local identifiability in our context that draws an interesting connection between identifiability and the output reachability of the derivative system.

Corollary 2.1: Given the assumptions of Theorem 2.1, the parameter vector θ is locally identifiable if and only if

$$\begin{aligned} &\operatorname{rank}\{[P_1\mathcal{Y}_\theta(t_1) \ \ldots \ P_1\mathcal{Y}_\theta(t_J) \ \ P_2\mathcal{Y}_\theta(t_1) \ \ldots \ P_2\mathcal{Y}_\theta(t_J) \\ &\qquad \ldots \ P_p\mathcal{Y}_\theta(t_1) \ \ldots \ P_p\mathcal{Y}_\theta(t_J)]\} = K. \end{aligned}$$

Proof: It follows from [17] that the parameter vector is locally identifiable if and only if the Fisher information matrix $I(\theta)$ is invertible. By Theorem 2.1 part 2.) $I(\theta)$ is invertible if and only if the above rank condition is satisfied.

III. EXAMPLE

Biosensors such as instruments by the BIAcore company allow for the monitoring of protein-protein interactions in real time using an optical detection principle based on surface plasmon resonance (SPR) technology (see, e.g. [2], [3]). In the experiments one of the proteins (ligand) is coupled to a sensor chip and the second protein (analyte) is flowed across the surface coupled ligand using a micro-fluidic device. The SPR response reflects a change in mass concentration at the detector surface as molecules bind or

$$\mathcal{F}_\theta(\mathcal{X}_\theta) := \left[\begin{array}{c} (-k_aR_{max}-k_T)x_{\theta,1}(t)+k_dx_{\theta,2}(t)+k_ax_{\theta,1}(t)x_{\theta,2}(t) \\ k_aR_{max}x_{\theta,1}(t)-k_dx_{\theta,2}(t)-k_ax_{\theta,1}(t)x_{\theta,2}(t) \\ \hline -R_{max}x_{\theta,1}(t)+x_{\theta,1}(t)x_{\theta,2}(t)+(-k_aR_{max}-k_T+k_ax_{\theta,2}(t))\frac{\partial x_{\theta,1}(t)}{\partial k_a}+(k_d+k_ax_{\theta,1}(t))\frac{\partial x_{\theta,2}(t)}{\partial k_a} \\ R_{max}x_{\theta,1}(t)-x_{\theta,1}(t)x_{\theta,2}(t)+(k_aR_{max}-k_ax_{\theta,2}(t))\frac{\partial x_{\theta,1}(t)}{\partial k_a}+(-k_d-k_ax_{\theta,1}(t))\frac{\partial x_{\theta,2}(t)}{\partial k_a} \\ \hline x_{\theta,2}(t)+(-k_aR_{max}-k_T+k_ax_{\theta,2}(t))\frac{\partial x_{\theta,1}(t)}{\partial k_d}+(k_d+k_ax_{\theta,1}(t))\frac{\partial x_{\theta,2}(t)}{\partial k_d} \\ -x_{\theta,2}(t)+(k_aR_{max}-k_ax_{\theta,2}(t))\frac{\partial x_{\theta,1}(t)}{\partial k_d}+(-k_d-k_ax_{\theta,1}(t))\frac{\partial x_{\theta,2}(t)}{\partial k_d} \\ \hline -x_{\theta,1}(t)+(-k_aR_{max}-k_T+k_ax_{\theta,2}(t))\frac{\partial x_{\theta,1}(t)}{\partial k_T}+(k_d+k_ax_{\theta,1}(t))\frac{\partial x_{\theta,2}(t)}{\partial k_T} \\ (k_aR_{max}-k_ax_{\theta,2}(t))\frac{\partial x_{\theta,1}(t)}{\partial k_T}+(-k_d-k_ax_{\theta,1}(t))\frac{\partial x_{\theta,2}(t)}{\partial k_T} \\ \hline -k_ax_{\theta,1}(t)+(-k_aR_{max}-k_T+k_ax_{\theta,2}(t))\frac{\partial x_{\theta,1}(t)}{\partial R_{max}}+(k_d+k_ax_{\theta,1}(t))\frac{\partial x_{\theta,2}(t)}{\partial R_{max}} \\ k_ax_{\theta,1}(t)+(k_aR_{max}-k_ax_{\theta,2}(t))\frac{\partial x_{\theta,1}(t)}{\partial R_{max}}+(-k_d-k_ax_{\theta,1}(t))\cdot\frac{\partial x_{\theta,2}(t)}{\partial R_{max}} \end{array}\right].$$

dissociate from the sensor chip. The resulting acquired data can be used to estimate the kinetic constants of protein-protein interactions.

A notorious problem in conducting such experiments is the presence of mass-transport (see e.g. [18]). The following compartmental model, written in standard form Φ has been suggested (see e.g. [18], [19]) to estimate the kinetic parameters of an interaction in the presence of mass transport,

$$\dot{x}_\theta(t) = f_\theta(x_\theta) + g_{\theta,1}(x_\theta)u_1(t), \quad x_\theta(t_0) = x_0, \tag{9}$$

$$y_\theta(t) = h_\theta(x_\theta), \quad t \geq t_0, \tag{10}$$

where

$$x_\theta(t) := \left[\begin{array}{c} x_{\theta,1}(t) \\ x_{\theta,2}(t) \end{array}\right] := \left[\begin{array}{c} C_S(t) \\ R(t) \end{array}\right], \quad x_0 := \left[\begin{array}{c} 0 \\ 0 \end{array}\right],$$

$$g_{\theta,1}(x_\theta) := \left[\begin{array}{c} k_T \\ 0 \end{array}\right], \quad u_1(t) := C_A(t), \quad h_\theta(x_\theta) := x_{\theta,2}(t),$$

$$f_\theta(x_\theta) := \left[\begin{array}{c} (-k_aR_{max}-k_T)x_{\theta,1}(t)+k_dx_{\theta,2}(t) \\ +k_ax_{\theta,1}(t)x_{\theta,2}(t) \\ \hline k_aR_{max}x_{\theta,1}(t)-k_dx_{\theta,2}(t) \\ -k_ax_{\theta,1}(t)x_{\theta,2}(t) \end{array}\right].$$

Here $C_S(t)$ is the concentration of analyte on the sensor surface, $R(t)$ is the measured SPR response in resonance units (RU) that are proportional to the mass accumulated on the surface, k_a and k_d are the kinetic association and dissociation constants of the interaction, respectively, k_T is a parameter that indicates the influence of mass transport on the kinetics, R_{max} is the maximum analyte binding capacity in RU, $C_A(t)$ is the concentration value of analyte in the flow cell which can be controlled as an input in the experiments, and the initial SPR response is assumed to be zero. The unknown parameter vector to be estimated in the experiments is $\theta := \begin{bmatrix} k_a & k_d & k_T & R_{max} \end{bmatrix}^T$.

The first step of the proposed method is the calculation of the derivative system by Theorem 2.1. The derivative system Φ' of the system Φ is represented by

$$\dot{\mathcal{X}}_\theta(t) = \mathcal{F}_\theta(\mathcal{X}_\theta) + \mathcal{G}_{\theta,1}(\mathcal{X}_\theta)\mathcal{U}_1(t), \quad \mathcal{X}_\theta(t_0) = \mathcal{X}_0, \tag{11}$$

$$\mathcal{Y}_\theta(t) = \mathcal{H}_\theta(\mathcal{X}_\theta), \quad t \geq t_0. \tag{12}$$

where

$$\begin{aligned}
\mathcal{X}_\theta(t) &:= \begin{bmatrix} x_\theta^T(t) & \frac{\partial x_\theta^T(t)}{\partial \theta_1} \dots \frac{\partial x_\theta^T(t)}{\partial \theta_4} \end{bmatrix}^T \\
&= \begin{bmatrix} x_\theta^T(t) & \frac{\partial x_\theta^T(t)}{\partial k_a} & \frac{\partial x_\theta^T(t)}{\partial k_d} & \frac{\partial x_\theta^T(t)}{\partial k_T} & \frac{\partial x_\theta^T(t)}{\partial R_{max}} \end{bmatrix}^T; \\
\mathcal{X}_\theta(t_0) &:= \begin{bmatrix} x_\theta^T(t_0) & \frac{\partial x_\theta^T(t_0)}{\partial \theta_1} & \dots & \frac{\partial x_\theta^T(t_0)}{\partial \theta_4} \end{bmatrix}^T \\
&= [0 \;\; 0 \mid 0 \;\; 0 \mid 0 \;\; 0 \mid 0 \;\; 0 \mid 0 \;\; 0]^T; \\
\mathcal{G}_{\theta,1}(\mathcal{X}_\theta) &:= [k_T \;\; 0 \mid 0 \;\; 0 \mid 0 \;\; 0 \mid 1 \;\; 0 \mid 0 \;\; 0]^T; \\
\mathcal{H}_\theta(\mathcal{X}_\theta) &:= \begin{bmatrix} \frac{\partial x_{\theta,2}(t)}{\partial k_a} & \frac{\partial x_{\theta,2}(t)}{\partial k_d} & \frac{\partial x_{\theta,2}(t)}{\partial k_T} & \frac{\partial x_{\theta,2}(t)}{\partial R_{max}} \end{bmatrix}^T,
\end{aligned}$$

$\mathcal{U}_1(t) := u_1(t) = C_A(t)$, $t \geq t_0$ and $\mathcal{F}_\theta(\mathcal{X}_\theta)$ (see next page).

We simulate the SPR experiments with the parameters $k_a = 10^{-3}$, $k_d = 0.01$, $R_{max} = 700$. The input of the system is assumed to be

$$C_A(t) = \begin{cases} 1000, & \text{for } t \in [300, 1000], \\ 0, & \text{otherwise}, \end{cases}$$

and the starting time is $t_0 = 0$. As we mentioned earlier, k_T is a parameter whose value indicates the influence of mass transport on the kinetic interaction. In our study, the range of k_T is from 10^{-5} to 10^5, where a small (large) value indicates a large (small) influence of mass transport on the data.

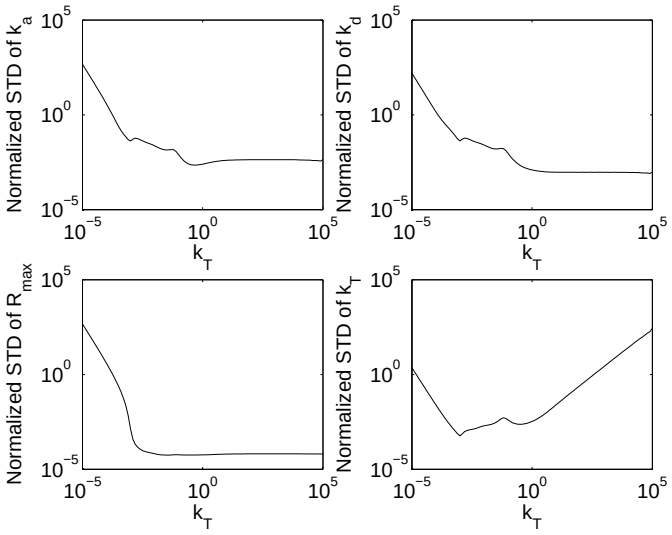

Fig. 1. Normalized standard deviations (STDs) of k_a, k_d, k_T and R_{max} as functions of the parameter k_T.

Using a numerical algorithm provided by Matlab [20], we obtain the outputs of the derivative system Φ' represented by (11)-(12). We then calculate the Fisher information matrix for the data sets generated by the nonlinear system Φ represented by (9)-(10) with white Gaussian measurement noise. We assume that the output of the experiment is uniformly sampled at $t_j = 300 + (j-1)T$, where T is the sampling interval and $j = 1, \ldots, J$. Let the noise variance be $\sigma_{i,j} = 1$ for $i = 1, \ldots, p$, $j = 1, \ldots, J$. In the following discussion, we fix $T = 1$, $J = 1400$.

Fig. 1 plots the best achievable normalized standard deviations (STDs), (i.e. the square roots of the corresponding entries of the inverse of the Fisher information matrix) of k_a, k_d, k_T and R_{max} (STD(k_a)/k_a *etc.*) as functions of the parameter k_T in log scale. It can be seen that the best achievable normalized standard deviations of all the four parameters are large when k_T is very small (near 10^{-5}). This confirms that the kinetic parameters are difficult to estimate in the presence of significant mass transport effects, if the other experimental conditions remain unchanged. On the other hand, when k_T is very large (near 10^5), i.e. the mass transport effect is small, the best achievable normalized standard deviations of k_a, k_d, and R_{max} are reasonably small, while the best achievable normalized standard deviation of k_T becomes very large.

REFERENCES

[1] E. N. Bruce, *Biomedical Signal Processing and Signal Modeling.* New York: Wiley, 2001.

[2] 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.

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

[4] V. V. Fedorov and P. Hackl, *Model-oriented design of experiments*. Springer Verlag, 1997.

[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] R. J. Ober, Z. Lin, H. Ye, and E. S. Ward, "Achievable accuracy of parameter estimation for multidimensional NMR experiments," *Journal of Magnetic Resonance*, vol. 157, pp. 1–16, 2002.

[8] 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.

[9] P. Tichavsky, C. H. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Processing*, vol. 46, pp. 1386–1396, May 1998.

[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] A. Isidori, *Nonlinear Control Systems*, 3rd ed. New York: Springer-Verlag, 1995.

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

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

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

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

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

[18] P. Schuck, "Kinetics of ligand binding to receptor immobilized in a polymer matrix, as detected with an evanescent wave biosensor. I. A computer simulation of the influence of mass transport," *Biophysical Journal*, vol. 70, pp. 1230–1249, 1996.

[19] R. J. Fisher, M. Fivash, J. Casas-Finet, S. Bladen, and K. L. McNitt, "Real-time BIAcore measurements of *escherichia coli* single-stranded DNA binding (SSB) protein to polydeoxythymidic acid reveal single-state kinetics with steric cooperativity," *Methods: A companion to Methods in Enzymology*, vol. 6, pp. 121–133, 1994.

[20] *Matlab*, MathWorks, 2000.