Maximum Likelihood Modeling Of Orbits Of Nonlinear ODEs

Torsten Söderström, Torbjörn Wigren and Emad Abd-Elrady

Department of Systems and Control, Information Technology, Uppsala Univ., PO Box 337, SE-751 05 Uppsala, SWEDEN. e-mail: torsten.soderstrom@it.uu.se, torbjorn.wigren@it.uu.se, emad.abd-elrady@it.uu.se

Abstract

This report treats a new approach to the problem of periodic signal estimation. The idea is to model the periodic signal as a function of the state of a second order nonlinear ordinary differential equation (ODE). This is motivated by Poincare theory which is useful for proving the existence of periodic orbits for second order ODEs. The functions of the right hand side of the nonlinear ODE are then parameterized, and a maximum likelihood algorithm is developed for estimation of the parameters of these unknown functions from the measured periodic signal. The approach is analyzed by derivation and solution of a system of ODEs that describes the evolution of the Cramer-Rao bound over time. The proposed methodology reduces the number of estimated unknowns at least in cases where the actual signal generation resembles that of the imposed model. This in turn is expected to result in an improved accuracy of the estimated parameters.

Index Terms—Adaptive Filtering, Nonlinear Systems, Ordinary Differential Equations, Periodic Orbits, Spectrum Estimation, System Identification

I. **Introduction**

The modeling of periodic signals has widespread applications. Examples include vibration analysis and overtone analysis in power networks. Periodic signal analysis is also closely tied to the measurement of linearity of electronic power amplifiers and other devices, using sinusoidal inputs.

As a consequence, the field of periodic signal analysis has been widely studied. One of the most widespread signal processing methods ever, the periodogram method in combination with fast Fourier transform (FFT) techniques, forms a baseline against which other methods can be compared. See e.g. Stoica and Moses (1997) for detailed algorithms and performance analysis issues. Parametric methods for line spectra are directly applicable to the periodic signal estimation problem. Many references (e.g. Rife and Boorstyn, 1976; Kumaresan and Tufts, 1982; Porat and Friedlander, 1987; Stoica and Nehorai, 1989; Li and Stoica, 1996; Stoica and Moses, 1997) discuss a number of such methods including autoregressive moving average (ARMA) model based methods, the nonlinear least squares method, the high order Yule-Walker method, the Pisarenko method, the MUSIC method, the ESPRIT method, the APES method as well as more direct parametric approaches. Theoretical results on expected performance are available in Stoica and Moses (1997). Further extensions can be found in e.g. Händel and Tichavsky (1994) and Tichavsky and Händel (1997). The last reference also treats chirp signals, i.e. it describes estimation of frequencies as well as corresponding frequency rates. The harmonic relation between the frequencies of the spectral components is explicitly exploited in the comb filter approach of Nehorai and Porat (1986) and in algorithms for recursive modeling of static nonlinearities and a fundamental frequency (Wigren and Händel ,1996; Abd-Elrady, 2002, 2004). The reason for this exploitation is that, in general, the achievable estimation accuracy should improve when more prior information is imposed on the algorithm. Note that in the single frequency case, a considerable interest has been focused on low complexity methods (Fitz, 1994).

The present report is inspired by one possible model for the generation of periodic signals, namely nonlinear ordinary differential equations (ODEs). There is a rich theory on the subject as outlined

in e.g. Khalil (1996). The focus here will be on periodic orbits and their properties. Some of the strongest results of the theory concern ODEs with two state variables. The reason for this is that in R^2 closed periodic orbits that do not intersect themselves divide the space into one part interior to the orbit and one part exterior to the orbit. This is the important Jordan curve theorem, see Khalil (1996). The mathematical consequence of this is that there are several powerful theorems on the existence of periodic solutions to ODEs in R^2 - hence it seems to be advantageous to base estimation algorithms on second order ODEs. The most well known theoretical results include the Poincare-Bendixon theorem, the Bendixon criterion and the Poincare map (Khalil, 1996). It has recently been found (Wigren and Söderström, 2003a) that an ODE order of 2 is enough for modeling quite a wide class of periodic signals. This is a main motivation for the approach of Wigren *et al.* (2003b, c) and of this report, where second order ODEs are used as models. This fact is practically important when the low complexity methods of Wigren *et al*. (2003b, c) are used for initialization of e.g. the maximum likelihood method. The reason is that the use of a higher than order 2 ODE model in these low complexity methods would require a numerical computation of at least the second derivative of the signal. With an order 2 ODE model the first derivative is all that is required.

The signal model of the report is obtained by introducing a parameterization of the right hand side of a general second order ODE, and by defining the periodic signal to be modeled as a function of the states of this ODE. A polynomial parameterization is utilized in the present report. Further contributions include the derivation of a maximum likelihood algorithm and a theoretical analysis of performance aspects using the Cramer-Rao bound (CRB). The above aspects are also illustrated in a simulation study. Initial aspects of the present work can be found in Wigren *et al*. (2003b,c). Wigren *et al.* (2003b) presents recursive algorithms based on Kalman and extended Kalman filters, while Wigren *et al.* (2003c) uses an off-line approach to formulate a least squares algorithm suitable for high SNR scenarios. None of these papers touch on optimal algorithms or performance analysis issues.

What are the advantages of the approach taken? First, many systems that generate periodic signals are best described by nonlinear ODEs. Examples include tunnel diodes, pendulums, biological predator-prey systems and frequency synthesizers, see Khalil (1996). Many of these systems are described by second order ODEs with polynomial right hand sides and it can be expected that there are then good opportunities to obtain highly accurate models by estimation of only a few parameters. The parsimony principle, e.g. Söderström and Stoica (1989), therefore suggests that the achievable accuracy would be improved by the proposed methods, as compared e.g. to the periodogram or other methods that do not impose the same amount of prior information on the solution. Even if the data generation is more complex than the proposed model, it could be expected that addition of a few parameters in the right hand side of the ODE should allow for accurate modeling and the above conclusion would again be valid. The results of Wigren and Söderström (2003a) supports this conjecture further.

The report is organized as follows. Section II introduces the details on the model, including a definition of the parameterization. Section III discusses the maximum likelihood algorithm, while section IV presents the CRB. The report ends with a simulation study and conclusions in sections V and VI, respectively.

The conditions of the report, are ordered C1), C2), and so on.

II **The ODE model and its parameterization**

A. Measurements and modeled signals

The starting point is the discrete time measured signal $z(t)$, where

$$z(t) = y(t) + e(t). \tag{1}$$

Here $y(t)$ is the signal to be modeled, $e(t)$ is the discrete time measurement noise and t denotes discrete time. It is assumed here that $y(t)$ is periodic, i.e.

C1) $y(t+T) = y(t)$, $\forall t \in R$, $0 < T < \infty$.

Furthermore, $e(t)$ is assumed to be zero mean Gaussian white noise, i.e.

C2) $e(t) \in N(0, \sigma^2)$, $E[e(t)e(t+kT_s)] = \delta_{k,0}\sigma^2$

where $E[.]$ denotes the expectation operator and where T_S is the sampling period.

B. *Model structures*

As stated above, the main idea of the report is to model the generation of the signal $y(t)$ by means of an ordinary differential equation of order two, as shown in (2)

$$\begin{pmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \end{pmatrix} = \begin{pmatrix} f_1(x_1(t), x_2(t), \boldsymbol{\theta}_1) \\ f_2(x_1(t), x_2(t), \boldsymbol{\theta}_2) \end{pmatrix}$$
$$z(t) = \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + e(t). \qquad (2)$$

In (2) $\begin{pmatrix} x_1(t) & x_2(t) \end{pmatrix}^T$ is the state vector and $\begin{pmatrix} c_1 & c_2 \end{pmatrix}$ is the vector containing the *selected* output weighting factors that generate $y(t)$ when multiplied with the state vector. Furthermore

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1^T & \boldsymbol{\theta}_2^T \end{pmatrix}^T \qquad (3)$$

is the unknown parameter vector(s). Note that it is proved in Wigren and Söderström (2003a) that an ODE order of 2 is sufficient for modeling of a relatively wide class of periodic signals.

At this point it is highly relevant to pose the question whether the model structure (2) may be too general, i.e. to consider model structure selection for the problem at hand. In Wigren *et al.* (2003b) it is assumed that the second order ODE

$$\frac{d^2y}{dt^2} = f_2\left(y, \frac{dy}{dt}, \boldsymbol{\theta}_2\right) \tag{4}$$

generates the periodic *signal that is actually measured*. This fact allows the state variable selection

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y \\ dy/dt \end{pmatrix}. \tag{5}$$

This results in the state space model

$$\begin{aligned} \begin{pmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \end{pmatrix} &= \begin{pmatrix} x_2(t) \\ f_2\left(x_1(t), x_2(t), \boldsymbol{\theta}_2\right) \end{pmatrix} \\ z(t) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + e(t). \end{aligned} \tag{6}$$

The model (6) depends only on the parameters of the second right hand side function of (2), a fact that should improve the performance of algorithms based on this model. The analysis of Wigren and Söderström (2003a) also suggests that one right hand function is sufficient. However, this choice of model structure is by no means a necessity. Future developments may very well result in further motivation for the use of (2). In order to obtain results of general validity, the treatment of this report will therefore deal with both (2) and (6).

C. *Parameterization*

A natural approach is now to expand the right hand side of the state equations of (2) as well as the second state equation of (6) in terms of a polynomial model. The parameterizations are given by

$$f_1\left(x_1(t), x_2(t), \boldsymbol{\theta}_1\right) = \sum_{l=0}^{L_1}\sum_{m=0}^{M_1}\theta_{1,l,m}x_1^l(t)x_2^m(t)$$

$$f_2\left(x_1(t), x_2(t), \boldsymbol{\theta}_2\right) = \sum_{l=0}^{L_2}\sum_{m=0}^{M_2}\theta_{2,l,m}x_1^l(t)x_2^m(t) \qquad (7)$$

$$f_1\left(x_1(t), x_2(t)\right) = x_2(t)$$

$$f_2\left(x_1(t), x_2(t), \boldsymbol{\theta}_2\right) = \sum_{l=0}^{L_2}\sum_{m=0}^{M_2}\theta_{2,l,m}x_1^l(t)x_2^m(t), \qquad (8)$$

where (7) corresponds to (2) and (8) corresponds to (6).

Remark 1: No scale factor problems are expected with the parameterization of (6). The reason is that the transformation $y' = k_0 y$ transforms (4) to

$$\frac{d^2y'}{dt^2} = k_0 f_2\left(\frac{1}{k_0}y', \frac{1}{k_0}\frac{dy'}{dt}, \boldsymbol{\theta}_2\right). \qquad (9)$$

Since the polynomial model is a general function expansion, the function of (9) can be modeled equally well as the function of (4).

D. Discretization

In order to formulate complete discrete time models, the continuous time ODE models (2), (7) and (6), (8) need to be discretized. This is done by exploiting an Euler forward numerical integration scheme. For simplicity, the discretization interval is selected to be equal to the sampling period, resulting in

$$x_1(t+T_S) = x_1(t) + T_S \sum_{l=0}^{L_1} \sum_{m=0}^{M_1} \theta_{1,l,m} x_1^l(t) x_2^m(t)$$

$$x_2(t+T_S) = x_2(t) + T_S \sum_{l=0}^{L_2} \sum_{m=0}^{M_2} \theta_{2,l,m} x_1^l(t) x_2^m(t) \qquad (10)$$

and

$$x_1(t+T_S) = x_1(t) + T_S x_2(t)$$

$$x_2(t+T_S) = x_2(t) + T_S \sum_{l=0}^{L_2} \sum_{m=0}^{M_2} \theta_{2,l,m} x_1^l(t) x_2^m(t) \qquad (11)$$

respectively. These parameterized models can be more compactly written, by the introduction of the quantities

$$\boldsymbol{\varphi}_1(x_1(t), x_2(t)) = T_S \begin{pmatrix} 1 & \dots & x_2^{M_1}(t) & \dots & x_1^{L_1}(t) & \dots & x_1^{L_1}(t) x_2^{M_1}(t) \end{pmatrix}^T$$

$$\boldsymbol{\varphi}_2(x_1(t), x_2(t)) = T_S \begin{pmatrix} 1 & \dots & x_2^{M_2}(t) & \dots & x_1^{L_2}(t) & \dots & x_1^{L_2}(t) x_2^{M_2}(t) \end{pmatrix}^T \qquad (12)$$

$$\boldsymbol{\theta}_1 = \begin{pmatrix} \theta_{1,0,0} & \dots & \theta_{1,0,M_1} & \dots & \theta_{1,L_1,0} & \dots & \theta_{1,L_1,M_1} \end{pmatrix}^T$$

$$\boldsymbol{\theta}_2 = \begin{pmatrix} \theta_{2,0,0} & \dots & \theta_{2,0,M_2} & \dots & \theta_{2,L_2,0} & \dots & \theta_{2,L_2,M_2} \end{pmatrix}^T. \qquad (13)$$

The two models that follow from (10) and (11) then become

$$\begin{aligned} x_1(t+T_S) &= x_1(t) + \boldsymbol{\varphi}_1^T(x_1(t), x_2(t))\boldsymbol{\theta}_1 \\ x_2(t+T_S) &= x_2(t) + \boldsymbol{\varphi}_2^T(x_1(t), x_2(t))\boldsymbol{\theta}_2 \\ z(t) &= x_1(t) + e(t) \end{aligned} \qquad (14)$$

$$
\begin{aligned}
x_1(t+T_S) &= x_1(t) + T_S x_2(t) \\
x_2(t+T_S) &= x_2(t) + \boldsymbol{\varphi}_2^T\left(x_1(t), x_2(t)\right)\boldsymbol{\theta}_2 \\
z(t) &= x_1(t) + e(t).
\end{aligned} \qquad (15)
$$

III **The maximum likelihood method**

Before proceeding with the development of the criterion function, it is noted that (14) and (15) can be simultaneously treated by a use of the model

$$
\begin{aligned}
&\mathbf{x}(t+T_S) = \mathbf{F}\mathbf{x}(t) + \boldsymbol{\Phi}\left(\mathbf{x}(t), T_S\right)\boldsymbol{\theta}, \qquad \mathbf{x}(t_0) = \mathbf{x}_{t_0} \\
&z(t) = \mathbf{H}^T\mathbf{x}(t) + e(t)
\end{aligned} \qquad (16)
$$

.

This follows by the selections

$$
\begin{aligned}
&\mathbf{F} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&\boldsymbol{\Phi}\left(\mathbf{x}(t), T_S\right) = \begin{pmatrix} \boldsymbol{\varphi}_1^T\left(\mathbf{x}(t), T_S\right) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_2^T\left(\mathbf{x}(t), T_S\right) \end{pmatrix} \\
&\boldsymbol{\theta}^T = \begin{pmatrix} \boldsymbol{\theta}_1^T & \boldsymbol{\theta}_2^T \end{pmatrix}
\end{aligned} \qquad (17)
$$

and

$$
\begin{aligned}
&\mathbf{F} = \begin{pmatrix} 1 & T_S \\ 0 & 1 \end{pmatrix} \\
&\boldsymbol{\Phi}\left(\mathbf{x}(t), T_S\right) = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\varphi}_2^T\left(\mathbf{x}(t), T_S\right) \end{pmatrix} \\
&\boldsymbol{\theta} = \boldsymbol{\theta}_2
\end{aligned} \qquad (18)
$$

respectively.

Proceeding with the statement of the maximum likelihood criterion it is observed that the measurement disturbance $e(t)$ in (16) is assumed to be Gaussian, i.e. C2) is assumed to hold. Note also

that the initial values of the states need to be estimated together with the unknown parameter vector for the problem to be meaningful. It follows that the likelihood function equals

$$p\left(\mathbf{Z}^N|\boldsymbol{\theta},\mathbf{x}_{t_0}\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\left(z(t_0)-\mathbf{H}^T\mathbf{x}_{t_0}\right)^2}{2\sigma^2}}\frac{1}{(2\pi)^{N/2}\sigma^N}\prod_{k=1}^{N}e^{-\frac{\left(z(t_0+kT_S)-\mathbf{H}^T\mathbf{x}\left(t_0+kT_S,\boldsymbol{\theta},\mathbf{x}_{t_0}\right)\right)^2}{2\sigma^2}} \qquad (19)$$

where $\mathbf{Z}^N$ denotes the set of all measurements and N denotes the number of measurements. The dependence on the parameter vector and the initial value enter implicitly in (19), via the modeled states $\mathbf{x}\left(t_0+kT_S,\boldsymbol{\theta},\mathbf{x}_{t_0}\right)$ at the sampling instances. The generation of the model states is perhaps best understood by referring to the criterion minimization strategy that is outlined in the end of this section. There the model (16) is iterated to generate a complete state trajectory (of model states). This is done for each fixed set of initial values $\mathbf{x}_{t_0}$ and parameters $\boldsymbol{\theta}$, that corresponds to one specific iteration of the criterion minimization procedure.

Taking logarithms, scaling and changing signs, it is straightforward to see that the maximization of (19) is equivalent to the following minimization problem

$$\left(\hat{\boldsymbol{\theta}}_{ML}^T\quad\hat{\mathbf{x}}_{t_0}^T\right)^T=\arg\min_{\boldsymbol{\theta},\mathbf{x}_{t_0}}\bar{L}\left(\boldsymbol{\theta},\mathbf{x}_{t_0}\right) \qquad (20)$$

where

$$\bar{L}\left(\boldsymbol{\theta},\mathbf{x}_{t_0}\right)=\left(z(t_0)-\mathbf{H}^T\mathbf{x}_{t_0}\right)^2+\sum_{k=1}^{N}\left(z\left(t_0+kT_S\right)-\mathbf{H}^T\mathbf{x}\left(t_0+kT_S,\boldsymbol{\theta},\mathbf{x}_{t_0}\right)\right)^2 \qquad (21)$$

Typically , the minimization of the criterion has to be carried out in several steps since (20) seldom has a unique minimum point and since a dense grid search in a high dimensional parameter space is generally not feasible. Note that if a grid search would be applied, (16) would have to be iterated from t_0 to t_0+NT_S once for each grid point, using the parameters and initial values defined by the grid point.

One tentative multistep criterion minimization procedure could follow the steps:

1. Use any of the low complexity algorithms of Wigren *et al.* (2003b,c) (least squares method or the Kalman filter), possibly followed by the extended Kalman filter (Wigren *et al.*, 2003b), to compute an initial estimate.
2. Possibly perform a grid search around this initial estimate in order to refine it further.
3. Perform a final gradient or Gauss-Newton iterative search, using the refined initial estimate as initial values to the iteration.

Obviously, there are many ways to vary this theme. It is e.g. possible to compute gradients by forming differences numerically between two trajectories of (16). Alternatively, sensitivity derivatives could be computed analytically and integrated.

IV **The Cramer Rao Bound**

The accuracy of the estimates obtained will depend on the evolution of the trajectories of the nonlinear ODE that describes the system. The observability of the parameters will vary with the state of the ODE during the orbit. The approach taken to compute the CRB in this report allows for handling of orbits that are only asymptotically periodic, a situation that naturally arises because of transient phenomena. These effects are modeled by the inclusion of the initial values of the ODE in the CRB computation.

The model used in the CRB computation is selected as the most general of the models used for estimation, i.e. (2) in combination with (7) The CRB for the model (6) in combination with (8) follows by exclusion of the relevant matrix blocks of the obtained result for (2) and (6). The model used for the development of the CRB is hence

$$\begin{pmatrix} \dfrac{dx_1(t,\boldsymbol{\theta},\boldsymbol{\psi})}{dt} \\ \dfrac{dx_2(t,\boldsymbol{\theta},\boldsymbol{\psi})}{dt} \end{pmatrix} = \begin{pmatrix} f_1\left(x_1(t,\boldsymbol{\theta},\boldsymbol{\psi}), x_2(t,\boldsymbol{\theta},\boldsymbol{\psi}), \boldsymbol{\theta}_1\right) \\ f_2\left(x_1(t,\boldsymbol{\theta},\boldsymbol{\psi}), x_2(t,\boldsymbol{\theta},\boldsymbol{\psi}), \boldsymbol{\theta}_2\right) \end{pmatrix}$$

$$z(t,\boldsymbol{\theta},\boldsymbol{\psi}) = \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} x_1(t,\boldsymbol{\theta},\boldsymbol{\psi}) \\ x_2(t,\boldsymbol{\theta},\boldsymbol{\psi}) \end{pmatrix} + e(t) \qquad (22)$$

with the initial values

$$\begin{pmatrix} x_1(t_0) \\ x_2(t_0) \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \quad = \boldsymbol{\psi}. \tag{23}$$

The measurements are assumed to be discrete time and available at times $t_1 + T_S, \ldots, t_1 + NT_S$.

The objective is now to compute (see e.g. Söderström and Stoica (1989))

$$\mathbf{CRB}^{-1} = E\left[\begin{pmatrix} \left((\log L)_{\boldsymbol{\theta}_1}\right)^T \\ \left((\log L)_{\boldsymbol{\theta}_2}\right)^T \\ \left((\log L)_{\boldsymbol{\psi}_1}\right)^T \\ \left((\log L)_{\boldsymbol{\psi}_2}\right)^T \end{pmatrix} \begin{pmatrix} (\log L)_{\boldsymbol{\theta}_1} & (\log L)_{\boldsymbol{\theta}_2} & (\log L)_{\boldsymbol{\psi}_1} & (\log L)_{\boldsymbol{\psi}_2} \end{pmatrix}\right]. \tag{24}$$

As usual, the CRB is evaluated for the true parameter vector. The subscripts are used to denote partial differentiation. Since the measurement noise is Gaussian by C2), the likelihood function, L, is up to a constant given by

$$-\log L(\boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{1}{2\sigma^2} \sum_{k=1}^{N} \left(z(t_1 + kT_S) - c_1 x_1(t_1 + kT_S, \boldsymbol{\theta}, \boldsymbol{\psi}) - c_2 x_2(t_1 + kT_S, \boldsymbol{\theta}, \boldsymbol{\psi})\right)^2. \tag{25}$$

The expectation of (24) can then be evaluated, by exploiting that for the true parameter vector, the output error is white Gaussian noise. The components of the expectation of (24) are dependent on the partial derivatives of the states with respect to the unknowns (the sensitivity derivatives). This follows directly from (24). The computation of the sensitivity derivatives can be performed by partial differentiation of both sides of the original differential equation (22) with respect to the unknowns. The result is a new set of ODEs that need to be integrated together with (22), in order to compute the quantities needed in the computation of the components of (24). The initial values needed for the solution follow by partial differentiation of (23). The calculations are straightforward but tedious. The results are displayed in detail in the appendix.

The calculation of the CRB can now be summarized.

1. Solve the system of differential equations given by (22) together with (39)-(46), using the initial conditions (23), (47) and (48). This system is complete and can be solved with any high accuracy numerical routine for solving ODEs. The solution is computed on $\left[t_0, t_{end}\right] \supseteq \left[t_1 + T_S, t_1 + NT_S\right]$.
2. Compute the elements of the Fischer information matrix (29)-(38) from the computed trajectories.
3. Invert (24) to finalize the computation.

V **Simulation study**

The focus of this simulation study is on accuracy. Thus, other important issues like convergence, local minima and the use of multiple algorithms for initialization purposes (outlined above) are not covered. The two main purposes are to study the theoretical accuracies achievable by means of the CRB *and to* compare the maximum likelihood method performance to this bound. In so doing, the following model of the Van der Pol oscillator was used

$$\begin{pmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \end{pmatrix} = \begin{pmatrix} x_2(t) \\ -x_1(t) + 2\left(1 - \left(x_1(t)\right)^2\right)x_2(t) \end{pmatrix}. \qquad (26)$$

The model structure given by (6) was used, and the parameters $\theta_{0,1}$, $\theta_{1,0}$ and $\theta_{2,1}$ were estimated, together with the initial values. The remaining parameters were fixed to 0. This means that 3 parameters plus 2 initial values are estimated by the maximum likelihood algorithm and that 3 parameters plus 2 initial values are included in the CRB calculation. The reason for restricting the number of parameters was a need to restrict the run time of the simulations, convergence of the steepest descent algorithm (27) required more than 2000 iterations (this is not an uncommon number in practice, see Luenberger (1973)). The components corresponding to $\boldsymbol{\theta}_1$ in (24) disappear since here the CRB was computed from the restricted model (6) here. The state initial values were $x_1(0) = x_2(0) = 0.5$. The sampling time in all experiments was $T_S = 0.01s$.

The maximum likelihood criterion (20) was minimized with the following steepest descent algorithm

$$\begin{pmatrix} \boldsymbol{\theta}_{n+1} \\ \mathbf{x}_{t_0,n+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}_{n} \\ \mathbf{x}_{t_0,n} \end{pmatrix} - \frac{\left(\nabla\bar{L}\left(\boldsymbol{\theta}_n, \mathbf{x}_{t_0,n}\right)\right)^T \left(\nabla\bar{L}\left(\boldsymbol{\theta}_n, \mathbf{x}_{t_0,n}\right)\right)}{\left(\nabla\bar{L}\left(\boldsymbol{\theta}_n, \mathbf{x}_{t_0,n}\right)\right)^T \mathbf{H}\left(\boldsymbol{\theta}_n, \mathbf{x}_{t_0,n}\right)\left(\nabla\bar{L}\left(\boldsymbol{\theta}_n, \mathbf{x}_{t_0,n}\right)\right)} \nabla\bar{L}\left(\boldsymbol{\theta}_n, \mathbf{x}_{t_0,n}\right). \quad (27)$$

Here $\mathbf{H}\left(\boldsymbol{\theta}, \mathbf{x}_{t_0}\right)$ denotes the Hessian of $L\left(\boldsymbol{\theta}, \mathbf{x}_{t_0}\right)$ and $\nabla L\left(\boldsymbol{\theta}, \mathbf{x}_{t_0}\right)$ is the gradient of $L\left(\boldsymbol{\theta}, \mathbf{x}_{t_0}\right)$. The reader is referred to Luenberger (1973) , p. 150 and p.154 for further details. As stated above, the focus is on accuracy. Hence, to avoid problems with local minima of the criterion function, the algorithm was initiated close to the true parameter vector $\left(\boldsymbol{\theta}_0^T \quad \mathbf{x}_{t_0}^T\right)^T$.

Example 1. This example evaluates the CRB and the performance of the maximum likelihood method as a function of the SNR of the data. To evaluate the performance 50 Monte Carlo experiments, each using a data length of 2000 samples, were conducted for a number of different SNRs. The algorithm (27) was initiated according to

$$\begin{pmatrix} \boldsymbol{\theta} \\ \mathbf{x}_{t_0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}_0 \\ \mathbf{x}_{t_0} \end{pmatrix} - 5\boldsymbol{\sigma}_{CRB}. \quad (28)$$

Note again that each iteration of (27) requires a solution of the ODE and the corresponding sensitivity derivatives that appear in the appendix.

The results of the evaluation appear in Fig. 1 - Fig. 5. The CRB is plotted solid while the performance of the maximum likelihood method is plotted dashed. It can be seen that the (relevant) parameters are estimated with accuracies that are close to the CRB. This is to be expected, considering the fact that the maximum likelihood method is known to reach the CRB asymptotically under mild conditions (Note that this has not been formally proved for the method studied in this report). It can also be observed that the closeness to the CRB begins to deteriorate below SNR:s of 10 dB. The numerical values do however indicate that the algorithm continues to provide useful estimates even below 0 dB. Finally, the CRBs roll off linearly with the SNR in a loglog plot. This is consistent with the fact that the standard deviation enters as a factor in the CRB.

Example 2. This example evaluates the CRB and the performance of the maximum likelihood method as a function of the number of samples available for estimation. As compared to Example 1, the following differences apply. First, all results are evaluated for a fixed SNR of 30 dB. Secondly, the data lengths were chosen so as to include an integer number of complete periods. The reason for this is to provide estimates that are balanced with respect to the state space signal energy available for estimation. Unless this would be the case, the maximum likelihood method can be expected to be biased for finite data sets, thus distorting the intended accuracy assessment.

The results of the evaluation appears in Fig. 6- Fig. 10. It can be observed that the accuracy of the estimated parameters approach the CRB when the available amount of data increases. This is consistent with the conjecture that the maximum likelihood method in this case should approach the CRB. The performance for the initial values is worse. It can even be observed that the accuracy of the algorithm for the initial values gets worse when the data set is increased above a certain number of samples. This is believed to be a result of the fact that the effect of the initial values on the signals decay with time - hence the signal energy available for estimation of the initial values are located at the beginning of the data set. Therefore the overall signal to noise ratio for the estimation of these parameters starts to decrease after a certain period of time since only noise enters the estimation algorithm as far as the initial values are concerned. The result could then be a drift of these parameters. Note also that the initial values are not expected to be unbiased since the signal energy related to the initial values are finite, even when the data length turns to infinity. Hence, efficiency of the maximum likelihood method for these parameters cannot be expected.

Some further comments related to the simulations are in order. First, it is very important to apply the ODE solvers in a correct manner. One problem encountered was the fact that the inaccuracy of the MATLAN ODE solvers increased when the high order ODE's related to the CRB and the maximum likelihood gradient were solved. Numerical instability was another problem. The problems were solved by explicit control of the required inaccuracy, and by the use of another solver than ODE45 for the CRB and gradient calculations.

VI **Conclusions**

The report has presented a novel approach to the modeling of periodic signals. The main idea is to model the signal as being generated by a second order nonlinear ordinary differential equation with periodic orbits.

The estimated quantities are parameters that describe the right hand sides of the differential equation. In the present setting, a linear in the parameters polynomial model was used. Based on this model a maximum likelihood method was derived. The Cramer Rao bound for the selected model structure was also derived. It is computed via the solution of a large system of ordinary differential equations. The algorithm was tested in a simulation study. A comparison to the CRB was also performed. The algorithms seems to give a performance very close to the CRB, approaching the CRB when the number of samples increases. This fact can probably be proved, although it is beyond the scope of the present report.

Many other open topics for further research exist. Of particular importance is the derivation of methods that test estimated models for the existence of periodic orbits in certain regions of R^2. Other topics include treatment of parameterizations other than the polynomial one and an analysis of the asymptotic performance of the proposed methods.

VII **References**

Abd-Elrady, E. (2002). An adaptive grid point RPEM algorithm for harmonic signal modeling. *Proc. 15:th IFAC World Congress,* Barcelona, Spain.

Abd-Elrady, E. (2004). A nonlinear approach to harmonic signal modeling.. *Signal Processing*, **84**, 163-195.

Fitz, M. P. (1994). Further results in the fast estimation of a single frequency. *IEEE Trans. Communications,* **42**, 862-864.

Händel, P. and P. Tichavsky (1994). Adaptive estimation for periodic signal enhancement and tracking. *Int. J. Adaptive Control, Signal Processing,* **8**, 447-456.

Khalil, H. K. (1996). *Nonlinear Systems, Second Edition.* Prentice Hall, Upper Saddle River, NJ.

Kumaresan, R. and D. W. Tufts (1982). Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Trans. Acoust. Speech, Signal Processing,* **30**, 833-840.

Li, J. and P. Stoica (1996). An adaptive filtering approach to spectral estimation and SAR imaging. *IEEE Trans. Signal Processing,* **44**, 1469-1484.

Luenberger, D. G. (1973). *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA.

Nehorai, A. and B. Porat (1986). Adaptive comb filtering for harmonic signal enhancement. *IEEE Trans. Acoust., Speech, Signal Processing*, **34**, 1124-1138.

Porat, B. and B. Friedlander (1987). On the accuracy of the Kumaresan-Tufts method for estimating complex damped exponentials. *IEEE Trans. Acoust., Speech, Signal Processing,* **35**, 231-235.

Rife, D. C. and R. R. Boorstyn (1976). Multiple tone parameter estimation from discrete-time observations., *Bell Syst. Tech. J.,* **55**, 1389-1410.

Stoica, P. and R. Moses (1997). *Introduction To Spectral Analysis.* Prentice Hall, Upper Saddle River, NJ.

Stoica, P. and A. Nehorai (1989). Statistical analysis of two nonlinear estimators of sine wave parameters in the colored noise case. *Circ. Syst. Signal Proc.*, **8**, 3-15.

Söderström, T. and P. Stoica (1989). *System Identification.* Prentice Hall Int., Hemel Hempstead, UK.

Tichavsky, P. and P. Händel (1997). Recursive estimation of frequencies and frequency rates of multiple cisoids in noise. *Signal Processing,* **58**, 117-129.

Wigren, T. and P. Händel (1996). Harmonic signal modeling using adaptive nonlinear function estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing,* Atlanta, GA, USA.

Wigren, T. and T. Söderström (2003a). An ODE order of 2 is sufficient for modeling of many periodic signals. *Submitted to IEEE Trans. Automat. Contr.*

Wigren, T., E. Abd-Elrady and T. Söderström (2003b), "Harmonic signal analysis with Kalman filters using periodic orbits of nonlinear ODEs. *ICASSP 2003*, Hongkong, China, 669-672.

Wigren, T., E. Abd-Elrady and T. Söderström (2003c). Least squares harmonic signal analysis using periodic orbits of ODEs. *Proc. IFAC SYSID 2003, Rotterdam, the Netherlands,* 1584-1589.

A. **Appendix - Derivation of the details of the CRB**

The elements of the CRB are first evaluated, exploiting the symmetry of (24). The elements become, using the whiteness of the noise

$$E\left[\left((\log L)_{\boldsymbol{\theta}_1}\right)^T\left((\log L)_{\boldsymbol{\theta}_1}\right)\right]=$$

$$\sum_{k=1}^{N}\left(\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}^T \quad \left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}^T\right)\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}\end{pmatrix}\frac{1}{\sigma^2} \qquad (29)$$

$$E\left[\left((\log L)_{\boldsymbol{\theta}_2}\right)^T\left((\log L)_{\boldsymbol{\theta}_1}\right)\right]=$$

(30)

$$\sum_{k=1}^{N}\left(\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}^{T}\quad\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}^{T}\right)\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}\end{pmatrix}\frac{1}{\sigma^2}$$

$$E\left[\left(\left(\log L\right)_{\boldsymbol{\psi}_1}\right)^{T}\left(\left(\log L\right)_{\boldsymbol{\theta}_1}\right)\right]=$$

$$\sum_{k=1}^{N}\left(\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\quad\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\right)\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}\end{pmatrix}\frac{1}{\sigma^2}\qquad(31)$$

$$E\left[\left(\left(\log L\right)_{\boldsymbol{\psi}_2}\right)^{T}\left(\left(\log L\right)_{\boldsymbol{\theta}_1}\right)\right]=$$

$$\sum_{k=1}^{N}\left(\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}\quad\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}\right)\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}\end{pmatrix}\frac{1}{\sigma^2}\qquad(32)$$

$$E\left[\left(\left(\log L\right)_{\boldsymbol{\theta}_2}\right)^{T}\left(\left(\log L\right)_{\boldsymbol{\theta}_2}\right)\right]=$$

$$\sum_{k=1}^{N}\left(\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}^{T}\quad\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}^{T}\right)\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}\end{pmatrix}\frac{1}{\sigma^2}\qquad(33)$$

$$E\left[\left(\left(\log L\right)_{\boldsymbol{\psi}_1}\right)^{T}\left(\left(\log L\right)_{\boldsymbol{\theta}_2}\right)\right]=$$

$$\sum_{k=1}^{N}\left(\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\quad\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\right)\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}\end{pmatrix}\frac{1}{\sigma^2}\qquad(34)$$

$$E\left[\left(\left(\log L\right)_{\boldsymbol{\psi}_2}\right)^T\left(\left(\log L\right)_{\boldsymbol{\theta}_2}\right)\right]=$$

$$\sum_{k=1}^{N}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2} & \left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}\end{pmatrix}\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}\end{pmatrix}\frac{1}{\sigma^2} \qquad (35)$$

$$E\left[\left(\left(\log L\right)\right)_{\boldsymbol{\psi}_1}^T\left(\left(\log L\right)_{\boldsymbol{\psi}_1}\right)\right]=$$

$$\sum_{k=1}^{N}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1} & \left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\end{pmatrix}\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\end{pmatrix}\frac{1}{\sigma^2} \qquad (36)$$

$$E\left[\left(\left(\log L\right)\right)_{\boldsymbol{\psi}_2}^T\left(\left(\log L\right)_{\boldsymbol{\psi}_1}\right)\right]=$$

$$\sum_{k=1}^{N}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2} & \left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}\end{pmatrix}\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}\end{pmatrix}\frac{1}{\sigma^2} \qquad (37)$$

$$E\left[\left(\left(\log L\right)\right)_{\boldsymbol{\psi}_2}^T\left(\left(\log L\right)_{\boldsymbol{\psi}_2}\right)\right]=$$

$$\sum_{k=1}^{N}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2} & \left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}\end{pmatrix}\begin{pmatrix}c_1\\c_2\end{pmatrix}\begin{pmatrix}c_1 & c_2\end{pmatrix}\begin{pmatrix}\left(x_1\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}\\\left(x_2\left(t_1+kT_S,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}\end{pmatrix}\frac{1}{\sigma^2}. \qquad (38)$$

This completes the derivation of the different elements of the Fischer information matrix. It remains to compute the partial derivatives of the states with respect to the unknowns, i.e. the sensitivity derivatives. Towards this end, it is noted that a straightforward partial differentiation of (22) results in

$$\frac{d\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}}{dt}=\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{x_1}\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}$$
$$+\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{x_2}\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}+\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{\boldsymbol{\theta}_1} \quad (39)$$

$$\frac{d\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}}{dt}=\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{x_1}\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}$$
$$+\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{x_2}\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2} \quad (40)$$

$$\frac{d\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}}{dt}=\left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{x_1}\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1}$$
$$+\left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{x_2}\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_1} \quad (41)$$

$$\frac{d\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}}{dt}=\left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{x_1}\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}$$
$$+\left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{x_2}\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\theta}_2}+\left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{\boldsymbol{\theta}_2} \quad (42)$$

$$\frac{d\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}}{dt}=\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{x_1}\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}$$
$$+\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{x_2}\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1} \quad (43)$$

$$\frac{d\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}}{dt}=\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{x_1}\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}$$
$$+\left(f_1\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_1\right)\right)_{x_2}\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2} \quad (44)$$

$$\frac{d\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}}{dt} = \left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{x_1}\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1}$$
$$+\left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{x_2}\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_1} \quad (45)$$

$$\frac{d\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}}{dt} = \left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{x_1}\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}$$
$$+\left(f_2\left(x_1\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right),\boldsymbol{\theta}_2\right)\right)_{x_2}\left(x_2\left(t,\boldsymbol{\theta},\boldsymbol{\psi}\right)\right)_{\boldsymbol{\psi}_2}. \quad (46)$$

The initial conditions to these equations follow by differentiation of (23):

$$\left(x_1\left(t_0\right)\right)_{\boldsymbol{\theta}_1} = \left(x_2\left(t_0\right)\right)_{\boldsymbol{\theta}_1} = \mathbf{0}$$
$$\left(x_1\left(t_0\right)\right)_{\boldsymbol{\theta}_2} = \left(x_2\left(t_0\right)\right)_{\boldsymbol{\theta}_2} = \mathbf{0} \quad (47)$$

$$\begin{aligned}\left(x_1\left(t_0\right)\right)_{\boldsymbol{\psi}_1} &= 1\\ \left(x_2\left(t_0\right)\right)_{\boldsymbol{\psi}_1} &= 0\\ \left(x_1\left(t_0\right)\right)_{\boldsymbol{\psi}_2} &= 0\\ \left(x_2\left(t_0\right)\right)_{\boldsymbol{\psi}_2} &= 1.\end{aligned} \quad (48)$$

This completes the derivation of the CRB.

Figure Captions

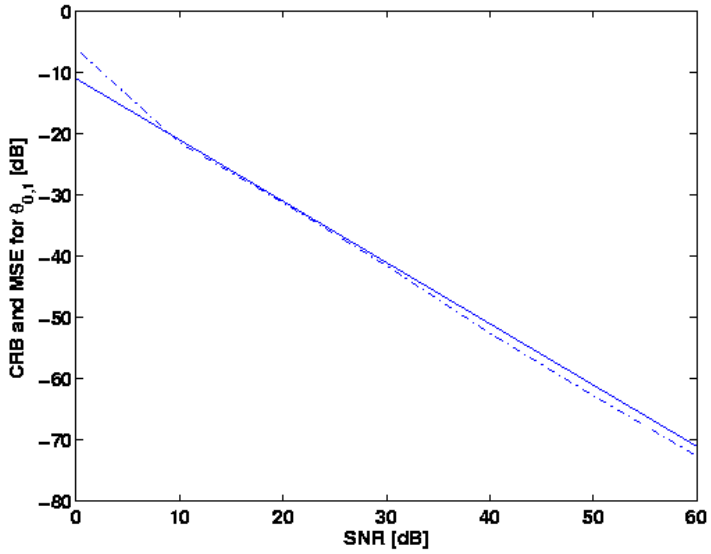

Figure 1: CRB (solid) and performance of the maximum likelihood method (dashes) for the parameter $\theta_{0,1}$.

SNR is varied.

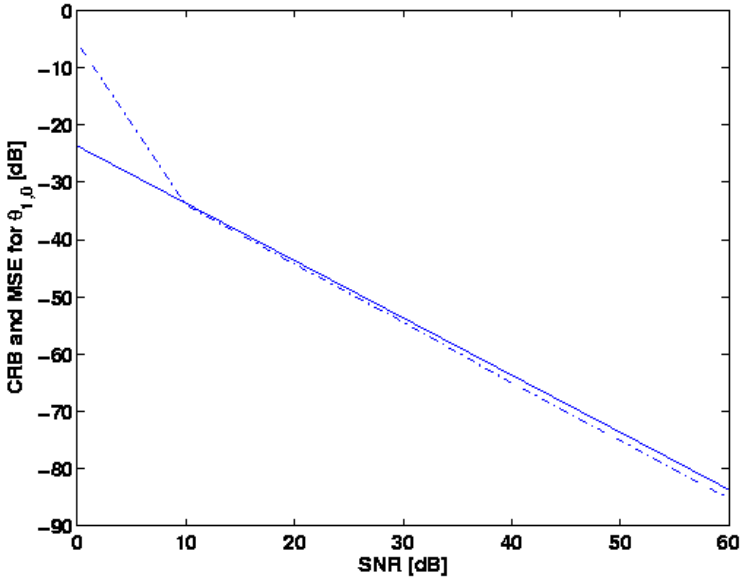

Figure 2: CRB (solid) and performance of the maximum likelihood method (dashes) for the parameter $\theta_{1,0}$.

SNR is varied.

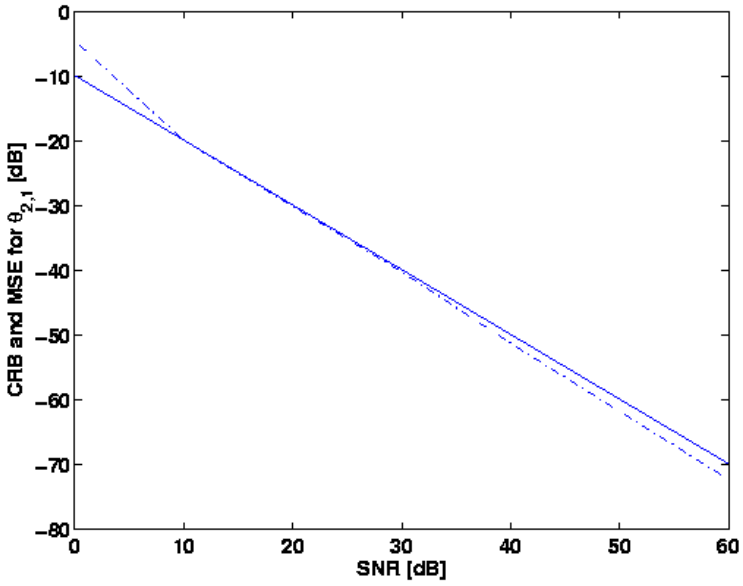

Figure 3: CRB (solid) and performance of the maximum likelihood method (dashes) for the parameter $\theta_{2,1}$. SNR is varied.

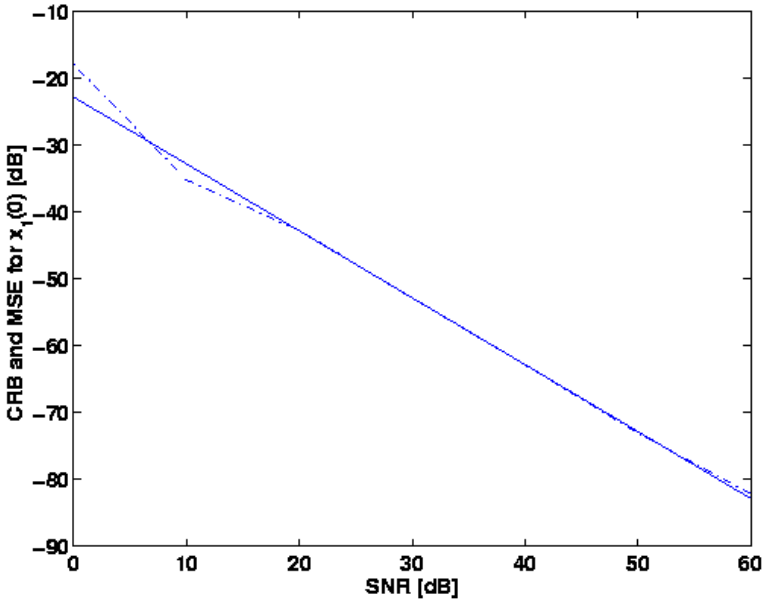

Figure 4: CRB (solid) and performance of the maximum likelihood method (dashes) for the initial value $x_1(0)$. SNR is varied.

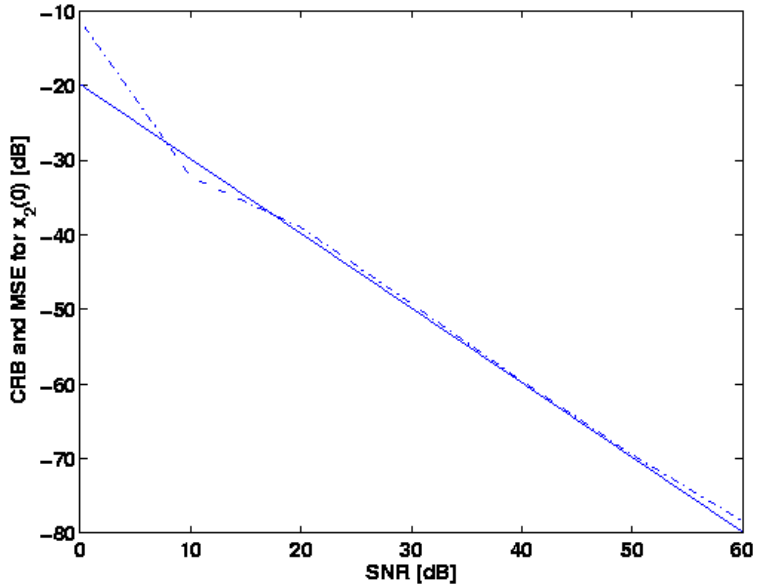

Figure 5: CRB (solid) and performance of the maximum likelihood method (dashes) for the parameter $x_2(0)$. SNR is varied.

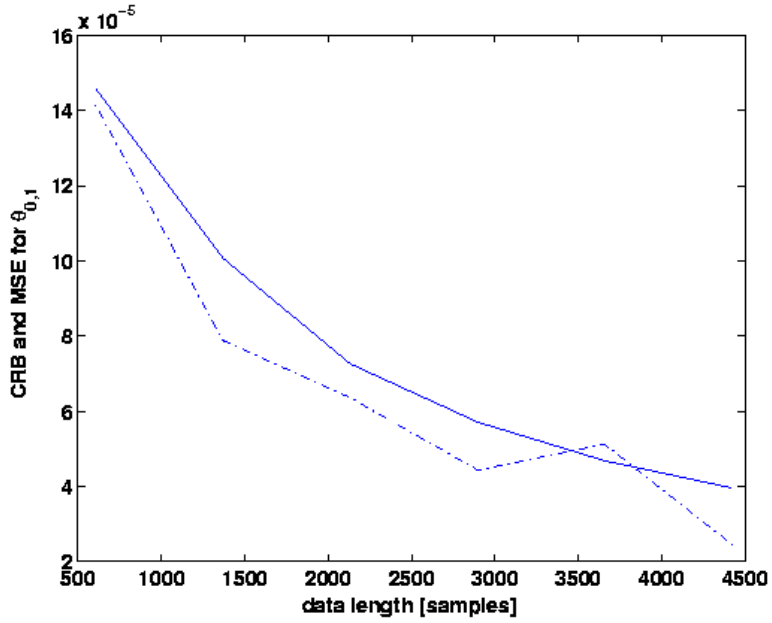

Figure 6: CRB (solid) and performance of the maximum likelihood method (dashes) for the parameter $\theta_{0,1}$. Data length is varied.

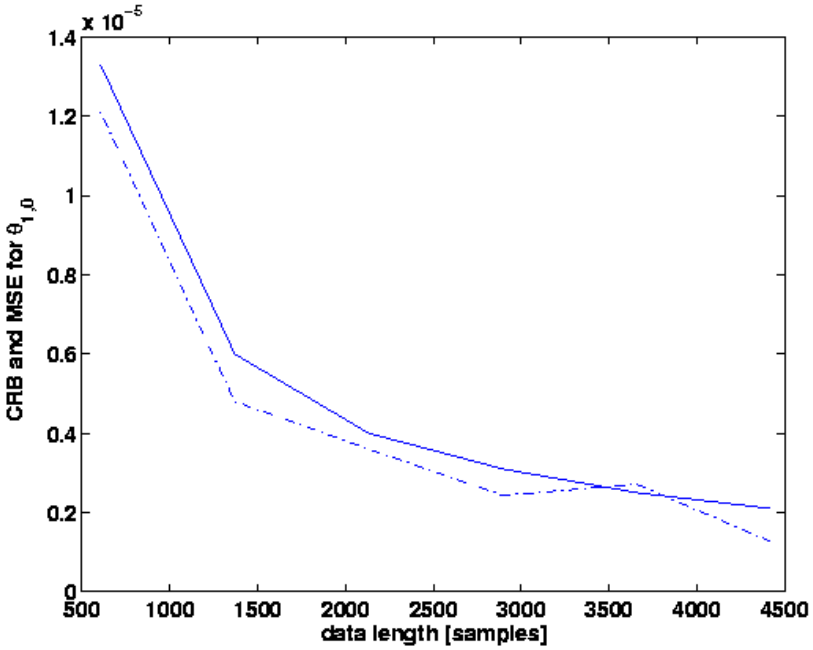

Figure 7: CRB (solid) and performance of the maximum likelihood method (dashes) for the parameter $\theta_{1,0}$.

Data length is varied.

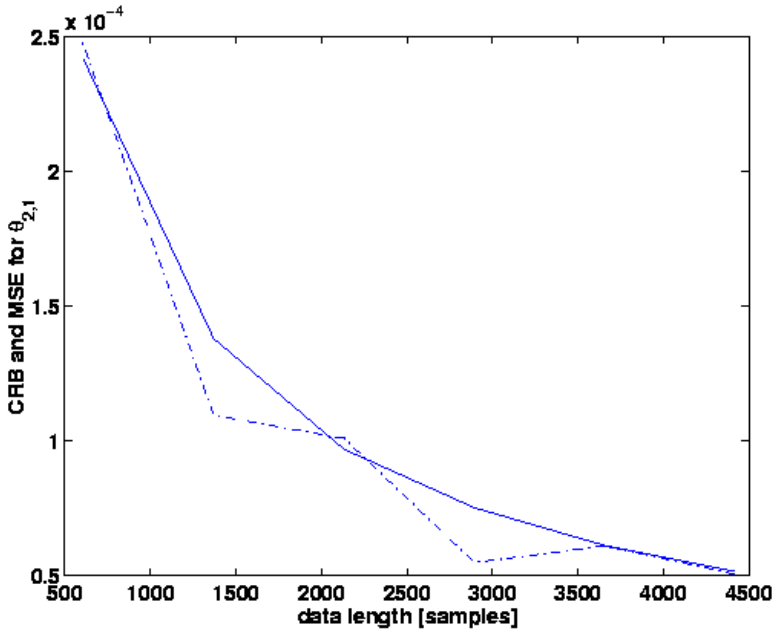

Figure 8: CRB (solid) and performance of the maximum likelihood method (dashes) for the parameter $\theta_{2,1}$.

Data length is varied.

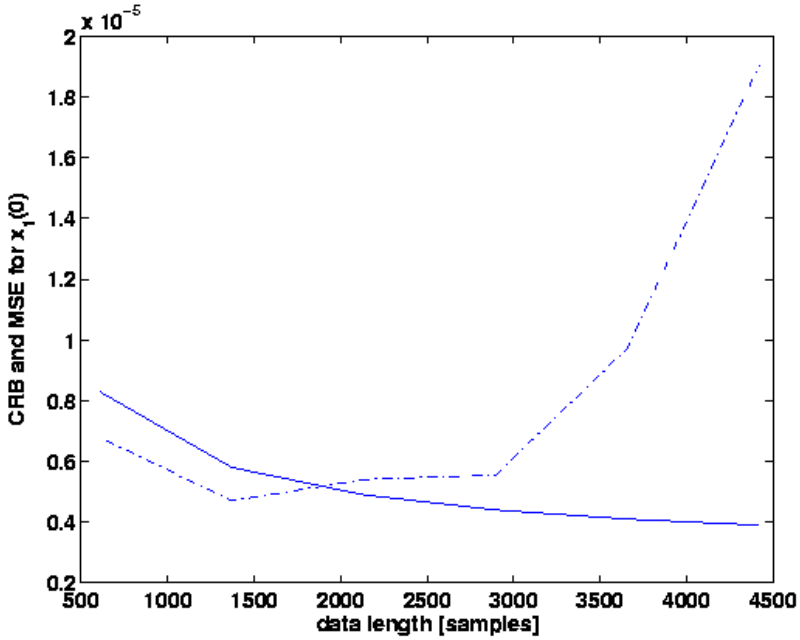

Figure 9: CRB (solid) and performance of the maximum likelihood method (dashes) for the initial value $x_1(0)$. Data length is varied.

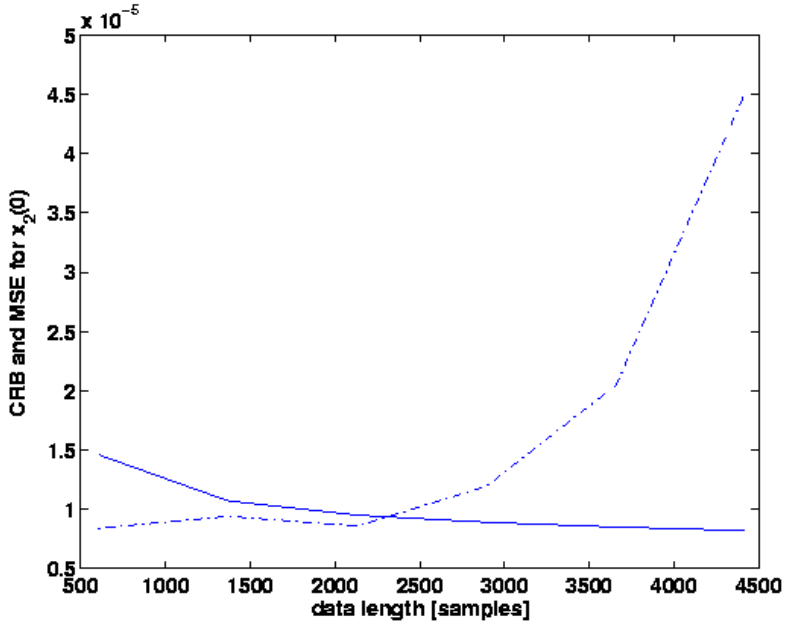

Figure 10: CRB (solid) and performance of the maximum likelihood method (dashes) for the parameter $x_2(0)$. Data length is varied.