Periodic signal analysis by maximum likelihood modeling of orbits of nonlinear ODEs

Torsten Söderström, Torbjörn Wigren*, Emad Abd-Elrady

Division of Systems and Control, Department of Information Technology, Uppsala University, PO Box 337, SE-75105 Uppsala, Sweden

Received 29 January 2004; received in revised form 12 August 2004; accepted 24 November 2004

Abstract

This paper 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 Poincaré 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 by a multivariate polynomial in the states, where each term is multiplied by an unknown parameter. A maximum likelihood algorithm is developed for estimation of the unknown parameters, 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. This allows the theoretically achievable accuracy of the proposed method to be assessed in the ideal case where the signals can be exactly described by the imposed model. 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.

Keywords: Nonlinear systems; Ordinary differential equations; Periodic orbits; Spectrum estimation; System identification

1. 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 Boorstin, 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. Härdel and Tichavsky (1994) and Tichavsky and Härdel (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 paper 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 $\mathbb{R}^2$ 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 $\mathbb{R}^2$—hence it seems to be advantageous to base estimation algorithms on second-order ODEs. The most well-known theoretical results include the Poincare–Bendixson theorem, the Bendixson criterion and the Poincare map (Khalil, 1996). It has recently been found (Wigren and Söderström, 2003) that an ODE order of 2 is enough for modeling quite a wide class of periodic signals. This is the main motivation for the approach of Wigren, Abd-Elrady and Söderström (2003a, b) and of this paper, where second-order ODEs are used as models. This fact is practically important when the low-complexity methods of Wigren et al. (2003a, b) 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, numerical computation of the first derivative of the measured signal is all that is required.

The signal model of the paper 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 paper. Further contributions include the derivation of a maximum likelihood algorithm and a theoretical analysis of performance aspects using the Cramer–Rao bound (CRB). It should be noted that the CRB is limited to a study of accuracy in the ideal case where the system (or signal) can be exactly described by the imposed model, by at least one value of the parameter vector. The CRB hence covers one aspect of the analysis of the performance of the proposed method, by assessment of the theoretically achievable accuracy. A practically more important aspect is perhaps that of under-modeling, where the system (or signal) cannot be exactly described by the imposed model. This aspect has e.g. been treated in Wigren et al. (2003b), example 4, where an ideal nonlinear pendulum is modeled by the proposed methodology with good results. There also the effect of different model orders is investigated and illustrated. In Abd-Elrady, Söderström and Wigren (2004a), a real acoustic signal is modeled to assess e.g. undermodeling. The issue of overparameterization seems to be more complicated and requires future work on the identifiability properties of the proposed model. An introductory example is however available in Wigren et al. (2003b). A possible alternative to the approach taken in this and related papers could be to exploit modulating function methods; see e.g. Benhadj-Braiek and Rotella (1990), Co and Ydstie (1990), Co and Ungarala (1997) and Daniel-Berhe and Unbehauen (1998).

One inevitable consequence of the selected model structure is that the accuracy study by means of the CRB is not being made against the conventional parameters like frequency, phase and amplitude; rather the accuracy of the parameters of the polynomial expansion of the model is studied. Note that there is no direct relationship between the parameters of the ODE model and the conventional frequency, phase and amplitude parameters encountered in the modeling of periodic signals. The above aspects are also illustrated in a simulation study. Initial aspects of the present work can be found in Wigren et al. (2003a, b). Wigren et al. (2003a) presents recursive algorithms based on Kalman and extended Kalman filters, while Wigren et al. (2003b) 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 on 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 (2003) supports this conjecture further.

The paper is organized as follows. Section 2 introduces the details on the model, including a definition of the parameterization. Section 3 discusses the maximum likelihood algorithm, while Section 4 presents the CRB. The paper ends with a simulation study and a discussion of user choices in Sections 5 and 6, respectively. Conclusions appear in Section 7.

The conditions of the paper are ordered (C1), (C2), and so on.
2. The ODE model and its parameterization

In the paper, both continuous and discrete time models appear. In order to get a consistent notation, the time \( t \) is to be interpreted as a continuous time quantity. Whenever discrete time models are referred to, the time \( t \) denotes the sampling instance.

2.1. Measurements and modeled signals

The starting point is the measured signal \( z(t) \) (at sampling instance \( t \)), where

\[
z(t) = y(t) + e(t).
\]

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

\[
(C1) \ y(t + T) = y(t), \quad \forall t \in \mathbb{R}, \ 0 < T < \infty.
\]

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

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

where \( E[\cdot] \) denotes the expectation operator and \( T_S \) is the sampling period.

2.2. Model structures

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

\[
\frac{dx_1(t)}{dt} = f_1(x_1(t), x_2(t), \theta_1), \quad \frac{dx_2(t)}{dt} = f_2(x_1(t), x_2(t), \theta_2),
\]

\[
z(t) = (c_1 \ c_2) \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + e(t).
\]

In (2) \( (x_1(t) \ x_2(t))^T \) is the state vector and \( (c_1 \ c_2) \) is the vector containing the selected output weighting factors that generate \( y(t) \) when multiplied with the state vector. Furthermore,

\[
\theta = (\theta_1^T \ \theta_2^T)^T
\]

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

At this point, it is highly relevant to pose the question of 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. (2003a) it is assumed that the second-order ODE

\[
\frac{d^2y}{dt^2} = f_2 \bigg( y, \frac{dy}{dt}, \theta_2 \bigg)
\]

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

\[
\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y \\ \frac{dy}{dt} \end{bmatrix}.
\]

This results in the state space model

\[
\begin{bmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \end{bmatrix} = \begin{bmatrix} x_2(t) \\ f_2(x_1(t), x_2(t), \theta_2) \end{bmatrix},
\]

\[
z(t) = (1 \ 0) \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + e(t).
\]

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 (2003) also suggests that one right-hand function is sufficient. However, this choice of model structure is by no means a necessity. Future developments may also result in further motivation for the use of (2). In order to obtain results of general validity, the treatment of this paper will therefore deal with both (2) and (6).

2.3. 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

\[
f_1(x_1(t), x_2(t), \theta_1) = \sum_{l=0}^{L_1} \sum_{m=0}^{M_1} \theta_{1,l,m} x_1^l(t) x_2^m(t),
\]

\[
f_2(x_1(t), x_2(t), \theta_2) = \sum_{l=0}^{L_2} \sum_{m=0}^{M_2} \theta_{2,l,m} x_1^l(t) x_2^m(t)
\]

and

\[
f_1(x_1(t), x_2(t)) = x_2(t),
\]

\[
f_2(x_1(t), x_2(t), \theta_2) = \sum_{l=0}^{L_2} \sum_{m=0}^{M_2} \theta_{2,l,m} x_1^l(t) x_2^m(t),
\]

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

\[
\frac{d^2y'}{dt^2} = k_0 f_2 \bigg( \frac{1}{k_0} y', \frac{1}{k_0} \frac{dy'}{dt}, \theta_2 \bigg).
\]

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

2.4. 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

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

\[
x_2(t + T_S) = x_2(t) + T_S \sum_{l=0}^{L_2} \sum_{m=0}^{M_2} \theta_{2,l,m} x_1^l(t)x_m^m(t) \tag{10}
\]

and

\[
x_1(t + T_S) = x_1(t) + T_S x_2(t),
\]

\[
x_2(t + T_S) = x_2(t) + T_S \sum_{l=0}^{L_2} \sum_{m=0}^{M_2} \theta_{2,l,m} x_1^l(t)x_m^m(t) \tag{11}
\]

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

\[
\varphi_1(x_1(t), x_2(t), T_S) = T_S(1 \ldots x_2^{M_1}(t) \ldots x_1^{L_1}(t) \ldots x_1^{L_1}(t)x_2^{M_1}(t))^T,
\]

\[
\varphi_2(x_1(t), x_2(t), T_S) = T_S(1 \ldots x_2^{M_2}(t) \ldots x_1^{L_2}(t) \ldots x_1^{L_2}(t)x_2^{M_2}(t))^T \tag{12}
\]

and

\[
\theta_1 = (\theta_{1,0,0} \ldots \theta_{1,0,M_1} \ldots \theta_{1,L_1,0} \ldots \theta_{1,L_1,M_1})^T
\]

\[
\theta_2 = (\theta_{2,0,0} \ldots \theta_{2,0,M_2} \ldots \theta_{2,L_2,0} \ldots \theta_{2,L_2,M_2})^T \tag{13}
\]

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

\[
x_1(t + T_S) = x_1(t) + \varphi_1^T(x_1(t), x_2(t), T_S)\theta_1,
\]

\[
x_2(t + T_S) = x_2(t) + \varphi_2^T(x_1(t), x_2(t), T_S)\theta_2,
\]

\[
z(t) = c_1x_1(t) + c_2x_2(t) + e(t) \tag{14}
\]

and

\[
x_1(t + T_S) = x_1(t) + T_S x_2(t),
\]

\[
x_2(t + T_S) = x_2(t) + \varphi_2^T(x_1(t), x_2(t), T_S)\theta_2,
\]

\[
z(t) = x_1(t) + e(t) \tag{15}
\]

3. 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 using the model

\[
x(t + T_S) = Fx(t) + \Phi(x(t), T_S)\theta, \quad x(t_0) = x_0,
\]

\[
z(t) = Hx(t) + e(t). \tag{16}
\]

This follows by the selections

\[
F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},
\]

\[
\Phi(x(t), T_S) = \begin{pmatrix} \varphi_1^T(x(t), T_S) & 0 \\ 0 & \varphi_2^T(x(t), T_S) \end{pmatrix},
\]

\[
\theta^T = (\theta_1^T, \theta_2^T),
\]

\[
H = (c_1 \ c_2) \tag{17}
\]

and

\[
F = \begin{pmatrix} 1 & T_S \\ 0 & 1 \end{pmatrix},
\]

\[
\Phi(x(t), T_S) = \begin{pmatrix} 0 & \varphi_2^T(x(t), T_S) \end{pmatrix},
\]

\[
\theta = \theta_2, \tag{18}
\]

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 \( x_0 = x(t_0) \) of the states need to be non-stochastic and estimated together with the unknown non-stochastic parameter vector for the problem to be meaningful with Gaussian distributions. It follows that the likelihood function equals

\[
p(Z^N, \theta, x_0) = \frac{1}{\sqrt{2\pi \sigma}^N} e^{-(z(t_0) - Hx_0)^2 / 2\sigma^2}
\]

\[
\times \frac{1}{(2\pi)^{N/2}\sigma^N} \prod_{k=1}^{N} e^{-(z(t_0 + kT_S) - Hx_0 - kT_S, \theta, x_0))}^2 / 2\sigma^2, \tag{19}
\]

where \( Z^N \) 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 \( x(t_0 + kT_S, \theta, x_0) \) at the sampling instances. The generation of the model states is perhaps best understood by referring to the criterion minimization strategy that is outlined at the end of this section. There model (16) is iterated to generate a complete state trajectory (of model states). This is done for the fixed set of initial values \( x_0 \) and parameters \( \theta \) that correspond 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:

\[
(\hat{\theta}_M L, \hat{x}_M)^T = \arg \min_{\theta, x_0} \bar{L}(\theta, x_0), \tag{20}
\]
\[ \bar{L}(\theta, x_{0}) = (z(t_{0}) - Hx_{0})^2 + \sum_{k=1}^{N} (z(t_{0} + kT_{S}) - Hx(t_{0} + kT_{S}, \theta, x_{0}))^2. \]  

Typically, the minimization of the criterion \[ \bar{L}(\theta, x_{0}) \] has to be carried out in several steps since (21) 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 \( t_{0} + N T_{S} \) once for each grid point, using the parameters and initial values defined by the grid point.

One tentative multi-step criterion minimization procedure could follow the steps:

1. Use any of the low-complexity algorithms of Wigren et al. (2003a, b) (least-squares method or the Kalman filter), possibly followed by the extended Kalman filter (Wigren et al., 2003a), to compute initial estimates of \( \theta \) and \( x_{0} \).
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.

### 4. 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 paper 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

\[ \begin{pmatrix} \frac{dx_{1}(t, \theta, \psi)}{dt} \\ \frac{dx_{2}(t, \theta, \psi)}{dt} \end{pmatrix} = \begin{pmatrix} f_{1}(x_{1}(t, \theta, \psi), x_{2}(t, \theta, \psi), \theta_{1}) \\ f_{2}(x_{1}(t, \theta, \psi), x_{2}(t, \theta, \psi), \theta_{2}) \end{pmatrix}, \]

\[ z(t, \theta, \psi) = (c_1 \ c_2) \begin{pmatrix} x_{1}(t, \theta, \psi) \\ x_{2}(t, \theta, \psi) \end{pmatrix} + e(t), \]  

with the initial values

\[ \begin{pmatrix} x_{1}(t_{0}) \\ x_{2}(t_{0}) \end{pmatrix} = \begin{pmatrix} \psi_{1} \\ \psi_{2} \end{pmatrix} = \psi. \]  

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)

\[ \text{CRB}^{-1} = E \left[ \begin{pmatrix} (\log L(\theta_{1})^{T} \\ (\log L(\theta_{2})^{T} \\ (\log L(\theta_{1})^{T} \\ (\log L(\theta_{2})^{T} \times (\log L(\theta_{1}) (\log L(\theta_{2}) (\log L(\theta_{1}) (\log L(\theta_{2}) \right]. \]  

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

\[ - \log L(\theta, \psi) = \frac{1}{2\sigma^{2}} \sum_{k=1}^{N} (z(t_{1} + kT_{S}) - c_{1}x_{1}(t_{1} + kT_{S}, \theta, \psi) - c_{2}x_{2}(t_{1} + kT_{S}, \theta, \psi))^{2}. \]  

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). 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 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 \([t_{0}, t_{\text{end}}] \supseteq [t_{1} + T_{S}, t_{1} + NT_{S}]\).
2. Compute the elements of the Fisher information matrix (29)–(38) from the computed trajectories.
3. Invert (24) to finalize the computation.

### 5. 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 three main purposes are to study the theoretical accuracies achievable by means of the CRB, to compare the maximum likelihood method performance to this bound and to provide a comparison between some previous standard periodic signal estimation methods and the approach of this and related papers. In so doing, the following model of the Van der Pol oscillator was used:

\[
\begin{pmatrix}
\frac{dx_1(t)}{dt} \\
\frac{dx_2(t)}{dt}
\end{pmatrix} = \begin{pmatrix}
x_2(t) \\
-x_1(t) + 2(1 - (x_1(t))^2)x_2(t)
\end{pmatrix}.
\]  

(26)

Example 1. This example evaluates the CRB and the performance of the maximum likelihood method as a function of the SNR of the data. 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 \(\theta_1\) in (24) disappear since here the CRB was computed from the restricted model (6). The state initial values were \(x_1(0) = x_2(0) = 0.5\). The sampling period in all experiments was \(T_S = 0.01\) s.

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

\[
\begin{pmatrix}
\theta_{n+1} \\
x_{0,n+1}
\end{pmatrix} = \begin{pmatrix}
\theta_n \\
x_{0,n}
\end{pmatrix} - \frac{(\nabla L(\theta_n, x_{0,n}))(\nabla L(\theta_n, x_{0,n}))(\nabla^2 L(\theta_n, x_{0,n}))(\nabla L(\theta_n, x_{0,n}))(\nabla L(\theta_n, x_{0,n})))}{(\nabla^2 L(\theta_n, x_{0,n})).}
\]

(27)

Here \(\nabla^2 L(\theta, x_0)\) denotes the Hessian of \(L(\theta, x_0)\) and \(\nabla L(\theta, x_0)\) is the gradient of \(L(\theta, x_0)\). The reader is referred to Luenberger (1973, pp. 150, 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 \(\theta_{0.1}^T x_0^T\).

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

\[
\begin{pmatrix}
\theta \\
x_0
\end{pmatrix} = \begin{pmatrix}
\theta_0 \\
x_0
\end{pmatrix} - 5\sigma_{\text{CRB}}.
\]  

(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 Figs. 1–5, which also contain 95% confidence intervals for the Monte Carlo simulation. The confidence intervals are computed according to Söderström and Stoica (1989), Appendix B9.

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 paper). It can also be observed that the closeness to the CRB begins to deteriorate below an SNR of 10 dB. Finally, the CRBs roll off linearly with the SNR in a log–log 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 Figs. 6–10. It can be observed that the accuracy of the estimated parameters approaches 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 is 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 for the corresponding parameter estimates. Note also that the initial values are not expected to be unbiased since the signal energy related to the initial
values is finite, even when the data length turns to infinity. Hence, efficiency of the maximum likelihood method for these parameters cannot be expected.

**Example 3.** This example compares a variant of the least-squares algorithm for estimation of the parameters of the ODE model (6), to the standard ESPRIT and periodogram methods (e.g. Stoica and Moses, 1997). The variant of least squares used is described in detail in Abd-Elrady and Schoukens (2004), and is not described further here. The measured signal was given by the first state of the Van der Pol oscillator with white Gaussian noise added. The number of samples was either 1000 or 10 000 and the sampling period was 0.1 s. The least-squares algorithm was set up for estimation of a second-order ODE model, i.e. $L = M = 2$ was used. To get a fair comparison in terms of performance and to be able to relate to advantages of model structures, the same number of unknown parameters (10) was used for ESPRIT and the Periodogram method. In detail, the ESPRIT method was evaluated by estimation of 10 frequency components. The amplitudes and phases of these components were estimated by the least-squares method. The estimated model for the periodogram method was obtained by retaining the 10 most powerful frequency bins and neglecting the rest (that consequently represent the model error). The mean square error (MSE) was calculated as the
Fig. 11. The nonlinear ODE approach (solid), the ESPRIT method (dash-dotted) and the periodogram (dashed) vs. the SNR. $N = 1000$.

Fig. 12. The nonlinear ODE approach (solid), the ESPRIT method (dash-dotted) and the periodogram (dashed) vs. the SNR. $N = 10\,000$. The deviations of the solid line for high SNRs can be explained by statistical variations—no Monte Carlo simulations were used in Example 3.

When the high-order ODEs 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 different ODE solvers for the ODE model and the sensitivity derivatives.

6. User choices

There are several user choices involved when periodic signals are to be modeled by nonlinear ODEs. The more important user choices include the selection of an algorithm, a choice that affects the computational complexity and the obtained accuracy. The sampling period is another crucial issue, considering the fact that the continuous time ODE is discretized with an Euler integration method (Vidyasagar, 1978). Finally, a systematic approach to model order selection is needed.

Algorithms that have been previously described in the literature include recursive algorithms (Wigren et al., 2003a) as well as off-line approaches as in this paper and in Wigren et al. (2003b), Abd-Elrady and Schoukens (2004), and Abd-Elrady, Söderström and Wigren (2004b).

Two recursive algorithms are presented in Wigren et al. (2003a). The Kalman filter variant exploits a differentiating filter to differentiate the signal, thereby making it possible to use measured and differentiated data in the regression vector. The consequence of this is that the filter avoids problems with false local minimum points, at the expense of an increased sensitivity to noise and slightly biased estimates. See Wigren et al. (2003a) for details. The other algorithm of that paper is an extended Kalman filter (EKF) that does not need a differentiated signal to construct the regression vector. Instead, also the regression vector is estimated. The result is a highly nonlinear filter that has significantly better performance than the Kalman filter. Since the EKF may converge to false local minima, one good approach is sometimes to use the Kalman filter to provide estimates that can be used as initial values for the EKF. The computational complexities of both methods are relatively low.

The basic off-line algorithms are the least-squares algorithm and the maximum likelihood algorithm, the latter being described in detail in this paper. The maximum likelihood algorithm is by far the most complex of the algorithms discussed, but it is also the one that can perform the best (close to the theoretical limit under ideal conditions). The computational complexities of the least-squares algorithms are all low. The basic least-squares algorithm is similar to the recursive Kalman filter in that it requires differentiation of the measured periodic signal for its operation; see Wigren et al. (2003b) for details. It is more sensitive to noise than the Kalman filter, since it does not model the differentiation step in a state space model. Therefore, two refinements have been developed to enhance the performance of the least-squares algorithm. The first refinement is described in...
Abd-Elrady et al. (2004b) and it implements the Markov estimator. The second refinement (Abd-Elrady and Schoukens, 2004) exploits a spectral analysis method to first find the fundamental frequency and a few basic Fourier coefficients of the signal. As a consequence, more accurate estimates of the measured periodic signal and its derivative can be used in order to estimate the nonlinear ODE. The effect is a significantly improved accuracy, also in quite noisy situations.

A proper choice of sampling period needs to consider two effects—noise sensitivity and discretization errors. The noise sensitivity impact is a result of the well-known fact that numerical differentiation amplifies noise. The discretization error results from the fact that the continuous time ODE is discretized with a numerical integration method in order to obtain the discrete time models used for estimation of the parameters of the model. Normally the discretization error increases when the sampling period is increased. The above effects were studied in Wigren et al. (2003b), Example 1. There the performance of the least-squares algorithm was plotted against the sampling period and it was observed that there exists an optimal choice of sampling period, where the bias error is minimal. This is consistent with the behaviour of two main contributions to the error, as functions of the sampling period. In Abd-Elrady and Söderström (2004), a theoretical study of the bias caused by noise and discretization error was undertaken. The least-square algorithm was studied. That analysis shows that the bias error \( \theta_B \) caused by additive measurement noise, and the bias error \( \theta_D \) caused by Euler discretization obey \( \| \theta_B \| \propto 1/(T^2 \overline{SNR}) \) and \( \| \theta_D \| \propto T_S^2 \), respectively, for high SNRs. This indicates that there is indeed an optimal choice of sampling period. The reader is referred to Abd-Elrady and Söderström (2004) for further details.

It remains to address the selection of model order, i.e. the degrees used in the multi-variate polynomial in the states. Model order selection is complicated and it is affected by effects such as overparameterization and undermodeling. Using too many parameters (overparameterization) normally means that the obtained accuracy is reduced; cf. Wigren et al. (2003b), Example 3. However, the effect of overparameterization becomes more severe in cases where the underlying optimization problem becomes singular because of an excessive number of parameters. In such situations algorithms may fail completely, which is a reason why a detailed understanding of overparameterization is central. Such a detailed analysis is however beyond the scope of this paper and it is left for future research. In practical situations, the main problem is most often undermodeling, i.e. the imposed model is not capable of describing the system (signal) completely. The methods discussed below for model order selection are all capable of handling undermodeling effects.

The most straightforward method to select model order is perhaps to exploit plots of the measured signal together with simulated signals obtained by the estimated model. In this particular setting it is particularly suitable to use phase plane plots as in Wigren et al. (2003a, b) or periodograms to assess frequency domain properties (Wigren et al., 2003a). Another useful method is to calculate the sum of the squared prediction errors for different selections of the model order. When this performance measure is plotted against the model order, a suitable model order can often be selected where the performance measure levels out, or where it starts to increase; cf. Wigren et al. (2003b).

Note that the difference between models (2) and (6) is not very important as such. The complexity and obtainable accuracy depends more on the number of parameters used for modeling, than on the models themselves. The two models thus offer two possibilities for modeling and the required number of parameters for each structure can be expected to depend on the application.

7. Conclusions

The paper 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 coefficients of polynomials 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 algorithm seems to give a performance very close to the CRB, approaching the CRB when the number of samples increases. A treatment of a number of user choices, such as the choice of sampling period and model order, was also given.

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. The practically very important issue of model order selection seems to require further research on identifiability properties to find conditions that may result in an overparameterized model.

Appendix A. 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,

\[
E[(\log L)_{\theta_1} (\log L)_{\theta_1}^T] = \sum_{k=1}^{N} ((x_1(t_1 + kT_S, \theta, \psi)_t)_{\theta_1}^T (x_2(t_1 + kT_S, \theta, \psi)_{\theta_1})^T ) \times \left( \begin{array}{cc} c_1 & c_2 \end{array} \right) \left( \begin{array}{c} c_1 \ \psi \end{array} \right) \frac{1}{\sigma^2}, \quad (29)
\]
This completes the derivation of the different elements of the Fisher 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

\[
\begin{align*}
\frac{\partial}{\partial \theta_1} & \left( f_1(x(t, \theta, \psi), x_2(t, \theta, \psi), \theta_1) \right)_{x_1} = \frac{\partial}{\partial \theta_1} \left( f_1(x(t, \theta, \psi), x_2(t, \theta, \psi), \theta_1) \right)_{x_1}, \\
\frac{\partial}{\partial \theta_2} & \left( f_1(x(t, \theta, \psi), x_2(t, \theta, \psi), \theta_1) \right)_{x_2} = \frac{\partial}{\partial \theta_2} \left( f_1(x(t, \theta, \psi), x_2(t, \theta, \psi), \theta_1) \right)_{x_2}.
\end{align*}
\]
\[
\frac{d(x_2(t, \theta, \psi))}{dt} = (f_2(x_1(t, \theta, \psi), x_2(t, \theta, \psi), \theta_2)x_1(t, \theta, \psi))_2
\]
\[
+ (f_2(x_1(t, \theta, \psi), x_2(t, \theta, \psi), \theta_2)x_2(t, \theta, \psi))_2 \times (x_2(t, \theta, \psi))_2.
\]
(46)

The initial conditions to these equations follow by differentiation of (23):
\[
(x_1(t_0))_{\theta_1} = (x_2(t_0))_{\theta_1} = 0,
\]
\[
(x_1(t_0))_{\theta_2} = (x_2(t_0))_{\theta_2} = 0,
\]
(47)
\[
(x_1(t_0))_{\psi_1} = 1,
\]
\[
(x_2(t_0))_{\psi_1} = 0,
\]
\[
(x_1(t_0))_{\psi_2} = 0,
\]
\[
(x_2(t_0))_{\psi_2} = 1.
\]
(48)

This completes the derivation of the CRB.

References


Torsten Söderström was born in Malmö, Sweden, in 1945. He received the M.Sc. degree (‘civilingenjör’) in engineering physics in 1969 and the Ph.D. degree in automatic control in 1973, both from Lund Institute of Technology, Lund, Sweden. He is a Fellow of IEEE.

In the period 1967–1974 he held various teaching positions at the Lund Institute of Technology. Since 1974, he has been with Division of Systems and Control, Uppsala University, Uppsala, Sweden, where he is a professor of automatic control.

(with P Stoica) and 'Discrete-time Stochastic Systems'; Prentice-Hall, 1994; second edition, Springer-Verlag, 2002. In 1981 he was, with coauthors, given an Automatica Paper Prize Award. Within IFAC he has served in several capacities including vice-chairman of the TC on Modelling, Identification and Signal Processing, 1993–99, IPC chairman of the IFAC SYSID’94 Symposium, Council member 1996–2002, Executive Board member 1999–2002 and Awards Committee Chair 1999–2002. Within Automatica he was an associate editor 1984–91, guest associate editor or guest editor for four special issues, and is the editor for the area of System Parameter Estimation since 1992.

Torbjörn Wigren was born in Mora, Sweden, in 1961. Currently, he is an adjunct professor of automatic control at Uppsala University, Sweden, a position he holds since 2000. He is employed by Ericsson AB, at the 3G cellular radio network development division, where he is technically responsible for positioning and navigation technology. Performance analysis and design of assisted GPS (A-GPS) functionality for the new WCDMA cellular system are key aspects of his current tasks. Before March 2001, Dr. Wigren was with SAABTech AB (former CelsiusTech AB), working with multi-sensor tracking, non-cooperative air-target recognition and sensor data fusion for automotive safety applications. He has also been employed by Ericsson Radio Systems AB, R&D-division, and by Bofors Aerotronics AB. In 2002, he was appointed a docent in electrical engineering (automatic control) at Uppsala University. He is a member of AIAA and a senior member of the IEEE. His research interests are focused on nonlinear systems, navigation and sensor data fusion. Together with Cecilia he has two children, Anna 13 and Tomas 10. He enjoys running, long distance skating and catamaran sailing (HC17).

Emad Abd-Elrady was born in Cairo, Egypt, in 1970. He received the B.Sc. and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, in 1993 and 1997, respectively. In 2005, he received the Ph.D. degree in electrical engineering from Uppsala University, Uppsala, Sweden. His research interests include adaptive filtering, system identification and nonlinear systems.