2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

On the Accuracy of Parameter Estimation for Continuous Time Nonlinear Systems from Sampled Data

Diego S. Carrasco, Lennart Ljung, Graham C. Goodwin and Juan C. Agüero

Abstract— **This paper deals with the issue of estimating the parameters in a continuous-time nonlinear dynamical model from sampled data. We focus on the issue of bias-variance trade-offs. In particular, we show that the bias error can be significantly reduced by using a particular form of sampled data model based on truncated Taylor series. This model retains the conceptual simplicity of models based on Euler integration but has much improved accuracy as a function of the sampled period.**

I. Introduction

Many (and arguably most) physical systems are described by continuous time differential equation models. Indeed, the parameters within these equations may be of direct physical interest and hence of importance in any estimation study. On the other hand, experimental data collected from a system is almost always in sampled form. This raises the issue as to how one can best estimate the parameters in a continuous time model from sampled data.

Even in the case of linear systems, the sampled response of a continuous time system is a highly nonlinear function of the continuous parameters. For example, the sampled data state transition matrix takes the form $e^{A\Delta}$ where A, Δ are respectively the continuous system matrix and the sampling period. The nonlinear case is even more difficult since no closed form exists for the sampled data model.

At fast sampling rates, it is tempting to use an Euler approximation for the discrete model. Thus, say that the continuous system is described by

$$\dot{x}(\tau) = f(x(\tau)) + g(x(\tau))u(\tau), \quad \tau \in \mathbb{R} \tag{1}$$
$$\tilde{y}(\tau) = h(x(\tau)) \tag{2}$$

then an approximate sampled data model (for a zero oder hold input and sample period Δ) is

$$x(t\Delta + \Delta) = x(t\Delta) + \Delta \left\{ f(x(t\Delta)) + g(x(t\Delta))u(t\Delta) \right\} \tag{3}$$
$$\tilde{y}(t\Delta) = h(x(t\Delta)), \quad t \in \mathbb{N} \tag{4}$$

This model is obtained by simple Euler integration and has local truncation error [1] of order Δ^2 and global truncation error [1] of order Δ.

Our interest in the current paper is in cases where the system has a well defined relative degree. This case has been extensively studied for linear dynamical systems. An issue of importance is that the corresponding sampled data model has extra zero dynamics with associated "sampling zeros". In the linear case, these sampling zeros have been given a precise asymptotic characterisation (see [2]). Moreover, these extra zeros play a key role in the accuracy of sampled data models. For example, it has been shown in [3] that it is crucial that these extra zeros be included in the approximate model if one wants to have a model for which the *relative errors* converge to zero as $\Delta \to 0$.

The nonlinear case is more difficult. Indeed, it is only recently that insights have been obtained into the "sampling zero dynamics" of nonlinear systems. For example, it has been shown in [4] (see also [5]) that by using a particular form of truncated Taylor series we can obtain an (approximate) sampled data model for a continuous time nonlinear system having three key properties, namely:

(i) the sampled data model has extra zero dynamics which are identical to the sampling zero dynamics for a linear system of the same relative degree,
(ii) the model has local and global truncation errors in the output of order Δ^{r+1} and Δ^r respectively, where r is the relative degree of the system,
(iii) the model depends only on $f(\cdot)$ and $g(\cdot)$ in a simple fashion.

Points (ii) and (iii) above actually provide the core motivation for the work described in the current paper. We argue that use of this particular sampled data model allows one to obtain a much improved bias-variance tradeoff in system identification than would be achieved if an Euler integration based model were to be used.

II. Preliminaries

Throughout the paper we limit attention to the single-input single-output case. We consider a nonlinear system of order n described as follows:

$$\dot{x}(\tau, \theta_o) = f(x(\tau, \theta_o)) + g(x(\tau, \theta_o))u(\tau) \tag{5}$$
$$\tilde{y}(\tau, \theta_o) = h(x(\tau, \theta_o)) \tag{6}$$

where $x(\tau)$ is the state evolving in an open subset $\mathcal{M} \subset \mathbb{R}^n$, and where the vector fields $f(\cdot)$, $g(\cdot)$, and the output function $h(\cdot)$ are analytic. The system is assumed to have a vector of true parameters θ_o and uniform relative degree r when $x(\tau) \in \mathcal{M}$.

Remark 1: If an anti-aliasing filter is used, then assume that this has been incorporated into the above model. □

D.S. Carrasco, G.C. Goodwin and J.C. Agüero are with the Priority Research Centre for Complex Dynamic Systems and Control, The University of Newcastle, Australia. {diego.carrascoyanez,graham.goodwin,juan.aguero} @newcastle.edu.au

L. Ljung is with the Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden. ljung@isy.liu.se

We will be interested in the problem of estimating θ_o from a finite sample of observations where the sample period is Δ. We assume that the samples satisfy

$$y_t = \tilde{y}_t(\theta_o) + v_t, \quad t \in \mathbb{N}^+ \tag{7}$$

where y_t denotes $y(t\Delta)$ and $\{v_t\}$ is a stationary stochastic process.

***Remark** 2:* In the sequel, for simplicity of exposition, we will take $\{v_t\}$ to be an i.i.d gaussian sequence of mean zero and variance σ^2. The extension to the case where $\{v_t\}$ is a coloured noise sequence is straightforward and will not be pursued. □

We introduce an (approximate) sampled model having output $\hat{y}_t(\theta)$ parameterised by θ. We assume that the approximate model has truncation error of order Δ^m. This is expressed as

$$\tilde{y}_t(\theta_o) - \hat{y}_t(\theta_o) = E(\Delta, \theta_o) \in \mathcal{O}(\Delta^m), \quad \forall t \in \mathbb{N} \tag{8}$$

Our goal in the current paper is to quantify the impact that the truncation error $E(\Delta, \theta_o)$ has on the estimation of θ_o. Before proceeding, we describe (in the next section) two possible choices for $\hat{y}_t(\theta_o)$.

III. Approximate Sampled Data Models

Our analysis will cover any model for which an error quantification of the form (8) is available. Two possible choices are discussed below. We assume throughout that the input is generated by a zero-order hold, i.e. $u(\tau) = u_t, \tau \in [t\Delta, t\Delta + \Delta)$. First, we provide the following definitions:

***Definition** 1:* We define $T = N \cdot \Delta$ as the time horizon over which the approximate model will be used, where Δ is the length of the time discretisation and N is the number of steps. We also define the following notation for any state $z_{i,t} = z_i(t\Delta) = z_i[t]$ and $z_i^+ = z_{i,t+1}, \forall t \in \mathbb{N}$. □□□

***Definition** 2 (Global Modified Truncation Error (see [5]):* Consider a dynamical system with states $(x_1, \ldots, x_n)$ and an associated approximate model with states $(\hat{x}_1, \ldots, \hat{x}_n)$. The *global modified truncation error* of the approximate model is said to be $(\Delta^{m_1}, \ldots, \Delta^{m_n})$ if, for initial state errors

$$\hat{x}_1[k] - x_1[k] \in \mathcal{O}(\Delta^{\bar{m}_1}) \tag{9}$$

$$\vdots$$

$$\hat{x}_n[k] - x_n[k] \in \mathcal{O}(\Delta^{\bar{m}_n}) \tag{10}$$

for any $\bar{m}_i \geq m_i, i = 1, \ldots, n$ then after N steps, where $N = \lfloor T/\Delta \rfloor$, we have that

$$\hat{x}_1[k+N] - x_1[k+N] \in \mathcal{O}(\Delta^{m_1}) \tag{11}$$

$$\vdots$$

$$\hat{x}_n[k+N] - x_n[k+N] \in \mathcal{O}(\Delta^{m_n}) \tag{12}$$

□□□

***Remark** 3:* The main difference between local and global (modified) truncation errors is that the first only holds for one step while the latter holds for finitely many N steps. Note though, that only the global errors allow us to characterise (8). □

A. Euler Integration

The simplest possible (approximate) model is obtained from (5)–(6) by Euler integration. This leads to the model (3)–(4) which we rewrite here as:

$$\hat{x}_{t+1} = x_t + \Delta \{f(\hat{x}_t, \theta) + g(\hat{x}_t, \theta) \cdot u_t\} \tag{13}$$

$$\hat{y}_t(\theta) = h(\hat{x}_t, \theta), \quad t \in \mathbb{N} \tag{14}$$

It is well known [1] that the global truncation error for Euler integration is of order Δ, i.e. $E(\Delta, \theta_o) \in \mathcal{O}(\Delta)$.

B. Truncated Taylor Series

Here we depart from the usual approach in numerical analysis [1] and consider a particular realisation of the nonlinear state space model having a particular relevance to control. A key result from nonlinear systems theory [6] is that a nonlinear system of degree n and uniform relative degree r can be described by a model in Normal Form. In this case, we write the model as:

$$\dot{z}_1(\tau) = z_2(\tau) \tag{15}$$

$$\vdots$$

$$\dot{z}_{r-1}(\tau) = z_r(\tau) \tag{16}$$

$$\dot{z}_r(\tau) = b(\zeta, \eta) + a(\zeta, \eta) \cdot u(\tau) \tag{17}$$

$$\dot{\eta}(\tau) = c(\zeta, \eta) \tag{18}$$

where the output is $\tilde{y}(\tau) = z_1(\tau)$ and $z_i(\tau), i = 2, \ldots, r$ are its first $r-1$ derivatives. Also ζ, η and a, b, c are well defined vectors and functions respectively (see [4] for further details).

The approximate sampled data model of interest here is obtained by using a truncated Taylor series expansion of the states in normal form.

***Definition** 3 (Approximate Sampled Data Model):* Using the procedure described in [4], we can obtain an approximate sampled-data model for (5)–(6) using (16)–(19), given by

$$\hat{z}_1^+ = \hat{z}_1 + \Delta\hat{z}_2 + \frac{\Delta^2}{2}\hat{z}_3 + \ldots + \frac{\Delta^{r-1}}{(r-1)!}\hat{z}_r + \frac{\Delta^r}{r!}\left\{b(\hat{\zeta},\hat{\eta}) + a(\hat{\zeta},\hat{\eta})\cdot u(\tau)\right\}\big|_{\tau=t\Delta} \quad (19)$$

$$\hat{z}_2^+ = \hat{z}_2 + \Delta\hat{z}_3 + \ldots + \frac{\Delta^{r-2}}{(r-2)!}\hat{z}_r + \frac{\Delta^{r-1}}{(r-1)!}\left\{b(\hat{\zeta},\hat{\eta}) + a(\hat{\zeta},\hat{\eta})\cdot u(\tau)\right\}\big|_{\tau=t\Delta} \quad (20)$$

$$\vdots$$

$$\hat{z}_{r-1}^+ = \hat{z}_{r-1} + \Delta\hat{z}_r + \frac{\Delta^2}{2}\left\{b(\hat{\zeta},\hat{\eta}) + a(\hat{\zeta},\hat{\eta})\cdot u(\tau)\right\}\big|_{\tau=t\Delta} \quad (21)$$

$$\hat{z}_r^+ = \hat{z}_r + \Delta\left\{b(\hat{\zeta},\hat{\eta}) + a(\hat{\zeta},\hat{\eta})\cdot u(\tau)\right\}\big|_{\tau=t\Delta} \quad (22)$$

$$\hat{\eta}^+ = \hat{\eta} + \Delta c(\hat{\zeta},\hat{\eta})\big|_{\tau=t\Delta} \quad (23)$$

where the output is $\hat{y}_k = \hat{z}_{1,k}$. (See [4] for further details). □□□

It has been shown [4], [5] that the local and global (modified) truncation errors *in the output* for this model are of order Δ^{r+1} and Δ^r respectively, where r is the relative degree of the system. This implies that $E(\Delta,\theta_o) \in \mathcal{O}(\Delta^r)$ for this specific model.

Remark 4: Note that for $r > 1$, the model in Section III-B has smaller output truncation errors than the model in Section III-A. We will discuss the implications of this observation in Section V. □

IV. Maximum Likelihood Estimation

We now return to the problem of estimating of θ_o. We assume we are given observations $\{y_0,\ldots,y_N\}$. We also assume that the initial state is known, say the origin. Since, $\{v_t\}$ is an i.i.d. gaussian sequence, the maximum likelihood cost function can be written as

$$J_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}[y_t - \hat{y}_t(\theta)]^2 \quad (24)$$

We then define

$$\hat{\theta}_N = \arg\min_\theta J_N(\theta) \quad (25)$$

We will also be interested in the asymptotic estimate θ_* defined by

$$\theta_* = \arg\min_\theta \lim_{N\to\infty} J_N(\theta) = \lim_{N\to\infty}\hat{\theta}_N \quad (26)$$

The existence of θ_* requires the assumption that there exist well defined limits:

$$J_N(\theta) \to J_\infty(\theta) \quad (27)$$

$$\hat{\theta}_N \to \theta_* \quad (28)$$

Note also that θ_o satisfies

$$\theta_o = \arg\min_\theta \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}[y_t - \tilde{y}_t(\theta)]^2 \quad (29)$$

when the model structure corresponds to the true system, i.e. $\hat{y}_t(\theta) = \tilde{y}_t(\theta)$

In addition we assume that

$$J''_\infty(\theta) \geq cI, \quad c > 0, \quad \forall\theta\in\mathcal{B} \quad (30)$$

where $\mathcal{B}$ is a neighbourhood that includes θ_0 and θ_*.

Our intention is to quantify the estimation error $\theta_o - \hat{\theta}_N$. For this purpose we will establish the following key result:

Theorem 1: Subject to standard regularity conditions and assuming the output truncation error satisfies $E(\Delta,\theta) \in \mathcal{O}(\Delta^m)$, then the estimation error $\theta_o - \hat{\theta}_N$ can be written as

$$\theta_o - \hat{\theta}_N = C(N)\left\{\frac{1}{\sqrt{N}}\cdot A(N) + B(N)\right\} \quad (31)$$

where $C(N) \in \mathcal{O}(1)$ is such that

$$\lim_{N\to\infty} C(N) \leq \sup_{\theta\in\mathcal{B}}\|J''_\infty(\theta)^{-1}\| \leq I\cdot 1/c \quad (32)$$

and where $A(N)$ converges to a normal random variable having zero mean and covariance

$$\Omega = \lim_{N\to\infty}\frac{1}{N}\left(\sum_{t=1}^{N}\hat{\Phi}_t\hat{\Phi}_t^T\right)\sigma^2 \quad (33)$$

where

$$\hat{\Phi}_t = \frac{\partial\hat{y}_t}{\partial\theta}\bigg|_{\theta_o} \quad (34)$$

and $B(N)$ is of order Δ^m.

Proof: By definition we have

$$\hat{\theta}_N = \arg\min J_N(\theta) = \arg\min\frac{1}{N}\sum_{t=1}^{N}[y_t - \hat{y}_t(\theta)]^2 \quad (35)$$

It follows that

$$J'_N(\hat{\theta}_N) = 0 \quad (36)$$

Then we can write

$$J'_N(\theta_0) = J'_N(\theta_0) - J'_N(\hat{\theta}_N) = J''_N(\xi)[\theta_0 - \hat{\theta}_N] \quad (37)$$

where we have used the mean value theorem. Note that

$$J''_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}2v_t\left(\frac{-\partial^2\hat{y}_t}{\partial\theta^2}\right)\bigg|_\theta + \frac{2}{N}\sum_{t=1}^{N}\left[\{\tilde{y}_t - \hat{y}_t(\theta)\}\left(\frac{-\partial^2\hat{y}_t}{\partial\theta^2}\right)\bigg|_\theta + \left(\frac{\partial\hat{y}_t}{\partial\theta}\right)^2\bigg|_\theta\right] \quad (38)$$

where

$$J''_{\infty}(\theta) = \lim_{N\to\infty} J''_N(\theta)$$
$$= \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} 2\left[\{\tilde{y}_t - \hat{y}_t(\theta)\}\left(\frac{-\partial^2 \hat{y}_t}{\partial\theta^2}\right)\bigg|_{\theta} + \left(\frac{\partial \hat{y}_t}{\partial\theta}\right)^2\bigg|_{\theta}\right] \in \mathcal{O}(1) \tag{39}$$

For convenience, we define

$$C(N) = J''_N(\xi)^{-1} \tag{40}$$

and (33) follows from the assumptions. From (38)

$$(\theta_0 - \hat{\theta}_N) = C(N)\left[\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t(\theta_0))\right]\hat{\Phi}_t(\theta_0) \tag{41}$$

$$= C(N)\frac{1}{N}\sum_{t=1}^{N}[v_t + \tilde{y}_t(\theta_0) - \hat{y}_t(\theta_0)]\hat{\Phi}_t(\theta_0) \tag{42}$$

$$= C(N)\left\{\frac{1}{\sqrt{N}}\left[\frac{1}{\sqrt{N}}\sum_{t=1}^{N} v_t\hat{\Phi}_t(\theta_0)\right] + \frac{1}{N}\sum_{t=1}^{N}\left[\tilde{y}_t(\theta_0) - \hat{y}_t(\theta_0)\right]\hat{\Phi}_t(\theta_0)\right\} \tag{43}$$

We now define

$$A(N) = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} v_t\hat{\Phi}_t(\theta_0) \tag{44}$$

$$B(N) = \frac{1}{N}\sum_{t=1}^{N}\left[\tilde{y}_t(\theta_0) - \hat{y}_t(\theta_0)\right]\hat{\Phi}_t(\theta_0) \tag{45}$$

A standard result, e.g. [7], [8], shows that $A(N)$ converges to a normal random variable as stated in the theorem. On the other hand, it is easy to see that

$$|B(N)| = \left|\frac{1}{N}\sum_{t=1}^{N}(\tilde{y}_t(\theta_0) - \hat{y}_t(\theta_0))\hat{\Phi}_t(\theta_0)\right| \tag{46}$$

$$\leq \sqrt{\frac{1}{N}\sum_{t=1}^{N}|\tilde{y}_t(\theta_0) - \hat{y}_t(\theta_0)|^2}\sqrt{\frac{1}{N}\sum_{t=1}^{N}|\Phi_t|^2} \tag{47}$$

$$\leq KE(\Delta, \theta_0) \tag{48}$$

which completes the proof.

■

V. Finite Data Length Interpretation

The quantification given in Section IV is asymptotic in N. In practice one can use this result to motivate a quantification of the error for finite N as:

$$MSE \approx \left|\lim_{N\to\infty} C(N)\right|^2 \cdot |B(N)|^2 + \frac{1}{N}\cdot \text{trace}\left\{\lim_{N\to\infty} C(N)\left(\frac{1}{N}\sum_{t=1}^{N}\hat{\Phi}_t^T\hat{\Phi}_t\right)C(N)^T\sigma^2\right\} \tag{49}$$

where $|B(N)|^2 \in \mathcal{O}(\Delta^{2m})$.

Moreover, we have seen that, for fixed N, $E(\Delta, \theta_*) \in \mathcal{O}(\Delta)$ for the model of Section III-A and $E(\Delta, \theta_*) \in \mathcal{O}(\Delta^r)$ for the model of Section III-B.

The expression (50) shows that $|B(N)|^2$ represents a lower bound on the achievable MSE no matter how much data we collect. In particular, there is no point choosing an N such that the second term (variance) is much smaller that the inescapable bias term.

Thus, if one can collect as much data as one likes, then it is preferable to use the model presented in Section III-B since the bias error is smaller than if the Euler model were to be used for a given sampling period.

VI. Conclusions

This paper has examined the question of bias-variance tradeoffs when estimating the parameters of continuous time models using sampled data. We have shown that when approximate sampled data models are used then bias errors result which place an inescapable lower bound on the estimation accuracy no matter how much data is collected. We have also shown that this lower bound can be reduced by using models other than those based on simple Euler integration.

VII. Acknowledgments

The authors gratefully acknowledge input from Dr. Juan Yuz.

References

[1] J. Butcher, *The numerical analysis of ordinary differential equations: Runge-Kutta and general linear methods*. Wiley-Interscience New York, NY, USA, 1987.

[2] K. Astrom, P. Hagander, and J. Sternby, "Zeros of sampled systems," *Automatica*, vol. 20, no. 1, pp. 31 – 38, 1984.

[3] G. Goodwin, J. Yuz, and J. Agüero, "Relative error issues in sampled data models," in *IFAC World Congress, Seoul, Korea*, 2008.

[4] J. Yuz and G. Goodwin, "On sampled-data models for nonlinear systems," *Automatic Control, IEEE Transactions on*, vol. 50, no. 10, pp. 1477–1489, 2005.

[5] D. Carrasco and G. Goodwin, "On the Accuracy of Sampled Data Models for Nonlinear Systems," *Submitted for publication. Available at http://www.cdsc.org.au/Publications/pdf/accuracy.pdf.*

[6] A. Isidori, *Nonlinear control systems*. Springer Verlag, 1995.

[7] L. Ljung, "Convergence analysis of parametric identification methods," *Automatic Control, IEEE Transactions on*, vol. 23, no. 5, pp. 770 – 783, Oct. 1978.

[8] H. White, *Estimation, inference and specification analysis*. Cambridge University Press, 1996.