# EM Identification of continuous–time state space models from fast sampled data

Juan I. Yuz, Jared Alfaro, Juan C. Agüero and Graham C. Goodwin

***Abstract*— In this paper we apply the Expectation-Maximization (EM) algorithm to the identification of continuous-time state-space models from fast sampled data. We modify the standard EM formulation, using a parametrization of the sampled-data model in incremental form. This model recovers the underlying continuous-time system when the sampling period goes to zero. Also, the use of the incremental model parametrization shows better numerical behavior for fast sampling rates. We also consider the case of non–uniform sampling and a robust identification procedure that can be applied in the time or frequency domain.**

## I. Introduction

State space models are widely used to represent multi-variable dynamic systems in many areas including control [1], and econometrics [2]. In particular, for multivariable systems, state-space models provide a concise and flexible representation [3]. In this paper, we are interested in identifying continuous-time state-space models from sampled data. We use maximum likelihood estimation to identify the continuous system parameters. However, the main difficulty in identifying a system expressed in state space form is that the likelihood function is non-convex. Thus, algorithms such as Expectation-Maximization (EM) [4], [5], [6], [7], have to be applied.

To estimate parameters in continuous-time state-space models, one usually first obtains the discrete-time matrices and then, in a second step, a transformation to recover the continuous-time matrices is applied. This operation may involve the computation of the logarithm of the system matrix or the use of Padé-like approximations [8], [9]. Also, different derivative approximations can be used to discretize the model, however, they have an impact on the quality of the estimates (see, for example, [10]).

In this paper we parametrize the discrete-time model in incremental form (also known as $\delta$-operator form [11], [12]). Incremental models are obtained by simply reparametrizing the discrete model. The use of this kind of model is known to provide improved numerical properties. It also gives a framework which allows one to unify discrete- and continuous-time results in estimation and control when fast sampling rates are utilized (see, for example, [13], [14], [15], [16], [17], [18]). Our approach allows us to readily consider non-uniformly sampled data. Continuous-time system Identification from non-uniform sampled-data has been also considered in [19] in the frequency domain using an approximate output reconstruction with B-splines. Another identification procedure for non-uniform sampling was proposed in [20]. The method proposed in that work is based on a least squares approach where the states are estimated by using Kalman filtering in shift operator form. In our approach we develop a procedure based on maximum likelihood and the system is parametrized in incremental form.

We apply the EM algorithm to maximize the likelihood function associated with the incremental model to obtain an estimate of the continuous-time model parameters. EM is a recursive two step procedure where, in a first step, a function of the state sequence estimate is obtained assuming that the parameters are known (E-step). Then, in a second step, that function is maximized with respect to the parameters to obtain a new estimate (M-step). This yields a new parameter estimate to be used in the next iteration of the algorithm [21]. We present the modifications required in both steps of EM when dealing with non-uniform fast-sampled data and incremental models.

Juan I. Yuz and Jared Alfaro are with the Electronic Engineering Department, Universidad Técnica Federico Santa María, Av. España 1680 Valparaíso, Chile. Email: juan.yuz@usm.cl, jalfaro@elo.utfsm.cl

Juan C. Agüero and Graham C. Goodwin are with the Centre of Complex Dynamics Systems and Control (CDSC), *The* University *of* Newcastle, Australia. Email: {juan.aguero, graham.goodwin}@newcastle.edu.au

## II. Problem statement

### A. Continuous-time system description

In this paper we consider a general multiple-input multiple-output (MIMO), linear time-invariant (LTI) continuous system subject to stochastic disturbances. To describe such a continuous system by a meaningful mathematical model we should use a stochastic differential equation (SDE) model [22]:

$$dx(t) = A_c\, x(t)\, dt + B_c\, u(t)\, dt + dw(t) \tag{1}$$

$$dz(t) = C_c\, x(t)\, dt + D_c\, u(t)\, dt + dv(t) \tag{2}$$

where $u(t) \in \mathbb{R}^{n_u}$, $x(t) \in \mathbb{R}^{n}$, $y(t) \in \mathbb{R}^{n_y}$ are the input, the system state, and the output signal respectively; the system matrices are $A_c \in \mathbb{R}^{n\times n}$, $B_c \in \mathbb{R}^{n_u\times n}$, $C_c \in \mathbb{R}^{n_y\times n}$, and $D_c \in \mathbb{R}^{n_y\times n_u}$; and the incremental state disturbance $dw(t)$ and incremental measurement disturbance $dv(t)$ are stochastic processes that are independent (in time), zero mean, and have Gaussian distribution (i.e., $w(t)$ and $v(t)$ are Wiener processes) such that

$$E\left\{ \begin{bmatrix} dw(t) \\ dv(t) \end{bmatrix} \begin{bmatrix} dw(s) \\ dv(s) \end{bmatrix}^T \right\} = \begin{cases} \begin{bmatrix} Q_c & 0 \\ 0 & R_c \end{bmatrix} dt & ;\, t = s \\ 0 & ;\, t \neq s \end{cases} \tag{3}$$

ISBN 978-963-311-370-7

where $Q_c \in \mathbb{R}^{n\times n}$ is a positive semi-definite matrix, and $R_c \in \mathbb{R}^{n_y\times n_y}$ is a positive definite matrix. The initial state of the system is assumed independent of $dw(t)$ and $dv(t)$, and Gaussian distributed having mean $\mu_o$ and covariance $\Sigma_o$.

The SDE description in (1)-(2) can be *formally* expressed as the following state-space form

$$\frac{dx(t)}{dt} = A_c\, x(t) + B_c\, u(t) + \dot{w}(t) \tag{4}$$

$$\frac{dz(t)}{dt} = C_c\, x(t) + D_c\, u(t) + \dot{v}(t) \tag{5}$$

where the process noise $\dot{w}(t) \in \mathbb{R}^n$ and the measurement noise $\dot{v}(t) \in \mathbb{R}^{n_y}$ are the *formal* derivatives of the Wiener processes $w(t)$ and $v(t)$, respectively. Such processes are referred to as *continuous-time white noise* (CTWN) processes with zero mean, Gaussian distribution, and covariance structure:

$$E\left\{\begin{bmatrix}\dot{w}(t)\\ \dot{v}(t)\end{bmatrix}\begin{bmatrix}\dot{w}(s)\\ \dot{v}(s)\end{bmatrix}^T\right\} = \begin{bmatrix}Q_c & 0\\ 0 & R_c\end{bmatrix}\delta(t-s) \tag{6}$$

where $\delta(\cdot)$ is the Dirac delta function.

*Remark 1:* A key observation in the continuous-time noise model in (6), or in (3), is that the matrices $Q_c$ and $R_c$ correspond to *spectral densities* of the noise processes [12].

In this paper, our prime interest is in obtaining an estimate of the system parameter

$$\theta^c = \{A_c, B_c, C_c, D_c, Q_c, R_c\} \tag{7}$$

from a finite set of input-output samples $\{u_k = u(t_k), y_k = y(t_k)\}$, where $k \in \{0,\ldots,N\}$, i.e., we assume that input updates and output samples are synchronized. To include in our analysis the case of non-uniform sampled data we assume that the sampling interval depends on the discrete-time index $k$:

$$\Delta_k = t_{k+1} - t_k > 0 \quad ; \text{ for all } k \in 0,\ldots,N-1 \tag{8}$$

*Assumption 1:* We assume that the sampling intervals $\{\Delta_k\}$ are uniformly bounded by a positive constant $\bar{\Delta}$, i.e., there exists

$$\bar{\Delta} \triangleq \max_k \Delta_k \tag{9}$$

such that $0 < \Delta_k \leq \bar{\Delta}$, for all $k$.

### B. Sampled-data model

We assume that the continuous-time input to the system, $u(t)$, is generated from an input sequence, $u_k$, by a zero-order hold (ZOH) device:

$$u(t) = u_k \qquad ; t_k \leq t < t_{k+1} \tag{10}$$

The sampling process of the system output has to be dealt with carefully. From (5), we see that the output has a *pure CTWN component*. Instantaneous sampling of this output would lead to a sequence having infinite variance. We assume that an *integrate and reset* filter (IRF) is included at the system output before instantaneous sampling [23], [12].

$$\bar{y}(t_{k+1}) = \frac{1}{t_{k+1}-t_k}\int_{t_k}^{t_{k+1}} \frac{dz_\tau}{d\tau}d\tau = \frac{1}{\Delta_k}\int_{t_k}^{t_{k+1}} dz(\tau) \tag{11}$$

$$= \frac{z(t_{k+1}) - z(t_k)}{\Delta_k} \tag{12}$$

An additional advantage of this sampling strategy is that, as the sampling rate is increased, the sampled output $\bar{y}$ is consistent with the continuous-time output:

$$\bar{y}(t) = \lim_{\Delta_k\to 0}\frac{z(t_{k+1}) - z(t_k)}{\Delta_k} = \frac{dz(t)}{dt} \tag{13}$$

*Lemma 1:* Consider the continuous-time state-space model (1)-(2), where the input is generated by ZOH device (10) and the output is sampled after the IRF (11). Then the following incremental discrete-time model has the same second order output properties as the sampled output of the continuous-time system:

$$d\,x_k^+ = A_k^\delta x_k\Delta_k + B_k^\delta u_k\Delta_k + dw_k^+ \tag{14}$$

$$\bar{y}_{k+1}\Delta_k = d\,z_k^+ = C_k^\delta x_k\Delta_k + D_k^\delta u_k\Delta_k + dv_k^+ \tag{15}$$

where the increments are defined as

$$d\,f_k^+ = f_{k+1} - f_k \tag{16}$$

The matrices are given by

$$A_k^\delta = \frac{e^{A_c\Delta_k} - I}{\Delta_k} \tag{17}$$

$$B_k^\delta = \left[\frac{1}{\Delta_k}\int_0^{\Delta_k} e^{A_c\eta}\,d\eta\right] B_c \tag{18}$$

$$C_k^\delta = C_c\left[\frac{1}{\Delta_k}\int_0^{\Delta_k} e^{A_c\eta}\,d\eta\right] \tag{19}$$

$$D_k^\delta = D_c + C_c\left[\frac{1}{\Delta_k}\int_0^{\Delta_k}\int_0^{\xi} e^{A\eta}\,d\eta\,d\xi\right] B_c \tag{20}$$

and the covariance structure of the noise vector is given by

$$E\left\{\begin{bmatrix}d\,w_\ell^+\\ d\,v_\ell^+\end{bmatrix}\begin{bmatrix}d\,w_k^+\\ d\,v_k^+\end{bmatrix}\right\} = \begin{bmatrix}Q_k^\delta & S_k^\delta\\ (S_k^\delta)^T & R_k^\delta\end{bmatrix}\Delta_k\,\delta_K[\ell-k] \tag{21}$$

where

$$\begin{bmatrix}Q_k^\delta & S_k^\delta\\ (S_k^\delta)^T & R_k^\delta\end{bmatrix} = \frac{1}{\Delta_k}\int_0^{\Delta_k} M\begin{bmatrix}Q_c & 0\\ 0 & R_c\end{bmatrix}M^T d\eta \tag{22}$$

where

$$M = \begin{bmatrix} e^{A_c\eta} & 0\\ C_c\displaystyle\int_0^{\eta} e^{A_c\xi}\,d\xi & I\end{bmatrix} \tag{23}$$

*Proof:* The details of the proof can be found, for example, in [12]. A numerically stable implementation to obtain the different integrals of matrix exponential in (17)-(22) is presented in [24]. ■

*Remark 2:* The incremental model closely resembles the continuous-time representation (1)-(2). In fact, as the maximum sampling period $\bar{\Delta}$ goes to zero, the incremental model

(14)-(15) converges to the stochastic differential equation representation.

As mentioned in Remark 1, one of the key facets of the parametrization of the discrete-time system by the incremental model (14)-(22) is that the noise processes are described by spectral densities. This is easy to see in the uniform sampling case where, for example, the spectrum of the discrete-time white noise process $\{\frac{dw_k^+}{\Delta}\}$ is given by [12]

$$\Phi_{\tilde{w}}^{d}(\omega) = \Delta \sum_{\ell=-\infty}^{\infty} \frac{Q^\delta}{\Delta}\delta_K[\ell]e^{-j\omega\ell\Delta} = Q^\delta \qquad (24)$$

*Lemma 2:* The discrete-time model (14)-(22) can also be represented in terms of the $q$- or forward-shift operator

$$q\,x_k = x_{k+1} = A_k^q x_k + B_k^q u_k + \tilde{w}_k \qquad (25)$$

$$\bar{y}_{k+1} = C_k^q x_k + D_k^q u_k + \bar{v}_k \qquad (26)$$

where the covariance structure of the noise vector is given by

$$E\left\{\begin{bmatrix}\tilde{w}_\ell\\ \bar{v}_\ell\end{bmatrix}\begin{bmatrix}\tilde{w}_k\\ \bar{v}_k\end{bmatrix}\right\} = \begin{bmatrix}Q_k^q & S_k^q\\ (S_k^q)^T & R_k^q\end{bmatrix}\delta_K[\ell-k] \qquad (27)$$

The transformation between the matrices in the incremental model (17)-(22) and the shift form in (25)-(27) is given by

$$Q_k^q = \Delta_k\, Q_k^\delta \qquad S_k^q = S_k^\delta \qquad R_k^q = \frac{1}{\Delta_k}R_k^\delta \qquad (28)$$

$$A_k^q = I + \Delta_k\, A_k^\delta \qquad B_k^q = \Delta_k B_k^\delta \qquad (29)$$

$$C_k^q = C_k^\delta \qquad D_k^q = D_k^\delta \qquad (30)$$

*Proof:* The shift operator model and the matrix transformation readily follow by substituting the increments defined in (16) into (14). ■

The shift operator model (25)-(26) is the most commonly used when identifying discrete-time state-space models. However, this kind of model may give poor results when attempting to recover the underlying continuous-time system. In particular, from (34) and (28)-(30), we have that

$$\lim_{\Delta_k\to 0} A_k^q = I \qquad \lim_{\Delta_k\to 0} B_k^q = 0 \qquad (31)$$

$$\lim_{\Delta_k\to 0} Q_k^q = 0 \qquad \lim_{\Delta_k\to 0} R_k^q = \infty \qquad (32)$$

independently of the underlying continuous-time description.

On the other hand, the exact sampled-data incremental model given by (14)-(22) provides a better way to highlight the relationship between continuous- and discrete-time state-space system matrices. In fact, if we consider the (time-varying) parameter

$$\theta_k^\delta = \{A_k^\delta, B_k^\delta, C_k^\delta, D_k^\delta, Q_k^\delta, R_k^\delta, S_k^\delta\} \qquad (33)$$

we have that, for every $k$,

$$\lim_{\bar{\Delta}\to 0} \theta_k^\delta = \lim_{\Delta_k\to 0} \theta_k^\delta = \theta^c \qquad (34)$$

where $\bar{\Delta}$ is the maximum sampling interval and $\theta_c$ is the continuous-time parameter (7).

As a consequence, for the problem of interest in this paper, we can use the exact **time-varying** sampled-data model (14)-(22), to identify the continuous (time-invariant) system parameters from fast-sampled data.

*Remark 3:* Note that, from the limiting result in (34), if $S_c = 0$ then $S_k^\delta$ will converge to zero for fast sampling rates, and, thus, it can be assumed to be zero in order to simplify the calculations.

## III. The EM algorithm

The likelihood function is the conditional probability of the data $Y$ given the parameter vector $\theta$. When the measurements are Gaussian distributed, it is common to maximize the logarithm of the likelihood function

$$\ell(\theta) = \log p(Y|\theta) \qquad (35)$$

EM is an iterative method used to maximise the log-likelihood function. If we choose some variable $X$ as *hidden data*, the log-likelihood can be decomposed as:

$$\log p(Y|\theta) = \mathcal{Q}(\theta,\hat{\theta}_i) - \mathcal{H}(\theta,\hat{\theta}_i) \qquad (36)$$

where $\hat{\theta}_i$ is an available estimate of the parameter $\theta$, and

$$\mathcal{Q}(\theta,\hat{\theta}_i) = E\{\log p(X,Y|\theta)|Y,\hat{\theta}_i\} \qquad (37)$$

$$\mathcal{H}(\theta,\hat{\theta}_i) = E\{\log p(X|Y,\theta)|Y,\hat{\theta}_i\} \qquad (38)$$

Using Jensen's inequality it is easy to show that $\mathcal{H}(\theta,\hat{\theta}_i) \le \mathcal{H}(\hat{\theta}_i,\hat{\theta}_i)$ [4]. As a consequence, if one finds a value of $\theta$ that increases the function $\mathcal{Q}(\theta,\hat{\theta}_i)$, then one can generate an iterative procedure to maximize the likelihood function (35). For state space models, the *natural* choice for the hidden variables $X$ es the state sequence.

The steps of the EM algorithm are thus: (i) Start with an initial estimate of the system parameter $\hat{\theta}_0$. (ii) Obtain the function $\mathcal{Q}(\theta,\hat{\theta}_i)$, defined in (37), which is the expected value of the complete data $(X,Y)$ given the observed data $Y$ and an available estimate $\hat{\theta}_i$. This step is known as the *E-step*. (iii) Maximize the function $\mathcal{Q}(\theta,\hat{\theta}_i)$ with respect to the parameter $\theta$. This yields a new parameter estimate, i.e.

$$\hat{\theta}_{i+1} = \arg\max_{\theta} \mathcal{Q}(\theta,\hat{\theta}_i) \qquad (39)$$

This step is known as the *M-step*. (iv) Finally, go to step (ii), increasing the index $i \to i+1$, and iterate until convergence (or some predefined level of accuracy) is achieved.

The EM algorithm is known to converge to a stationary point of the likelihood function which, for most problems, corresponds to a local maximum [4].

*Remark 4:* The EM algorithm has been previously applied to identify continuous-time state space models from sampled data. However, previous approaches have had two key restrictions, namely, (i) uniform sampling is necessary, and (ii) procedurally, one first estimates the shift domain model parameters as in [6], and then, in a second stage, these parameters are transformed to continuous-time. Our approach is different: we use the incremental form and identify the (reparametrized) discrete-time system parameter $\theta_k^\delta$, given in (33). Based on the fast sampling assumption, we know that

$\theta_k^\delta \to \theta^c$ and, thus, we directly obtain an estimate of the continuous parameters.

We next review the two steps of the EM algorithm applied to the time-varying discrete-model (25)-(27).

### A. The E-step for shift-operator models

For shift operator models, the function (37) can be expressed as

$$\begin{aligned}
-2\cdot\mathcal{Q}(\theta,\hat{\theta}_i) &= L_0 + N\log\det Q_k^q + N\log\det R_k^q \\
&+ \log\det\Sigma_0 + \text{trace}\left[\Sigma_0^{-1}E\left\{(x_0-\mu_0)(x_0-\mu_0)^T\right\}\right] \\
&+ \sum_{k=1}^{N}\text{trace}\Big[(Q_k^q)^{-1}\Big(E\left\{x_k x_k^T\right\} - E\left\{x_k z_{k-1}^T\right\}[A_k^q, B_k^q]^T \\
&\qquad - [A_k^q, B_k^q]E\left\{x_k z_{k-1}^T\right\}^T \\
&\qquad + [A_k^q, B_k^q]E\left\{z_{k-1} z_{k-1}^T\right\}[A_k^q, B_k^q]^T\Big)\Big] \\
&+ \sum_{k=0}^{N-1}\text{trace}\Big[(R_k^q)^{-1}\Big(E\left\{y_k y_k^T\right\} - E\left\{y_k z_k^T\right\}[C_k^q, D_k^q]^T \\
&\qquad - [C_k^q, D_k^q]E\left\{y_k z_k^T\right\}^T \\
&\qquad + [C_k^q, D_k^q]E\left\{z_k z_k^T\right\}[C_k^q, D_k^q]^T\Big)\Big] \qquad (40)
\end{aligned}$$

where $z_k = \begin{bmatrix} x_k \\ u_k \end{bmatrix}$ and $L_0$ accounts for constant terms. All of the expected values are conditioned to the data $Y$.

The expected values can be obtained by using a standard Kalman smoother. In [6] the Rauch-Tung-Streibel (RTS) [25] algorithm is implemented.

In frequency-domain formulations of EM, this smoothing is replaced by simple matrix operations (see, for example, [7], [9]).

### B. The M-step for shift-operator models

When uniform sampling is used then the M-step is relatively straightforward. In this step, we obtain the parameter vector that maximizes the auxiliary function $\mathcal{Q}(\theta,\hat{\theta}_i)$. The new parameter vector $\hat{\theta}_{i+1}$ can then be obtained by differentiating $\mathcal{Q}(\theta,\hat{\theta}_i)$ and setting the derivative to zero, to find its maximum. (Explicit expressions can be found, for example, in [21, Chapter 6] or [6]). However, we face a difficulty when non-uniform sampling is used since the discrete parameter is time-varying and is a nonlinear function of the underlying continuous system parameters. Hence, maximization with respect to the continuous parameters will, in general, be a non-convex problem. We will show in Section IV-B below how we can simplify the M-step when Assumption 1 holds and we use the incremental model parametrization.

## IV. EM ALGORITHM FOR FAST-SAMPLED SYSTEMS

In this section, we present the equivalent incremental form of the EM algorithm.

### A. E-Step: the Kalman smoother in incremental form

Here we rewrite the, so called, RTS Kalman smoother implementation originally presented in [25] and later rederived by several authors (see, for example, [21]). In that work an extension is also presented to continuous-time systems. In [26], the result is rederived in a purely continuous framework where equivalent (but simpler) expressions are derived. Here we present a *sampled-data* version of the Kalman smoother, i.e., a discrete time formulation that explicitly includes the sampling interval $\Delta_k$.

We first describe an alternative form of the RTS Kalman smoother. This gives the estimate $\hat{x}_{k|N} = E\{x_k|Y_k\}$ and its associated error covariance matrix $P_{k|N}$, based on the Kalman filter estimate $\hat{x}_{k|k}$ and the associated error covariance matrix $P_{k|k}$.

*Lemma 3:* The RTS smoother equations can be expressed in the following alternative form:

$$\hat{x}_{k|N} = \hat{x}_{k|k} + P_{k|k}(A_k^q)^T w_k \qquad (41)$$
$$P_{k|N} = P_{k|k} - P_{k|k}(A_k^q)^T\Omega_k A_k^q P_{k|k} \qquad (42)$$

where $w_k$ and $\Omega_k$ satisfy the following respective recursive equations:

$$w_{k-1} = (I - (C_k^q)^T(K_k^q)^T)(A_k^q)^T w_k + \epsilon_{m,k} \qquad (43)$$
$$\begin{aligned}\Omega_{k-1} &= (I - (C_k^q)^T(K_k^q)^T)(A_k^q)^T\Omega_k A_k^q(I - K_k^q C_k^q) \\ &\quad + (P_{k|k-1})^{-1}K_k^q C_k^q \qquad (44)\end{aligned}$$

where the backward recursions are initialized by $w_N = 0$ and $\Omega_N = 0$. The innovations $\epsilon_k$ and modified innovations $\epsilon_{m,k}$ are respectively given by:

$$\epsilon_k = y_k - C_k^q\hat{x}_{k|k-1} - D_k^q u_k \qquad (45)$$
$$\epsilon_{m,k} = (P_{k|k-1})^{-1}K_k^q\epsilon_k \qquad (46)$$

*Proof:* See, for example, [27] and [28]. ■

The following additional covariance matrices are also required for the E-step [5], [6].

The above alternative implementation of the RTS filter appears, for example, in [27, Chapter 13]. However, no reference to sampling is made. In the next result, we present the incremental form of the RTS Kalman smoother. This form of the smoother explicitly includes the sampling period.

*Lemma 4:* Consider the incremental model given in (14)-(22). The corresponding incremental RTS Kalman smoother is given by

$$d\,\hat{x}_{k|k}^+ = [A_k^\delta\hat{x}_{k|k} + B_k^\delta u_k]\Delta_k + K_{k+1}^\delta\epsilon_{k+1}\Delta_{k+1} \qquad (47)$$
$$\begin{aligned}d\,P_{k|k-1}^+ &= [A_k^\delta P_{k|k} + P_{k|k}(A_k^\delta)^T + Q_k^\delta]\Delta_k \\
&\quad - K_{k+1}^\delta C_{k+1}^\delta P_{k|k}\Delta_{k+1} + \Delta_k^2 A_k^\delta P_{klk}(A_k^\delta)^T \\
&\quad - \Delta_k\Delta_{k+1}(K_{k+1}^\delta C_{k+1}^\delta(A_k^\delta P_{k|k} + P_{k|k}(A_k^\delta)^T \\
&\qquad + \Delta_k A_k^\delta P_{klk}(A_k^\delta)^T) \qquad (48)\end{aligned}$$
$$K_k^\delta = P_{k|k-1}(C_k^\delta)^T(\Delta_k C_k^\delta P_{k|k-1}(C_k^\delta)^T + R_k^\delta)^{-1} \qquad (49)$$
$$P_{k|k} = (I - \Delta_k K_k^\delta C_k^\delta)P_{k|k-1} \qquad (50)$$

combined with the incremental form of the RTS smoother (41)-(46), i.e.

$$\hat{x}_{k|N} = \hat{x}_{k|k} + P_{k|k}(I + \Delta_k A_k^\delta)^T w_k \quad (51)$$

$$P_{k|N} = P_{k|k} - P_{k|k}(I + \Delta_k A_k^\delta)^T \Omega_k (I + \Delta_k A_k^\delta) P_{k|k} \quad (52)$$

where

$$-d\, w_{k-1}^+ = (A_k^\delta - K_k^\delta C_k^\delta + \Delta_k A_k^\delta K_k^\delta C_k^\delta)^T w_k \Delta_k + \epsilon_{m,k} \quad (53)$$

$$\begin{aligned} -d\,\Omega_{k-1}^+ = {} & [(A_k^\delta - K_k^\delta C_k^\delta)^T \Omega_k + \Omega_k (A_k^\delta - K_k^\delta C_k^\delta) \\ & + (P_{k|k-1})^{-1} K_k^\delta C_k^\delta]\Delta_k \\ & + [(A_k^\delta - K_k^\delta C_k^\delta)^T \Omega_k (A_k^\delta - K_k^\delta C_k^\delta)^T \Omega_k \\ & - (A_k^\delta K_k^\delta C_k^\delta)^T \Omega_k - \Omega_k A_k^\delta K_k^\delta C_k^\delta]\Delta_k^2 \\ & - [(A_k^\delta - K_k^\delta C_k^\delta)^T \Omega_k A_k^\delta K_k^\delta C_k^\delta \\ & + (A_k^\delta K_k^\delta C_k^\delta)^T \Omega_k (A_k^\delta - K_k^\delta C_k^\delta)]\Delta_k^3 \\ & + [(A_k^\delta K_k^\delta C_k^\delta)^T \Omega_k A_k^\delta K_k^\delta C_k^\delta]\Delta_k^4 \end{aligned} \quad (54)$$

where the modified innovations are given by:

$$\epsilon_{m,k} = (C_k^\delta)^T (\Delta_k C_k^\delta P_{k|k-1} (C_k^\delta)^T + R_\delta)^{-1} \epsilon_k \Delta_k \quad (55)$$

*Proof:* The incremental form of the filter is obtained by manipulating the Kalman filter equations. The derivation can be found, for example, in [13]. An important observation is that the Kalman gain has to be scaled by the corresponding sampling period in order to achieve convergence to the continuous filter (see Lemma 5 below), i.e.

$$K_k^q = \Delta_k\, K_k^\delta \quad (56)$$

To obtain the incremental form of the RTS Kalman smoother, we substitute the matrix transformations in (28)-(30) into (41)-(46). Equations (53)-(54) are the obtained by rearranging terms. For example, from (43), we have that:

$$\begin{aligned} w_{k-1} &= (I_n - (C_k^q)^T (\Delta_k K_k^\delta)^T)(I_n + \Delta_k A_k^\delta)^T w_k + \epsilon_{m,k} \\ &= (I_n + \Delta_k [(A_k^\delta)^T - (C_k^q)^T (K_k^\delta)^T] \\ &\quad - \Delta_k^2 [(C_k^q)^T (K_k^\delta)^T (A_k^\delta)^T]) w_k + \epsilon_{m,k} \end{aligned}$$

Hence

$$w_{k-1} - w_k = (A_k^\delta - K_k^\delta C_k^\delta - \Delta_k A_k^\delta K_k^\delta C_k^\delta)^T w_k \Delta_k + \epsilon_{m,k} \quad (57)$$

which corresponds to (53). Equation (54) is similarly obtained. ■

The above lemma describes an incremental form of the RTS smoother which explicitly includes the sampling period. The following result shows that the RTS smoother converges to the corresponding continuous-time smoother as $\bar{\Delta} \to 0$.

*Lemma 5:* As the sampling rate is increased, i.e., the bound $\bar{\Delta}$ goes to zero, the incremental form of the RTS smoother converges to the continuous-time form:

$$d\,\hat{x}_f(t) = A_c \hat{x}_f(t) dt + B_c u(t) dt + K_c(t) d\epsilon(t) \quad (58)$$

$$d\,P_f(t) = [A_c P_f(t) + P_f(t) A_c^T + Q_c - K_c(t) C_c P_f(t)] dt \quad (59)$$

$$K_c(t) = P_f(t) C_c^T (R_c)^{-1} \quad (60)$$

subject to $\hat{x}_f(0) = \mu_o$ and $P_f(0) = \Sigma_0$, together with

$$\hat{x}_s(t) = \hat{x}_f(t) + P_f(t) w(t) \quad (61)$$

$$P_s(t) = P_f(t) - P_f(t)\Omega(t) P_f(t) \quad (62)$$

$$-d\,w(t) = (A_c^T - C_c^T K_c(t)^T) w(t) dt + d\epsilon_m(t) \quad (63)$$

$$\begin{aligned} -d\,\Omega(t) = {} & [(A_c - K_c(t) C_c)^T \Omega(t) \\ & + \Omega(t)(A_c - K_c(t) C_c) + (P_f(t))^{-1} K_c(t) C_c] dt \end{aligned} \quad (64)$$

subject to $w(T_f) = 0$ and $\Omega(T_f) = 0$, where $T_f = N\Delta$ is the final time instant of the considered smoothing interval. The innovations and modified innovation signal are respectively given by

$$d\epsilon(t) = dz(t) - C_c \hat{x}_f(t) dt - D_c u(t) dt \quad (65)$$

$$d\epsilon_m(t) = C_c^T R_c^{-1} d\epsilon(t) \quad (66)$$

*Proof:* The proof readily follows from the incremental form of the smoother presented in Lemma 4. When the maximum sampling interval $\bar{\Delta}$ goes to zero, all the sampling intervals $\Delta_k$ also go to zero. ■

*Remark 5:* The continuous and discrete smoother equations appear to be unstable when using the innovations as driving input (see equations (43), (53), and (63)). However, when considering the relationship between available data $(u, y)$ as input and smoothed estimates $\hat{x}$ as output, this is stable (see, for example, [29], [30]).

### *B. M-step for the incremental model*

In this section we reformulate the second step of the EM algorithm by using the incremental model, which is more appropriate for fast sampling rates.

The transformations between the matrices in the incremental and shift-operator models are given in (28)-(30). Then we have the following closed form result for the maximization step:

*Lemma 6:* Consider the use of fast non-uniform sampling subject to Assumption 1. When using incremental models, the system parameters that maximize the auxiliary function $\mathcal{Q}(\theta, \hat{\theta}_i)$ are given by

$$[\hat{A}_c, \hat{B}_c] = (\Psi - \Upsilon)\, \Gamma^{-1} \quad (67)$$

$$[\hat{C}_c, \hat{D}_c] = \Lambda \cdot \Pi^{-1} \quad (68)$$

$$\hat{Q}_c = \frac{1}{N}\left(\Phi - (\Psi - \Upsilon)\, \Gamma^{-1} (\Psi - \Upsilon)^T\right) \quad (69)$$

$$\hat{R}_c = \frac{1}{N}\left(\Omega - \Lambda\, \Pi^{-1} \Lambda^T\right) \quad (70)$$

where

$$\Phi \triangleq \sum_{k=1}^{N} \Delta_k^{-1} E\left\{(x_k - x_{k-1})(x_k - x_{k-1})^T\right\} \quad (71)$$

$$\Omega \triangleq \sum_{k=0}^{N-1} \Delta_k y_k y_k^T \quad (72)$$

$$\Psi \triangleq \sum_{k=1}^{N} E\left\{x_k z_{k-1}^T\right\} \quad (73)$$

$$\Gamma \triangleq \sum_{k=1}^{N} \Delta_k E\left\{z_{k-1} z_{k-1}^T\right\} \tag{74}$$

$$\Lambda \triangleq \sum_{k=0}^{N-1} \Delta_k E\left\{y_k z_k^T\right\} \tag{75}$$

$$\Pi \triangleq \sum_{k=0}^{N-1} \Delta_k E\left\{z_k z_k^T\right\} \tag{76}$$

$$\Upsilon \triangleq \sum_{t=1}^{N} E\left\{x_{t-1} z_{t-1}^T\right\} \tag{77}$$

and we use the definition $z_k = \begin{bmatrix} x_k \\ u_k \end{bmatrix}$

*Proof:* The above expressions are obtained by replacing the definitions (28)-(30) into the M-step described in (40) and then maximizing the auxiliary function $\mathcal{Q}(\theta, \hat{\theta}_i)$ with respect to the incremental model matrices.

A key observation is that, with fast sampling rates, $\delta_k^\delta \to \delta^c$, i.e., we can assume time-invariant parameters provided we make $\Delta_k$ explicit!

The details of the proof are similar to those presented in [31] and can be found in [28]. ■

*Remark 6:* From the E-step we have available the optimal estimate of the system state given the output data $Y = \{y_0, \dots, y_N\}$. This is given by the RTS Kalman smoother. The matrices (71)-(77) are then given by

$$\Phi = \sum_{k=1}^{N} \Delta_k^{-1} \Big( (\hat{x}_k - \hat{x}_{k-1})(\hat{x}_k - \hat{x}_{k-1})^T + P_{k|N} - M_{k|N} - M_{k|N}^T + P_{k-1|N} \Big) \tag{78}$$

$$\Omega = \sum_{k=0}^{N-1} \Delta_k y_k y_k^T \tag{79}$$

$$\Psi = \sum_{k=1}^{N} \left( x_k \hat{z}_{k-1}^T + \begin{bmatrix} M_{k|N} & 0 \end{bmatrix} \right) \tag{80}$$

$$\Gamma = \sum_{k=1}^{N} \Delta_k \left( \hat{z}_{k-1} \hat{z}_{k-1}^T + \begin{bmatrix} P_{k-1|N} & 0 \\ 0 & 0 \end{bmatrix} \right) \tag{81}$$

$$\Upsilon = \sum_{k=1}^{N} \left( \hat{x}_{k-1} \hat{z}_{k-1}^T + \begin{bmatrix} P_{k-1|N} & 0 \end{bmatrix} \right) \tag{82}$$

$$\Lambda = \sum_{k=0}^{N-1} \Delta_k y_k \hat{z}_k^T \tag{83}$$

$$\Pi = \sum_{k=0}^{N-1} \Delta_k \left( \hat{z}_k \hat{z}_k^T + \begin{bmatrix} P_{k|N} & 0 \\ 0 & 0 \end{bmatrix} \right) \tag{84}$$

*Remark 7:* Two important points need to be made about the expressions given above, namely (i) they directly lead to the continuous parameters and (ii) the sampling period sequence $\{\Delta_k\}$ appears explicitly.

## V. Robust dual-domain EM

For state space models, the *natural* choice for the *hidden variables* $X$ is the state sequence. However, the choice of the *hidden variables* is highly dependent on the application. For example, EM has been used when there are missing measurements in the data [32].

The EM algorithm can also be applied to maximize the likelihood function when the data is transformed to the frequency domain using the discrete Fourier transform (DFT) [7], [9]. In [33], frequency domain maximum likelihood is applied on a restricted bandwidth in order to achieve robustness to the assumptions about the system model. This is particularly important for sampled-data models where, unmodelled high frequency dynamics (even beyond the Nyquist rate $\frac{\pi}{\Delta}$) have an impact on the exact sampled-data model due to the *aliasing* effect.

More recently [34] it has been shown that, instead of discarding some frequency components in the likelihood function, these can be included as *hidden variables* in the EM algorithm. This approach allows us to preserve the properties of the maximum likelihood estimates. The basic idea is to define a (non-invertible) matrix transformation

$$\vec{g} = E\vec{y} \tag{85}$$

where $\vec{y}$ is the available (time-domain) data, $E$ is a (non invertible) matrix, and $\vec{g}$ is the data that we *trust*. Note that (85) may correspond to missing measurements in vector $\vec{y}$, or, if $E$ includes the DFT matrix transformation, to consider only some frequency domain components.

The *hidden variable* can then be chosen as $X = \begin{bmatrix} \vec{h} \\ \vec{x} \end{bmatrix}$, where $\vec{h} = E_\perp \vec{y}$, and $\vec{x}$ is the state sequence. Matrix $E_\perp$ is chosen such that the matrix $F = \begin{bmatrix} E \\ E_\perp \end{bmatrix}$ is non-singular.

The methodology presented in [34] in the context of discrete-time state-space models can be equally applied to incremental models. Thus, for fast sampling rates, it can be exploited to identify continuous state-space models such that robustness is achieved, for example, to high frequency undermodeling.

## VI. Examples

In this section we present examples to illustrate the application of the EM algorithm for incremental models at fast sampling rates.

*Example 1:* Consider a state-space continuous-time system as in (1)-(2), where the matrices are given by

$$A_c = \begin{bmatrix} -6.00 & -2.50 \\ 2.00 & 0 \end{bmatrix} \qquad B_c = \begin{bmatrix} 2.00 \\ 0 \end{bmatrix} \tag{86}$$

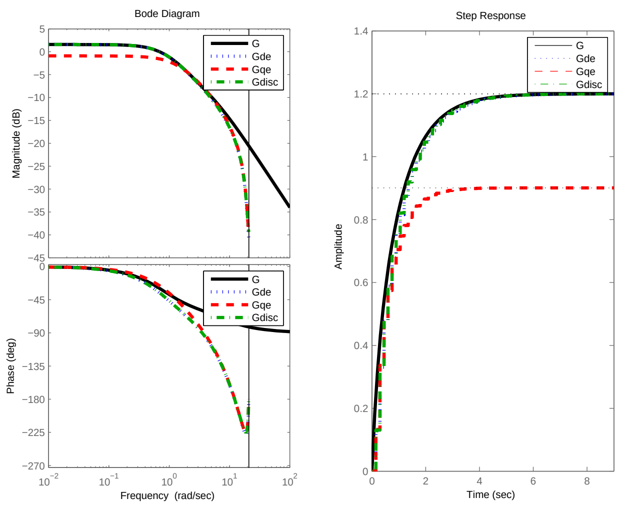
$$C_c = \begin{bmatrix} 1.00 & 1.50 \end{bmatrix} \qquad D_c = 0 \tag{87}$$

The spectral densities of the continuous-time process and measurement noise are respectively given by

$$Q_c = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix} \qquad R_c = 0.1 \tag{88}$$

The corresponding transfer function from the deterministic input $u(t)$ to the output is given by

$$G(s) = \frac{2(s+3)}{s^2 + 6s + 5} \tag{89}$$




(a) Frequency response (b) Step response

Fig. 1. Comparison between real and estimated systems for $\Delta = 0.15[s]$. The thick black line corresponds to the true continuous-time system $G$. The green dot-dashed line corresponds to the exact sampled-data model $G_{disc}$, whereas the red dashed line is the estimated shift-operator model $G_{qe}$, and the blue dotted line is the estimated incremental model $G_{de}$ (Example 1).

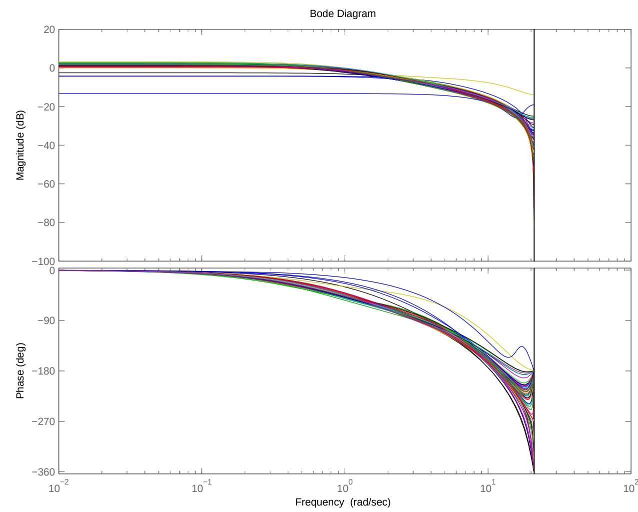
To compare the EM algorithm using incremental and shift-operator models we consider a uniform sampling period $\Delta = 0.15$.

To simulate data from the associated continuous-time system we use the exact sampled-data model obtained in Section II-B. The continuous-time input is generated using a pseudo-random binary sequence of length $2047 = 2^{11} - 1$, passed through a ZOH of fixed sampling period $\Delta = 0.15$. The continuous-time horizon is $T_f = 300$. Thus we have $N = 2000$ data points.

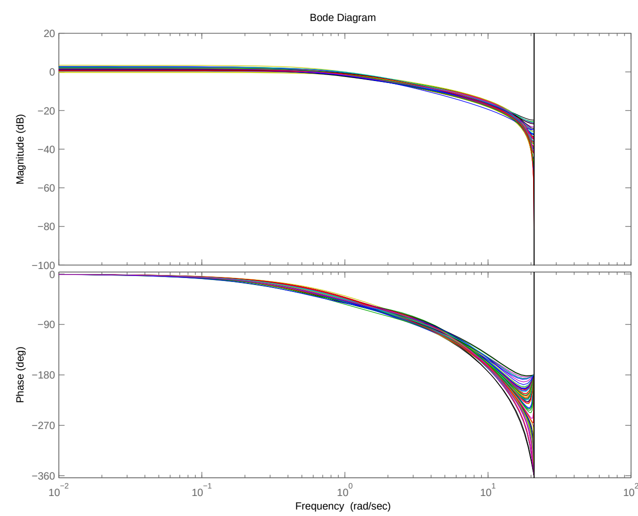
Figure 1 shows the frequency and the step response of the true and identified systems for a sampling period $\Delta = 0.15$. The thick black line corresponds to the true continuous-time system $G$. The green dot-dashed line corresponds to the exact sampled-data model $G_{disc}$, whereas the red dashed line is the estimated shift-operator model $G_{qe}$, and the blue dotted line is the estimated incremental model $G_{de}$. The estimated models are the **average** of the models obtained for 60 realizations of the noise processes.

We see that the incremental model can hardly be distinguished from the exact discrete-model at the resolution of the plot. Figure 1(b) shows a clear bias in the DC gain for the model estimated using EM for shift-operator model. This is consistent with the Bode diagram in Figure 1(a). On the other hand, the incremental model presents a better accuracy in the time constant and the DC gain estimation.

The problem with the shift operator model arises from the presence of several outliers in the estimated models. These outliers arise for numerical issues and are much less common when using the EM algorithm for incremental models. This can be seen in Figure 2, where the Bode plots of the estimated (shift-operator and incremental) models are shown for each of the realizations.




(a) EM for shift-operator models




(b) EM for incremental models

Fig. 2. Frequency response for the models estimated for 60 different noise realizations when $\Delta = 0.15$ (Example 1).

*Example 2:* We consider the same continuous-time system in Example 1. However in this case we assume non-uniform sampled data.

The sampling instants are generated (only once) as $t_k = \Delta(k + \eta_k)$, where we consider the smallest sampling period in the previous example, i.e., $\Delta = 0.006$, and $\eta_k$ is a random perturbation, uniformly distributed on the interval $[-0.25; 0.25]$. This means that there is a variation of $\pm 25\%$ in the *nominal* sampling period. However, note that these sampling instants are exactly known by the algorithm.

The results are shown in Figure 3. The left hand side plot shows the frequency response of the true system $G$ and the (average) estimated system $G_e$. The estimated continuous-time system is obtained from the matrices obtained in Lemma 6. The plot on the right hand side of each figure shows the estimated systems corresponding to each of 20 realization of the experiment. We can see that, in both cases, the proposed implementation of EM gives a an excellent estimate of the

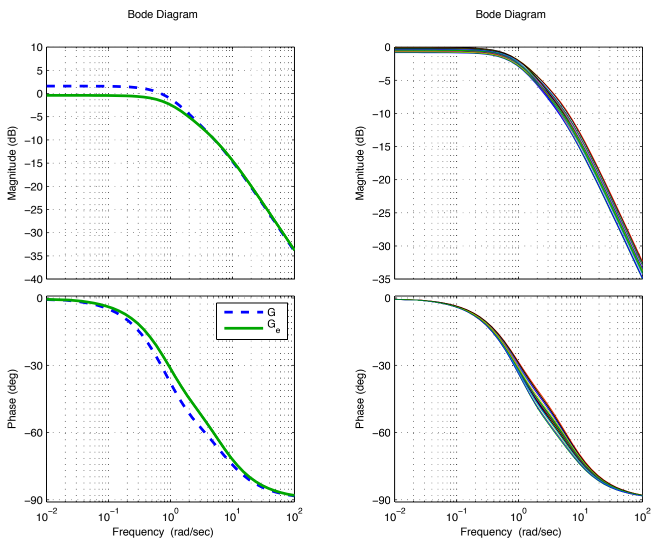



Fig. 3. System identification with nonuniform fast-sampled data for $\pm 25\%$ variation of the *nominal* sampling interval (Example 2). The left hand side plot compares the true and estimated (averaged) model. The right hand side plot shows the result the estimated model for 20 different realizations.

continuous-time system.

## VII. Conclusions

In this paper we have shown the use of the EM algorithm to identify continuous-time state-space models from sampled data. The sampling intervals are not assumed constant and, as a consequence, a discrete time-varying system has been used in the development. Assuming that the sampling rate is fast (i.e., all the sampling intervals are *small*) the discrete-time description has been parametrized in incremental form, in order to estimate the continuous-time system parameters. The results show that the use of incremental models in the EM formulation provides excellent numerical behaviour and a direct way to estimate continuous state-space models.

## References

[1] G. Goodwin, S. Graebe, and M. Salgado, *Control System Design*. New Jersey: Prentice Hall, 2001.

[2] E. Hannan, "The identification and parameterization of Armax and state space forms," *Econometrica*, vol. 44, no. 4, pp. 713–723, 1976.

[3] J. Durbin and S. Koopman, *Time series analysis by state space methods*. Oxford University Press, 2005.

[4] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from imcomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[5] R. Shumway and D. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," Division of Statistics, University of California, Davis, Tech. Rep., 1981.

[6] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667–1682, 2005.

[7] J. Agüero, J. Yuz, and G. Goodwin, "Frequency domain identification of MIMO state space models using the EM algorithm," in *European Control Conference - ECC'07*, Kos, Greece, 2007.

[8] T. McKelvey and A. Helmersson, "State space parameterization of multivariable linear systems using tridiagonal matrix form," in *35th IEEE Conference on Decision and Control*, 1996.

[9] A. Wills, B. Ninnesss, and S. Gibson, "Maximum likelihood estimation of state space models from frequency domain data," *Automatic Control, IEEE Transactions on*, vol. 54, no. 1, pp. 19 – 33, Jan 2009.

[10] E. Larsson and T. Söderström, "Identification of continuous-time ar processes from unevenly sampled data," *Automatica*, vol. 38, pp. 709–718, Jan 2002.

[11] R. Middleton and G. Goodwin, *Digital Control and Estimation. A Unified Approach*. Englewood Cliffs, New Jersey: Prentice Hall, 1990.

[12] A. Feuer and G. Goodwin, *Sampling in Digital Signal Processing and Control*. Boston: Birkhäuser, 1996.

[13] M. Salgado, R. Middleton, and G. Goodwin, "Connection between continuous and discrete Riccati equations with applications to Kalman filtering," *Control Theory and Applications, IEE Proceedings D*, vol. 135, no. 1, pp. 28–34, 1988.

[14] G. Goodwin, R. Middleton, and H. Poor, "High-speed digital signal processing and control," *Proceedings of the IEEE*, vol. 80, no. 2, pp. 240–259, 1992.

[15] K. Premaratne, S. Touset, and E. Jury, "Root distribution of delta-operator formulated polynomials," *IEE Proceedings in Control Theory and Applications*, vol. 147, no. 1, pp. 1–12, January 2000.

[16] M. Mansour, "Stability and robust stability of discrete-time systems in the $\delta$-transform," in *Fundamentals of Discrete-Time Systems: A Tribute to Prof. Eliahu I. Jury*. Albuquerque, USA: TSI Press, 1993.

[17] P. Suchomski, "Numerically robust delta-domain solutions to discrete-time Lyapunov equations," *Systems and Control Letters*, vol. 47, no. 4, pp. 319–326, 2002.

[18] V. Kadirkamanathan and S. Anderson, "Maximum-likelihood estimation of delta-domain model parameters from noisy output signals," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, Part 1, pp. 3765 – 3770, Aug 2008.

[19] J. Gillberg and L. Ljung, "Frequency domain identification of continuous-time output error models, part ii: Non-uniformly sampled data and b-spline output approximation," *Automatica*, vol. 46, no. 1, pp. 11–18, Jan 2010.

[20] F. Ding, L. Qiu, and T. Chen, "Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems," *Automatica*, vol. 45, no. 2, pp. 324–332, Jan 2009.

[21] R. Shumway and D. Stoffer, *Time Series Analysis and its Applications (2nd ed.)*. Springer, 2006.

[22] B. Øksendal, *Stochastic Differential Equations. An Introduction with Applications*, 6th ed. Springer, 2003.

[23] K. Åström, *Introduction to Stochastic Control Theory*. New York: Academic Press, 1970.

[24] L. Ljung and A. Wills, "Issues in sampling and estimating continuous-time models with stochastic disturbances," *Automatica*, vol. 46, no. 5, pp. 925 – 931, 2010.

[25] H. Rauch, F. Tung, and C. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA journal*, vol. 3, no. 8, pp. 1445–1450, Jan 1965.

[26] T. Kailath and P. Frost, "An innovations approach to least-squares estimation–part ii: Linear smoothing in additive white noise," *Automatic Control, IEEE Transactions on*, vol. 13, no. 6, pp. 655 – 660, 1968.

[27] A. Bryson and Y. Ho, *Applied optimal control: optimization, estimation, and control*. Washington, D.C.: Hemisphere Publishing Corp. – John Wiley & Sons, Jan 1975.

[28] J. Yuz, J. Alfaro, J. Agüero, and G. Goodwin, "Identification of continuous-time state space models from nonuniform fast-sampled data," *(Submitted)*, 2010.

[29] S. Chirarattananon and B. Anderson, "The fixed-lag smoother as a stable finite-dimensional linear system," *Automatica*, vol. 7, no. 6, pp. 657–669, Jan 1971.

[30] J. Meditch, "A survey of data smoothing for linear and nonlinear dynamic systems," *Automatica*, vol. 9, no. 2, pp. 151–162, Jan 1973.

[31] S. Gibson, "Maximum likelihood estimation of multivariable dynamic systems via the EM algorithm," Ph.D. dissertation, School of Electrical Engineering and Computer Science, The University of Newcastle, Australia, 2003.

[32] G. Goodwin and A. Feuer, "Estimation with missing data," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 5, no. 3, pp. 220–244, 1999.

[33] G. Goodwin, J. Agüero, J. Welsh, J. Yuz, G. Adams, and C. Rojas, "Robust identification of process models from plant data," *Journal of Process Control*, vol. 18, no. 9, pp. 810–820, 2008.

[34] J. Agüero, J. Yuz, G. Goodwin, and W. Tang, "Identification of state-space systems using a dual time-frequency domain approach," *(submitted)*, 2010.