Correspondence

Maximum-Likelihood Estimation of Delta-Domain Model Parameters From Noisy Output Signals

Visakan Kadirkamanathan, *Member, IEEE*, and Sean R. Anderson

***Abstract*—Fast sampling is desirable to describe signal transmission through wide-bandwidth systems. The delta-operator provides an ideal discrete-time modeling description for such fast-sampled systems. However, the estimation of delta-domain model parameters is usually biased by directly applying the delta-transformations to a sampled signal corrupted by additive measurement noise. This problem is solved here by expectation-maximization, where the delta-transformations of the true signal are estimated and then used to obtain the model parameters. The method is demonstrated on a numerical example to improve on the accuracy of using a shift operator approach when the sample rate is fast.**

***Index Terms*—Delta operator, EM algorithm, expectation-maximization, fast sampling, system identification.**

I. Introduction

High-speed signal transmission in digital communications and control systems has become common in recent years due to advances in technology that allow rapid signal processing and storage of the large quantities of associated data. Communication receivers and control systems require processing of these fast-sampled signals in estimating the communication channel or identifying system dynamics. The fast sampling of signals is desirable for wide bandwidth systems, however, commonly used discrete-time methods for signal processing are prone to ill-conditioning at fast-sample rates [1], which requires careful consideration. This contribution addresses the problem of parameter estimation for fast-sampled linear time-invariant systems from noisy measurement signals.

The term *fast sampling* refers here to a signal that is sampled rapidly compared to the decay rate of the dominant system time constant. The successive values of such fast-sampled signals are numerically similar, which often leads to ill-conditioning of matrices containing signals transformed by the shift operator [2], [3]. The δ-operator transformations of a signal, by contrast, tend toward the signal derivatives as sample rate increases: The δ-operator is a forward difference operator that transforms a sampled signal to a first order discrete-time representation of gradient information [4]–[6]. Differing orders of δ transformation at a sample point are unlikely to be numerically similar, which typically gives an improvement in the numerical properties of signal processing algorithms at fast-sample rates compared with the shift operator [1].

Manuscript received December 1, 2006; revised February 5, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhengyuan (Daniel) Xu.

V. Kadirkamanathan is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, U.K. (e-mail: visakan@sheffield.ac.uk).

S. R. Anderson is with the Centre for Signal Processing in Neuroimaging and Systems Neuroscience, Department of Psychology, University of Sheffield, Sheffield, S10 2TP, U.K. (e-mail: s.anderson@sheffield.ac.uk).

Digital Object Identifier 10.1109/TSP.2008.920443

The δ-operator has been widely studied in the adaptive signal processing context to improve the numerical conditioning of recursive estimation algorithms such as least mean square (LMS) [7], lattice [8], and Levinson [9] algorithms. In an offline modeling context the δ-operator can be used to provide an exact discrete-time model of a system, or in the direct identification of continuous-time systems where the estimated parameters tend to the continuous-time values with an increasing sample rate [10]–[13]. In each context, it is common to use an observed signal to form the δ-transformations of the signal.

Additive measurement noise on the output signal biases the estimation of input–output (I/O) model parameters when using a linear regression technique such as least squares (LS), motivating prefiltering [14], or noise modeling [15]. The bias is exacerbated further in the identification of systems using the δ-operator because of the (derivative-like) amplification of high frequency measurement noise when forming the δ transformations. A naive attempt to model directly from the noisy transformed signal, e.g., using LS, would lead to severe parameter bias. Therefore, previous approaches to modeling systems in the δ-domain have focused on either low-pass filtering the observed signal before parameter estimation or attempting to compensate for the bias as part of the parameter estimation stage, such as bias compensating least squares [16].

The contrasting approach to systems modeling taken here is to regard the δ-transformations of the true output signal as unobserved, or hidden, quantities. This motivates a state-space representation of the model, which has a canonical form where the δ-transformations of the true output signal correspond to the system states. The state-space model structure naturally leads to the formation of a state and parameter estimation problem. The framework chosen here to solve the estimation problem is maximum likelihood: the algorithm thus derived is based on the principle of expectation-maximization (EM) [17]. This is advantageous because EM is a well-known algorithm, often used in channel identification [18], [19] as well as system identification [20]–[22] and convergence behavior is well understood [23], [24].

In Section II, a brief background is given to the EM algorithm, which highlights the main concepts and provides a basis for the following algorithm development. The system model structure is presented in Section III and the identification algorithm for the model is derived in Section IV. An example of the identification procedure is given in Section V, which demonstrates the performance of the proposed algorithm and compares the results to an equivalent shift operator approach; it is shown that the δ approach performs accurately across a range of sample frequencies whilst the shift approach is adversely affected by fast-sample rates. The conclusions are made in Section VI.

II. Background on the EM Algorithm

The EM algorithm is an iterative method of providing a maximum likelihood estimate of a set of parameters θ, which are dependent on both a hidden data set $\mathcal{X}$ and an observed data set $\mathcal{Y}$ [17], [20]. The maximum-likelihood estimate of the parameters may be obtained from the log-likelihood function

$$L(\theta) = \log p(\mathcal{Y}|\theta) = \log \int_{\mathcal{X}} p(\mathcal{X}, \mathcal{Y}|\theta) d\mathcal{X}. \tag{1}$$

It is possible to obtain a lower bound on $L(\theta)$ by introducing the distribution of the hidden data conditioned upon *any* available estimate

of the parameter vector; in the iterative procedure of EM the available estimate can be taken to be that at the k^{th} iteration, $\hat{\theta}_k$, so that first

$$L(\theta) = \log \int_{\mathcal{X}} p(\mathcal{X}|\mathcal{Y}, \theta_k) \frac{p(\mathcal{X}, \mathcal{Y}|\theta)}{p(\mathcal{X}|\mathcal{Y}, \theta_k)} d\mathcal{X}. \quad (2)$$

Noting that $p(\mathcal{X}|\mathcal{Y}, \theta_k) > 0$ and that $\int_{\mathcal{X}} p(\mathcal{X}|\mathcal{Y}, \theta_k) d\mathcal{X} = 1$, Jensen's inequality may be applied to (2) to yield the expression

$$L(\theta) \geq \int_{\mathcal{X}} p(\mathcal{X}|\mathcal{Y}, \theta_k) \log \frac{p(\mathcal{X}, \mathcal{Y}|\theta)}{p(\mathcal{X}|\mathcal{Y}, \theta_k)} d\mathcal{X} \quad (3)$$

which, after rearranging, leads to the desired expression of the lower bound on $L(\theta)$

$$L(\theta) \geq F(\theta, \hat{\theta}_k) = \int_{\mathcal{X}} p(\mathcal{X}|\mathcal{Y}, \theta_k) \log p(\mathcal{X}, \mathcal{Y}|\theta) d\mathcal{X} - \int_{\mathcal{X}} p(\mathcal{X}|\mathcal{Y}, \theta_k) \log p(\mathcal{X}|\mathcal{Y}, \theta_k) d\mathcal{X}. \quad (4)$$

Since $F(\theta, \hat{\theta}_k)$ provides a lower bound on $L(\theta)$, any value of θ that increases $F(\theta, \hat{\theta}_k)$ must also increase $L(\theta)$. Hence, the task of maximizing $L(\theta)$ directly becomes one of iteratively maximizing $F(\theta, \hat{\theta}_k)$.

The second term in $F(\theta, \hat{\theta}_k)$ is not a function of θ, which means that it may be neglected when performing the maximization. This leads to the definition of the first step in the EM algorithm, the expectation problem, expressed as

$$Q(\theta, \hat{\theta}_k) = \int_{\mathcal{X}} p(\mathcal{X}|\mathcal{Y}, \theta_k) \log p(\mathcal{X}, \mathcal{Y}|\theta) d\mathcal{X} = \mathbb{E}[\log p(\mathcal{X}, \mathcal{Y}|\theta)|\mathcal{Y}, \hat{\theta}_k] \quad (5)$$

where $\mathbb{E}[.]$ denotes the expected value. The definition of the function $Q(\theta, \hat{\theta}_k)$ leads to the expression of the second step in the EM algorithm, that is, the maximization problem

$$\hat{\theta}_{k+1} = \arg\max_{\theta} Q(\theta, \hat{\theta}_k). \quad (6)$$

The two steps of the EM algorithm are repeated until some metric of convergence is satisfied; properties of convergence of the log-likelihood function and the parameter estimates are discussed in [23] and [24].

III. System Representation

A single-input single-output (SISO) linear time-invariant system can be described in I/O form using the δ-operator as

$$\delta^n z_t + a_{n-1}\delta^{n-1} z_t + \ldots + a_0 z_t = b_m \delta^m u_t + \ldots + b_1 \delta u_t + b_0 u_t + e_t \quad (7)$$

where

$$\delta = \frac{q-1}{T} \quad (8)$$

and $z_t \in \mathbb{R}$ is the system output at time t, $u_t \in \mathbb{R}$ is the system input, $e_t \sim N(0, \sigma^2)$ is a zero mean Gaussian white process noise signal, T is the sample time and q is the forward shift operator, i.e., $q u_t = u_{t+1}$.

The modeling task is to estimate the parameter vector θ, which is typically performed using the one-step-ahead prediction model corresponding to (7)

$$\delta^n z_t = \phi_t \theta + \varepsilon_t \quad (9)$$

where ε_t is the residual model error and

$$\phi_t = [\, \psi_t \quad \upsilon_t \,] \quad (10)$$

$$\theta = [\, \alpha \quad \beta \,]^T \quad (11)$$

$$\psi_t = [\, \delta^0 z_t \quad \ldots \quad \delta^{n-1} z_t \,] \quad (12)$$

$$\upsilon_t = [\, \delta^0 u_t \quad \ldots \quad \delta^m u_t \,] \quad (13)$$

$$\alpha = [\, -a_0 \quad \ldots \quad -a_{n-1} \,] \quad (14)$$

$$\beta = [\, b_0 \quad \ldots \quad b_m \,]. \quad (15)$$

An estimate of the parameter vector θ can be obtained via a technique such as maximum-likelihood estimation. However, it is usually the case that the target data z_t is corrupted by measurement noise and hence unavailable. Therefore, it is appropriate to introduce an augmented I/O model to describe such a system, which makes use of a canonical state-space form.

The system dynamics described in (7) can be mapped into a state equation, where the state vector contains the hidden data $\delta^k z_t$, for $k = 0, \ldots, n-1$,

$$\delta \mathbf{x}_t = A\mathbf{x}_t + B\mathbf{u}_t + We_t \quad (16)$$

where $A \in \mathbb{R}^{n \times n}$ is the state transition matrix, $B \in \mathbb{R}^{n \times (m+1)}$ is the input matrix, $W \in \mathbb{R}^n$, $\mathbf{x}_t \in \mathbb{R}^n$ is the system state and

$$\mathbf{x}_t = \psi_t^T \quad (17)$$

$$\mathbf{u}_t = \upsilon_t^T \quad (18)$$

$$A = \begin{bmatrix} 0 & I \\ \alpha & \end{bmatrix} \quad (19)$$

$$B = [\, 0 \quad \beta^T \,]^T \quad (20)$$

$$W = [\, 0 \quad \ldots \quad 0 \quad 1 \,]^T. \quad (21)$$

The mapping from the hidden to observed output signal is described by the measurement equation

$$y_t = C\mathbf{x}_t + \upsilon_t \quad (22)$$

where

$$C = [\, 1 \quad 0 \quad \ldots \quad 0 \,] \quad (23)$$

and $C \in \mathbb{R}^n$, $y_t \in \mathbb{R}$ is the observed system output and $\upsilon_t \sim N(0, \lambda^2)$ is zero-mean Gaussian white measurement noise.

In combination (16) and (22) provide a state-space representation of the I/O system to be identified from measurement noise corrupted output signals. The system considered is SISO, but (as is typical) the state-space representation is readily extended to include multiple-input multiple-output (MIMO) descriptions.

IV. Maximum-Likelihood Estimation

The aim of this section is to formulate the parameter estimation problem for the δ-domain model defined in (16) and (22), in terms of the well-known EM algorithm [17], [20]. This draws on the theory defined above in Section II on the background of the EM algorithm.

Lemma 1: The expected complete data log-likelihood function from (5), $Q(\theta, \hat{\theta}_k) = \mathbb{E}[\log p(\mathcal{X}, \mathcal{Y}|\theta)|\mathcal{Y}, \hat{\theta}_k]$, is defined for the δ-domain state-space model, described in (16) and (22), as

$$Q(\theta, \hat{\theta}_k) = \mathbb{E}\left[-\sum_{t=1}^{N} \frac{1}{2} (\delta \bar{x}_t - \phi_t \theta)^T \sigma^{-2} (\delta \bar{x}_t - \phi_t \theta) + c \right] \quad (24)$$

where c contains additional terms that are not dependent on θ and $\mathbf{x}_t$ is partitioned according to $\mathbf{x}_t = [\hat{\mathbf{x}}_t^T \quad \bar{x}_t]^T$.

Proof: The observed data set is defined as $\mathcal{Y} = \{y_1, \ldots, y_N\}$ and the hidden data set as $\mathcal{X} = \{\delta \mathbf{x}_1, \ldots, \delta \mathbf{x}_N\}$, where N is the number

of samples. The joint distribution of hidden and observed data can be expressed as

$$p(\mathcal{X},\mathcal{Y}|\theta)=\prod_{t=1}^{N}p(\delta\mathbf{x}_t|\mathbf{x}_t)\prod_{t=1}^{N}p(y_t|\mathbf{x}_t) \tag{25}$$

where $p(y_t|\mathbf{x}_t) \sim N(C\mathbf{x}_t,\lambda^2)$. Furthermore, the distribution $p(\delta\mathbf{x}_t|\mathbf{x}_t)$ can be partitioned so that

$$p(\delta\mathbf{x}_t|\mathbf{x}_t)=p(\delta\bar{x}_t|\mathbf{x}_t)p(\delta\tilde{\mathbf{x}}_t|\mathbf{x}_t) \tag{26}$$

where from (9)

$$p(\delta\bar{x}_t|\mathbf{x}_t)\sim N(\phi_t\theta,\sigma^2) \tag{27}$$

$$p(\delta\tilde{\mathbf{x}}_t|\mathbf{x}_t)=\delta([I\quad 0]\mathbf{x}_t) \tag{28}$$

where $\delta(.)$ denotes the Dirac delta function and I is the identity matrix. Substitution of (26) into (25) leads to

$$p(\mathcal{X},\mathcal{Y}|\theta)=\prod_{t=1}^{N}p(\delta\bar{x}_t|\mathbf{x}_t)\prod_{t=1}^{N}p(\delta\tilde{\mathbf{x}}_t|\mathbf{x}_t)\prod_{t=1}^{N}p(y_t|\mathbf{x}_t). \tag{29}$$

Substitution of (29) into (5), with appropriately parameterized normal distributions, leads directly to (24). ∎

Theorem 1: The parameter estimate that maximizes the expected complete data log-likelihood function $Q(\theta,\hat{\theta}_k)$ is

$$\hat{\theta}_{k+1}=\left(\sum_{t=1}^{N}\mathbb{E}\left[\phi_t^T\phi_t|\mathcal{Y},\hat{\theta}_k\right]\right)^{-1}\sum_{t=1}^{N}\mathbb{E}\left[\phi_t^T\delta\bar{x}_t|\mathcal{Y},\hat{\theta}_k\right] \tag{30}$$

where

$$\mathbb{E}\left[\phi_t^T\phi_t\right]=\begin{bmatrix}\mathbb{E}\left[\psi_t^T\psi_t\right] & \mathbb{E}\left[\psi_t^T\right]\upsilon_t\\ \upsilon_t^T\mathbb{E}[\psi_t] & \upsilon_t^T\upsilon_t\end{bmatrix} \tag{31}$$

$$\mathbb{E}\left[\phi_t^T\delta\bar{x}_t\right]=\begin{bmatrix}\mathbb{E}\left[\psi_t^T\delta\bar{x}_t\right]\\ \upsilon_t^T\mathbb{E}[\delta\bar{x}_t]\end{bmatrix}. \tag{32}$$

Proof: Expanding (24) leads to the quadratic expression in θ

$$Q(\theta,\hat{\theta}_k)=\mathbb{E}\left[-\sum_{t=1}^{N}\frac{1}{2}\sigma^{-2}\left(\delta\bar{x}_t^T\delta\bar{x}_t-2\theta^T\phi_t^T\delta\bar{x}_t+\theta^T\phi_t^T\phi_t\theta\right)^T\right]+c \tag{33}$$

Taking the partial derivative of $Q(\theta,\hat{\theta}_k)$ with regard to θ leads to

$$\frac{\partial}{\partial\theta}Q(\theta,\hat{\theta}_k)=\mathbb{E}\left[\sum_{t=1}^{N}\sigma^{-2}\left(\phi_t^T\delta\bar{x}_t-\phi_t^T\phi_t\theta\right)\right] \tag{34}$$

where the matrix $\phi_t^T\phi_t$ is guaranteed positive definite by the presence of the process noise in (16); hence, $\phi_t^T\phi_t$ is invertible. Setting $(\partial)/(\partial\theta)Q(\theta,\hat{\theta}_k)=0$ and solving for θ leads directly to (30). Taking the 2^{nd} partial derivative of the Q function with regard to θ leads to

$$\frac{\partial}{\partial^2\theta}Q(\theta,\hat{\theta}_k)=-\mathbb{E}\left[\sum_{t=1}^{N}\sigma^{-2}\phi_t^T\phi_t\right] \tag{35}$$

which is negative definite, verifying that $\hat{\theta}_{k+1}$ is located at a maximum, completing the proof. ∎

Remark 1: Numerically the estimation of the model parameters $\hat{\theta}_{k+1}$ is affected by the condition number of the matrix $\Phi_k^{(\delta)}=\sum_{t=1}^{N}\mathbb{E}[\phi_t^T\phi_t|\mathcal{Y},\hat{\theta}_k]$ in (30). In the shift-operator version of this algorithm the parameter estimates will be dependent on the analogous term $\Phi_k^{(q)}$; for fast-sampled systems the matrix $\Phi_k^{(q)}$ will typically be ill-conditioned because of the numerical similarity between successive samples, that is $\lim_{T\to 0}\mathrm{cond}(\Phi_k^{(q)})=\infty$, where "cond" indicates condition number. By contrast numerical ill-conditioning is less likely to occur when using the δ-operator to model a fast-sampled system because the δ-transformations of a signal tend to the signal derivatives with increasing sample rate, which are unlikely to be numerically similar [1]. Hence, the estimation of model parameters benefits numerically from being performed in the δ-domain at fast-sample rates as will be demonstrated in Section V.

The maximization of the function $Q(\theta,\hat{\theta}_k)$ requires calculation of the conditional expectations defined in (31) and (32); these are obtained from the application of the following state-space model in the Kalman smoother

$$\mathbf{x}_{t+1}=F_k\mathbf{x}_t+G_k\mathbf{u}_t+TWe_t \tag{36}$$

where $F_k=I+TA_k$, $G_k=TB_k$, the subscript k denotes that the system matrices are dependent on $\hat{\theta}_k$, and the measurement equation is defined in (22). Application of this model in one of the well-known Kalman smoothing algorithms [25], [26] provides the information needed to obtain the smoothed state estimate $\hat{\mathbf{x}}_t$, the smoothed state covariance P_t, and hence the expectations defined in (31) and (32), which are

$$\mathbb{E}[\psi_t|\mathcal{Y},\hat{\theta}_k]=\hat{\mathbf{x}}_t^T \tag{37}$$

$$\mathbb{E}\left[\psi_t^T\psi_t|\mathcal{Y},\hat{\theta}_k\right]=P_t+\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t^T \tag{38}$$

$$\mathbb{E}[\delta\bar{x}_t|\mathcal{Y},\hat{\theta}_k]=\frac{1}{T}\left[\hat{x}_{t+1}^{(n)}-\hat{x}_t^{(n)}\right] \tag{39}$$

$$\mathbb{E}\left[\psi_t^T\delta\bar{x}_t|\mathcal{Y},\hat{\theta}_k\right]=\frac{1}{T}\left[\mathbf{m}_t+\hat{\mathbf{x}}_t\hat{x}_{t+1}^{(n)}-\mathbf{p}_t-\hat{\mathbf{x}}_t\hat{x}_t^{(n)}\right] \tag{40}$$

$$\mathbf{m}_t=M_t(1:n,n) \tag{41}$$

$$\mathbf{p}_t=P_t(1:n,n) \tag{42}$$

where $x_t^{(n)}$ denotes the n^{th} element of the vector $\mathbf{x}_t$, $M_t=\mathrm{cov}(\mathbf{x}_t,\mathbf{x}_{t+1})$, $\mathrm{cov}(,)$ denotes covariance and the notation $A(i:j,k)$ denotes for a matrix A the vector formed from the elements in rows i to j of column k. The covariance matrix M_t is obtained from a nonstandard augmentation of the backward pass (described fully in [27]), where

$$M_t=[P_tJ_{t-1}^T+J_t(M_{t+1}^T-AP_t)J_{t-1}^T]^T \tag{43}$$

for $t=N,\ldots,2$, where J is the backward Kalman gain. The recursion is initialized using the expression

$$M_N=[(I-K_NC)AP_{N-1}]^T \tag{44}$$

where K is the forward Kalman gain.

Algorithm 1: The EM algorithm for estimating the parameters of the δ-domain model defined in (16) and (22) is as follows.

1) Initialize the parameter estimate $\hat{\theta}_0$.
2) Increment counter k and calculate the expectation quantities defined in (31) and (32) using the augmented Kalman smoothing algorithm, with the state-space model (36) parameterized by $\hat{\theta}_k$.
3) Maximize $Q(\theta,\hat{\theta}_k)$ using (30) to obtain the updated parameter estimate $\hat{\theta}_{k+1}$.
4) Test the convergence of the algorithm using an appropriate stopping condition and then terminate if satisfied or go to Step 2 and repeat.

The estimation procedure presented here is by construction a special case in the generic form of the EM algorithm, hence convergence of the iterative procedure outlined above to a local maximum of the likelihood function $L(\theta^*)$ follows from the standard proofs developed in [17] and [23] and more recently in the system identification context in [21]. For completeness, the convergence is proved here.

Theorem 2: The sequence of log-likelihood values generated by Algorithm 1 are guaranteed to be nondecreasing, that is

$$L(\hat{\theta}_{k+1}) \geq L(\hat{\theta}_k) \quad \text{for } k = 1, 2, \ldots \tag{45}$$

and are assured to converge to a stationary point of the likelihood function $L(\theta^*)$.

Proof: The proof draws on the standard format presented in [17] and [23]; from (4) observe that $L(\theta)$ is lower bounded by $F(\theta, \hat{\theta}_k)$ and that

$$F(\theta, \hat{\theta}_k) = Q(\theta, \hat{\theta}_k) - R(\theta, \hat{\theta}_k) \tag{46}$$

where

$$R(\theta, \hat{\theta}_k) = \int_{\mathcal{X}} p(\mathcal{X}|\mathcal{Y}, \theta_k) \log p(\mathcal{X}|\mathcal{Y}, \theta_k) d\mathcal{X}. \tag{47}$$

From (46)

$$F(\theta, \hat{\theta}_{k+1}) - F(\theta, \hat{\theta}_k) = [Q(\theta, \hat{\theta}_{k+1}) - Q(\theta, \hat{\theta}_k)] - [R(\theta, \hat{\theta}_{k+1}) - R(\theta, \hat{\theta}_k)]. \tag{48}$$

By definition of the maximization step in Algorithm 1 $Q(\theta, \hat{\theta}_{k+1}) - Q(\theta, \hat{\theta}_k) \geq 0$; furthermore, the term $R(\theta, \hat{\theta}_{k+1}) - R(\theta, \hat{\theta}_k)$ is known to be the Kullback–Leibler distance between the density functions $p(\mathcal{X}|\mathcal{Y}, \theta)$ and $p(\mathcal{X}|\mathcal{Y}, \hat{\theta})$ [28], which is ≥ 0 with equality only if the two distributions are equal. Hence, $F(\theta, \hat{\theta}_{k+1}) - F(\theta, \hat{\theta}_k) \geq 0$, which implies that $L(\hat{\theta}_{k+1}) - L(\hat{\theta}_k) \geq 0$, completing the first part of the proof.

The convergence of the likelihood function to a stationary point is assured if the conditions of Theorem 2 of [23] are met, namely that $Q(\theta, \hat{\theta})$ is continuous in both arguments, which is the case here. ■

V. Numerical Example

An example identification problem is presented in this section, where the parameter estimation procedure derived above was applied to signals generated from a third-order δ-domain system. The results focus on showing the performance of the method when increasing sampling frequency, and are compared to a least squares approach and an equivalent q-domain approach.

A. Data Generation

The discrete-time test system was generated by mapping the same continuous-time system to the δ-domain at varying sample rates. The continuous-time system was

$$G(s) = \frac{1}{s^3 + 0.5s^2 + 0.4s + 0.1} \tag{49}$$

where s is the Laplace operator. The mapping to the δ-domain resulted in a model with higher input order (due to sampling zeros), which was of the form

$$G(\delta) = \frac{b_2\delta^2 + b_1\delta + b_0}{\delta^3 + a_2\delta^2 + a_1\delta + a_0}. \tag{50}$$

The δ-domain system was identified from the I/O u_t and y_t data generated from simulating the system as described in (16) and (22), and the analogous procedure was carried out for the q-domain case (i.e., data generation and identification were performed in the q-domain). The input was defined as a sum-of-sinusoids signal, with 20 sinusoids evenly distributed over the range [0,0.5) Hz. The process noise was defined to have a signal-to-noise ratio (SNR) of 20 dB relative to the input signal and was varied according to experimental conditions in the case of measurement noise (e.g., from 10 to 60 dB). In order to investigate the performance of the estimation algorithm at varying sampling rates, the system was simulated across a range of sampling frequencies from 5 to 30 Hz. The system was simulated for 150 s at each sampling frequency. The simulations were repeated at each sampling frequency 50 times to give an empirical estimate of the consistency of the algorithm performance.

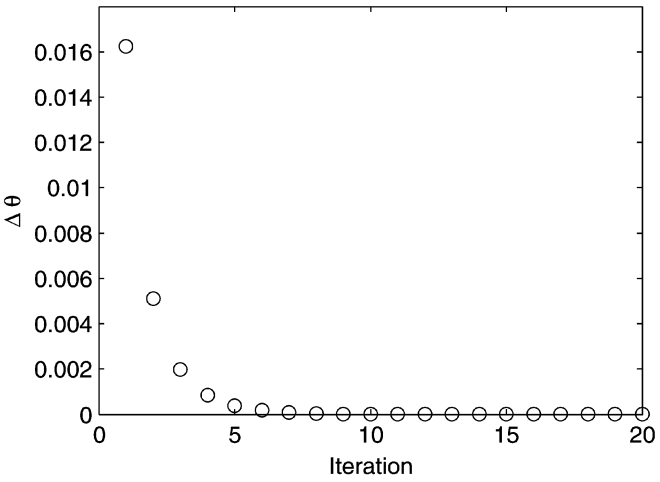

Fig. 1. Convergence of the parameter estimates to a steady value.

B. Parameter Initialization and Convergence

The parameter estimates were initialized using least squares, after low-pass filtering the data with cutoff frequency $\omega_c = 0.5$ Hz (using a third-order Butterworth filter). Smoothing in the EM algorithm was always performed on the raw signals using the updated parameters. Parameter convergence of the EM algorithm was assessed using the metric

$$\Delta\theta_k = \frac{(\theta_k - \theta_{k-1})^T(\theta_k - \theta_{k-1})}{\theta_k^T\theta_k}. \tag{51}$$

The EM algorithm was run in each case for 20 iterations; the convergence of one typical experimental trial is shown in Fig. 1.

C. Modeling Results

The identification results demonstrate that the EM procedure derived here improves the accuracy of parameter estimates in comparison to LS; application of LS results in significant bias for both q and δ as discussed in the introduction, which is shown in Fig. 2. The bias in parameter estimates increases with sampling rate for LS in both q and δ. The results demonstrate that the q-domain EM identification algorithm is also adversely affected by increasing sampling rate, shown in Fig. 2. In contrast, it is apparent that the δ-domain EM approach performs well across all sampling frequencies. One possible explanation for this is that the matrix inverted during parameter estimation grows progressively more ill-conditioned as sample rate increases (as shown for typical data sets in Fig. 3).

The effect of ill-conditioning due to fast sampling on the estimation of q- and δ-domain model parameters is discussed in [1], from a perspective of least-squares, which is similar to the estimation step in the EM algorithm presented here. The q-domain algorithmic problem may

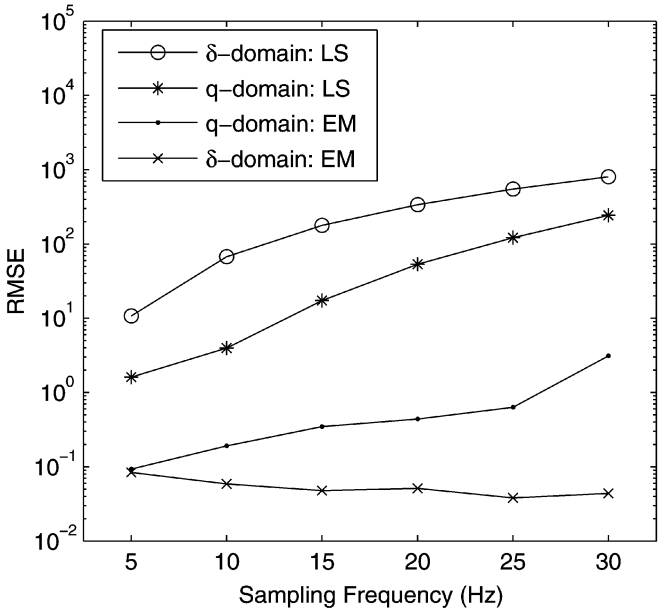

Fig. 2. Averaged RMSE (across all 50 trials) of the estimated δ-domain and q-domain model parameters from the application of both LS and the EM algorithm.

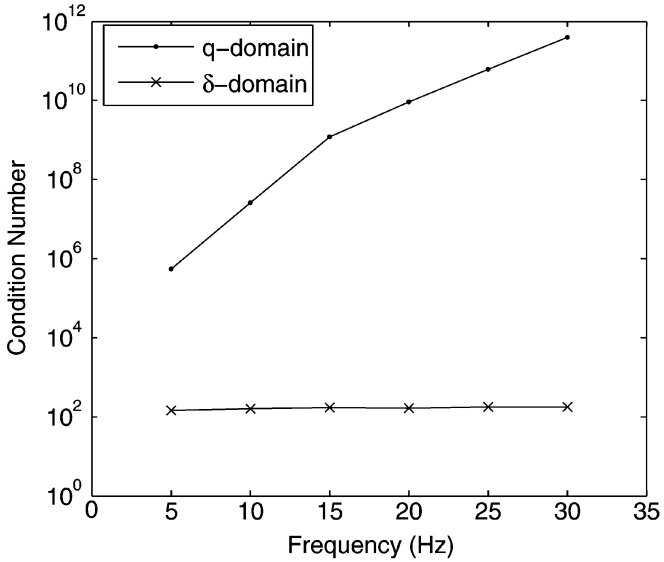

Fig. 3. Condition number of the matrix inverted during parameter estimation for both the q- and δ-domain models at the final EM iteration, for a typical data set at an SNR of 20 dB.

lie in the fact that if signal estimation is performed successfully, then the sequential values of the estimated output signal will be very similar due to fast sampling; this leads to the ill-conditioning problem at the parameter estimation stage, because $y(t) \approx y(t-1) \approx \ldots$ The δ version of the algorithm does not suffer from this ill-conditioning problem, which is also demonstrated in Fig. 3, because the δ transformations are unlikely to be numerically similar.

To illustrate the performance of each EM algorithm and LS, for both q and δ, at a single fast-sampling frequency the simulations were performed fixing sample frequency to 25 Hz and varying the SNR of the measurement noise signal for 10 to 60 dB in steps of 10 dB. The results, shown in Fig. 4, demonstrate that the EM δ-operator algorithm is significantly more accurate than the equivalent q-operator approach,

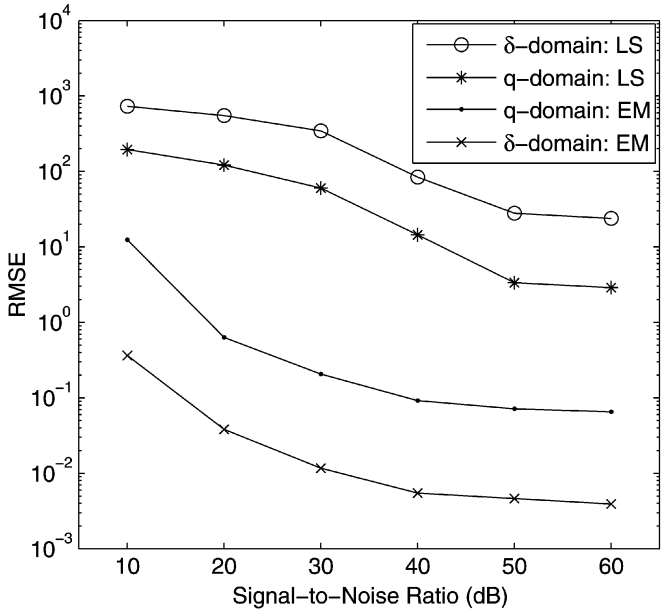

Fig. 4. Averaged RMSE (across all 50 trials) of the estimated δ-domain and q-domain model parameters from the application of the EM algorithm and LS when varying SNR at a fixed sample rate of 25 Hz.

which both improve on LS. As would be expected the accuracy of each estimation method improves as the SNR increases. The results imply that use of the EM δ-operator algorithm can lead to improved parameter estimation compared to the shift operator for the modeling scenario considered here, that is, identification from a fast-sampled output signal corrupted by additive Gaussian white measurement noise.

The complexity of the LS estimate is on the order of $\mathcal{O}(Nr^3)$ where $r = n + m$ is the number of model parameters. By comparison the complexity of the EM approach proposed here is dominated by the parameter estimation step, which is of equivalent order to that of LS; hence, it is of order $\mathcal{O}(Nr^3)$ per iteration. Hence, the EM algorithm is more complex than the LS approach. However, the proposed EM estimation algorithm is an offline technique and the additional complexity may be of little practical importance; convergence for each example data set used here was on a time scale of seconds. This is coupled with a significant improvement in accuracy (in comparison to LS), which is likely to offset the disadvantage of the increase in complexity in a typical modeling scenario.

VI. Conclusion

This contribution has treated the problem of identifying δ-domain model parameters from output signals that are corrupted by measurement noise. The novel approach to identification taken here has been the formulation of an estimation problem where the δ-domain transformations of the true signal are directly estimated and are then used to obtain the model parameters. The estimation problem has been solved using a statistical signal processing approach based on the EM algorithm. Implementation of the derived algorithm leads to maximum-likelihood estimates of the parameters in a procedure that is guaranteed to be convergent to a local maximum. The approach is shown here to lead to significantly more accurate parameter estimates than an equivalent shift operator method when fast sampling.

References

[1] G. C. Goodwin, R. H. Middleton, and H. V. Poor, "High-speed digital signal processing and control," in *Proc. IEEE*, 1992, vol. 80, no. 2, pp. 240–259.

[2] H. Fan, "A structural view of asymptotic convergence speed of adaptive IIR filtering algorithms—Part 1: Infinite precision implementation," *IEEE Trans. Signal Process.*, vol. 41, no. 4, pp. 1493–1517, Apr. 1993.
[3] Q. Li and H. Fan, "On the properties of information matrices of delta-operator based adaptive signal processing algorithms," *IEEE Trans. Signal Process.*, vol. 45, no. 10, pp. 2454–2467, Oct. 1997.
[4] R. H. Middleton and G. C. Goodwin, "Improved finite word length characteristics in digital control using the delta operator," *IEEE Trans. Autom. Control*, vol. 31, no. 11, pp. 1015–1021, Nov. 1986.
[5] R. H. Middleton and G. C. Goodwin, *Digital Control and Estimation: A Unified Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
[6] C. Neuman, "Properties of the delta operator model of dynamic physical systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 1, pp. 296–301, Jan.–Feb. 1993.
[7] A. Feuer and R. Middleton, "Conditioning of LMS algorithm with fast sampling," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1978–1981, Aug. 1995.
[8] P. De and H. Fan, "A delta least squares lattice algorithm for fast sampling," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2396–2406, Sep. 1999.
[9] H. Fan and X. Liu, "Delta Levinson and Schur-type RLS algorithms for adaptive signal processing," *IEEE Trans. Signal Process.*, vol. 42, no. 7, pp. 1629–1639, Jul. 1994.
[10] T. Soderstrom, H. Fan, B. Carlsson, and S. Bigi, "Least squares parameter estimation of continuous-time ARX models from discrete-time data," *IEEE Trans. Autom. Control*, vol. 42, no. 5, pp. 659–673, 1997.
[11] H. Fan, T. Soderstrom, M. Mossberg, B. Carlsson, and Z. Yuanjie, "Estimation of continuous-time AR process parameters from discrete-time data," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1232–1244, May 1999.
[12] E. K. Larsson, M. Mossberg, and T. Soderstrom, "An overview of important practical aspects of continuous-time ARMA system identification," *Circuits, Syst., Signal Process.*, vol. 25, no. 1, pp. 17–46, 2006.
[13] S. R. Anderson and V. Kadirkamanathan, "Modelling and identification of non-linear deterministic systems in the delta-domain," *Automatica*, vol. 43, no. 11, pp. 1859–1868, 2007.
[14] L. Ljung, "Initialisation aspects for subspace and output error identification methods," presented at the 5th Eur. Control Conf., Cambridge, U.K., 2003.
[15] E. Karlsson, T. Soderstrom, and P. Stoica, "The Cramer-Rao bound for noisy input-output systems," *Signal Process.*, vol. 80, pp. 2421–2447, 2000.
[16] H. Fan, T. Soderstrom, and Y. Zhou, "Continuous-time AR process parameter estimation in presence of additive white noise," *IEEE Trans. Signal Process.*, vol. 47, no. 12, pp. 3392–3398, Dec. 1999.
[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Roy. Stat. Soc. B*, vol. 39, no. 1, pp. 1–38, 1977.
[18] C. H. Aldana, E. d. Cardevalho, and J. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the EM algorithm," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3280–3292, Dec. 2003.
[19] R. A. Iltis and S. Kim, "Geometric derivation of EM and generalized successive interference cancellation algorithms with applications to CDMA channel estimation," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1367–1377, May 2003.
[20] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Anal.*, vol. 3, no. 4, pp. 253–264, 1982.
[21] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667–1682, 2005.
[22] M. Dewar and V. Kadirkamanathan, "A canonical space-time state space model: state and parameter estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4862–4870, Oct. 2007.
[23] C. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, 1983.
[24] R. A. Boyles, "On the convergence of the EM algorithm," *J. Roy. Stat. Soc. B*, vol. 45, no. 1, pp. 47–50, 1983.
[25] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *Amer. Inst. Aeronaut. Astronaut. J.*, vol. 3, pp. 1445–1450, 1965.
[26] D. C. Fraser and J. E. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Trans. Autom. Control*, vol. 14, no. 4, pp. 387–390, 1969.
[27] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and its Applications*. Berlin, Germany: Springer, 2000.
[28] C. R. Rao, *Linear Statistical Inference and its Application*. New York: Wiley, 1965.

An Efficient Structure for the Design of Variable Fractional Delay Filters Based on the Windowing Method

J. Selva

***Abstract*—A variable fractional delay (VFD) filter is presented that constitutes an efficient alternative to the usual Farrow structure. The filter is constructed by multiplying the signal to interpolate by a window with proper truncation properties, and then applying the Sampling Theorem. The resulting interpolator involves less arithmetic operations than the existing variants of the Farrow structure. Besides, if its coefficients are numerically optimized, its performance is close to that of the FIR equi-ripple filter. The final part of the correspondence includes a description of the VFD filter implementation corresponding to this interpolator.**

***Index Terms*—Bandlimited signals, Farrow structure, interpolation, Lagrange polynomial, Lagrange-type variable fractional delay (VFD) filter, VFD digital filter.**

I. Introduction

The design of variable fractional delay (VFD) filters is a signal processing field in which the requirements imposed on a given interpolator are very stringent [1], [2]. Besides delivering accurate approximations to a given signal from its own samples, the interpolator must only employ arithmetic operations and be parallelizable. Currently, the design of VFD filters is dominated by the so-called Farrow structure in [3]. This structure can be regarded as an extension of an FIR filter that introduces a fixed delay. In short, since the coefficients of this kind of FIR filter are smooth functions of the fractional delay, Farrow's idea in [3] was to approximate them using polynomials, but switching the order in which the filtering and polynomial evaluation operations are carried out. The resulting interpolator (Farrow structure) has been optimized in various ways during the last two decades [4]–[7].

Though efficient, the Farrow structure implicitly requires the evaluation of one polynomial for each sample involved in the interpolation. The purpose of this correspondence is to present a new VFD filter structure that only requires one polynomial evaluation, plus a number of arithmetic operations proportional to the filter length. Its derivation is based on the classical windowing method which has been successfully used in other interpolation problems [8]–[10]. The interpolation method in this paper is different to the "generalized interpolation" approach in [11] and [12]. This last approach is based on prefiltering the

Manuscript received June 11, 2007; revised January 27, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chong-Meng Samson See. This work has been supported by the Spanish Ministry of Education and Science (MEC), Generalitat Valenciana (GV), and by the University of Alicante (UA) under the following projects/programmes: TEC2005-06863-C02-02, HA2007-075 and "Ramón y Cajal" (MEC); ACOMP07-087 and GV07/214 (GV); and GRE074P (UA).

The author is with the Department of Physics, Systems Engineering and Signal Theory (DFISTS), University of Alicante, E-03080 Alicante, Spain (e-mail: jesus.selva@ua.es).

Digital Object Identifier 10.1109/TSP.2008.920148