A numerical study of time and frequency domain maximum likelihood estimation

Ramón A. Delgado[1] Juan I. Yuz[1] Juan C. Agüero[2]
Graham C. Goodwin[2]

[1]Department of Electronic Engineering
Universidad Técnica Federico Santa María, Valparaíso, Chile

[2]ARC Centre for Complex Dynamics Systems and Control
School of Electrical Engineering and Computer Science,
The University of Newcastle, Australia

SYSID 2009

Motivation

In system identification, the likelihood function is defined as the conditional probability density function (PDF) of the data given the parameters, i.e.

$$l(\beta) = p(Y|\beta) \tag{1}$$

where

- Y represent the data, that can be in the time or in the frequency domain, and
- β represent the parameters to be estimated.

Introduction

In the time domain we consider the following SISO linear system model, that can be expresed equivalently in this two forms:

$$y_t = G(q)u_t + H(q)w_t \quad (2)$$

$$\begin{aligned} x_{t+1} &= A\,x_t + B\,u_t + K\,w_t & (3) \\ y_t &= C\,x_t + D\,u_t + w_t & (4) \end{aligned}$$

In (3) different assumptions can be made regarding the initial state x_0:

- (T1) x_0 is assumed to be zero, or
- (T2) x_0 is assumed to be a deterministic parameter to be estimated, or
- (T3) x_0 is assumed to be a random vector.

β contains θ, σ_w^2 and x_0 (or its mean and covariance).

Introduction

We can translate the problem to the frequency domain using the DFT, that is given by

$$Y_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y_t z_k^{-t}$$

Then from the state space model (3)-(4) we obtain:

$$\begin{aligned} z_k X_k - \frac{z_k}{\sqrt{N}}\alpha &= A\, X_k + B\, U_k + K\, W_k \quad (5) \\ Y_k &= C\, X_k + D\, U_k + W_k \quad (6) \end{aligned}$$

where $\alpha = x_0 - x_N$, and different assumptions can be made about this term:

- (F1) α is assumed to be zero (i.e. periodicity in the state),
- (F2) α is estimated as a deterministic parameter, or,
- (F3) α is assumed to be a random variable.

β contains θ, σ_w^2 and α (or its mean and covariance).

Objective

Different likelihood functions are obtained depending on the assumptions made about x_0 and α.

Our interest is on simulation studies which compare the cases described above.

Time Domain Maximum Likelihood

For the sake of simplicity, we represent the system response using block matrices. The system response can be rewritten as

$$\vec{y} = \Gamma x_0 + \Lambda \vec{u} + \Omega \vec{w} \tag{7}$$

where

$$\vec{y} = [y_0, \ldots, y_{N-1}]^T \tag{8}$$

$$\vec{u} = [u_0, \ldots, u_{N-1}]^T \tag{9}$$

$$\vec{w} = [w_0, \ldots, w_{N-1}]^T \tag{10}$$

Time Domain Maximum Likelihood

and

$$\Gamma = \begin{bmatrix} C & CA & \cdots & CA^{N-1} \end{bmatrix}^T \tag{11}$$

$$\Lambda = \begin{bmatrix} D & 0 & \dots & 0 \\ CB & D & \dots & 0 \\ \vdots & & \ddots & \vdots \\ CA^{N-2}B & CA^{N-3}B & \dots & D \end{bmatrix} \tag{12}$$

$$\Omega = \begin{bmatrix} I & 0 & \dots & 0 \\ CK & I & \dots & 0 \\ \vdots & & \ddots & \vdots \\ CA^{N-2}K & CA^{N-3}K & \dots & I \end{bmatrix} \tag{13}$$

Time Domain Maximum Likelihood

Initial state as random variable

Assumption (1)

The initial state x_0 and the noise vector $\vec{w}$ are independent and Gaussian distributed. They are thus uncorrelated and jointly Gaussian distributed:

$$\begin{bmatrix} x_0 \\ \vec{w} \end{bmatrix} \backsim N_r \left(\begin{bmatrix} \mu_{x_0} \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{x_0} & 0 \\ 0 & \sigma_w^2 I \end{bmatrix} \right) \tag{14}$$

Time Domain Maximum Likelihood

Likelihood function when the initial state is considered as a random variable

Lemma (T3)

Consider the system (2) and the Assumption 1. Then the time domain (negative log-)likelihood function is given by

$$L_{T3}(\theta, \sigma_w^2, \mu_{x_0}, \Sigma_{x_0}) = -\log p_{\vec{y}}(\vec{y}|\theta, \sigma_w^2, \mu_{x_0}, \Sigma_{x_0})$$

$$= \frac{1}{2}\left[N \log(2\pi) + \log\det \Sigma_{\vec{y}} + (\vec{y} - \mu_{\vec{y}})^T \Sigma_{\vec{y}}^{-1} (\vec{y} - \mu_{\vec{y}})\right] \qquad (15)$$

where $\mu_{\vec{y}}$ *and* $\Sigma_{\vec{y}}$ *are the conditional mean and covariance matrix for the output data given the parameters, i.e.*

$$\mu_{\vec{y}} = \Lambda \vec{u} + \Gamma \mu_{x_0} \qquad (16)$$

$$\Sigma_{\vec{y}} = \Gamma \Sigma_{x_0} \Gamma^T + \sigma_w^2 \Omega \Omega^T \qquad (17)$$

Time Domain Maximum Likelihood

The Initial state is considered a deterministic parameter to be estimated.

Corollary (T2)

If the initial state x_0 is considered as a deterministic parameter to be estimated, then the corresponding likelihood function is given by

$$L_{T2}(\theta, \sigma_w^2, x_0) = \frac{N}{2}\left(\log(2\pi) + \log\sigma_w^2 + \frac{1}{N\sigma_w^2}\sum_{t=0}^{N-1}\epsilon_t^2\right) \tag{18}$$

where ϵ_t is the prediction error given by

$$\epsilon_t = \frac{y_t - G(q)\,u_t - F(q)\,s_t}{H(q)} \tag{19}$$

Time Domain Maximum Likelihood

The initial state is assumed to be zero.

Corollary (T1)

If the initial state x_0 is assumed to be zero, then the corresponding likelihood function is given by

$$L_{T1}(\theta, \sigma_w^2) = \frac{N}{2}\left(\log(2\pi) + \log\sigma_w^2 + \frac{1}{N\sigma_w^2}\sum_{t=0}^{N-1}\epsilon_t^2\right) \qquad (20)$$

where ϵ_t is the prediction error given by

$$\epsilon_t = \frac{y_t - G(q)u_t}{H(q)} \qquad (21)$$

Frequency Domain Maximum Likelihood

We can write the output DFT sequence $\{Y_k\}$ in vector form as:

$$\vec{Y} = F_D \alpha + G_D \vec{U} + H_D \vec{W} \tag{22}$$

where $\vec{Y}$, $\vec{U}$, and $\vec{W}$ are the DFT's corresponding to $\vec{y}$, $\vec{u}$, and $\vec{w}$, (for example $\vec{Y} = M_F\, \vec{y}$, where M_F is the Fourier matrix.

Frequency Domain Maximum Likelihood

and

$$F_D = \begin{bmatrix} F_0 & \cdots & F_{N-1} \end{bmatrix}^T \tag{23}$$

$$G_D = \begin{bmatrix} G_0 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & G_{N-1} \end{bmatrix} \tag{24}$$

$$H_D = \begin{bmatrix} H_0 & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & H_{N-1} \end{bmatrix} \tag{25}$$

Frequency Domain Maximum Likelihood

α is considered as a random variable

Assumption (2)

The term α and the noise vector $\vec{w}_t$ are jointly Gaussian distributed having mean $(\mu_\alpha^T, 0)^T$, and joint covariance matrix

$$\Sigma_{\left[\begin{smallmatrix}\alpha\\ w\end{smallmatrix}\right]} = E\left\{\begin{bmatrix}\alpha-\mu_\alpha\\ \vec{w}\end{bmatrix}\begin{bmatrix}\alpha-\mu_\alpha\\ \vec{w}\end{bmatrix}^T\right\} = \begin{bmatrix}\Sigma_\alpha & \Sigma_{\alpha\vec{w}}\\ \Sigma_{\alpha\vec{w}}^T & \sigma_w^2 I_N\end{bmatrix} \quad (26)$$

Equivalence between ML in time and frequency domain [1]

$$\alpha = (I - A^N)x_0 - \sum_{t=0}^{N-1} A^{N-1-t}(Bu_t + Kw_t) \quad (27)$$

Frequency Domain Maximum Likelihood

Lemma (F3)

Consider the frequency domain representation of the linear system (2), given in (5)-(6) and the Assumption 2. Then the frequency domain (negative log-) likelihood function, is given by

$$
\begin{aligned}
L_{F3}(\theta, \sigma_w^2, \mu_\alpha, \Sigma_{\left[\begin{smallmatrix}\alpha\\ w\end{smallmatrix}\right]}) &= -\log p_{\vec{Y}_R}(\vec{Y}_R | \theta, \sigma_w^2, \mu_\alpha, \Sigma_{\left[\begin{smallmatrix}\alpha\\ w\end{smallmatrix}\right]}) \\
&= L_0 + \log\det \Sigma_{\vec{Y}_R} + (\vec{Y}_R - \mu_{\vec{Y}_R})^T \Sigma_{\vec{Y}_R}^{-1} (\vec{Y}_R - \mu_{\vec{Y}_R}) \qquad (28)
\end{aligned}
$$

where the term L_0 accounts for unimportant constants and where

$$
\begin{aligned}
\mu_{\vec{Y}_R} &= M_T G_D \vec{U} + M_T F_D \mu_\alpha \\
\Sigma_{\vec{Y}_R} &= M_T \begin{bmatrix} F_D & H_D M_F \end{bmatrix} \begin{bmatrix} \Sigma_\alpha & \Sigma_{\alpha\vec{w}} \\ \Sigma_{\alpha\vec{w}}^T & \sigma_w^2 I_N \end{bmatrix} \begin{bmatrix} F_D^H \\ M_F^H H_D^H \end{bmatrix} M_T^H
\end{aligned}
$$

F_D, G_D, and H_D are defined in (23)-(25), and ^{H} denotes conjugate-transpose.

Frequency Domain Maximum Likelihood

Corollary (F2)

If the term α is considered as a deterministic parameter to be estimated, then the corresponding (negative log-)likelihood function is given by

$$\begin{aligned} L_{F2}(\theta,\sigma_w^2,\alpha) &= -\log p_{\vec{Y}_R}(\vec{Y}_R|\theta,\sigma_w^2,\alpha) \\ &= L_0 + N\log\sigma_w^2 + \sum_{k=0}^{N-1}\left[\log(|H_k|^2) + \frac{1}{\sigma_w^2}|E_k|^2\right] \end{aligned} \tag{29}$$

where L_0 accounts for unimportant constants, and E_k is given by

$$E_k = \frac{Y_k - G_k - F_k\alpha}{H_k} \tag{30}$$

Frequency Domain Maximum Likelihood

Corollary (F1)

If the term α is assumed to be zero, then the corresponding (negative log-) likelihood function is given by

$$\begin{aligned} L_{F1}(\theta, \sigma_w^2) &= -\log p_{\vec{Y}_R}(\vec{Y}_R|\theta, \sigma_w^2) \\ &= L_0 + N \log \sigma_w^2 + \sum_{k=0}^{N-1} \left[\log(|H_k|^2) + \frac{1}{\sigma_w^2} |E_k|^2 \right] \end{aligned} \tag{31}$$

where L_0 accounts for unimportant constants, and E_k is given by

$$E_k = \frac{Y_k - G_k}{H_k} \tag{32}$$

Numerical examples

For the examples we consider a simple model in state-space form:

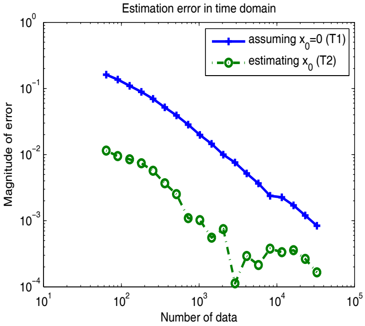
$$x_{t+1} = a\,x_t + b\,u_t + k\,w_t \quad (33)$$

$$y_t = x_t + w_t \quad (34)$$

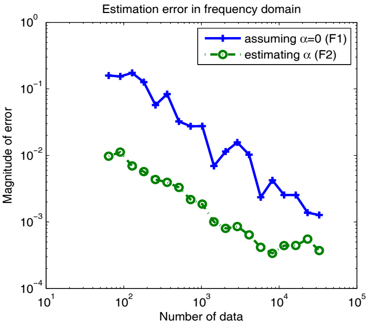
- u_t: zero mean Gaussian with $\sigma_u^2 = 1$.
- w_t: zero mean Gaussian with $\sigma_w^2 = 0.01$.
- $x_0 = 3$.
- $a = 0.75$, $b = 0.5$, and $k = 1$.

The system was simulated over N data points, from $N = 2^6$ to 2^{15}. For each data lenght N we performed $1,000$ Monte-Carlo simulations using different seeds of noise. We optimize the likelihood function using `fminunc` and initialize the routine with `n4sid`.

Numerical examples



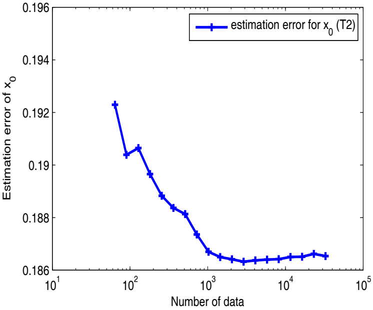
(a) Estimation error $||\theta_0 - \hat{\theta}||$ for (T1) and (T2)



(b) Estimation error $||\theta_0 - \hat{\theta}||$ for (F1) and (F2)

Figure: Time and frequency domain estimation using a random input.

Numerical examples



(a) Error in estimating the initial condition, x_0

Figure: Estimation error of x_0 in the time domain

Numerical examples

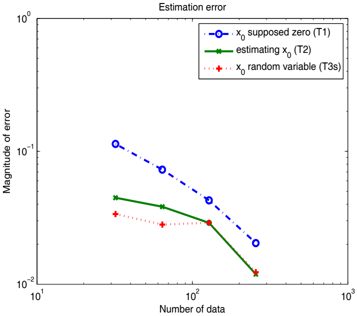
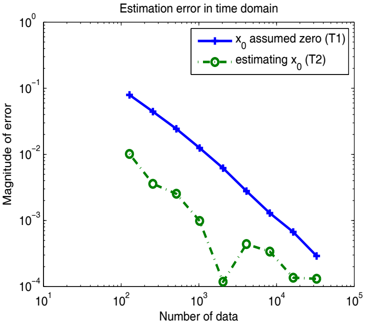
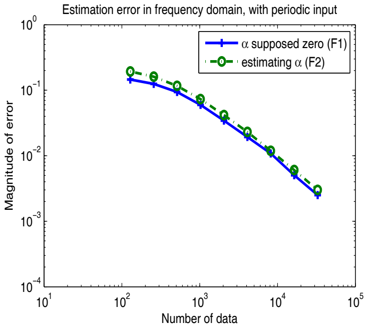


Figure: Estimation error, comparison of cases (T3s), (T2) and (T1).

Numerical examples



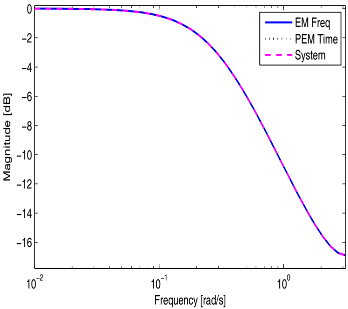
(a) Error in estimating the parameter vector in time domain, with periodic input



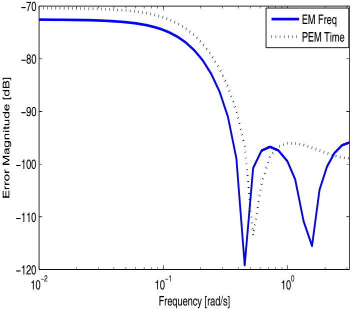
(b) Error in estimating the parameter vector in frequency domain, with periodic input

Figure: Time and frequency domain estimation using a periodic input

Numerical examples
Recent Work



(a) Bode diagram comparing the estimated models with the true system



(b) Bode diagram comparing the estimation error

Figure: Estimation for state space models using the EM algorithm for the case (F3) when α is considered as a random variable.

Conclusions

- The results show that the assumptions made about x_0 and α play a key role in the estimation results.
- For short data set, the impact of these assumptions is more important. For long data set all the different approaches give similar results.
- The choice of the input plays an important role on the validity of the assumptions on x_0 and α.
- Time and Frequency domain maximum likelihood estimation are equivalent when consistent assumptions are made in both domains.

Thanks for your attention.
Any questions?

References

J.C. Agüero, J.I. Yuz, G.C. Goodwin, and R.A. Delgado.
On the equivalence of time and frequency domain maximum likelihood estimation.
Provisionally accepted in Automatica, 2009.

Assumption for (T3s)

To implement (T3s), we consider the stationary response of the system when the input is a random sequence.

$$\mu_{x_0} = \frac{b}{1-a}\mu_u \tag{35}$$

$$\Sigma_{x_0} = \sigma_{x_0}^2 = \frac{\sigma_u^2\, b^2 + \sigma_\omega^2\, k^2}{1-a^2} \tag{36}$$

where σ_u^2 is the variance of the input signal, σ_ω^2 is the variance of the noise, and μ_u is the mean of the input signal.