# On the Equivalence of Time and Frequency Domain Maximum Likelihood Estimation [⋆]

Juan C. Agüero [a] , Juan I. Yuz [b] , Graham C. Goodwin [a] , and Ramón A. Delgado [b]

[a] *ARC Centre for Complex Dynamics Systems and Control (CDSC)*
*School of Electrical Engineering and Computer Science, The University of Newcastle, Australia*

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

**Abstract**

Maximum likelihood estimation has a rich history. It has been successfully applied to many problems including dynamical system identification. Different approaches have been proposed in the time and in the frequency domains. In this paper we discuss the relationship between these approaches and we establish conditions under which the different formulations are equivalent for finite length data. A key point in this context is how initial (and final) conditions are considered and how they are introduced in the likelihood function.

*Key words:* Maximum likelihood estimation, frequency domain identification, system identification.

## 1 Introduction

Maximum Likelihood (ML) estimation methods have become a popular approach to dynamic system identification [10,40,18]. Different approaches have been proposed in the time and frequency domains [19,17,21,33,34]. A commonly occurring question is how time- and frequency-domain versions of ML estimation are related. Some insights into the relationship between the methods have been given in past literature. However, to the best of the authors knowledge, there has not previously been a comprehensive account of the equivalence between the two approaches, in particular, for finite length data. For example, [17,21,23] have shown that, when working in the frequency domain, an extra term arises in the likelihood function that depends on the noise model. This term vanishes *asymptotically* for long data sets when considering uniformly spaced frequency points over the full bandwidth $[-\pi, \pi]$ (see, for example, [34]).

In [34], Box-Jenkins identification has been analyzed. Extensions to identification in closed loop have also been presented [34]. Also the case of reduced bandwidth estimation has been considered. A surprising result, in this context, is that, for processes operating in open loop, the commonly used frequency domain ML method requires exact knowledge of the noise model in order to obtain consistent estimates for the plant parameters. On the other hand, it is well known that the commonly used ML in the time domain (for systems driven by a quasi-stationary input and Gaussian white noise that are mutually uncorrelated) provides consistent estimates for the transfer function from input to output irrespective of possible under-modelling of the transfer function from noise to output [18]. This fact suggests that there could be key differences between the time- and frequency-domain approaches in the usual formats. In the current paper we will see that the apparent differences are a result of inconsistent formulations rather than fundamental issues between the use of time or frequency domain data. In particular, we establish in this paper that the domain chosen to describe the available data (i.e., time or frequency) does not change the result of the estimation problem (see also [39,19]). Instead, it is the choice of the likelihood function, i.e., which parameters are to be estimated and what data is assumed available, that leads to perceived differences in the estimation problems. This issue has previously been highlighted for

[⋆] This paper was not presented at any IFAC meeting. Corresponding author: Tel. +61-2-49216351, Fax +61-2-49601712. Work supported by CDSC and FONDECYT-Chile through grant 11070158.

*Email addresses:* `juan.aguero@newcastle.edu.au` (Juan C. Agüero), `juan.yuz@usm.cl` (Juan I. Yuz), `graham.goodwin@newcastle.edu.au` (Graham C. Goodwin), `ramon.delgado@alumnos.usm.cl` (Ramón A. Delgado).

time domain methods in the statistics literature where, for example, the way in which initial conditions are considered defines different likelihood functions and, thus, different estimation problems (see e.g. [36, chapter 22]). More specifically, for dynamic system identification, the time domain likelihood function is different depending on the assumptions made regarding the initial state ($x_0$), e.g.

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

On the other hand, if we convert the data to the frequency domain by applying the discrete Fourier transform (DFT) then a term arises which depends on the difference between the initial and final states, $\alpha = x_0 - x_N$. Different assumptions can be made about this term in the frequency domain, e.g.

(**F1**) $\alpha$ is assumed to be zero (equivalent to assuming periodicity in the state)
(**F2**) $\alpha$ is estimated as a deterministic parameter (as in, e.g., [1,34]), or
(**F3**) $\alpha$ is considered as a *hidden* random variable.

In this paper we show that the case when the term $\alpha$ is considered as a random variable is the most general. In fact, we show that each of the six cases described above, i.e., (**T1**)–(**T3**) and (**F1**)–(**F3**), can be obtained by making particular assumptions regarding the statistical properties of the random variable $\alpha$ and $x_0$. In particular, our analysis shows that the same solution is obtained (in the time and in the frequency domain, and using finite data) only if the properties of $\alpha$ as a random variable are chosen such that they are consistent with the system dynamics and the way the initial state is considered (Theorem 21 in Section 4.2). Throughout the paper we assume that the system is operating in open loop. Closed loop data can be treated in a similar fashion by the inclusion of additional terms (see for example Remark 2 below).

## 2 Time Domain Maximum Likelihood

### 2.1 Time-domain model and data

We consider the following Single-Input Single-Output (SISO) linear system model:

$$y_t = G(q)u_t + H(q)w_t \tag{1}$$

where $\{u_t\}$ and $\{y_t\}$ are the (sampled time-domain) input and output signals, respectively, and $\{w_t\}$ is zero mean Gaussian noise with variance $\sigma_w^2$. $G(q,\theta)$ and $H(q,\theta)$ are rational functions in the forward shift operator $q$. We also assume that: *(i)* $G(q)$ and $H(q)$ are stable, with no poles on the unit circle; *(ii)* $H^{-1}(q)$ is stable (i.e., $H(q)$ is minimum phase, with no zeros on the unit circle); and *(iii)* $\lim_{q\to\infty} H(q) = 1$. The transfer function description of the system in (1) can equivalently be represented in state space form as:

$$x_{t+1} = A\,x_t + B\,u_t + K\,w_t \tag{2}$$
$$y_t = C\,x_t + D\,u_t + w_t \tag{3}$$

The system parameter vector $\theta$ contains the coefficients of the transfer functions $G(q)$ and $H(q)$ in (1). This parameter vector also uniquely defines the matrices $A, B, C, D, K$ in the state space representation in (2)–(3) for controllable or observable canonical forms [16,22,5]. The two alternative models are related by

$$G(q) = C(qI - A)^{-1}B + D \tag{4}$$
$$H(q) = C(qI - A)^{-1}K + 1 \tag{5}$$

The initial conditions in (2) summarize the past of the system prior to time $t = 0$. To include the effect of initial conditions on the system response, we note that the solution of the state-space model (2)–(3) can be written as

$$\begin{aligned} y_t &= F(q)s_t + G(q)u_t + H(q)w_t \qquad (6)\\ &= CA^t\,x_0 + \left[\sum_{\ell=0}^{t-1} CA^{t-1-\ell}B\,u_\ell\right] + Du_t \\ &\quad + \left[\sum_{\ell=0}^{t-1} CA^{t-1-\ell}K\,w_\ell\right] + w_t \qquad (7)\end{aligned}$$

where the additional term $s_t$ captures the effect of an initial state $x_0$ on the system response. If we interpret $s_t$ as a Kronecker delta function, i.e., $s_t = x_0\delta_K[t]$, then the transfer functions in (6) are given by (4), (5), and

$$F(q) = C(qI - A)^{-1}q \tag{8}$$

For the sake of simplicity, we represent the system response using block matrices. Equation (7) can then be rewritten as

$$\vec{y} = \Gamma x_0 + \Lambda\vec{u} + \Omega\vec{w} \tag{9}$$

where

$$\vec{y} = [y_0, \ldots, y_{N-1}]^T \tag{10}$$
$$\vec{u} = [u_0, \ldots, u_{N-1}]^T \tag{11}$$
$$\vec{w} = [w_0, \ldots, w_{N-1}]^T \tag{12}$$

and

$$\Gamma = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{bmatrix}, \Lambda = \begin{bmatrix} D & 0 & \dots & 0 \\ CB & D & \dots & 0 \\ \vdots & & \ddots & \vdots \\ CA^{N-2}B & CA^{N-3}B & \dots & D \end{bmatrix} \tag{13}$$

$$\Omega = \begin{bmatrix} I & 0 & \dots & 0 \\ CK & I & \dots & 0 \\ \vdots & & \ddots & \vdots \\ CA^{N-2}K & CA^{N-3}K & \dots & I \end{bmatrix} \tag{14}$$

### *2.2 Time domain maximum likelihood*

The time-domain likelihood function is defined as the conditional probability density function (PDF) of the data given the parameters, i.e.

$$\ell(\beta) = p_{\vec{y}}(y_0, \dots, y_{N-1}|\beta) = p_{\vec{y}}(\vec{y}|\beta) \tag{15}$$

where $\mathcal{D} = \{y_0, \dots, y_{N-1}\}$ is the available data given in the time domain, and the vector $\beta$ contains the parameters to be estimated: the system parameters in $\theta$, the noise covariance $\sigma_w^2$, and the initial state $x_0$ (when considered as a deterministic parameter) or its mean $\mu_{x_0}$ and covariance matrix $\Sigma_0$ (when $x_0$ is considered as a random variable). Note that, as stated in the introduction, different likelihood functions will arise from (15), depending on the assumptions made about the initial state $x_0$.

We examine the three time-domain problems corresponding to assumptions (**T1**)–(**T3**) described in the introduction. The next lemma considers the time-domain likelihood function when the initial state $x_0$ is considered as a random variable (i.e., case (**T3**)).

**Lemma 1 (T3)** *Consider the system (1) (or its equivalent form as in (2)–(3)). Assume that the initial state $x_0$ is a random vector, Gaussian distributed, independent of the noise process $\{w_t\}$, with mean $\mu_{x_0}$ and covariance matrix $\Sigma_0$. Then the time domain (negative log-) likelihood function is given by*

$$\begin{aligned} L_{T3}(\theta, \sigma_w^2, \mu_{x_0}, \Sigma_0) &= -\log p_{\vec{y}}(\vec{y}|\theta, \sigma_w^2, \mu_{x_0}, \Sigma_0) \\ &= \frac{1}{2}\left[N\log(2\pi) + \log\det\Sigma_{\vec{y}} + (\vec{y}-\mu_{\vec{y}})^T\Sigma_{\vec{y}}^{-1}(\vec{y}-\mu_{\vec{y}})\right] \end{aligned} \tag{16}$$

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

$$\mu_{\vec{y}} = \Lambda\vec{u} + \Gamma\mu_{x_0} \tag{17}$$

$$\Sigma_{\vec{y}} = \Gamma\Sigma_0\Gamma^T + \sigma_w^2\Omega\Omega^T \tag{18}$$

**PROOF.** The likelihood of the data given the parameter vector $\theta$ and the random vector $x_0$ can be obtained from (9):

$$\vec{y} = \Lambda\vec{u} + \begin{bmatrix}\Gamma & \Omega\end{bmatrix}\begin{bmatrix} x_0 \\ \vec{w}\end{bmatrix} \tag{19}$$

Using the PDF of a transformation of random variables (see, for example, [15, page 34]), we have that the PDF for $\vec{y}$ is readily obtained from the affine transformation in (19), i.e.

$$p_{\vec{y}}(\vec{y}) = \frac{\exp\left\{-\frac{1}{2}(\vec{y}-\mu_{\vec{y}})^T\Sigma_{\vec{y}}^{-1}(\vec{y}-\mu_{\vec{y}})\right\}}{\sqrt{(2\pi)^N\det\Sigma_{\vec{y}}}} \tag{20}$$

where the mean and covariance are given by equations (17) and (18). The negative log-likelihood function is readily obtained from (20). □

**Remark 2** *The previous lemma describes the likelihood function when the initial condition is considered as a random vector. Note that (16) includes the system parameter vector $\theta$ and also the mean $\mu_{x_0}$ and covariance matrix $\Sigma_0$ of the initial state $x_0$. These quantities, $\mu_{x_0}$ and $\Sigma_0$, can also be considered as* given, *i.e., introducing prior knowledge to the estimation problem. Alterntively, the properties of $x_0$ can also be expressed in terms of the system parameters: for example, if we assume that the input $\{u_t\}$ and noise $\{w_t\}$ are i.i.d., mutually uncorrelated, zero mean random processes having variances $\sigma_u^2$ and $\sigma_w^2$, respectively*[1]*, and acting on the system from $t = -\infty$, then $\mu_{x_0} = 0$ and the covariance $\Sigma_0$ is the solution of the Lyapunov equation:*

$$\Sigma_0 = A\Sigma_0A^T + \sigma_u^2BB^T + \sigma_w^2KK^T \tag{21}$$

*Notice that, for closed loop data, extra terms appear on the right hand side of* (21)*. In fact, for systems operating in closed loop, $\Sigma_0$ depends on the cross-covariance between $u_t$ and $w_t$, the cross-covariance between $x_t$ and $w_t$, and the cross-covariance between $u_t$ and $x_t$. These quantities depend on the structure of the controller, and are typically not estimated in closed loop identification in the prediction error framework (see [18]).* ▽▽▽

[1] We assume that the second order properties of the input are known for $t < 0$. For sake of simplicity, we constrain the analysis to the case when the input is white noise. However, the extension to inputs having rational spectrum is straightforward.

The following two corollaries cover the cases when the initial state is considered as a deterministic parameter (**T2**), and when the initial state is assumed to be equal to zero (**T1**).

**Corollary 3 (T2)** *If the initial state $x_0$ is considered as a deterministic parameter to be estimated, then the corresponding (negative log-) likelihood function is given by*

$$L_{T2}(\theta, \sigma_w^2, x_0) = \frac{N}{2}\log(2\pi) + \frac{N}{2}\log\sigma_w^2 + \frac{1}{2\sigma_w^2}\sum_{t=0}^{N-1}\epsilon_t^2 \tag{22}$$

*where $s_t = \delta_K(t)x_o$ and $\epsilon_t$ is the* prediction error

$$\epsilon_t = \frac{y_t - G(q)u_t - F(q)s_t}{H(q)} \tag{23}$$

**PROOF.** The proof follows from Lemma 1 considering an initial state with mean $\mu_{x_0} = x_0$ and a zero covariance matrix, i.e., $\Sigma_0 = 0$. Substituting into (17)–(18), we have

$$\mu_{\vec{y}} = \Lambda\vec{u} + \Gamma x_0 \tag{24}$$

$$\Sigma_{\vec{y}} = \sigma_w^2\Omega\Omega^T \tag{25}$$

Moreover, from (14), we have that $\Omega^{-1}$ is related to the inverse of the noise transfer function, $H^{-1}(q)$, thus:

$$\Omega^{-1}(\vec{y} - \mu_{\vec{y}}) = \Omega^{-1}(\vec{y} - \Gamma x_0 - \Lambda\vec{u}) = [\epsilon_0, \ldots, \epsilon_{N-1}]^T \tag{26}$$

where $\epsilon_t$ are the prediction errors defined in (23). Substituting (25) and (26) in (16), the likelihood function (22) is obtained. □

**Corollary 4 (T1)** *If the initial state $x_0$ is assumed to be equal to zero, then the corresponding likelihood function is given by*

$$L_{T1}(\theta, \sigma_w^2) = \frac{N}{2}\log(2\pi) + \frac{N}{2}\log\sigma_w^2 + \frac{1}{2\sigma_w^2}\sum_{t=0}^{N-1}\epsilon_t^2 \tag{27}$$

*where $\epsilon_t = (y_t - G(q)u_t)/H(q)$ is the* prediction error.

**PROOF.** Follows immediately from Corollary 3, by taking $x_0 = 0$. □

**Remark 5** *In the statistics literature the estimation algorithms for ARMA models are described by the following two cases (see e.g. [27,37,36]):* ***exact*** *Maximum Likelihood, when $x_0$ is considered as a random variable with zero mean and variance given in* (21)*, and* ***approximate*** *Maximum Likelihood, when $x_0$ is considered as a parameter (In particular, the case $x_0 = 0$ is known as* conditional *Maximum Likelihood).* ▽▽▽

**Remark 6** *Note that the different likelihood functions obtained in the time- and frequency domains can be concentrated on the parameters of interest. For system identification purposes, we are usually only interested in system parameter vector $\theta$, that defines the transfer functions $G(q)$ and $H(q)$ in (1) (and the state-space model matrices $A$, $B$, $C$, $D$, and $K$ in (2)–(3)).* ▽▽▽

## 3 Frequency Domain Maximum Likelihood

### *3.1 Frequency domain model and data*

In this section we consider the situation where the data has been transformed from time to frequency domain. We review the statistical properties of transformed data when using the DFT. Later, in Section 3.2, we examine the impact that this transformation has on maximum likelihood estimation.

To transfer the estimation problem into the frequency-domain we apply the discrete Fourier transform (DFT) to the data (see e.g. [11, chapter V]):

$$Y_k = \frac{1}{\sqrt{N}}\sum_{t=0}^{N-1} y_t z_k^{-t} = \Re\{Y_k\} + j\Im\{Y_k\} \tag{28}$$

where $z_k = e^{j\omega_k}$, $\omega_k = \frac{2\pi}{N}k$, and $\Re\{Y_k\}$ and $\Im\{Y_k\}$ represent the real and imaginary part of $Y_k$, respectively. The DFT defined in (28) can be expressed in matrix form as follows:

$$\vec{Y} = M_F\,\vec{y} \tag{29}$$

where $\vec{y}$ is given in (10), and

$$\vec{Y} = [Y_0, \ldots, Y_{N-1}]^T \tag{30}$$

and the matrix $M_F$ is the Fourier matrix given by [2, page 214]:

$$M_F = \frac{1}{\sqrt{N}}\left[z_{k-1}^{-(i-1)}\right] \tag{31}$$

where $z_k = e^{j\frac{2\pi}{N}k}$ and the notation $[a_{ik}]$ denotes a matrix having $a_{ik}$ in the $i$-th row and $k$-th column. Note that $z_k^N = 1$, and, thus, $z_k^{N-\ell} = z_k^{-\ell}$. Moreover, the matrix $M_F$ in (31) is Hermitian ($M_F = {M_F}^H$), non-singular ($\det M_F \neq 0$), and unitary ($M_F {M_F}^H = I$) [2].

A difficulty associated with forming the likelihood function in the frequency domain is that $Y_k$ in (28) is a *complex* random variable. Some of the issues arising from having a complex random variable $z = x + jy$, where $x, y \in \mathbb{R}^n$ are real random variables, are outlined below.

(**CV1**) The probability density function (PDF) of $z$ cannot, in general, be written in the traditional way since complex numbers are not an ordered set, i.e., $F_Z(z) = P(Z < z)$ is not meaningful [30,31]. Thus, the PDF of $z$ has to be understood as the joint PDF of the real and imaginary parts [6], i.e.

$$p_Z(z) = p_{X,Y}(x, y) = \left.\frac{\partial^2 F(u, v)}{\partial u \partial v}\right|_{(u,v)=(x,y)} \quad (32)$$

where $F(x, y) = P(X \leq x, Y \leq y)$ is the probability distribution of the random variables $X$ and $Y$ evaluated at $(x, y)$.
(**CV2**) The PDF $p_{X,Y}(x, y)$ can be written as a function $p_Z(z)$ which depends on $z$ [24–26]. This is because $x$ and $y$ can always be written in terms of $z$ (and $z^*$). In fact, $x = \frac{1}{2}(z + z^*)$, $y = \frac{1}{2j}(z - z^*)$ where $z^*$ is the complex conjugate of $z$.
(**CV3**) The terms *proper*, *circular*, or *circular symmetric* are different names used for complex Gaussian random variables. The imaginary and real part of this type of complex random variables are uncorrelated, and the covariance of the real and imaginary parts are the same.
(**CV4**) For a proper complex Gaussian random variable $z$, the PDF $p_Z(z)$ has the *usual form* [29]:

$$p_Z(z) = \frac{\exp\{-[z - \mu]^H \Sigma^{-1} [z - \mu]\}}{\pi^n \det \Sigma} \quad (33)$$

where $\mu$ and $\Sigma$ are the mean and covariance matrix of $z$ (More details are given in Appendix A.1).
(**CV5**) If $W_k \in \mathbb{C}^n$ is proper then any affine transformation of $W_k$ (e.g. $AW_k + b$, where $A \in \mathbb{C}^{m\times n}$, and $b \in \mathbb{C}^n$ are constants) is also proper [25, Lemma 3].

We will be interested in analyzing Maximum Likelihood estimation when the available time-domain data is transformed to the frequency domain by applying the DFT defined as in (28). It is well known that, if $\{Y_k\}$ is the DFT sequence of a time sequence $\{y_t\}$, then:

(**DFT1**) If $\{y_t\}$ is zero mean i.i.d (real) Gaussian sequence with variance $P$. We write $y_t \sim N_r(0, P)$. Then $\{Y_k\}$ given by (28) is an independent zero mean real-complex proper Gaussian sequence having variance $P$. We write $Y_k \sim N_{r-c}(0, P)$. The random variable $Y_k$ is characterized by the following Gaussian distribution [14,17,25,30,31,41]:

$$p(Y_k) = \begin{cases} \dfrac{\exp\{-\frac{1}{2} Y_k^T P^{-1} Y_k\}}{\sqrt{(2\pi)^n \det P}} & ; Y_k \text{ real} \\ \dfrac{\exp\{-Y_k^H P^{-1} Y_k\}}{\pi^n \det P} & ; Y_k \text{ complex} \end{cases} \quad (34)$$

where $^T$ denotes transpose and $^H$ denotes conjugate-transpose. It is important to note that $Y_k$ is a real Gaussian random variable for $k = 0$ and for $k = \frac{N}{2}$ (when $N$ is an even number), and, otherwise, it is a complex Gaussian random variable.
(**DFT2**) The DFT of an i.i.d sequence (not necessarily having a Gaussian PDF but with finite second order moments and mixing) is asymptotically a proper Gaussian i.i.d sequence. This is basically a consequence of the Central Limit Theorem [3,17,33].
(**DFT3**) The PDF of the sequence $Y_0, \cdots, Y_{N-1}$ is singular. This is a consequence of the deterministic relationship that exists between the components of this sequence [14]. If we start with a real vector of length N and we use a linear invertible transformation then we still only have N independent components in the new domain. In fact, $Y_{N-k} = Y_k{}^*$, where $^*$ denotes complex conjugation. This means that only $L = \lfloor \frac{N}{2} \rfloor$ random variables are necessary to define a non-singular PDF, where $\lfloor \frac{N}{2} \rfloor$ denotes the smallest integer greater than or equal to $\frac{N}{2}$. Note that the random variable $Y_L$ is real if $N$ is an even number and is complex if $N$ is an odd number.

In equation (34), and in the sequel, where the sense is clear from the context, we will use the function $p(\cdot)$ to denote a PDF where the form of the PDF is characterized by the arguments. Otherwise, we will use a subscript on $p(\cdot)$ to define the appropriate function.

### *3.2 Frequency domain maximum likelihood*

In this section we study the impact of using the DFT to translate the maximum likelihood estimation problem to the frequency domain. If we apply the DFT (defined in (28)) to the state space model (2)–(3) we obtain:

$$z_k X_k + z_k \tfrac{1}{\sqrt{N}}(x_N - x_0) = A\, X_k + B\, U_k + K\, W_k \quad (35)$$

$$Y_k = C\, X_k + D\, U_k + W_k \quad (36)$$

It has previously been pointed out [1,21,35] that a key difference between time and frequency domain representation of the system is the presence of the extra term $\alpha$, where

$$\alpha = x_0 - x_N \quad (37)$$

The DFT representation (35)–(36) can also be written in terms of transfer functions as (see e.g. [21]):

$$Y_k = F_k \alpha + G_k U_k + H_k W_k \quad (38)$$

where

$$F_k = C(z_k I - A)^{-1} \tfrac{z_k}{\sqrt{N}} = F(z_k) \tfrac{1}{\sqrt{N}} \quad (39)$$

$$G_k = C(z_k I - A)^{-1} B + D = G(z_k) \quad (40)$$

$$H_k = C(z_k I - A)^{-1} K + 1 = H(z_k) \quad (41)$$

and $F(\cdot)$, $G(\cdot)$, and $H(\cdot)$ are defined in (8), (4), and (5), respectively.

Note that the output DFT sequence, $\{Y_k\}$, in (38) can be written in vector form as:

$$\vec{Y} = F_D\alpha + G_D\vec{U} + H_D\vec{W} \tag{42}$$

where the matrices $F_D$, $G_D$, and $H_D$ are defined by

$$F_D = [F_0, \ldots, F_{N-1}]^T \tag{43}$$
$$G_D = \mathrm{diag}\{G_0, \ldots, G_{N-1}\} \tag{44}$$
$$H_D = \mathrm{diag}\{H_0, \ldots, H_{N-1}\} \tag{45}$$

and where $F_k$, $G_k$, and $H_k$ are defined in (39)–(41).

**Remark 7** *Note that the numerator of $F_k\alpha$ in (38) is a polynomial whose coefficients depend on the initial and final state of the system. These coefficients are proportional to $\frac{1}{\sqrt{N}}$. In [17,21] it has been suggested that when the noise signal $w_t$ is bounded, this extra term can be neglected as the number of data points $N$ increases [18, pages 31–33]. However, in most of the analysis in System Identification the noise signal is* **not** *considered as a bounded signal (for example, Gaussian noise is not bounded) and, in addition, by neglecting this term systematic errors are introduced for a finite sample. In [32], an Errors in Variables framework was utilized in order to cope with this difficulty. In [35,33] it was pointed out that, in order to improve the small sample behaviour of the estimates, the numerator coefficients of $F_k\alpha$ can be considered as extra parameters to be estimated.* ▽▽▽

The frequency domain representation for the data above can be used to obtain an associated frequency domain likelihood function. However, a difficulty is the fact that the joint PDF of the sequence of complex DFTs $\{Y_0, \ldots, Y_{N-1}\}$ is degenerate (see Property DFT.3 in Section 3.1). This introduces some difficulties when calculating the conditional probability. In particular, to obtain a PDF, we have to restrict the frequency components up to $L = \lfloor \frac{N}{2} \rfloor$, that is

$$\ell_F(\theta) = p(Y_0, \ldots, Y_L|\theta) \tag{46}$$

Equation (29) represents a (complex) linear transformation from time to frequency domain, which gives the complex (or exponential) Fourier series coefficients $\{Y_k\}$ associated with the time sequence $\{y_t\}$. In the sequel it will be of interest to utilize a **real** linear transformation from time to frequency domain. The real discrete Fourier transform (RDFT) (see, for example, [13,14,7]) is given by the coefficients of the trigonometric Fourier series associated with $\{y_t\}$. The RDFT transforms a real time-domain vector of length $N$ to a real frequency-domain vector of length $N$. Such a real transformation allows us to translate the PDF of the time-domain data to the frequency domain, and also establish a clear equivalence between time and frequency domain Maximum Likelihood estimation.

**Lemma 8** *Given the time-domain vector $\vec{y}$ (defined in* (10)*), there is a real unitary matrix transformation $M_R$ that gives a real valued frequency domain representation $\vec{Y}_R$:*

$$\vec{Y}_R = M_R\vec{y} \tag{47}$$

*where*

$$\vec{Y}_R = \begin{cases} [Y_0, \sqrt{2}\Re\{Y_1\}, \sqrt{2}\Im\{Y_1\}, \ldots, Y_L]^T & \text{if } N \text{ is even} \\ [Y_0, \sqrt{2}\Re\{Y_1\}, \sqrt{2}\Im\{Y_1\}, \\ \quad \ldots, \sqrt{2}\Re\{Y_L\}, \sqrt{2}\Im\{Y_L\}]^T & \text{if } N \text{ is odd} \end{cases} \tag{48}$$

**PROOF.** See Appendix A.2. □

**Remark 9** *The real matrix transformation in Lemma 8 allows us to write the likelihood function in the frequency domain as*

$$\ell_F(\beta) = p(Y_0, \ldots, Y_L|\beta) = p_{\vec{Y}_R}(\vec{Y}_R|\beta) \tag{49}$$

*where $\{Y_0, \ldots, Y_L\}$ is the transformed frequency-domain data, and where the vector $\beta$ contains the parameters to be estimated, i.e., the system parameters in $\theta$, the noise covariance $\sigma_w^2$, and the initial and final state difference $\alpha$. Note that, as stated in the introduction, different likelihood functions will arise from (49), depending on the assumptions made regarding the term $\alpha$.* ▽▽▽

In Lemma 10 below we consider the likelihood function for the case (**F3**) where $\alpha$ is a random variable. We will show that the other cases, (**F1**) and (**F2**) in the frequency domain, and (**T1**)–(**T3**) in the time domain can then be obtained as special cases.

**Lemma 10 (F3)** *Consider the frequency domain representation of the linear system (1), given in (35)–(36) (or, equivalently, in transfer function form (38)). Assume that the term $\alpha$ is a random vector, jointly Gaussian distributed and correlated with the noise process $\{w_t\}$, and having mean $\mu_\alpha$, and joint covariance matrix is*

$$\Sigma_{\left[\substack{\alpha \\ \vec{w}}\right]} = E\left\{ \begin{bmatrix} \alpha - \mu_\alpha \\ \vec{w} \end{bmatrix} \begin{bmatrix} \alpha - \mu_\alpha \\ \vec{w} \end{bmatrix}^T \right\} = \begin{bmatrix} \Sigma_\alpha & \Sigma_{\alpha\vec{w}} \\ \Sigma_{\alpha\vec{w}}^T & \sigma_w^2 I_N \end{bmatrix} \tag{50}$$

*Then the frequency-domain (negative log-) likelihood function, i.e.*

$$L_{F3}(\theta, \sigma_w^2, \mu_\alpha, \Sigma_{\left[\substack{\alpha \\ \vec{w}}\right]}) = -\log p_{\vec{Y}_R}(\vec{Y}_R|\theta, \sigma_w^2, \mu_\alpha, \Sigma_{\left[\substack{\alpha \\ \vec{w}}\right]}) \tag{51}$$

*can be expressed as*

$$L_{F3}(\theta, \sigma_w^2, \mu_\alpha, \Sigma_{\left[\begin{smallmatrix}\alpha\\ \vec{w}\end{smallmatrix}\right]}) = L_0 + \log\det \Sigma_{\vec{Y}_R} \\ + (\vec{Y}_R - \mu_{\vec{Y}_R})^T \Sigma_{\vec{Y}_R}^{-1} (\vec{Y}_R - \mu_{\vec{Y}_R}) \tag{52}$$

*where the term $L_0$ accounts for unimportant constants and where*

$$\mu_{\vec{Y}_R} = M_T G_D \vec{U} + M_T F_D \mu_\alpha \tag{53}$$

$$\Sigma_{\vec{Y}_R} = M_T \begin{bmatrix} F_D & H_D M_F \end{bmatrix} \begin{bmatrix} \Sigma_\alpha & \Sigma_{\alpha\vec{w}} \\ \Sigma_{\alpha\vec{w}}^T & \sigma_w^2 I_N \end{bmatrix} \begin{bmatrix} F_D^H \\ M_F^H H_D^H \end{bmatrix} M_T^H \tag{54}$$

*where $F_D$, $G_D$, and $H_D$ are defined in (43)–(45), $\vec{U}$ is the DFT of the input sequence $\{u_t\}$, and $M_T$ is the (unitary) matrix transformation between the DFT and the RDFT (see Appendix A.2).*

**PROOF.** Since the term $\alpha$ and the noise sequence $\{w_t\}$ are assumed jointly Gaussian distributed:

$$\begin{bmatrix} \alpha \\ \vec{w} \end{bmatrix} \sim N_r \left( \begin{bmatrix} \mu_\alpha \\ 0 \end{bmatrix} ; \begin{bmatrix} \Sigma_\alpha & \Sigma_{\alpha\vec{w}} \\ {\Sigma_{\alpha\vec{w}}}^T & \sigma_w^2 I_N \end{bmatrix} \right) \tag{55}$$

From (42) we have that the RDFT of the output can be written as:

$$\vec{Y}_R = M_T G_D \vec{U} + M_T \begin{bmatrix} F_D & H_D M_F \end{bmatrix} \begin{bmatrix} \alpha \\ \vec{w} \end{bmatrix} \tag{56}$$

where $M_T = M_R {M_F}^H$ (see Appendix A.2). Thus, in the frequency domain, the RDFT of the output sequence is also Gaussian distributed:

$$\vec{Y}_R \sim N_r \left( \mu_{\vec{Y}_R}, \Sigma_{\vec{Y}_R} \right) \tag{57}$$

where $\mu_{\vec{Y}_R}$ and $\Sigma_{\vec{Y}_R}$ are given in (53) and (54), respectively. The negative log-likelihood function is then given by

$$-\log p(\vec{Y}_R) = \frac{1}{2} \Big( N \log(2\pi) + \log\det \Sigma_{\vec{Y}_R} \\ + (\vec{Y}_R - \mu_{\vec{Y}_R})^T \Sigma_{\vec{Y}_R}^{-1} (\vec{Y}_R - \mu_{\vec{Y}_R}) \Big) \tag{58}$$

Note that

$$\log\det \Sigma_{\vec{Y}_R} = \sum_{k=1}^{N} \log \lambda_k(\Sigma_{\vec{Y}_R}) \tag{59}$$

where $\lambda_k(\Sigma_{\vec{Y}_R})$ represent the eigenvalues of $\Sigma_{\vec{Y}_R}$. □

**Remark 11** *In Lemma 10 (F3) we have assumed that $\alpha$ and $\vec{w}$ are correlated. In order to establish the equivalence between time- and frequency-domain identification, one should make the same assumptions in both frameworks. Since different assumptions can be made regarding the data generating mechanism and the parameters to be estimated, one needs to carefully translate the assumptions made in one domain to the equivalent assumptions in the alternative domain. For example, the assumptions made in the time-domain regarding the nature of $x_0$, should be mapped to assumptions on $\alpha$ in the frequency domain by using equation* (37) *(see also* (70) *presented later). Note that these equations make it clear that $x_N$, and hence $\alpha$, depend upon $\vec{w}$. Thus, the cross-covariance between $\alpha$ and $\vec{w}$ will be, in general, non-zero. More will be said in Section 4.* ▽▽▽

**Remark 12** *Notice that, in order to obtain estimates that are robust, for example, to modelling errors, it is possible to consider only a reduced set of frequency-domain data in a specific bandwidth. This strategy has been previously utilized in the statistics literature in e.g. [12,38], and in the engineering literature in [8,21,34,9,42]. In particular, in [42] this approach is proposed to avoid the effect of under-modeling errors when using approximate sampled-data models. This method motivated by the ML principle can be developed for the two types of problem, namely, when $\alpha$ is deterministic (parameter) or stochastic (random variable).* ▽▽▽

**Corollary 13 (F2)** *If the term $\alpha$ is considered as a deterministic parameter to be estimated, then the corresponding (negative log-)likelihood function is given by*

$$L_{F2}(\theta, \sigma_w^2, \alpha) = -\log p_{\vec{Y}_R}(\vec{Y}_R | \theta, \sigma_w^2, \alpha) \\ = L_0 + N \log(\sigma_w^2) + \sum_{k=0}^{N-1} \left[ \log(|H_k|^2) + \frac{1}{\sigma_w^2} |E_k|^2 \right] \tag{60}$$

$$= L_0 + N \log(\sigma_w^2) + 2 \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} f_k \left[ \log(|H_k|^2) + \frac{1}{\sigma_w^2} |E_k|^2 \right] \tag{61}$$

*where $L_0$ accounts for unimportant constants, $H_k$ is as in* (41), *$E_k = (Y_k - G_k U_k - F_k \alpha)/H_k$, and $f_k$ is equal to* 0.5 *whenever $H_k$ (and $E_k$) is real and equal to* 1 *in all other cases.*

**PROOF.** The proof follows from Lemma 10, on setting $\mu_\alpha = \alpha$ and $\Sigma_\alpha = 0$. Note that $\Sigma_{\alpha\vec{w}} = 0$ is a consequence of assuming $\alpha$ to be a parameter. Note, in particular, that:

$$\mu_{\vec{Y}_R} = M_T G_D \vec{U} + M_T F_D \alpha \tag{62}$$

$$\Sigma_{\vec{Y}_R} = \sigma_w^2 M_T H_D {H_D}^H {M_T}^H \tag{63}$$

and, from (45), (59), and (63), we have that

$$\log\det \Sigma_{\vec{Y}_R} = \sum_{k=0}^{N-1} \log\left(\sigma_w^2 |H_k|^2\right) \tag{64}$$

Then, considering that $M_T\vec{Y} = \vec{Y}_R$, $M_T M_T^H = I$, we have that

$$\begin{aligned}
&[\vec{Y}_R - \mu_{\vec{Y}_R}]^T \Sigma_{\vec{Y}_R}^{-1} [\vec{Y}_R - \mu_{\vec{Y}_R}] \\
&= [\vec{Y}_R - M_T G_D \vec{U} - M_T F_D \alpha]^H \Sigma_{\vec{Y}_R}^{-1} \\
&\quad [\vec{Y}_R - M_T G_D \vec{U} - M_T F_D \alpha] \\
&= \frac{1}{\sigma_w^2} [M_T^H \vec{Y}_R - M_T^H M_T G_D \vec{U} - M_T^H M_T F_D \alpha]^H \\
&\quad [H_D H_D^H]^{-1} [M_T^H \vec{Y}_R - M_T^H M_T G_D \vec{U} - M_T^H M_T F_D \alpha] \\
&= \frac{1}{\sigma_w^2} [\vec{Y} - G_D \vec{U} - F_D \alpha]^H [H_D H_D^H]^{-1} [\vec{Y} - G_D \vec{U} - F_D \alpha] \\
&= \frac{1}{\sigma_w^2} \sum_{k=0}^{N-1} |E_k|^2 \qquad (65)
\end{aligned}$$

Finally, we obtain (61) by using Lemma 27 in the Appendix. □

**Corollary 14 (F1)** *If the term $\alpha$ is assumed to be equal to zero, then the corresponding (negative log-) likelihood function is given by*

$$\begin{aligned}
&L_{F1}(\theta, \sigma_w^2) \\
&= L_0 + N\log(\sigma_w^2) + \sum_{k=0}^{N-1} \left[\log(|H_k|^2) + \frac{1}{\sigma_w^2}|E_k|^2\right] \\
&= L_0 + N\log(\sigma_w^2) + 2\sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} f_k \left[\log(|H_k|^2) + \frac{1}{\sigma_w^2}|E_k|^2\right]
\end{aligned} \tag{66}$$

*where $E_k = (Y_k - G_k U_k)/H_k$ and $L_0$ accounts for constants.*

**PROOF.** Immediate from Corollary 13 making $\alpha = 0$.

□

**Remark 15** *The fact that the likelihood function (66) differs from the time domain likelihood function has been discussed in several earlier papers. For example, in [17,21,33] it has been remarked that the logarithmic term that depends on $H_k$ in* (60) *and* (66) *can be neglected in some special cases, for example, if $H_k$ is assumed given or known. Moreover, if we use equally spaced data points on the unit circle (i.e. $w_k = \frac{2\pi}{N}k$, $k = 0, \dots, N-1$), then, as $N$ goes to infinity, the sum of the logarithmic term converges to the following integral:*

$$\frac{1}{N}\sum_{k=0}^{N-1} \log|H_k|^2 \to \int_{-\pi}^{\pi} \log|H|^2 \tag{67}$$

*Since $H(0) = 1$, we have from Jensen's formula on the unit circle (see e.g. [4, page 74]), that this integral is zero. Hence, the cost function* (60) *and* (66) *is* asymptotically *equivalent to the one utilized in the time domain [21].*

*In addition, it has been shown in [34] that the sum on the left hand side of* (67) *is bounded by a term that depends on the dominant pole of $H(z_k)$. This analysis shows that time and frequency approaches asymptotically provide the same estimates. However, for finite N, the estimates obtained in the time and in the frequency domains will be, in general, different. Moreover, in [34] it is shown that the estimates for a Box-Jenkins model obtained in time and frequency domain have different properties.* ▽▽▽

**Remark 16** *Applying the DFT to* (6) *we obtain equation* (38)*. Then, applying* (**CV4**), (**CV5**), (**DFT1**), (**DFT3**) *and Lemma 27, we have that the likelihood function for the cases* (**F1**) *and* (**F2**) *are given as in Corollaries 13 and 14. This procedure has been the usual way to derive the likelihood function described in [18,33,21,34,19,1,42], where the term $\alpha$ is considered as an extra parameter to be estimated. However, this procedure is not useful to derive the likelihood function for the case when $\alpha$ is considered as a random variable that depends on the initial state $x_0$, the input signal $u_t$ and the noise $w_t$* (**F3**)*. This issue is also the core difficulty when attempting to establish the equivalence between the likelihood estimates developed in both domains (time and frequency).* ▽▽▽

Lemma 10, Corollary 13 and Corollary 14 show that the different frequency domain maximum likelihood estimation problems can be obtained by including constraints on the case where $\alpha$ is a random vector (**F3**). In the next section, we show that the time-domain maximum likelihood estimation problems (**T1**)–(**T3**) can also be obtained by making particular assumptions on the frequency domain ML case (**F3**). This result is valid for finite data length.

## 4 Equivalence between time and frequency domain maximum likelihood

### *4.1 General considerations*

We first show that the *values* of the time and frequency domain likelihood estimates are equal when consistent parameter definitions are used.

**Theorem 17** *Given a dynamic system model and a set of output measurements, the estimates obtained by maximizing the likelihood function are the same irrespective of the representation of the data, either in time- or frequency-domain. That is*

$$\arg\max_{\beta} p_{\vec{y}}(\vec{y}|\beta) = \arg\max_{\beta} p_{\vec{Y}_R}(\vec{Y}_R|\beta) \tag{68}$$

*where $\beta$ is a vector that contains the parameters to be estimated and equality holds with probability 1.*

**PROOF.** The result follows from the fact that a (unitary) real matrix transformation exists between the time and frequency domain representation of the data in (47). Thus, from Remark 9 in Section 3.2 and the PDF of a transformation of random variables (see, for example, [28, page 144]), we have

$$p_{\vec{Y}_R}(\vec{Y}_R|\beta) = \frac{1}{|\det M_R|} p_{\vec{y}}(\vec{y}|\beta) = p_{\vec{y}}(\vec{y}|\beta) \tag{69}$$

□

Theorem 17 shows that the probability of the data given the parameter vector $\beta$ takes the same *value* when the data is represented in the time domain or in the frequency domain. Hence, the same estimates are obtained when maximizing the corresponding likelihood function irrespective of whether a time- or frequency-domain representation is used for the data.

**Remark 18** *Theorem 17 shows that time and frequency domain ML estimation are, in fact, different ways of expressing the same problem provided the parameter vector $\beta$ is defined consistently. This means that each one of the maximum likelihood estimation problems in the time-domain,* (**T1**)*–*(**T3**)*, and in the frequency domain,* (**F1**)*–*(**F3**)*, has a corresponding counterpart in the other domain. In the next subsection we show that by considering $\alpha$ as a random vector all of the six different problems can be understood in a common framework. In fact, we show that, if we express $\alpha$ in terms of the initial state $x_0$, the deterministic input $\{u_t\}$, and the noise process $\{w_t\}$ then an equivalence can be established between time- and frequency-domain maximum likelihood approaches.* ▽▽▽

### *4.2 Linking time and frequency domain ML estimation problems*

In order to establish the equivalence between time and frequency ML estimation problems, we first clarify the underlying relationship between time and frequency-domain solutions of the state-space model for the system in (1). These solutions depend on different quantities, namely, the initial state $x_0$, and the difference between the initial and final state, $\alpha = x_0 - x_N$, respectively. The following lemma clarifies the relationship between the two approaches (see [20]).

**Lemma 19** *Consider the state space representation of the system (2)–(3). The time-domain solution of the system in (7) and the frequency-domain solution (38), where $\alpha = x_0 - x_N$, are related by the DFT (28) if, and only if, the final state $x_N$ is replaced by the time-domain solution given by:*

$$x_N = A^N x_0 + \sum_{t=0}^{N-1} A^{N-1-t}(B\,u_t + K\,w_t) \tag{70}$$

**Corollary 20** *Consider the vector representation of the time- and frequency-domain solutions of the system in (9) and (42), respectively, where the matrices $\Gamma$, $\Lambda$, and $\Omega$ are given in (13)–(14), and the matrices $F_D$, $G_D$, and $H_D$ are given in* (43)*–*(45)*. Then the following matrix equalities hold:*

$$\Gamma = M_F{}^H F_D(I - A^N) \tag{71}$$
$$\Lambda = M_F{}^H G_D M_F - M_F{}^H F_D M_u \tag{72}$$
$$\Omega = M_F{}^H H_D M_F - M_F{}^H F_D M_w \tag{73}$$

*where $M_F$ is the Fourier matrix (31), and we have defined the following block matrices:*

$$M_u = \begin{bmatrix} A^{N-1}B & \cdots & B \end{bmatrix} \tag{74}$$
$$M_w = \begin{bmatrix} A^{N-1}K & \cdots & K \end{bmatrix} \tag{75}$$

**PROOF.** The proof follows from Lemma 19, substituting $\alpha = x_0 - x_N$ and $x_N$ as in (70).

The core issue in the current paper is whether the estimates associated with the different likelihood functions, in the time and frequency domain, are the same or not. The following theorem clarifies this issue by using the relationship highlighted in Lemma 19 to define $\alpha$ as a random variable depending on the initial state $x_0$, the input $\{u_t\}$ and the noise $\{w_t\}$.

**Theorem 21 (Frequency domain equivalent of (T3))** *Consider the frequency-domain likelihood function in Lemma 10 where the random variable $\alpha$ is expressed in terms of $x_0$, $\{u_t\}$, and $\{w_t\}$ via the system dynamics, i.e.*

$$\begin{aligned}\alpha &= (I - A^N)x_0 - \sum_{t=0}^{N-1} A^{N-1-t}(Bu_t + Kw_t)\\ &= (I - A^N)x_0 - M_u\vec{u} - M_w\vec{w} \end{aligned} \tag{76}$$

| Problem | $\mu_\alpha$ | $\boldsymbol{\Sigma}_\alpha$ | $\boldsymbol{\Sigma}_{\alpha\vec{w}}$ | **Extra Constraint** |
|---|---|---|---|---|
| **T1** | $-M_u\vec{u}$ | $M_w M_w^T \sigma_w^2$ | $-M_w\sigma_w^2$ | none |
| **T2** | $(I-A^N)x_0 - M_u\vec{u}$ | $M_w M_w^T \sigma_w^2$ | $-M_w\sigma_w^2$ | none |
| **T3** | $(I-A^N)\mu_{x_0} - M_u\vec{u}$ | $(I-A^N)\Sigma_0(I-A^N)^T + M_w M_w^T \sigma_w^2$ | $-M_w\sigma_w^2$ | $\Sigma_0 = A\Sigma_0 A^T + \sigma_u^2 BB^T + \sigma_w^2 KK^T$ |
| **F1** | 0 | 0 | 0 | none |
| **F2** | $\alpha$ | 0 | 0 | none |

Table 1
Constraints imposed on (**F3**) in order to obtain the different estimation procedures.

*Then the associated frequency domain likelihood function of Lemma 10 is equivalent to the time domain likelihood function given in Lemma 1.*

**PROOF.** The proof follows by obtaining the marginal probability of the time domain data (given the parameters to be estimated) from the joint probability of the frequency domain data and the term $\alpha$ (given the parameters to be estimated), i.e., from (69) we have that

$$p_{\vec{y}}(\vec{y}|\beta) = p_{\vec{Y}_R}(\vec{Y}_R|\beta) = \int_\alpha p_{\vec{Y}_R,\alpha}(\vec{Y}_R,\alpha|\beta)d\alpha \qquad (77)$$

Note that in Lemma 10, $\vec{Y}_R$ is expressed in terms of $\alpha$, the DFT of the input, $\{U_k\}$, and the DFT of the noise sequence, $\{W_k\}$. From Lemma 19 we know that $\vec{Y}_R$ can be equivalently expressed in terms of $x_0$, $\{u_t\}$, and $\{w_t\}$ if, and only if, $x_N$ is considered as in (70) or, equivalently, if $\alpha$ is considered as in (76).

Thus, the joint PDF of $\vec{Y}_R$ and $\alpha$ can be obtained from

$$\begin{bmatrix}\vec{Y}_R \\ \alpha\end{bmatrix} = \begin{bmatrix} M_R\Lambda \\ -M_u\end{bmatrix}\vec{u} + \begin{bmatrix} M_R\Omega & M_R\Gamma \\ -M_w & (I-A^N)\end{bmatrix}\begin{bmatrix}\vec{w} \\ x_0\end{bmatrix} \qquad (78)$$

Hence,

$$\begin{bmatrix}\vec{Y}_R \\ \alpha\end{bmatrix} \sim N_r\left(\begin{bmatrix} M_R\Lambda\vec{u} + M_R\Gamma\mu_{x_0} \\ -M_u\vec{u} + (I-A^N)\mu_{x_0}\end{bmatrix};\right.$$
$$\left.\begin{bmatrix} M_R\Omega & M_R\Gamma \\ -M_w & (I-A^N)\end{bmatrix}\begin{bmatrix}\sigma_w^2 I_N & 0 \\ 0 & \Sigma_0\end{bmatrix}\begin{bmatrix} M_R\Omega & M_R\Gamma \\ -M_w & (I-A^N)\end{bmatrix}^T\right) \qquad (79)$$

The marginal distribution for $\vec{Y}_R$ is then readily obtained

$$\vec{Y}_R \sim N_r\left(M_R\Lambda\vec{u} + M_R\Gamma\mu_{x_0};\right.$$
$$\left.\sigma_w^2 M_R\Omega\Omega^T {M_R}^T + M_R\Gamma\Sigma_0\Gamma^T {M_R}^T\right) \qquad (80)$$

We notice that the PDF of $\vec{Y}_R$ will lead to the same likelihood function in (16) from the fact that $\vec{y} = M_R^T\vec{Y}_R$ and, thus

$${M_R}^T\vec{Y}_R \sim N_r\left(\Lambda\vec{u} + \Gamma\mu_{x_0}; \sigma_w^2\Omega\Omega^T + \Gamma\Sigma_0\Gamma^T\right) \qquad (81)$$

□

**Remark 22** *Another way of obtaining the result in Lemma 21 would be to rewrite (77) as*

$$p_{\vec{y}}(\vec{y}|\beta) = \int_\alpha p_{\vec{Y}_R|\alpha}(\vec{Y}_R|\alpha,\beta)p_\alpha(\alpha|\beta)d\alpha \qquad (82)$$

*Inside the integral, the PDF of $\alpha$ can be obtained from (76), however, the conditional probability $p_{\vec{Y}_R|\alpha}(\vec{Y}_R|\alpha,\beta)$ is not straightforward to obtain. This conditional probability resembles the one considered in Corollary 13. In that result, however, $\alpha$ is taken to be a* parameter. *By way of contrast, the conditional probability inside the integral in (82) considers $\alpha$ as a* random variable *having a specific relation to the initial state $x_0$, the input $\{u_t\}$, and the noise $\{w_t\}$. The conditional probability in this case has to be obtained (if required) from the joint Gaussian distribution described in (79).* ▽▽▽

**Remark 23** *Theorem 21 has shown that, provided one calculates the appropriate joint distribution for $\alpha$ and $\{w_t\}$, then the time- and frequency-domain maximum likelihood estimation problems are equivalent. Note that this equivalence holds true for finite data length.* ▽▽▽

**Remark 24** *Depending on the assumptions made on the initial state $x_0$, different joint distributions for $\alpha$ and $\{w_t\}$ are obtained which, in fact, correspond to the time-domain likelihood functions associated with the different cases* (**T1**)–(**T3**). *This shows that considering $\alpha$ as a random variable is the most general case among the six cases described in the introduction, namely* (**T1**)–(**T3**) *and* (**F1**)–(**F3**). *Table 4.2 summarizes the different constraints necessary to obtain the other methods presented in this paper, i.e.* (**T1**, **T2**, **T3**, **F1**, **F2**). ▽▽▽

**Remark 25** *The frequency domain ML developed in Theorem 21 will lead to consistent estimates of $G(q)$ for Box-Jenkins models of systems operating in open loop, irrespective of under-modeling in $H(q)$. This is a consequence of the equivalence to the time-domain maximum likelihood estimation problem.* ▽▽▽

The following lemma compares the covariance of the (dependent) random variables involved in the two domains i.e. $x_0$ and $\alpha$.

**Lemma 26** *Assume that the input $\{u_t\}$ and noise $\{w_t\}$ are i.i.d., mutually uncorrelated, zero mean random processes having variances $\sigma_u^2$ and $\sigma_w^2$, respectively. If the initial state of system (2) is assumed to have as covariance matrix the solution of the Lyapunov equation* (21)*, then the covariance of the term $\alpha$ is asymptotically (as N goes to infinity) twice the covariance of $x_0$, i.e. ,* $\lim_{N\to\infty} \Sigma_\alpha = 2\,\Sigma_0$.

**PROOF.** The term $\alpha$ can be obtained from (70):

$$\alpha = (I - A^N)x_0 - \sum_{t=0}^{N-1} A^{N-1-t}(Bu_t + Kw_t) \tag{83}$$

Thus, $E\{\alpha\} = 0$ and its covariance is given by

$$\Sigma_\alpha = (I - A^N)\Sigma_0(I - A^N)^T + \sigma_u^2 \sum_{t=0}^{N-1} A^t BB^T (A^T)^t + \sigma_w^2 \sum_{t=0}^{N-1} A^t KK^T (A^T)^t \tag{84}$$

As $N$ goes to infinity, $A^N$ goes to zero because the system is assumed to be stable, thus $\lim_{N\to\infty} \Sigma_\alpha = \Sigma_0 + \Sigma_1$ where

$$\Sigma_1 = \lim_{N\to\infty} \left[ \sigma_u^2 \sum_{t=0}^{N-1} A^t BB^T (A^T)^t + \sigma_w^2 \sum_{t=0}^{N-1} A^t KK^T (A^T)^t \right] \tag{85}$$

Matrix $\Sigma_1$ then satisfies the Lyapunov equation

$$\Sigma_1 = A\Sigma_1 A^T + \sigma_u^2 BB^T + \sigma_w^2 KK^T \tag{86}$$

This means that $\Sigma_1 = \Sigma_0$ and, as a consequence, we have that $\Sigma_\alpha = 2\Sigma_0$. □

Lemma 26 shows that, under certain conditions, $\mathrm{E}\{\alpha\} = 0$, and $\Sigma_\alpha = 2\Sigma_0$. If we recall the case (**F1**), we see that the constraint imposed on $\mu_\alpha$ is asymptotically correct, however, the constraints imposed on $\Sigma_\alpha$ and $\Sigma_{\alpha\vec{w}}$ are valid only for low noise power ($\sigma_w^2 \approx 0$) and for a deterministic zero mean input.

## 5 Numerical examples

In this section we present numerical examples to highlight the estimates obtained using time and frequency domain maximum likelihood. We consider a simple model expressed in state-space form:

$$x_{t+1} = ax_t + bu_t + kw_t \tag{87}$$

$$y_t = x_t + w_t \tag{88}$$

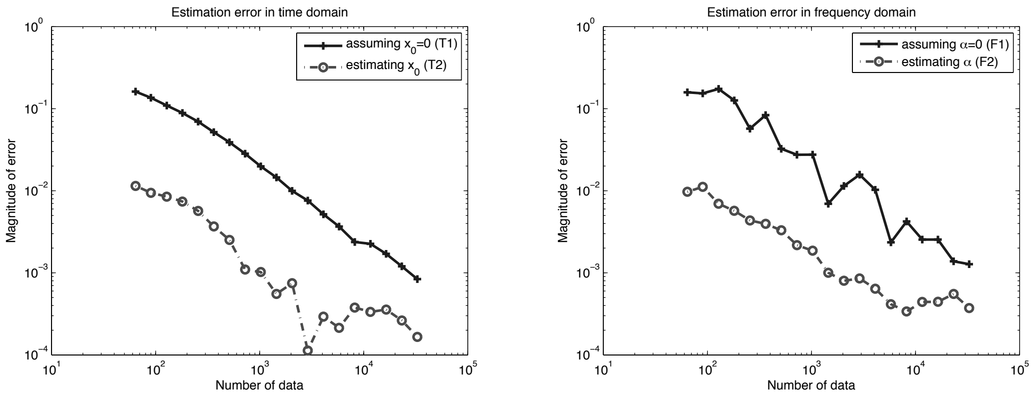
where $\{u_t\}$ is a zero mean Gaussian noise sequence with unit variance, $\{w_t\}$ is a zero mean Gaussian noise sequence with variance $\sigma_w^2 = 0.1$, and $\{x_t\}$ is the state vector. The true parameters are $a = 0.75$, $b = 0.5$, and $k = 1$. We consider an initial state $x_0 = 3$.

The system was simulated over $N$ data points, where $N$ ranges from $2^6$ to $2^{15}$. For each data length we run $N_{MC} = 1000$ Monte-Carlo simulations with different seeds of noise. To minimize the negative log-likelihood function we use the command `fminunc` from the optimization toolbox of MATLAB and we use the `n4sid` command to obtain an initial estimate of the parameters for the optimization routine.

Figure 1 shows the results obtained when applying the time-domain approaches (**T1**) and (**T2**), and the frequency domain approaches (**F1**) and (**F2**). Figure 1(a) shows the error $\|\theta - \hat{\theta}\|_2$, where $\theta = [a, b, k]^T$ is the *true* system parameter and $\hat{\theta}$ is the average of the estimates obtained from the Monte Carlo simulations. We can see that the error when estimating the initial condition (**T2**) is smaller than the error when the initial condition is assumed to be zero (**T1**). Similarly, Figure 1(b) shows that the error when estimating the term $\alpha$ as a deterministic parameter (**F2**) is smaller than the error when the term $\alpha$ is assumed to be zero (**F1**). However, in both figures we see that, as the data length is increased, the magnitude and the difference between the errors decrease.

Table 2 shows the empirical mean and variance of the parameter estimates obtained for (**T1**), (**T2**), (**F1**) and (**F2**). We show the results obtained for three different data lengths $N = 2^7$, $2^{10}$, and $2^{15}$. We can see that the standard deviation decreases for long data sets, and the mean approaches the true parameter value. We notice that the largest error corresponds to the parameter $k$ when the initial condition is not estimated. We also see that the largest standard deviation is in the parameter $k$. When we include $x_0$ (or $\alpha$) as an unknown parameter to be estimated, the quality of the estimates is greatly improved, especially, for short data lengths.

The results in Figure 1 show that, in order to obtain a good estimate of $\theta$, it is better to estimate the initial state $x_0$ (when working in the time domain) or the term $\alpha$ (in the frequency). However, assuming this extra parameter equal to zero may, in some cases, provide parameter



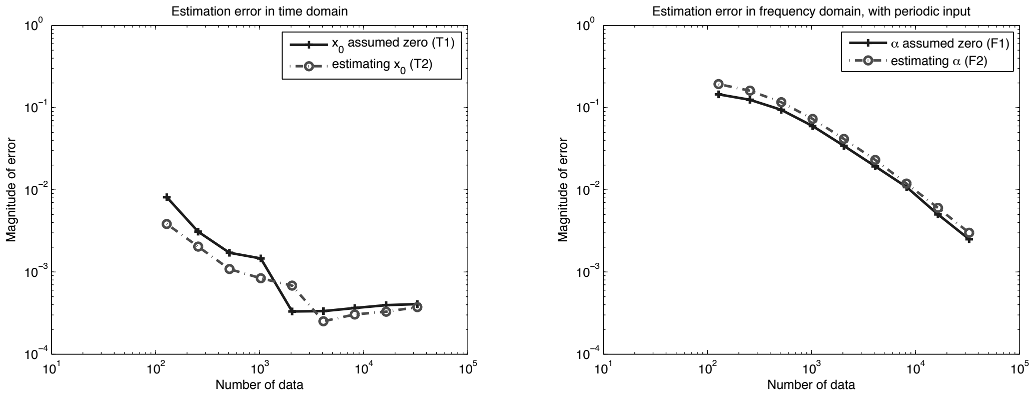
(a) Estimation error $||\theta_0 - \hat{\theta}||$ for (**T1**) and (**T2**) (b) Estimation error $||\theta_0 - \hat{\theta}||$ for (**F1**) and (**F2**)

Fig. 1. Time and frequency domain estimation results when using a random input.

| $\hat{\theta} \rightarrow$ | $a$ | | $b$ | | $k$ | |
|---|---|---|---|---|---|---|
| $\theta_0 \rightarrow$ | 0.75 | | 0.5 | | 1 | |
| $N$<br>↓ | (**T1**)<br>(**F1**) | (**T2**)<br>(**F2**) | (**T1**)<br>(**F1**) | (**T2**)<br>(**F2**) | (**T1**)<br>(**F1**) | (**T2**)<br>(**F2**) |
| $2^7$ | $0.743 \pm 0.041$<br>$0.689 \pm 0.071$ | $0.742 \pm 0.040$<br>$0.752 \pm 0.067$ | $0.499 \pm 0.031$<br>$0.527 \pm 0.067$ | $0.498 \pm 0.031$<br>$0.501 \pm 0.038$ | $0.891 \pm 0.071$<br>$0.842 \pm 0.129$ | $1.001 \pm 0.092$<br>$1.016 \pm 0.158$ |
| $2^{10}$ | $0.7500 \pm 0.0136$<br>$0.7537 \pm 0.0135$ | $0.7498 \pm 0.0136$<br>$0.7494 \pm 0.0135$ | $0.4994 \pm 0.0099$<br>$0.4916 \pm 0.0104$ | $0.4993 \pm 0.0099$<br>$0.4997 \pm 0.0102$ | $0.9802 \pm 0.0295$<br>$0.9766 \pm 0.0320$ | $1.0007 \pm 0.0307$<br>$1.0012 \pm 0.0310$ |
| $2^{15}$ | $0.74993 \pm 0.00226$<br>$0.74988 \pm 0.00230$ | $0.74993 \pm 0.00226$<br>$0.74998 \pm 0.00230$ | $0.49999 \pm 0.00179$<br>$0.50005 \pm 0.00171$ | $0.49999 \pm 0.00179$<br>$0.50001 \pm 0.00171$ | $0.99916 \pm 0.00523$<br>$0.99904 \pm 0.00519$ | $0.99985 \pm 0.00523$<br>$0.99999 \pm 0.00517$ |

Table 2

Estimation Results: Each cell shows the (empirical mean)±(empirical standard deviation) of the parameters



(a) $||\theta_0 - \hat{\theta}||$ for (**T1**) and (**T2**) when $E\{x_0\} = 0$ (b) $||\theta_0 - \hat{\theta}||$ for (**F1**) and (**F2**) when $E\{\alpha\} = 0$

Fig. 2. Results for time-domain estimation for $E\{x_0\} = 0$ and frequency domain estimation when $E\{x_0 - x_N\} = 0$.

estimates of similar quality. For example, Figure 2(a) shows the estimation results when applying a zero mean input sequence $\{u_t\}$, and using the part of the data such that initial transient has disappeared. We notice that, in this case, both time-domain approaches (**T1**) and (**T2**) provide similar results. On the other hand, Figure 2(b) shows the estimation results when the input $\{u_t\}$ is a periodic signal, whose period has been chosen such that an integer number of periods fit into the data length. In this case, both frequency domain approaches (**F1**) and (**F2**) give similar estimation results. In fact, a slightly smaller estimation error is obtained for (**F2**) which confirms that $\alpha = 0$, i.e., $x_N = x_0$, is a *good* assumption in this case.

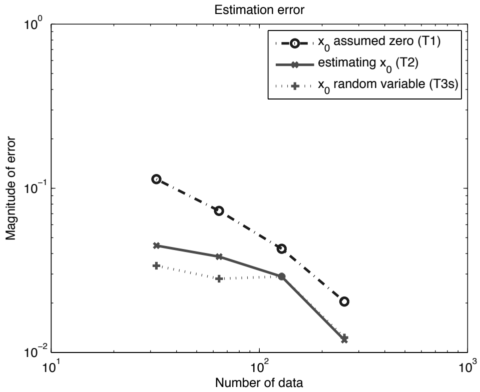


Fig. 3. Estimation error for time-domain approaches.

Finally, we consider the case when the initial condition is assumed as a random vector (**T3**). In particular, we consider that the system has reached its stationary response (see Remark 2). Thus, the initial state mean and covariance are $\mu_{x_0} = \frac{b}{1-a}\mu_u$ and $\Sigma_{x_0} = \sigma^2_{x_0} = \frac{\sigma^2_u\, b^2+\sigma^2_w\, k^2}{1-a^2}$, where $\sigma^2_u$ and $\mu_u$ are the input variance and mean, respectively, and $\sigma^2_w$ is the noise variance. Note that the mean and variance of $x_0$ depend on the system parameters. We refer to this approach as (**T3s**). Figure 3 shows the estimation results for the three possible time-domain approaches (**T1**), (**T2**), and (**T3s**). We consider $N = 2^5, \ldots,\ 2^8$, running $N_{MC} = 1000$ Monte-Carlo simulations for each data length. The input is a zero mean random Gaussian sequence having mean $\mu_u = 1.5$ and variance $\sigma^2_u = 0.625$. The corresponding initial state *true* mean is $\mu_{x_0} = 3$ and variance $\sigma^2_{x_0} = 1.5$. The results in Figure 3 show that (**T3s**) yields a lower estimation error than (**T1**) or (**T2**). Moreover, as the data length is increased the three approaches converge and provide a good estimate of the system parameters.

## 6 Conclusions

This paper has considered the equivalence between maximum likelihood estimation in the time and frequency domains for finite length data. It has been shown that both approaches are equivalent provided consistent assumptions are made regarding the unknown parameters in the problem. Also, it is apparent from the development presented here, that it is unsurprising that the methods lead to different results when inconsistent assumptions are made.

## Acknowledgements

We thank the Associate Editor and the reviewers for their helpful comments and suggestions.

## A Appendix

**Lemma 27** *Consider a DFT sequence $\{Y_k\}$, then the following quadratic cost equivalence holds:*

$$\sum_{k=0}^{N-1} g(|Y_k|^2) = 2\sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} f_k\, g(|Y_k|^2) \tag{A.1}$$

*where $g(\cdot)$ is a function, $f_k$ is equal to* 0.5 *whenever $Y_k$ is real and* 1 *in any other case.*

***PROOF.*** *Straightforward from the fact that $Y_k = {Y_{N-k}}^*$, and recalling that $Y_k$ is real for $k = 0$ and for $k = \frac{N}{2}$ (when $N$ is even).* □

### *A.1 Real and complex Gaussian distributions*

A random vector $x \in R^{n\times 1}$ is said to be normally distributed if its PDF is given by [24, page 36]:

$$p(x) = (2\pi)^{-\frac{n}{2}} \det\Sigma^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}[x-\mu]^T\Sigma^{-1}[x-\mu]\} \tag{A.2}$$

where $\mu$ and $\Sigma$ are the mean and covariance of $x$. The PDF of a complex circular multivariate random vector $z \in \mathbb{C}^{n\times 1}$, with $\mathrm{E}\{z\} = \mu \in \mathbb{C}^{n\times 1}$ and $\mathrm{E}\left\{[z-\mu][z-\mu]^H\right\} = \Sigma \in \mathbb{C}^{n\times n}$, is given by [24, page 77]:

$$p(z) = \frac{\exp\{-[z-\mu]^H\Sigma^{-1}[z-\mu]\}}{\pi^n \det\Sigma} \tag{A.3}$$

where ${}^H$ stands for conjugate-transpose. Alternatively, let $z = x + jy$, $x, y \in \mathbb{R}^{n\times 1}$ and define $\xi = [x, y]^T$, then

$$p(z) = \frac{\exp\{-\frac{1}{2}[\xi-\gamma]^T M^{-1}[\xi-\gamma]\}}{[2\pi]^n \det M^{\frac{1}{2}}} \tag{A.4}$$

where $\gamma = \begin{bmatrix}\Re\{\mu\}\\ \Im\{\mu\}\end{bmatrix}$ and $M = \frac{1}{2}Re\begin{bmatrix}\Sigma & j\Sigma\\ j\Sigma^H & \Sigma\end{bmatrix}$.

We also have that $\det M = 2^{-2n}(\det\Sigma)^2$.

### *A.2 Proof of Lemma 8*

Equation (48) can be expressed in matrix form as

$$\vec{Y}_R = \begin{bmatrix} 1 & 0 & 0 \ldots 0 & 0\\ 0 & \frac{\sqrt{2}}{2} & 0 \ldots 0 & \frac{\sqrt{2}}{2}\\ 0 & \frac{\sqrt{2}}{2j} & 0 \ldots 0 & \frac{-\sqrt{2}}{2j}\\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} \vec{Y} = M_T\,\vec{Y} \tag{A.5}$$

Note that, from (48), the last rows of matrix $M_T$ are a function of whether $N$ is an even or an odd number. Let us define $L = \lfloor \frac{N}{2} \rfloor$, i.e., the largest integer greater than or equal to $\frac{N}{2}$. If $N$ is odd, then the last two elements of $\vec{Y}_R$ are $\sqrt{2}\Re\{Y_L\}$ and $\sqrt{2}\Im\{Y_L\}$. On the other hand, if $N$ is even, $Y_L$ is real and, thus, the last row in the matrix $M_T$ has only a 1 in the $(L+1)$-th position.

We notice that the matrix $M_T$ in (A.5) is non-singular and unitary (i.e., $M_T {M_T}^H = I$). Thus, there exists a real unitary matrix transformation $M_R$, from $\vec{y}$ to $\vec{Y}_R$, obtained as:

$$\vec{Y}_R = M_T\,\vec{Y} = M_T\,M_F\,\vec{y} = M_R\,\vec{y} \tag{A.6}$$

where $M_F$ is the (unitary) Fourier matrix defined in (31). The matrix transformation given by $M_R$, as described above, corresponds to the trigonometric Fourier series representation:

$$y_t = \begin{cases} \frac{1}{\sqrt{N}} \Big( Y_0 + Y_L + 2 \sum_{\ell=1}^{L-1} \Re\{Y_\ell\} \cos(\omega_\ell t) \\ \qquad + \Im\{Y_\ell\} \sin(\omega_\ell t) \Big) \qquad \text{if } N \text{ is even} \\ \frac{1}{\sqrt{N}} \Big( Y_0 + 2 \sum_{\ell=1}^{L} \Re\{Y_\ell\} \cos(\omega_\ell t) \\ \qquad + \Im\{Y_\ell\} \sin(\omega_\ell t) \Big) \qquad \text{if } N \text{ is odd} \end{cases}$$

where the coefficients are given by $Y_0 = \sqrt{\frac{1}{N}} \sum_{t=0}^{N-1} y_t$, $\sqrt{2}\Re\{Y_\ell\} = \sqrt{\frac{2}{N}} \sum_{t=0}^{N-1} y_t \cos(\omega_\ell t)$, and $\sqrt{2}\Im\{Y_\ell\} = \sqrt{\frac{2}{N}} \sum_{t=0}^{N-1} y_t \sin(\omega_\ell t)$. □

## References

[1] J. C. Agüero, J. I. Yuz, G. C. Goodwin, Frequency domain identification of MIMO state space models using the EM algorithm, in: European Control Conference ECC, 2007.

[2] D. Bernstein, Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory, Princeton University Press, 2005.

[3] D. R. Brillinger, Time series: Data analysis and theory, SIAM, 2001.

[4] G. F. Carrier, M. Krook, C. E. Pearson, Functions of a complex variables: theory and technique, McGraw-Hill, 1966.

[5] M. Deistler, System identification- general aspects and structure, Model Identification and Adaptive Control: From windsurfing to telecommunications. Edited by G. C. Goodwin. (2000) 3–26.

[6] J. L. Doob, Stochastic Processes, John Wiley & Sons, 1953.

[7] O. Ersoy, Real discrete fourier transform, IEEE Transactions on Acoustics, Speech and Signal Processing 33 (4) (1985) 880–882.

[8] G. C. Goodwin, Some observations on robust estimation and control, in: 7th IFAC Symposium on Identification and System Parameter Estimation, York, UK, 1985.

[9] G. C. Goodwin, J. C. Agüero, J. S. Welsh, J. I. Yuz, G. J. Adams, C. R. Rojas, Robust identification of process models from plant data, Journal of Process Control 18 (2008) 810–820.

[10] G. C. Goodwin, R. Payne, Dynamic System Identification: Experiment design and data analysis, Academic Press, 1977.

[11] E. J. Hannan, Multiple time series, John Wiley and Sons, Inc., 1970.

[12] E. J. Hannan, P. M. Robinson, Lagged regression with unknown lags, Journal of the Royal Statistical Society. Series B (Methodological) 35 (2) (1973) 252–267.

[13] A. C. Harvey, Linear regression in the frequency domain, International Economic Review 19 (2) (1978) 507–512.

[14] W. Hirt, J. L. Massey, Capacity of the discrete-time Gaussian channel with intersymbol interference, IEEE Transactions on Information Theory 34 (3) (1988) 380–388.

[15] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, 1970.

[16] T. Kailath, Linear Systems, Prentice Hall, USA, 1980.

[17] L. Ljung, Some results on identifying linear systems using frequency domain data, in: Proceedings of the 32nd IEEE Conference on Decision and Control, 1993.

[18] L. Ljung, System Identification: Theory for the user, 2nd ed., Prentice Hall, 1999.

[19] L. Ljung, Frequency domain versus time domain methods in system identification - revisited, Lecture notes in control and information sciences 329 (2006) 277–291.

[20] T. McKelvey, Frequency domain identification, in: 12th IFAC Symposium on System Identification, Santa Barbara, USA, 2000.

[21] T. McKelvey, Frequency domain identification methods, Circuits systems signal processing 21 (1) (2002) 39–55.

[22] T. McKelvey, A. Helmersson, State space parameterization of multivariable linear systems using tridiagonal matrix form, Proceedings of the 35th IEEE Conference on Decision and Control, CDC (1996) 3654–3659.

[23] T. McKelvey, L. Ljung, Frequency domain maximum likelihood identification, in: 11th IFAC Symposium on System Identification, Fukuoka, Japan, 1997.

[24] K. S. Miller, Complex stochastic processes: An introduction to theory and applications, Addison-Wesley Publishing Company, INC., 1974.

[25] F. D. Neeser, J. L. Massey, Proper complex random processes with applications to information theory, IEEE Transactions on Information Theory 39 (4) (1993) 1293–1302.

[26] S. Olhede, On probability density functions for complex variables, IEEE Transactions on Information Theory 52 (3) (2006) 1212–1217.

[27] D. R. Osborn, Exact and approximate maximum likelihood estimators for vector moving average processes, Journal of the Royal Statistical Society, Series B 39 (1) (1977) 114–118.

[28] A. Papoulis, Probability, random variables, and stochastic processes, 3rd ed., McGraw-Hill, 1991.

[29] B. Picinbono, Random signals and systems, Englewood Cliffs (NJ) : Prentice Hall, 1993.

[30] B. Picinbono, On circularity, IEEE Transactions on Signal Processing 42 (12) (1994) 3473–3482.

[31] B. Picinbono, Second-order complex random vectors and normal distributions, IEEE Transactions on Signal Processing 44 (10) (1996) 2637–2640.

[32] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, H. Van Hamme, Parametric identification of transfer functions in the frequency domain-a survey, Automatic Control, IEEE Transactions on 39 (11) (1994) 2245–2260.

[33] R. Pintelon, J. Schoukens, System identification: A frequency domain approach, IEEE Press, 2001.

[34] R. Pintelon, J. Schoukens, Box-jenkins identification revisited–part i: Theory, Automatica 42 (1) (2006) 63–75.

[35] R. Pintelon, J. Schoukens, G. Vandersteen, Frequency domain system identification using arbitrary signals, Automatic Control, IEEE Transactions on 42 (12) (1997) 1717–1720.

[36] D. S. G. Pollock, A handbook of time-series analysis, signal processing and dynamics, Academic Press, 1999.

[37] G. C. Reinsel, Elements of Multivariable Time Series Analysis, Springer, 1997.

[38] P. M. Robinson, Stochastic difference equations with non-integral differences, Advances in Applied Probability 6 (3) (1974) 524–545.

[39] J. Schoukens, R. Pintelon, Y. Rolain, Time domain identification, frequency domain identification. equivalences! differences?, in: 2004 American Control Conference, 2004.

[40] T. Söderström, P. Stoica, System identification, Prentice-Hall International, 1989.

[41] A. Van den Bos, The real-complex normal distribution, IEEE Transactions on Information Theory 44 (4) (1998) 1670–1672.

[42] J. I. Yuz, G. C. Goodwin, Robust identification of continuous-time systems from sampled data, in: H. Garnier, L. Wang (eds.), Continuous-time Model Identification from Sampled Data, chap. 3, Springer, 2008, pp. 67–89.