Frequency Domain Identification

FREQUENCY DOMAIN IDENTIFICATION

T. McKelvey *,[1]

* *Dept. of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.*
tomas@isy.liu.se

Abstract: Techniques to identify parametric transfer functions from noisy frequency domain data are considered. A maximum-likelihood estimation method is presented which in parallel with the system transfer function also estimates a parametric noise transfer function. This leads to a consistent and efficient estimator. It is shown how the discrete Fourier transform can be applied to generate frequency domain data from sampled time domain data. For the finite data case the exact frequency domain expressions are derived relating the transfer function with the discret Fourier transformed data for both continuous and discrete time systems.

Keywords: Identification, spectral estimation, discrete Fourier transform, maximum-likelihood estimation

1. INTRODUCTION

Building mathematical models based on measured input and output signals of a dynamical system is known as system identification. Such models based on empirical information are important if the dynamical system is unknown or is only partially known and when it is in-feasible to derive a theoretical model from first principles. The availability of accurate models is important in order to derive high performance solutions, e.g., for model based control design or model based signal processing.

Almost all measurements originating from real world devices intrinsically belong to the time domain, i.e. are samples of continuous time signals. Consequently most system identification methods and the theory developed around them deals with how to determine parametric models from such time domain measurements (Ljung 1999, Söderström and Stoica 1989). The very basic and old technique of frequency analysis departs from the time domain technique.

In frequency analysis the system is assumed to be excited by a (sum of) pure sinusoidal signal(s). When the output has settled to a (sum of) stationary sinusoidal signal(s), the complex value of the transfer function at the specific excitation frequencies is determined by the discrete Fourier transform (DFT).

So called frequency analyzers are dedicated pieces of equipment which performs such experiments and deliver the result as a non-parametric transfer function. In a second step a parametrized transfer function model can be fitted to the transfer function data using some complex curve fitting technique (Levy 1959, Sanathanan and Koerner 1963, Whitfield 1986, Whitfield 1987, Pintelon *et al.* 1994). A discussion of techniques to fit parametric models to noisy frequency domain data is the scope of this paper. An early reference for the related time series problem is (Whittle 1951) wherein classical inferential procedures, e.g. maximum-likelihood estimation, were combined with spectral theory. See also (Hannan 1970, Robinson 1976). The book (Schoukens and Pintelon 1991) covers many aspects of frequency domain identification techniques and presents

[1] Supported by the Swedish Foundation for International Cooperation in Research and Higher Education (STINT). This work was completed while the author was visiting Dept. of Electrical and Computer Engineering, University of Newcastle, Australia.

a method for the frequency domain errors-in-variables problem.

Since the transformation of a signal from time to frequency domain using the discrete Fourier transform is nothing but a unitary transformation it might appear, at first sight, that nothing is gained.

However, an important difference arises when colored noises are present. Specifically, for such a colored signal $v(t)$ which have a covariance function $c(\tau) = \mathrm{E}\{v(t+\tau)v(t)\}$ (decaying to zero at an exponential rate), the vector of all time domain measurements have a full covariance matrix. The optimal estimation procedure, i.e. the maximum-likelihood method, requires the use the full covariance matrix since all time domain signals are statistically dependent (Hannan and Deistler 1988). The unitary transformation, represented by the DFT, however decouples this statistical dependence. That is, in the frequency domain each frequency point is asymptotically (as the number of time domain data points tends to infinity) independent of the others (Brillinger 1981).

Furthermore under very mild assumptions regarding the probability distribution of the time domain noise signal, the frequency domain data are all (asymptotically) normally distributed. Optimal estimators therefore only include a diagonal covariance matrix which simplifies the actual computations.

Prefiltering with the inverse of the filter describing the noise color will also make the time domain covariance matrix diagonal. This is also known as pre-whitening since the resulting noise is white. However this requires the knowledge of the noise color or it needs to be jointly estimated together with the rest of the unknown parameters. So-called prediction error methods (Åström 1980, Ljung 1999, Söderström and Stoica 1989) are based on prefiltering with a parametric dependent prefilter and are asymptotically equivalent to the method of maximum-likelihood. In case only non-parametric noise information is at hand, the design of whitening-filter is more demanding. Finally it is important to recognize that the unitary transformation does not change any asymptotic properties if all frequency domain data points are retained. The time-domain maximum-likelihood methods and the frequency domain counterpart both have equal asymptotic properties as pointed out by (Hannan 1970, Ljung and Glover 1981, Schoukens *et al.* 1997*a*) and others. The actual choice of one over the other must therefore be based on the finite data properties of each estimator and also on computational complexity.

A distinctive feature of frequency domain techniques is that the modeling of continuous time systems from sampled data can be done in a straightforward fashion if a certain class of band-limited excitation signals is employed (Robinson 1976, Schoukens *et al.* 1994). This is a great advantage in contrast with the rather involved time domain techniques which, even in the noise free case, are only approximate if a finite set of sampled data is available. See (Young 1981, Unbehauen and Rau 1990). A continuous time system with a time delay is rather difficult to model in the time domain since it cannot be described by a finite system of ordinary differential equations. In the frequency domain however a nice finite dimensional parametric description exists which lends itself to identification using parametric methods.

Frequency domain identification of nonlinear systems of Wiener-Hammerstein type have been reported in (Vandersteen *et al.* 1997, Vandersteen and Schoukens 1999). These techniques rely on a convergent Volterra representation of the nonlinear system.

Techniques based on higher order spectra (polyspectra) can be found in (Giannakis 1995, Tugnait 1998) which by assuming particular noise and input distributions generate consistent estimates.

1.1 *Problem formulation*

Let us assume that we are interested in obtaining a model of a system that can be described by the following linear time-invariant form

$$y(t) = G_0(q)u(t) + H_0(q)e(t) \qquad (1)$$

where $y(t)$, $u(t)$ and $e(t)$ are the real valued output, input and innovation signals, respectively. The operators $G_0(q)$ and $H_0(q)$ represent the discrete time linear transfer functions. We assume that H_0 is a stable monic filter. For a continuous time representation, exchange the delay operator q with the differentiation operator p. The noise signal $e(t)$ is assumed to be i.i.d. and zero mean with variance λ_0 and independent of the input signal $u(t)$. We will later in Section 3.5 discuss the closed loop case where u and e are dependent through a feedback controller.

Let us postulate the following frequency domain system equation

$$Y(\omega) = G_0(e^{i\omega})U(\omega) + H_0(e^{i\omega})E_0(\omega). \qquad (2)$$

where Y and U are the (weak) limits of the Fourier transform of y and u respectively and E_0 is the frequency domain innovation which is a zero mean complex normal random variable with frequency invariant variance λ_0. The complex functions $G_0(e^{i\omega})$ and $H_0(e^{i\omega})$ are the frequency functions of the linear operators G and H respectively. In Section 2 we will more formally justify (2). For a continuous time system just exchange the argument $e^{i\omega}$ for $i\omega$.

Assume the relation (2) can be sampled at a sequence of arbitrary frequencies in the set $\Omega_N = \{\omega_k\}_{k=1}^N$ yielding the set of measurements $Z^N = \{Y_k, U_k | k = 1, \ldots, N\}$ where $U_k = U(\omega_k)$ and

$Y_k = Y(\omega_k)$. The aim is to find a model of (1) and to do so we construct a model set by using a parametrized model structure

$$Y(\omega) = G_\theta(e^{i\omega})U(\omega) + H_\theta(e^{i\omega})E(\omega), \qquad (3)$$

where $G_\theta(z)$ and $H_\theta(z)$ are respectively the rational transfer functions of the system and noise which are parametrized by a real valued vector θ. Let $D_\mathcal{M}$ denote the set of valid parameters. We assume $H_\theta(z)$ is stable and inversely stable monic transfer function for all $\theta \in D_\mathcal{M}$. We impose no particular structure on how the parameters enter into $G_\theta(z)$ or $H_\theta(z)$ and this enables the use of various parametrizations such as fraction of polynomials or state-space models. Hence, parametrized *gray-box models* which are only partially unknown can also be used.

Given the parametrized model class and data Z^N, the estimate of the system is found by parametric optimization of some criterion function

$$\hat{\theta} = \arg\min_\theta V_N(\theta, Z^N) \qquad (4)$$

where V_N is a function which provides a metric on how to optimally fit the model (3) to the given data Z^N.

1.2 *Outline*

The paper is organized as follows. In the next section we provide the background to the frequency domain equation (2) for both the discrete time case as well as the continuous time case, with a key feature being the provision of insight into how the noise properties carry over to the frequency domain by using a probabilistic setup. In Section 3 an identification criterion V_N is derived which is based on the technique of maximum-likelihood. The asymptotic properties of such a criterion are analyzed and their relation with the time domain counterpart is discussed.

The aim of the paper is to provide a basic understanding for the frequency domain identification problem. In order to accommodate this we will restrict ourself to consider the case when the system and noise models G_0 and H_0 are finite dimensional and scalar. Extensions to the multivariable case are straightforward.

2. FROM TIME TO FREQUENCY DOMAIN

The exact noise free relationships between the discrete Fourier transform (DFT) of the input and outputs and the frequency response function of the linear system will be derived under rather general excitation conditions. It is well known that the Fourier transform of the noise free output is exactly the frequency response function multiplied by the transform of the input signal if the transform make use of data sequences of infinite length (Ljung 1999). In the finite length case an extra term appears which accounts for the history of the system prior to the measurement interval. As the number of data tends to infinity the extra term tends to zero at a rate proportional to $\frac{1}{\sqrt{N}}$. In the following analysis we will use state-space models to describe the finite-dimensional linear systems. A similar derivation using transfer function models can also be found in (Pintelon *et al.* 1997).

2.1 *Discrete Fourier Transform*

Assume a signal $s(t)$ is sampled at N equidistant time instances $t = kh, k = 0, 1, \ldots, N-1$ where h is the sampling interval. The N-point discrete Fourier transform (DFT) of the set $\{s(kh)\}_{k=0}^{N-1}$ is then defined as

$$S_N(\omega) \triangleq \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s(kh)e^{-i\omega k} \qquad (5)$$

where $\omega \in [-\pi, \pi]$ is the normalized angular frequency in radians per second. Hence ω/h is the unnormalized angular frequency. Unless h is explicitly included we assume $h = 1$ in the sequel. Predominately we will focus on the N distinct values of S_N given by the argument $\omega_k \triangleq \frac{2\pi k}{N}$ for $k = 0, 1, \ldots, N-1$. Notice that $e^{i\omega_k} = e^{-i\omega_{N-k}}$ and $S_N(\omega_k) = S_N^*(\omega_{N-k}) = S_N^*(-\omega_k)$ for $k = 1, \ldots, N-1$. Here X^* denotes the complex conjugate of X.

2.2 *Discrete time systems*

In this section a derivation of the DFT relation for a discrete time system is presented. The discussion of the influence of noise is deferred to Section 2.5. A noise free discrete time system of finite order admits a state-space realization

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \qquad (6)$$

where $x(t)$ is a state-vector of length n, A is a $n \times n$ real matrix, B and C^T are vectors of size n. If the realization is of minimal order, n is the McMillan degree of the linear system (Kailath 1980). The transfer function is given by

$$G(z) = C(zI - A)^{-1}B \qquad (7)$$

which is a fraction between two polynomials. The *frequency response function* at frequency ω is defined as the transfer function evaluated on the unit circle $G(e^{i\omega})$.

Consider the system described by (6) and assume N points of the input and output signals are available, i.e. $\{y(t), u(t)\}_{t=0}^{N-1}$. The history of the input up to time $t < 0$ is unknown but its impact on the future is captured by the state at time zero, $x(0) = x_0$. Assume $\det(e^{i\omega_k}I - A)$ is nonzero for all $k = 0, 1, \ldots, N-1$ which is equivalent to requiring the frequency response function to be finite for all ω_k or equivalently no poles of the system to be located at $e^{i\omega_k}$.

Let $Y_N(\omega)$ and $U_N(\omega)$ denote the N-point DFT of the output and input signals respectively. Then for $\omega_k = \frac{2\pi k}{N}$, $k = 0, 1 \ldots, N-1$ the following equation holds

$$Y_N(\omega_k) = G(e^{i\omega_k})U_N(\omega_k) + T(e^{i\omega_k})\frac{1}{\sqrt{N}} \quad (8)$$

where

$$\begin{aligned} T(z) &= zC(zI - A)^{-1}(I - A^N)(x_0 - x_p) \\ x_p &= (I - A^N)^{-1}\sum_{t=0}^{N-1} A^t B u(N-1-t). \end{aligned} \quad (9)$$

A proof is detailed in Appendix A.1.

The transient term $T(z)\frac{1}{\sqrt{N}}$, which picks up the transient effects of the unmatched initial condition, decays as $\frac{1}{\sqrt{N}}$ if the system is stable, i.e., the eigenvalues of the A matrix have a modulus strictly less than one or, equivalently, all poles lie strictly inside the unit circle. If the input signal is non-periodic then $U_N(\omega_k)$ will stay bounded as N increases and hence $G(e^{i\omega_k})U_N(\omega_k)$ will be the dominating term for large N. If the input is periodic with period time P and $N = PM$, i.e. the measurements are collected over M full periods, then $U_N(\omega)$ will at most have P non-zero values at $\omega = \frac{2\pi k}{P}$, $k = 0, \ldots, P-1$. The size of the non-zero values will grow at a rate of $\sqrt{M}$. In this case we obtain

$$\frac{|T(e^{i\omega_k})|\frac{1}{\sqrt{N}}}{|G(e^{i\omega_k})U_N(\omega_k)|} \sim O(\frac{1}{N})$$

The explicit form of $T(z)$ enables the possibility of estimating it along with the transfer function $G(z)$. This could be beneficial when the data record is short and when a non-periodic excitation signal is used. Furthermore note that $G(z)$ and $T(z)$ share the same denominator polynomial and only the n parameters of the numerator polynomial need to be determined. If the input excitation is such that $U_N(\omega_k)$ is constant for all frequencies it is however not possible to distinguish between $G(z)$ and $T(z)$.

Note that the relation (8) holds also for unstable systems. In such a case the size of transient term will grow exponentially as N increases and it is essential to include it in any open loop estimation scheme.

2.3 *Continuous time systems*

The output of a finite dimensional continuous time system can be described as the solution to a system of first order differential equations

$$\begin{aligned} \dot{x}(t) &= A_c x(t) + B_c u(t) \\ y(t) &= C_c x(t) \end{aligned} \quad (10)$$

where $x(t)$ is the size n state vector. The transfer function is

$$G_c(s) = C_c(sI - A_c)^{-1}B_c \quad (11)$$

and the continuous time *frequency response function* at frequency ω is defined as the transfer function evaluated along the imaginary axis, i.e. $G_c(i\omega)$.

To successfully identify a continuous time system from sampled data it is important to consider how the input signal excites the continuous time system. If the input is piecewise constant between the sampling instances then the continuous system has a discrete time representation which exactly describes the output signal at the sample points and hence the expressions (8) and (9) hold. The mapping which takes a continuous time system into a discrete one is called zero order hold (ZOH) sampling (Åström and Wittenmark 1984). A restriction is that this mapping is not bijective. Several continuous time systems are represented by the same discrete time one. Furthermore the inverse mapping is not defined for certain discrete time systems (Åström and Wittenmark 1984). The method of first identifying a discrete time model and then employ the inverse ZOH mapping might consequently fail. The correct approach is to parameterize a continuous time model and then via the ZOH mapping derive the discrete time model which is matched with the sampled data (Ljung 1999).

A different approach which is well suited for the frequency domain, is to excite the continuous time system using a band limited input with a zero spectrum for all frequencies on and above the Nyquist frequency ($\frac{\pi}{h}$). Furthermore, assume that the input signal is such that it admitts a band-limited Nh-periodic continuation outside the measurement interval, i.e. $u(t) = u(t + Nh)$ for all t. Fourier analysis then tells us that all such signals can be described as (for N odd)

$$u(t) = \sum_{k=-\frac{N-1}{2}}^{\frac{N-1}{2}} f_k e^{i\frac{2\pi k}{Nh}t} \quad (12)$$

where $u(t)$ is real if $f_k = f^*_{-k}$. Clearly all such real valued signals are composed of a sum of sinusoids with (normalized) frequencies constrained to belong to the finite set $\{\omega_k \,|\, \omega_k = \frac{2\pi k}{N} < \pi, k = 0, 1, \ldots\}$. This type of excitation signal is known as multi-sine excitation (Schoukens and Pintelon 1991).

Assume a continuous time system (10) from $t = 0$ and onwards is excited by an input signal which is band limited and has a band limited Nh-periodic extension outside the measurement interval. The input and output signals are sampled at N points with sampling interval h. No assumptions are made about the character of the input to the system prior to to tome $t = 0$. The input history is concisely represented by the initial state x_0. Also assume $\det(i\frac{\omega_k}{h}I - A_c) \neq 0$ for all ω_k. Then the following equation holds.

$$Y_N(\omega_k) = G_c(i\frac{\omega_k}{h})U_N(\omega_k) + T_c(e^{i\omega_k})\frac{1}{\sqrt{N}} \quad (13)$$

where

$$G_c(s) = C_c(sI - A_c)^{-1}B_c$$
$$T_c(z) = zC_c(zI - e^{A_ch})^{-1}(I - e^{A_chN})(x_0 - x_p)$$
$$x_p = (I - e^{A_chN})^{-1}\int_{\tau=0}^{Nh} e^{A_c\tau}B_cu(Nh-\tau)\,\mathrm{d}\tau \quad (14)$$

where $e^{A_c\tau}$ is the matrix exponential associated with the matrix A_c (Kailath 1980). The proof of (13) can be found in Appendix A.2. The relation (13) is quite similar to the discrete time version (8) with one important exception. Here $T_c(z)$ is a discrete time transfer function. However just as in the discrete time case $T_c(z)$ is completely defined by the system, the input and the initial condition at time $t = 0$. In (Pintelon *et al.* 1997) an approximate expression was derived for the transient term which made use of a continuous time transfer function.

2.4 *Discussion*

In both domains we have left out systems which have a direct feed-through term, which exists if the input instantaneously can influence the output. Most physical systems do not have such properties. However the inclusion of a direct feed-through term is straightforward and only involves an additional notational effort. By comparing (8) with (13) we conclude that the only differences between the frequency response function for discrete time systems and continuous time system is the argument,(i.e. $e^{i\omega}$ or $i\omega$) and the transient terms. The extension to the multivariable case is straightforward since the derivations made use of a state-space representation. In the general case when the system is infinite dimensional and stable, the equations (8) and (13) still hold. In this case the transient transfer function $T(z)$ does not admit any finite representation but the frequency response can be upper bounded (Ljung 1999).

2.5 *The noise*

Recall from (1) that we assume the noise to be described as $v(t) = H_0(q)e(t)$ where $H_0(q)$ is a stable linear filter and $e(t)$ is zero mean i.i.d. with variance λ_0. Denote by $E_N(\omega_k)$ the DFT of the noise signal $\{e(t)\}_{t=0}^{N-1}$. It is easy to establish that $E_N(\omega_k)$ is a zero mean random variable with the second moment properties

$$\mathbf{E}\{E_N(\omega_k)E_N^*(\omega_s)\} = \begin{cases} \lambda_0, & \omega_k = \omega_s \\ 0, & \omega_k \neq \omega_s \end{cases} \quad (15)$$

and $\mathbf{E}\{E_N(\omega_k)E_N(\omega_s)\} = 0$ for all $\omega_k, \omega_k > 0$ and $\omega_k = \omega_s \neq \pi$. If $e(t)$ is normally distributed or if $N \to \infty$ then $E_N(\omega_k)$ has a complex normal distribution (Brillinger 1981)

$$E_N(\omega_k) \in N^c(0, \lambda_0) \quad (16)$$

for $\omega_k \in \{\frac{2\pi k}{N}, k = 1 \dots, N-1\}$ and $\omega_k \neq \pi$. For $\omega_k = 0$ and $\omega_k = \pi$, $E_N(\omega_k)$ is real, zero mean, with variance λ_0 and normal distributed.

Using equation (8) the DFT of $v(t) = H_0(q)e(t)$ is conveniently described by

$$V_N(\omega_k) = H_0(e^{i\omega_k})E_N(\omega_k) + T_{H_0}(e^{i\omega_k})\frac{1}{\sqrt{N}} \quad (17)$$

where the last frequency function $T_{H_0}(z)$ is due to the "unmatched" initial condition of the noise filter and is a linear function of the of the innovations $e(t)$ for $t < N$. This implies it has zero mean for all frequencies. When $H_0(z)$ is finite dimensional it is also possible to derive an exact expression for the covariance of T_H. We refrain from doing so here and just note that for our purposes it suffices to use the uniform bound

$$C_{H_0} = \max_k |T_{H_0}(e^{i\omega_k})|^2 \quad (18)$$

which exists since $H_0(z)$ is assumed strictly stable. Since (by definition) $e(t)$ is a non-periodic signal $E_N(\omega)$ will remain bounded and $H_0(e^{i\omega_k})E_N(\omega_k)$ will be the dominating term in $V_N(\omega_k)$ when N is large. All together we obtain for the first and second moments:

$$\mathbf{E}\{V_N(\omega_k)\} = 0, \quad \forall\omega_k \quad (19)$$

$$\mathbf{E}\{V_N(\omega_k)V_N^*(\omega_s)\} = \begin{cases} |H_0(e^{i\omega_k})|^2\lambda_0 + \dfrac{\xi_1(\omega_k)}{N}, & \omega_k = \omega_s \\ \dfrac{\xi_2(\omega_k)}{N}, & \omega_k \neq \omega_s \end{cases} \quad (20)$$

where $|\xi_2(\omega_k)| \leq |\xi_1(\omega_k)| \leq C_H$. Asymptotically, as the number of data points tends to infinity, the term $T_{H_0}(e^{i\omega_k})\frac{1}{\sqrt{N}}$ tends to zero. Then $V_N(\omega_k)$ becomes complex normal distributed with zero mean and variance $|H_0(e^{i\omega_k})|^2\lambda_0$ and independent between frequencies. It is now clear that the DFT of the output signal asymptotically has the properties as postulated in (2). For a much more thorough treatment and relaxed assumptions we refer to (Brillinger 1981, Ljung 1999).

We make no distinction between the continuous time case and the discrete time case regarding the noise signal $v(t)$. For both cases we use a *discrete time model* to describe the noise properties for the sampled data. The validity of this approach can be argued as follows. Consider a continuous time noise signal described by a stochastic differential equation which is equidistantly sampled in time. The first and second order moments of the signal at the sampling instances can equally be described by a discrete time stochastic model of the same order as the continuous time stochastic model (Åström 1970).

2.6 *Periodic input and averaging techniques*

The expressions derived above hold for arbitrary excitation signals and we made no assumption about the history of the input signal prior to time $t = 0$. If on the other hand a stable system is excited with an P-periodic signal which has been applied prior to taking the measurements, say at time $t = -s$ for some integer s and the measurement record is a multiple of the period length, then the initial condition x_0 will approach x_p as s increases. The size of the transfer function $T(z)$ will therefore also decrease, see equations (9) and (14).

If measurements are collected over M full periods then $U_N(\omega)$ will at most have P non-zero values. However the size of the non-zero values will grow at a rate of $\sqrt{M}$. Consequently the noise free counterpart of $Y_N(\omega)$ at the non-zero excitation frequencies will also grow at the same rate. Since the noise signal is non-periodic, $V_N(\omega)$ will remain bounded for all frequencies and the signal to noise ratio at the excitation frequencies will increase with a factor M. For a given size of the measurement window $N = MP$ the choice between period length P and number of periods M involves a fundamental tradeoff between frequency resolution and the signal to noise ratio.

An alternative view is given by averaging the measurements over one period. Assume the input is periodic with period time P. A natural estimate of the noise free output is then

$$\hat{y}_M(t) = \frac{1}{M}\sum_{k=0}^{M-1} y_m(t+kP), \quad 0 \le t \le P-1 \tag{21}$$

and periodically continued for larger t. This gives a noise estimate

$$\hat{v}(t) = y_m(t) - \hat{y}(t)$$

from which the noise properties can be estimated. Furthermore we can use the averaged period as data $\{u(t), \hat{y}_M(t)\}_{t=0}^{N-1}$ in the model building session. This gives both a data size reduction as well as a noise level reduction. In the paper (Schoukens *et al.* 1997*b*) it is shown that such a non-parametric noise estimate based on averaging over only four periods can be used instead of the true ones and still preserve asymptotic optimality of a maximum-likelihood type estimator. See also Section 3.6. Finally it is worth mentioning that $\hat{y}_M(t)$ from (21) is a consistent (non-parametric) estimate as $M \to \infty$ of the noise free periodic output under very mild assumptions on the noise properties. Consequently $\hat{y}_M(t)$ and $u(t)$ can in a second step be used to fit a parametric model in either domain. See (McKelvey and Akçay 1995, McKelvey 1996, Forssell *et al.* 1999).

3. IDENTIFYING A PARAMETRIC MODEL

After providing the necessary information about the measurement signals we are now ready to use the frequency domain data in order to find a model of the underlying system. Certainly the estimate

$$\hat{G}(\omega_k) = \frac{Y_N(\omega_k)}{U_N(\omega_k)}$$

which is known as the empirical transfer function estimate (Ljung 1999) would be one possible alternative. Its simplicity is appealing but the variance at each frequency is asymptotically, c.f. (20), $|H_0(e^{j\omega_k})|^2\lambda_0/|U(\omega)|^2$ which does not decrease with an increasing N unless $u(t)$ is periodic. To reduce the variance the transforms can be smoothed which decreases the variance but introduces a bias in the estimate. Here we will approach the problem by instead fitting a parametrized finite order transfer function (3) to the noisy frequency domain data. If the model set is large enough such that the true system is included, the estimate will asymptotically be unbiased.

3.1 *Method of maximum-likelihood*

The method of maximum-likelihood is frequently used for many estimation problems (Kendall and Stuart 1967, Åström 1980). Most frequency domain identification techniques do not explicitly model the noise properties with a parametric model. Instead consistent estimates are obtained by output-error type algorithms (Tugnait 1998) or IV-type methods (McKelvey 1997) or by using a non-parametric noise model which is either known or estimated a-priori (Schoukens and Pintelon 1991, Schoukens *et al.* 1997*b*). Disregarding the correct noise properties leads to an increased variance of the estimated system. Here we will present the frequency domain maximum-likelihood estimator which explicitly models the unknown noise transfer function. To simplify notation in what follows let $G_{0,\omega} \triangleq G_0(e^{i\omega})$ and $G_{\theta,\omega} \triangleq G_\theta(e^{i\omega})$ and similarly for H. Notice that in this section the frequencies ω_k in the set Ω_N can be arbitrary and not constrained to the frequency grid implied by a DFT.

Recall the postulated identification setup given by equation (2) which was motivated in Section 2. Accordingly the samples Y_k of the output Fourier transform have a complex normal distribution with mean value $G_{0,\omega_k}U_k$ and variance $|H_{0,\omega_k}|^2\lambda_0$. The probability density function for each measurement Y_k is thus[2] (Brillinger 1981)

[2] Here we assume that the frequencies $\omega = 0$ and $\omega = \pi$ is not part of our set Ω_N since at these frequencies $Y(\omega)$ is real valued and has a slightly different probability density function. Including them will however not change any asymptotic properties.

$$\frac{1}{\pi|H_{0,\omega_k}|^2\lambda_0}\exp\left(-\frac{|Y_k - G_{0,\omega_k}U_k|^2}{|H_{0,\omega_k}|^2\lambda_0}\right) \quad (22)$$

Since the measurements are independent of each other the joint probability density function for Z^N is the product of the individual functions (22).

If instead of the true (but unknown) transfer functions G_0, H_0 and variance λ_0 we insert the parametrized counterparts G_θ, H_θ and λ of our model and consider the measurements Z^N fixed we obtain a parametrized likelihood function. By taking the negative logarithm of the parametrized likelihood function and removing terms which do not depend on the parameters (θ and λ) we obtain (Ljung 1994, McKelvey and Ljung 1997)

$$V_N(\theta,\lambda) = \frac{1}{N}\sum_{k=1}^{N}\Big[\log\left(|H_{\theta,\omega_k}|^2\lambda\right) + \frac{|Y_k - G_{\theta,\omega_k}U_k|^2}{|H_{\theta,\omega_k}|^2\lambda}\Big] \quad (23)$$

which for given values of θ and λ can be calculated. Minimizing V_N with respect to θ and the noise variance λ yields the *maximum-likelihood estimate*,

$$\hat{\theta}_N, \hat{\lambda}_N = \arg\min_{\theta,\lambda} V_N(\theta). \quad (24)$$

To reduce the number of free parameters, the variance λ can be eliminated from the criterion by analytical minimization, see (McKelvey and Ljung 1997).

If the data record is short and it is expected that the transient terms in (8) or (13) cannot be neglected then $G_{\theta,\omega}U_k$ should be exchanged for $G_{\theta,\omega}U_k + T_{\theta,\omega}\frac{1}{\sqrt{N}}$ in order to capture the transient effects. Note that the n parameters in T_θ which depend on the initial condition $x_0 - x_p$ cannot consistently be estimated since the effect of $T(z)$ on Y_k decays as $\frac{1}{\sqrt{N}}$.

In general the minimization of (23) can not be performed analytically and an iterative optimization strategy needs to be employed. Often Newton type methods (Dennis and Schnabel 1983) perform well for this class of problems.

3.2 *Asymptotic properties*

As the number of frequency points increases to infinity the ML criterion (23) converges to a limit function which can be described by an integral.

Denote by Ω the interval of the real line to which the set of sample frequencies belongs and let Ω_N (as before) denote the set of sample frequencies. Let us define

$$W_N(\omega) = \frac{|\{k\,|\,\omega_k < \omega, \omega_k \in \Omega_N\}|}{N} \quad (25)$$

where $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$. In probability theory $W_N(\omega)$ corresponds to a distribution function. We assume that the sequence of sample frequencies is such that $W_N(\omega)$ converges to a function $W(\omega)$ in all points of continuity of $W(\omega)$. By using a Stieltjes integral notation (Rudin 1987), an infinite sum can be written as an integral. In our case we have that as $N \to \infty$ (see (McKelvey and Ljung 1997) for the details)

$$V_N(\theta,\lambda) \to \bar{V}(\theta,\lambda) \triangleq \int_\Omega \frac{|G_{0,\omega} - G_{\theta,\omega}|^2\Phi_u(\omega) + \Phi_v(\omega)}{|H_{\theta,\omega}|^2\lambda} + \log|H_{\theta,\omega}|^2\lambda\,\mathrm{d}W(\omega), \quad (26)$$

where $\Phi_u(\omega) = |U(\omega)|^2$ and $\Phi_v(\omega) = |H_{0,\omega}|^2\lambda_0$. Under some regularity conditions on G_θ and H_θ the convergence in (26) is uniform with probability one. It then follows that the estimate $\hat{\theta}_N, \hat{\lambda}_N$ converges to values which minimize $\bar{V}(\theta,\lambda)$.

Assume the model is sufficiently flexible such that there exists a non-empty set Θ_* such that for all $\theta_* \in \Theta_*$

$$\int_\Omega |G_{0,\omega} - G_{\theta_*,\omega}|^2\Phi_u(\omega)\,dW(\omega) = 0 \quad (27)$$

$$\int_\Omega \left||H_{0,\omega}|^2 - |H_{\theta_*,\omega}|^2\right|^2 dW(\omega) = 0 \quad (28)$$

Then it can be shown that $\bar{V}(\theta,\lambda)$ is minimized by all $\theta \in \Theta_*$ and $\lambda = \lambda_0$. If the model structure is restricted such that the limiting set Θ_* is a singleton then $\hat{\theta}_N \to \theta_*$ with probability one as $N \to \infty$ which means that the estimator is consistent (in the sense of satisfying (27) and (28)).

By using a fixed noise model which does not depend on the parameters θ the limiting estimate is the minimizer of

$$\int_\Omega \frac{|G_{0,\omega} - G_{\theta,\omega}|^2\Phi_u(\omega)}{|H_\omega|^2}\,\mathrm{d}W(\omega) \quad (29)$$

If the true system G_0 is not part of the model class then an approximate model will result. In this case the estimated model will be the model which in a weighted mean square sense best approximates the transfer function of the system. As shown in (29) the weight is dependent on the spectrum of the excitation signal, the inverse of the assumed noise model and the distribution function of the frequency samples. Note that the true noise spectrum does not influence the limiting estimate.

If the frequency distribution function $W(\omega)$ is differentiable in the interior of the interval Ω the Stieltjes integrals simplifies into $\int_\Omega(\cdot)w(\omega)\,\mathrm{d}\omega$ where $w(\omega) = \frac{d}{d\omega}W(\omega)$. In this case $w(\omega)$ acts as a standard weighting function.

3.3 *Asymptotic variance*

Consider the case when a fixed noise model H_ω is used in the criterion (23) and assume the limiting set Θ_*, defined by (27) is a singleton

θ_*. Furthermore assume that $G'_{\omega,\theta}$ and $G''_{\omega,\theta}$ are Lipschitz continuous. Define

$$Q = \int_\Omega \frac{\Phi_v(\omega)\Phi_u(\omega)2\mathrm{Re}\{G'_{\omega,\theta_*}(G'_{\omega,\theta_*})^*\}}{|H_\omega|^4}\,\mathrm{d}W(\omega) \tag{30}$$

and

$$R = \int_\Omega \frac{\Phi_u(\omega)2\mathrm{Re}\{G'_{\omega,\theta_*}(G'_{\omega,\theta_*})^*\}}{|H_\omega|^2}\,\mathrm{d}W(\omega) \tag{31}$$

and assume $R > \delta I$ for some $\delta > 0$. Then the estimate given by minimizing (23) is asymptotically normally distributed (McKelvey and Ljung 1997)

$$\sqrt{N}(\hat{\theta}_N - \theta_*) \in AsN(0, P_\theta) \tag{32}$$

with covariance matrix $P_\theta = R^{-1}QR^{-1}$. If the noise model is chosen equal to the true one i.e.

$$\int_\Omega \left||H_{0,\omega}|^2 - |H_\omega|^2\right|^2 dW(\omega) = 0$$

then the size of the covariance is minimized[3] and is equal to $P_\theta^{opt} = \lambda_0 R^{-1}$. The estimator is then asymptotically *efficient*. If an independently parametrized noise model H_ξ is estimated together with G_θ then the variance result (32) still holds and the optimal variance P_θ^{opt} is obtained if the estimated noise model satisfies (28).

3.4 *Discussion*

It is interesting to compare the frequency domain ML-estimator (23) with its time-domain counterpart as described in (Ljung 1999, Söderström and Stoica 1989). Applying Parsevals formula to the time-domain equation reveals that the time domain ML estimator minimizes the function

$$\int_{-\pi}^{\pi} \frac{|Y_N(\omega_k) - G_{\theta,\omega_k}U_N(\omega_k)|^2}{|H_{\theta,\omega_k}|^2}\,\mathrm{d}\omega. \tag{33}$$

A few points are worth noticing. For a fixed known noise model $H_\theta(z) = H(z)$ the frequency domain ML-estimate (23) and the time domain estimate (33) are essentially the same. Whenever the noise model is estimated, the additional term

$$\frac{1}{N}\sum_{k=1}^{N} \log(|H_{\theta,\omega_k}|^2\lambda) \tag{34}$$

occurs in the criterion. In fact (34) is the determinant of the transformation which change variables from Y to E (output to innovations). In the time domain this transformation is triangular with 1's along the diagonal. Hence, this transformation has a determinant equal to 1, so it does not affect the ML criterion.

However if the frequencies ω_k are equidistantly distributed between $-\pi$ and π this additional term again becomes of no importance since, independently of $\theta \in D_\mathcal{M}$,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log(|H_{\theta,\omega}|^2\lambda)\,\mathrm{d}\omega = \log\lambda$$

for any stable and inversely stable monic transfer function. The sum in (34) is also $\log\lambda$ if $\omega_k = \frac{2\pi(k-1)}{N}$, $k = 1, \ldots, N$. This implies that the time-domain ML estimator and the frequency domain estimator are (asymptotically) identical if the frequency domain estimator obtains the data by DFT and all frequencies are used (see also (Schoukens *et al.* 1997*a*)). If the frequency domain estimator uses only a subset of frequencies the limiting criterion function will be weighted by the distribution function of the frequencies and it is vital to use the full ML-criterion (23) to guarantee consistency. If a good fit is required at a certain frequency band the input spectrum should be large for these frequencies or the input and output data could be bandpass filtered prior to estimation. For example if the model is intended for control design it is desired to find a low complexity model with a good fit around the desired crossover frequency. In the time domain it does not quite make sense to prefilter the data prior to the identification if a noise model is also estimated since the noise model will then undo the effect of the prefilter (Ljung 1999). In the frequency domain a prefiltering effect is obtained by only including a subset of frequencies in Ω_N corresponding to the frequency band where a good fit is desired. Here an estimate of a parametric noise model (to improve the variance properties) still makes sense if it can be expected that G_θ is flexible enough to provide an unbiased estimate (in the sense of (27)). It could also be expected that it suffices with a less complex noise model since only a part of the true noise transfer function needs to be accounted for.

3.5 *Closed loop*

Consider the case when the system operates in closed loop with a linear controller K. The input is then described by

$$u(t) = r(t) - K(q)y(t)$$

where $r(t)$ is an external reference signal. The output signal is just as before $y(t) = G_0(q)u(t) + H_0(q)e(t)$. Assume that U_k and Y_k are used in the criterion function (23). The dependence between U_k and Y_k make the criterion function converge to a limiting function different from (26). Straightforward calculations then reveal that

$$V_N(\theta,\lambda) \to \bar{V}(\theta,\lambda) \triangleq \int_\Omega \frac{|G_{0,\omega} - G_{\theta,\omega}|^2|R(\omega)|^2 + |1 + K_\omega G_{\theta,\omega}|^2\Phi_v(\omega)}{|1 + K_\omega G_{0,\omega}|^2|H_{\theta,\omega}|^2\lambda} + \log|H_{\theta,\omega}|^2\lambda\,\mathrm{d}W(\omega). \tag{35}$$

where $R(\omega)$ is the Fourier transform of $r(t)$. First consider the case when the frequencies ω_k are

[3] Minimized in the sense that $P_\theta - P_\theta^{opt}$ is a positive semidefinite matrix for all choices of fixed noise models.

uniform in $[-\pi, \pi]$. Hence the integral of the log term in (35) is zero. Also assume that the product $K(q)G_0(q)$ has no direct feed-trough term. Then $1 + K(q)G_0(q)$ is a monic transfer function which in turn implies that

$$\int_\Omega \frac{|1 + K_\omega G_{\theta,\omega}|^2 \Phi_v(\omega)}{|1 + K_\omega G_{0,\omega}|^2 |H_{\theta,\omega}|^2 \lambda} \mathrm{d}\omega$$

is minimized when the model equals the true system. Hence, whenever the correct noise model is used consistency is preserved also for the closed loop case. This property is analogous to the time domain case, see (Ljung 1999). If frequencies are drawn from a non-uniform distribution a slight modification is needed to the criterion function. By subtracting the term $\log|1 + K_{\omega_k} G_{\theta,\omega_k}|^2$ from the summand in (23) the estimator again become consistent. Note however that in this case we need to know the controller K and could thus have reformulated the closed loop problem to an open loop one, so called indirect identification.

3.6 *Some variants of the criterion*

Consider a system of the form

$$y(t) = \frac{b(q)}{a(q)}u(t) + \frac{1}{a(q)}e(t)$$

where $a(q)$ and $b(q)$ are polynomials. By disregarding λ and the log term the ML-criterion (23) simplifies into

$$\frac{1}{N}\sum_{k=1}^{N} |a(e^{i\omega_k}, \theta)Y_k - b(e^{i\omega_k}, \theta)U_k|^2 \qquad (36)$$

The criterion (36) is quadratic in the coefficients of the a and b polynomials and the minimizing parameters can be found by simple linear regression. Asymptotically this estimator equals the time domain ARX method (Ljung 1999). For continuous time models this estimator was proposed in (Levy 1959). If on the other hand

$$y(t) = \frac{b(q)}{a(q)}u(t) + e(t)$$

the use of the simple criterion (36) will lead to bias and an undesired weighting $|a(e^{i\omega})|^2$. A remedy to this problem was given in (Sanathanan and Koerner 1963) where an iterative procedure was suggested where at iteration m a weighted criterion was minimized

$$\frac{1}{N}\sum_{k=1}^{N} |a(e^{i\omega_k}, \theta)Y_k - b(e^{i\omega_k}, \theta)U_k|^2 W_k^{(m)} \qquad (37)$$

where $W_k^{(m)} = |\hat{a}^{(m-1)}(e^{i\omega_k})|^{-2}$ and $\hat{a}^{(m-1)}(z)$ is the estimated a polynomial from step $m-1$. The scheme is not guaranteed to converge to the true system (Whitfield 1987). A multivariable version is described in (de Callafon *et al.* 1996).

A number of subspace based methods have recently been developed for the frequency domain multivariable problem. A key feature is that they provide accurate state-space models for high order systems (McKelvey *et al.* 1996*a*, McKelvey *et al.* 1996*b*, Jacques *et al.* 1996, Liu *et al.* 1996, McKelvey 1997, Van Overschee and De Moor 1996). An accurate method for high order systems based on more standard curve fitting and polynomial matrix descriptions is given in (Bayard 1994). These simplified procedures are well suited to provide the optimal ML-estimator with good staring values for the iterative nonlinear optimization.

If the input signal also is corrupted with noise we obtain a dynamical errors-in-variables problem. Assuming the true system is given by $G_0(q) = b(q)/a(q)$, the input and output power spectrums of the noise, denoted by $\sigma_U^2(\omega)$ and $\sigma_Y^2(\omega)$ are known, the maximum-likelihood estimate is given by minimizing (Schoukens and Pintelon 1991)

$$\sum_{k=1}^{N} \frac{|b(e^{\omega_k}, \theta)U(\omega_k) - a(e^{\omega_k}, \theta)Y(\omega_k)|^2}{\sigma_U^2(\omega_k)|b(e^{\omega_k}, \theta)|^2 + \sigma_Y^2(\omega_k)|a(e^{\omega_k}, \theta)|^2} \qquad (38)$$

if the input and output noise sources are uncorrelated. For correlated noise a correction is subtracted from the denominator in (38). Recent extensions in (Schoukens *et al.* 1997*b*, Schoukens *et al.* 1999) show that the true input and output noise spectra can be exchanged by non-parametric estimates derived from four repeated experiments while still providing consistent estimates.

3.7 *Numerical issues for continuous time modeling*

It is well known that the parameter estimation problem is better conditioned for discrete time transfer functions than continuous time ones, since powers of $e^{i\omega}$ form a natural orthogonal basis (Bayard 1993). This property is amplified if the model order is high or if a large frequency band is used. By use of the bilinear transformation the continuous-time identification problem can be solved in the discrete domain without introducing any approximation errors but drastically improve the numerical conditioning (McKelvey *et al.* 1996*b*).

The bilinear transformation maps the complex values in the s domain to the z domain as

$$s \to \frac{2(z-1)}{T(z+1)}, \text{ with inverse } z \to \frac{2+sT}{2-sT}$$

The parameter T is a parameter which the user is free to specify under constraint that $2/T$ is not a pole of the continuous-time system (Al-Saggaf and Franklin 1988), and can be seen as a sort of sampling period.

If the continuous-time transfer function is given by $G_c(s)$ the bilinear transformation gives the discrete time transfer function

$$G(z) = G_c\left(\frac{2(z-1)}{T(z+1)}\right).$$

Stability and minimum-phase properties are preserved by the transformation.

An important feature of the bilinear transformation is that the frequency response is invariant if we pre-warp the frequency scale. Let the continuous-time transfer function be evaluated at $i\omega_k$ and let the bilinearly transformed discrete time transfer function be evaluated at $e^{i\omega_k^d}$. Then it holds that (Åström and Wittenmark 1984)

$$G_c(i\omega_k) = G_c\left(\frac{2(e^{i\omega_k^d}-1)}{T(e^{i\omega_k^d}+1)}\right) = G(e^{i\omega_k^d})$$

if $\tan(\omega_k^d/2) = \omega_k T/2$. Assume we have samples U_k and Y_k from a continuous time system at frequencies ω_k. Then they also are samples of the corresponding bilinearly transformed discrete time system at new frequencies ω_k^d

$$\omega_k^d = 2\,\mathrm{atan}(\frac{\omega_k T}{2}), \quad k = 1,\ldots,M, \tag{39}$$

where atan denotes the inverse of tan. After estimation of a discrete time transfer function $\hat{G}(z)$, the sought continuous time transfer function is obtained through the inverse map

$$\hat{G}_c(s) = \hat{G}\left(\frac{2+sT}{2-sT}\right).$$

4. CONCLUSIONS

It has been demonstrated that the frequency domain ML method asymptotically equals the time domain version in the case where the frequency data is obtained by DFT. An important question to pose then is in which cases can it be advantageous to use the frequency domain method? The following points can shed some light on the answer.

Partial modeling Often it is enough to find a model which accurately describes the true system in a limited frequency band. A low order model could thus be sufficient. In the frequency domain this is simply accomplished by only including the desired frequencies in the set Ω_N. In the time domain the raw data needs to be prefiltered and the influence of the initial conditions of the filters might distort the end results. Also it makes no sense to estimate a noise model in the time domain if a prefilter has been applied.

Continuous time systems If the experimental conditions are such that a band limited input can be used then the modeling in the frequency domain is straightforward as demonstrated in Section 2. Direct identification of continuous time models in the time domain requires difficult user choices of how to approximate higher order derivatives from sampled data. Systems with time delay are also straightforward to handle by using $G_c(s)e^{-\tau s}$ as the model structure where τ is the time delay.

Periodic input When a system is excited with a periodic input applied for a sufficiently long time prior to the actual measurement interval, then the effect of the initial condition is diminished even for a finite measurement interval. Hence the transient effect need not be included in the model. Periodic excitation also make it possible to first estimate a non-parametric noise model which can be used as fixed noise model in the ML-criterion, see (Schoukens *et al.* 1997*b*).

Merging Data If data is obtained by different experiments all frequency data can be merged into one estimation data set. By performing several experiments using different sampling rates a single wide band frequency domain data set can efficiently be assembled in order to estimate a continuous time model valid over a large frequency band (Schoukens *et al.* 1994).

ACKNOWLEDGMENT

I would like to thank Lennart Ljung, Johan Schoukens, Rik Pintelon and Petre Stoica who at various times have generously shared with me their expertise on this subject. Thanks also to Brett Ninness and Will Heath for providing valuable comments during the preparation of this manuscript.

5. REFERENCES

Al-Saggaf, U. M. and G. F. Franklin (1988). Model reduction via balanced realizations: An extension and frequency weighting techniques. *IEEE Trans. on Automatic Control* **33**(7), 687–692.

Åström, K. J. (1970). *Introduction to Stochastic Control Theory.* Academic Press. New York.

Åström, K. J. (1980). Maximum likelihood and prediction error methods. *Automatica* **16**, 551–574.

Åström, K. J. and B. Wittenmark (1984). *Computer Controlled Systems.* Prentice-Hall. Englewood Cliffs, New Jersey.

Bayard, D. S. (1993). Multivariable state-space identification in the delta and shift operators: Algorithms and experimental results. In: *Proc. American Control Conference, San Francisco, CA, June.* pp. 3038–3042.

Bayard, D. S. (1994). High-order multivariable transfer function curve fitting: Algorithms, sparse matrix methods and experimental results. *Automatica* **30**(9), 1439–1444.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* McGraw-Hill Inc., New York.

de Callafon, R. A., D. de Roover and P. M. J. Van den Hof (1996). Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions. In: *Proc. 35th IEEE Conference on Decision and Control.* Kobe, Japan.

Dennis, J. E. and R. B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall. Englewood Cliffs, New Jersey.

Forssell, U., F. Gustafsson and T. McKelvey (1999). Time-domain identification of dynamic errors-in-variables systems using periodic excitation signals. In: *Preprints of the 14th World Congress of IFAC, Beijing, P. R. China* (H.F. Chen and B. Wahlberg, Eds.). Vol. H. Elsevier Science. pp. 421–426.

Giannakis, G. B. (1995). Polyspectral and cyclostationary approaches for identification of closed-loop systems. *IEEE Trans. on Automatic Control* **AC-40**, 882–885.

Hannan, E. G. (1970). *Multiple Time Series.* John Wiley and Sons, Inc.

Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems.* Wiley.

Jacques, R. N., K. Liu and D. W. Miller (1996). Identification of highly accurate low order statespace models in the frequency domain. *Signal Processing, EURASIP* **52**(2), 195–208.

Kailath, T. (1980). *Linear Systems.* Prentice-Hall. Englewood Cliffs, New Jersey.

Kendall, M. G. and A. Stuart (1967). *The Advanced Theory of Statistics.* Vol. 2. second ed.. Griffin, London.

Levy, E. C. (1959). Complex curve fitting. *IRE Trans. on Automatic Control* **AC-4**, 37–44.

Liu, K., R. N. Jacques and D. W. Miller (1996). Frequency domain structural system identification by observability range space extraction. *ASME J. Dyn. Syst., Meas. and Control.*

Ljung, L. (1994). Building models from frequency domain data. In: *IMA Workshop on Adaptive Control and Signal Processing.* Minneapolis, Minnesota.

Ljung, L. (1999). *System Identification: Theory for the User.* second ed.. Prentice-Hall. Englewood Cliffs, New Jersey.

Ljung, L. and K. Glover (1981). Frequency domain versus time domain methods in system identification. *Automatica* **17**(1), 71–86.

McKelvey, T. (1996). Periodic excitation for identification of dynamic errors-in-variables systems operating in closed loop. In: *Proc. 13th World Congress of International Federation of Automatic Control.* Vol. J. San Francisco, California. pp. 155–160.

McKelvey, T. (1997). Frequency domain system identification with instrumental variable based subspace algorithm. In: *Proc. 16th Biennial Conference on Mechanical Vibration and Noise, DETC'97, ASME.* Sacramento, California, USA. pp. VIB–4252.

McKelvey, T. and H. Akçay (1995). Subspace based system identification with periodic excitation signals. *Systems & Control Letters* **26**(5), 349–361.

McKelvey, T. and L. Ljung (1997). Frequency domain maximum likelihood identification. In: *Proc. of the 11th IFAC Symposium on System Identification.* Vol. 4. Fukuoka, Japan. pp. 1741–1746.

McKelvey, T., H. Akçay and L. Ljung (1996*a*). Subspace-based identification of infinite-dimensional multivariable systems from frequency-response data. *Automatica* **32**(6), 885–902.

McKelvey, T., H. Akçay and L. Ljung (1996*b*). Subspace-based multivariable system identification from frequency response data. *IEEE Trans. on Automatic Control* **41**(7), 960–979.

Pintelon, R., J. Schoukens and G. Vandersteen (1997). Frequency domain system identification using arbitrary signals. *IEEE Trans. on Automatic Control* **42**(12), 1717–1720.

Pintelon, R., P. Guillaume, Y. Rolain, J. Schoukens and H. Van Hamme (1994). Parametric identification of transfer functions in the frequency domain – A survey. *IEEE Trans. of Automatic Control* **94**(11), 2245–2260.

Robinson, P. M. (1976). The estimation of linear differential equations with constant coefficients. *Econometrica* **44**(4), 751–764.

Rudin, W. (1987). *Principles of Mathematical Analysis.* third ed.. McGraw-Hill.

Sanathanan, C. K. and J. Koerner (1963). Transfer function synthesis as a ratio of two complex polynomials. *IEEE Trans. on Automatic Control* **8**, 56–58.

Schoukens, J. and R. Pintelon (1991). *Identification of Linear Systems: a Practical Guideline to Accurate Modeling.* Pergamon Press, London, UK.

Schoukens, J., G. Vandersteen, R. Pintelon and P. Guillaume (1999). Frequency-domain identification of linear systems using arbitrary excitations and a nonparametric noise model. *IEEE Trans. on Automatic Control* **AC-44**(2), 343–347.

Schoukens, J., R. Pintelon and H. Van Hamme (1994). Identification of linear dynamic systems using piecewise constant excitations: Use, misuse and alternatives. *Automatica* **30**(7), 1153–1169.

Schoukens, J., R. Pintelon and Y. Rolain (1997*a*). Study of conditional ML estimators in time and frequency domain system identification. In: *Proc. European Control Conference.* Brussels, Belgium.

Schoukens, J., R. Pintelon, G. Vandersteen and P. Guillaume (1997*b*). Frequency domain identification using non-parametric noise models estimated from a small number of data sets. *Automatica* **33**(6), 1073–1086.

Söderström, T. and P. Stoica (1989). *System Identification.* Prentice-Hall International. Hemel Hempstead, Hertfordshire.

Tugnait, J. K. (1998). Identification of multivariable stochastic linear systems via polyspectral analysis given noisy input-output time-

domain data. *IEEE Trans. on Automatic Control* **43**(8), 1084–1100.
Unbehauen, H. and G. P. Rau (1990). Continuous-time approaches to system identification – A survey. *Automatica* **26**(1), 23–35.
Van Overschee, P. and B. De Moor (1996). Continuous-time frequency domain subspace system identification. *Signal Processing, EURASIP* **52**(2), 179–194.
Vandersteen, G. and J. Schoukens (1999). Measurement and identification of nonlinear systems consisting of linear dynamic blocks and one static nonlinearity. *IEEE Trans. on Automatic Control* **44**(6), 1266–1271.
Vandersteen, G., Y. Rolain and J. Schoukens (1997). Non-parametric estimation of the frequency-response functions of the linear blocks of a Wiener-Hammerstein model. *Automatica* **33**(7), 1351–1355.
Whitfield, A. H. (1986). Transfer function synthesis using frequency response data. *Int. J. Control* **43**(5), 1413–1426.
Whitfield, A. H. (1987). Asymptotic behaviour of transfer function synthesis methods. *Int. J. Control* **45**(3), 1083–1092.
Whittle, P. (1951). *Hypothesis testing in time series analysis.* Thesis. Uppsala University, Almqvist and Wiksell, Uppsala. Hafner, New York.
Young, P. (1981). Parameter estimation for continous-time models – A survey. *Automatica* **17**(1), 23–39.

Appendix A. PROOFS

A.1 *Proof of* (8)-(9)

Consider the state-space equations (6). The state at time $t = N$ is given by

$$x(N) = A^N x(0) + \sum_{k=0}^{N-1} A^k Bu(N-1-k)$$

Now with the initial condition $x(0) = x_p$ for (6) where x_p is given by (9) then $x(N) = x_p$. Let $x^{\text{per}}(t)$ denote the state evolution of (6) when $x(0) = x_p$ and let $y^{\text{per}}(t) = Cx^{\text{per}}(t)$ denote the corresponding output. Define the transient output as $y^{\text{tra}}(t) = CA^t(x_0 - x_p)$. It is straightforward to verify that the original output $y(t)$ from the state-space system (6) with initial condition $x(0) = x_0$ can be decomposed as $y(t) = y^{\text{per}}(t) + y^{\text{tra}}(t)$. The DFT of the transient output is simply a finite geometric sum

$$\begin{aligned} Y_N^{\text{tra}}(\omega_k) &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} CA^t(x_0 - x_p)e^{-i\omega_k t} \\ &= \frac{1}{\sqrt{N}} e^{i\omega_k} C(e^{i\omega_k} I - A)^{-1}(I - A^N)(x_0 - x_p) \end{aligned}$$

where we used the property that $e^{i\omega_k N} = 1$. Denote the DFT of $x^{\text{per}}(t)$ as $X_N^{\text{per}}(\omega)$. The DFT of $x^{\text{per}}(t+1)$ at $\omega_k = \frac{2\pi k}{N}$ is given by

$$\frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x^{\text{per}}(t+1)e^{-i\omega_k t} =$$

$$\frac{1}{\sqrt{N}} e^{i\omega_k} \sum_{t=0}^{N-1} x^{\text{per}}(t+1)e^{-i\omega_k (t+1)} = e^{i\omega_k} X_N^{\text{per}}(\omega_k)$$

where the last equality follows from $x^{\text{per}}(N) = x^{\text{per}}(0)$. By taking the DFT of both sides of the state-space equations governing $x^{\text{per}}(t)$ we have

$$e^{i\omega_k} X_N^{\text{per}}(\omega_k) = AX_N^{\text{per}}(\omega_k) + BU_N(\omega_k)$$

By eliminating the state $X_N^{\text{per}}(\omega_k)$ we finally obtain for the DFT of $y^{\text{per}}(t)$

$$Y_N^{\text{per}}(\omega_k) = C(e^{i\omega_k} I - A)^{-1} BU_N(\omega_k)$$

which concludes the proof. □

A.2 *Proof of* (13)-(14)

This proof parallels the discrete time case with some obvious modifications. In the proof we assume $h = 1$, N is an odd integer and define $n_f \triangleq \frac{N-1}{2}$. The proof for the general case is analogous.

The evolution of the continuous time state variable is given by (Kailath 1980)

$$x(t) = e^{A_c t} x(0) + \int_{\tau=0}^{t} e^{A_c \tau} B_c u(t - \tau) \qquad \text{(A.1)}$$

Now if we assume that $x(0)$ in (10) is given by x_p defined by (14), straight forward calculations yields $x(N) = x(0) = x_p$. Define $x^{\text{per}}(t)$ to be the state response when $x(0)$ is given by x_p and similarly define $y^{\text{per}}(t) = C_c x^{\text{per}}(t)$. By introducing $y^{\text{tra}}(t) \triangleq C_c e^{A_c t}(x_0 - x_p)$ it is easy to verify that the measured output for $t = [0, N]$ is given by $y(t) = y^{\text{per}}(t) + y^{\text{tra}}(t)$ Obviously, the state x^{per} is also an N-periodic signal. By explicitly using the structure of the input given by (12) we obtain

$$x_p = \sum_{k=-n_f}^{n_f} (i\omega_k I - A_c)^{-1} B_c f_k$$

Straightforward calculations using the expression (A.1) with $x(0) = x_p$ yields

$$x^{\text{per}}(t) = \sum_{k=-n_f}^{n_f} (i\omega_k I - A_c)^{-1} B_c f_k e^{i\omega_k t}$$

and the DFT of the state is simply

$$X_N^{\text{per}}(\omega_k) = (i\omega_k I - A_c)^{-1} B_c U_N(\omega_k)$$

For the output we simply have $Y_N^{\text{per}}(\omega_k) = C_c X^{\text{per}}(\omega_k)$. As in the discrete time case the DFT of the transient output is a finite geometric sum

$$Y_N^{\text{tra}}(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} C_c e^{A_c t}(x_0 - x_p)e^{-i\omega_k t} =$$

$$\frac{1}{\sqrt{N}} e^{i\omega_k} C_c (e^{i\omega_k} I - e^{A_c})^{-1}(I - e^{A_c N})(x_0 - x_p)$$

□