Circuits Systems Signal Processing
Vol. 21, No. 1, 2002, Pp. 39–55

Frequency Domain Identification Methods*

Tomas McKelvey[1]

Abstract. Methods for estimating linear dynamical models from frequency data are studied, including the properties of frequency domain data generated by the discrete Fourier transform. The stochastic characteristics of the frequency domain data lead to a maximum likelihood (ML) formulation of the frequency domain estimation problem. Both discrete-time and continuous time models are discussed. Consistency and variance of the ML estimate are described, and the connection with simpler frequency domain estimation schemes as well as the time domain ML method is pointed out.
Key words: Estimation, system identification, linear systems, frequency functions, discrete Fourier transform, maximum likelihood, least squares, continuous-time systems, discrete-time systems.

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 partially unknown and when it is infeasible to theoretically derive a model from first principles. The availability of accurate models is important to derive high-performing solutions, e.g., for model-based control design or model-based signal processing.

Most measurements originating from real-world devices intrinsically belong to the time domain, i.e., a sequence of real values each tagged with a time stamp. Consequently, most system identification methods and the theory developed

* Received February 18, 2001. Supported by the Swedish Foundation for International Cooperation in Research and Higher Education (STINT). This work was partly completed while the author was visiting the Department of Electrical and Computer Engineering, University of Newcastle, Australia.

[1] Department of Signals and Systems, Chalmers University of Technology, SE-412 96, Göteborg, Sweden. E-mail: mckelvey@s2.chalmers.se

around them deals with how to determine models from such time domain measurements [9], [18]. However, in some application areas such as vibration analysis, it is common to subject the raw data to the Fourier transform before fitting them to parametric models. The very basic and old technique of frequency analysis is the classical example. In frequency analysis the linear dynamical system is excited by a pure sinusoidal signal. When the output has settled to a stationary sinusoidal signal, the complex value of the transfer function at the specific excitation frequency is determined by comparing the amplitudes and phases of the input and output signals, respectively. Repeating the experiment for many frequencies yields a nonparametric estimate of the system's frequency response. In a second step, a parametrized transfer function model can then be fitted to the transfer function data using some complex curve fitting technique [7]. Techniques to fit parametric models to frequency domain data is the scope of this paper. An early reference for the time series problem is by Whittle [21]. During the last decade the frequency domain techniques have received more attention in the system identification literature [16], [13], [9].

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 bandlimited excitation signals is employed. This is a great advantage in contrast to 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. A continuous-time system with a time delay is also rather difficult to model in the time domain because it cannot be described by a finite-dimensional system of ordinary differential equations. However, in the frequency domain, a nice finite-dimensional parametric description exists that lends itself to identification using parametric methods.

1.1. Outline

The properties of the discrete Fourier transform (DFT) transformed data is studied in Section 2. In particular expressions valid for finite data lengths and the stochastic properties of the transformed noise are covered. Section 3 presents the frequency domain maximum likelihood method and some asymptotic results regarding consistency, bias, and variance. The section is summarized by pointing out the relations to the time domain counterpart and some simpler frequency domain techniques.

2. From time to frequency domain

The exact relationships between the DFT of the input and outputs and the frequency response function of a finite-dimensional linear system under rather general excitation conditions will be discussed next. If an infinite amount of time (and data) is at our disposal, it is well known that the Fourier transform of the output is exactly the frequency response function multiplied by the transform of the input signal. In the finite data case, an extra term appears which accounts for the history of the system prior to the measurement interval. As the number of data samples N tends to infinity, the extra term goes to zero at a rate proportional to $\frac{1}{\sqrt{N}}$. In the 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 [14].

2.1. *Discrete Fourier transform*

Assume that a signal $s(t)$ is sampled at N equidistant time instances $t = kT_s$, $k = 0, 1, \dots, N-1$, where T_s is the sampling interval. The N-point DFT of the set $\{s(kT_s)\}_{k=0}^{N-1}$ is defined as

$$S_N(\omega) \triangleq \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s(kT_s)e^{-i\omega k}, \tag{1}$$

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

2.2. *Discrete-time systems*

The DFT relation for a discrete-time system is presented in this section. 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

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t), \end{aligned} \tag{2}$$

where $u(t)$ is the input, $y(t)$ is the output, $x(t)$ is a state vector of length n, A is an $n \times n$ real matrix, and B and C^T are vectors of length n. If the realization is of minimal order, n is the McMillan degree of the linear system [5]. The transfer function is given by

$$G(z) = \sum_{k=1}^{\infty} CA^{k-1}Bz^{-k} = C(zI - A)^{-1}B \triangleq \frac{\sum_{k=1}^{n} b_k z^{-k}}{1 + \sum_{k=1}^{n} a_k z^{-k}} \triangleq \frac{b(z)}{a(z)}. \tag{3}$$

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 (2) and assume that 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 $k = 0, 1, \ldots, N-1$.

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:

$$Y_N(\omega_k) = G(e^{i\omega_k})U_N(\omega_k) + T(e^{i\omega_k})\frac{1}{\sqrt{N}}, \tag{4}$$

where

$$\begin{aligned}
G(z) &= C(zI - A)^{-1}B = \frac{b(z)}{a(z)} \\
T(z) &= zC(zI - A)^{-1}(I - A^N)(x_0 - x_p) = \frac{\sum_{k=1}^{n} t_k z^{-k}}{a(z)} = \frac{t(z)}{a(z)} \\
x_p &= (I - A^N)^{-1}\sum_{t=0}^{N-1} A^t B u(N - 1 - t).
\end{aligned} \tag{5}$$

The proof is straightforward and is based on splitting the output into a sum of two signals, $y(t) = y^{\text{per}} + y^{\text{tra}}$. The signal y^{per} originates from an assumed system that operates in an N-periodic fashion, i.e., $u(t+N) = u(t)$ and $y^{\text{per}}(t) = y^{\text{per}}(t+N)$ for all t. This system is simply constructed by letting $x(0) = x_p$. The second signal is the transient response from $x_0 - x_p$, the difference in initial condition between the true system and the assumed periodic one. The full proof can be found in [11].

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 equivialently, all poles lie strictly inside the unit circle. The explicit form of $T(z)$ enables us to estimate it along with the transfer function $G(z)$. This could be beneficial when the data record is short and thereby reduces the bias in the estimate of G. The price is n more parameters to estimate and an increased variance. Furthermore, note that $G(z)$ and $T(z)$ share the same pole polynomial $a(z)$ and only n extra parameters need to be determined, the n coefficients t_k of the numerator polynomial $t(z)$. However, if the input excitation is such that $U_N(\omega_k)$ is constant for all frequencies, it is not possible to distinguish between $b(z)$ and $t(z)$.

2.3. Continuous-time systems

The output $y(t)$ of a finite-dimensional continuous-time system can be described as the solution to a system of first-order differential equations

$$\begin{aligned} \dot{x}(t) &= A_c x(t) + B_c u(t) \\ y(t) &= C_c x(t), \end{aligned} \tag{6}$$

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

$$G_c(s) = \sum_{k=1}^{\infty} C_c A_c^{k-1} B_c s^{-k} = C_c(sI - A_c)^{-1} B_c \triangleq \frac{\sum_{k=1}^{n} b_k^c s^{-k}}{1 + \sum_{k=1}^{n} a_k^c s^{-k}} \triangleq \frac{b_c(s)}{a_c(s)} \tag{7}$$

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, which we assume is known, excites the continuous-time system. If the input is piecewise constant between the sampling instances, then the continuous system has a discrete-time representation that exactly describes the output signal at the sample points, and hence the expressions (4) and (5) hold. The mapping takes a continuous-time system into a discrete one is called zero order hold (ZOH) sampling [9]. 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 [2]. The method of first identifying a discrete-time model and then employing the inverse ZOH mapping consequently might fail. The correct approach is to parameterize the model using a continuous-time system and then, via the ZOH mapping, derive the discrete-time model that is matched to the sampled data [9].

A second possibility, which is well suited for the frequency domain, is to excite the continuous-time system using a bandlimited input with a zero spectrum for all frequencies on and above the Nyquist frequency ($\frac{\pi}{T_s}$). In order to use a similar technique as in the discrete-time case, we assume the signal to have an NT_s-periodic continuation outside the measurement interval, i.e., $u(t) = u(t + NT_s)$ for all t. Fourier analysis then tells us that all such signals can be described as

$$u(t) = \begin{cases} \sum_{k=-\frac{N-1}{2}}^{\frac{N-1}{2}} f_k e^{i\frac{2\pi k}{NT_s}t}, & \text{for } N \text{ odd} \\ \sum_{k=-\left(\frac{N}{2}-1\right)}^{\frac{N}{2}-1} f_k e^{i\frac{2\pi k}{NT_s}t}, & \text{for } N \text{ even,} \end{cases} \tag{8}$$

where $u(t)$ is real if $f_k = f_{-k}^*$. Clearly, all such real signals are composed of a sum of sinusoids with (normalized) frequencies constrained to belong to the finite set $\{\omega_k \mid \omega_k = \frac{2\pi k}{N} < \pi,\ k = 0, 1, \dots\}$. This type of excitation signal is known as multi-sine excitation [16] and is a generalization of the classical frequency analysis which uses one frequency at a time.

Assume that a continuous-time system (6) from $t = 0$ to $t = NT_s$ is excited by an input signal of the form (8). The input and output signals are sampled at N points with sampling interval T_s. No assumptions are made about the character of the input for $t < 0$ and $t > N$. Also assume $\det(i\frac{\omega_k}{T_s}I - A_c) \neq 0$ for all ω_k. Then the following equation holds:

$$Y_N(\omega_k) = G_c\left(i\frac{\omega_k}{T_s}\right)U_N(\omega_k) + T_c(e^{i\omega_k})\frac{1}{\sqrt{N}}, \tag{9}$$

where

$$\begin{aligned}
G_c(s) &= C_c(sI - A_c)^{-1}B_c = \frac{b_c(s)}{a_c(s)} \\
T_c(z) &= zC_c(zI - e^{A_cT_s})^{-1}(I - e^{A_cT_sN})(x_0 - x_p) \triangleq \frac{\sum_{k=1}^{n} t_k^c z^{-k}}{a_{cd}(z)} \\
x_p &= (I - e^{A_cT_sN})^{-1}\int_{\tau=0}^{NT_s} e^{A_c\tau}B_c u(NT_s - \tau)\,d\tau
\end{aligned} \tag{10}$$

and e^{A_c} is the standard matrix exponential [5]. A derivation of the result above can be found in [11]. The relation (9) is quite similar to the discrete-time version (4) with one important exception. The transient term $T_c(z)$ is a discrete-time transfer function. However, just as in the discrete-time case, $T_c(z)$ is can be finitely parameterized.

2.4. *Discussion*

In both domains we have left out systems that 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 (3) with (7), we conclude that the only difference between the frequency response function for discrete-time systems and continuous-time systems is the argument, i.e., $e^{i\omega}$ or $i\omega$. However, the transient terms differ. The extension to the multivariable case is straightforward because the derivations make use of a state-space representation. In the general case when the system is infinite dimensional and stable, equations (4) and (9) still hold. In this case the transient transfer functions $T(z)$ do not admit any finite representation but have an upper bound [9].

2.5. *The noise*

To obtain high-quality estimation results, it is vital to also consider the errors that inevitably are present in the measured output signal. We assume that the errors enter additive to the measurement

$$y_m(t) = y(t) + v(t), \tag{11}$$

where $y_m(t)$ is the measured signal, $y(t)$ is the noise-free output from the system, and $v(t)$ is the noise. Here we will make a fairly standard assumption that the noise $v(t)$ can be represented by a white noise signal with a normal distribution filtered through a stable and inversely stable discrete-time linear system, i.e.,

$$v(t) = e(t) + \sum_{k=1}^{\infty} h(k)e(t-k) = H(q)e(t), \tag{12}$$

where the input $e(t)$ is a realization of a zero mean random variable with a normal distribution and variance λ. The noise transfer function is denoted by $H(z) = 1 + \sum_{k=1}^{\infty} h(k)z^{-k}$. By making the noise transfer function monic (i.e., $H(\infty) = 1$), the size of the noise signal is uniquely determined by the variance λ.

A complex (scalar) random variable x is called *complex normal* [3] $x \in N^c(m_x, \lambda_x)$ if the real and imaginary parts of x both have a normal distribution and $m_x = \mathbf{E}\{x\}$, $\mathbf{E}\{|x - m_x|^2\} = \lambda_x$, and $\mathbf{E}\{(x - m_x)^2\} = 0$. Here $\mathbf{E}\{\cdot\}$ denotes expectation. The probability density function for a complex normal variable x is given by [3]

$$p_X(x) = \frac{1}{\pi\lambda} \exp\left(\frac{|x - m_x|^2}{\lambda}\right). \tag{13}$$

Denote by $E_N(\omega_k)$ the DFT of the noise signal $\{e(t)\}_{t=0}^{N-1}$. It is well known that $E_N(\omega_k)$ is a random variable with a complex normal distribution [3]:

$$E_N(\omega_k) \in N^c(0, \lambda), \tag{14}$$

where $\omega_k \in \{\frac{2\pi k}{N}, k = 1, \dots, N-1\}$. For $\omega_k = 0$ (and possibly π), $E_N(0)$ is real, zero mean, and normal distributed with variance λ. The variables are independent between different frequencies.

Using equation (4) and (12), the DFT of $v(t)$ is conveniently described by

$$V_N(\omega_k) = H(e^{i\omega_k})E_N(\omega_k) + T_H(e^{i\omega_k})\frac{1}{\sqrt{N}}, \tag{15}$$

where the last frequency function $T_H(z)$ is due to the "unmatched" initial condition of the noise filter and is a linear function of the innovations $e(t)$ for $t < N$. This implies it has zero mean for all frequencies. When $H(z)$ 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 know that there exists a uniform bound

$$C_H = \max_k |T_H(e^{i\omega_k})|^2, \tag{16}$$

which exists because the noise filter H is assumed strictly stable.

Putting it all together, we have

$$\mathbf{E}\{V_N(\omega_k)\} = 0, \qquad \forall \omega_k \tag{17}$$

$$\mathbf{E}\left\{V_N(\omega_k)V_N^*(\omega_s)\right\} = \begin{cases} |H(e^{i\omega_k})|^2\lambda + \frac{\xi_1(\omega_k)}{N}, & \omega_k = \omega_s \\ \frac{\xi_2(\omega_k,\omega_s)}{N}, & \omega_k \neq \omega_s, \end{cases} \tag{18}$$

where

$$|\xi_2(\omega_k, \omega_s)| \leq |\xi_1(\omega_k)| \leq C_H.$$

Asymptotically, as the number of data points tends to infinity, the term $T_H(e^{i\omega_k})\frac{1}{\sqrt{N}}$ tends to zero. Then $V_N(\omega_k)$ becomes complex normal distributed with zero mean and variance $|H(e^{i\omega_k})|^2\lambda$, and the correlation between different frequencies is zero. For a much more thorough treatment and relaxed assumptions, we refer to [3], [9].

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 at the sampling instances. The validity of this approach can be argued as follows. Consider a sampled continuous-time noise signal described by a stochastic differential equation. The first- and second-order moments of the signal at the sampling instances can equally be modeled by a discrete-time stochastic model of the same order as the continuous-time stochastic model. See [1] for the details.

3. Identification methods

Based on the properties of the frequency domain data derived in the previous section, we first state the system identification problem formulation and then discuss some identification methods with an emphasis on the maximum likelihood method.

3.1. Problem formulation

Assume that a linear system, in the frequency domain, is described as

$$Y(\omega) = G_0(e^{i\omega})U(\omega) + H_0(e^{i\omega})E_0(\omega), \tag{19}$$

where Y and U are the (weak) limits of the Fourier transform of the output and input, respectively, and E_0 is the frequency domain innovations, which are zero mean complex normal random variables with variance λ_0 and independent between different frequencies and of $U(\omega)$. 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. Asymptotically, in N, equation (19) follows from (4) and (15)–(18). For a continuous-time system formulation, just exchange the argument $e^{i\omega}$ for $i\omega$ in G_0.

Assume that the relation (19) can be sampled at a sequence of 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 $Y_k = Y(\omega_k)$. If $U_k = 1$ for all frequencies, then Y_k is simply a noisy measurement of the system transfer function $G_0(e^{i\omega_k})$. The aim is to find a model of (19), and to do so we construct a model set by using a parameterized model structure

$$Y(\omega) = G_\theta(e^{i\omega})U(\omega) + H_\theta(e^{i\omega})E(\omega), \tag{20}$$

where $G_\theta(z)$ and $H_\theta(z)$, respectively are the rational transfer functions of the system and noise which are parameterized by a real-valued vector θ. Let the compact set $D_{\mathcal{M}}$ denote the set of valid parameters. For all $\theta \in D_{\mathcal{M}}$ we assume $H_\theta(z)$ and $G_\theta(z)$ are Lipschitz continuous and bounded functions for $|z| = 1$ and we assume $H_\theta(z)$ is a stable and inversely stable monic transfer function. Given the parameterized model class and data Z^N, the estimate of the system is found by parametric optimization of some criterion function

$$\hat{\theta} = \arg\min_\theta V_N(\theta, Z^N), \tag{21}$$

where V_N is a function that provides a metric on how to optimally fit the model (20) to the given data Z^N.

3.2. Maximum likelihood method

Most frequency domain identification techniques do not explicitly model the noise properties with a parametric model. Instead consistent estimates are obtained by using smoothed spectral estimates [19] or instrumental variable (IV)-type methods [10] or by using a nonparametric noise model which is either known or estimated a priori [16], [17]. Disregarding the correct noise properties leads to an increased variance of the estimates. Here we will focus on the frequency domain maximum likelihood (ML) estimator which explicitly models the unknown noise transfer function. The ML method is frequently used for many estimation problems [6]. 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.

Recall the postulated identification setup given by equation (19). Accordingly, the samples Y_k of the output Fourier transform have a complex normal distribution [3] 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

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

Here we assume that the frequencies 0 and π are not part of the set Ω_N because at these frequencies $Y(\omega)$ is real valued and has a slightly different probability density function. However, an inclusion of these points would not change any asymptotic properties. Because the measurements are independent of each other, the joint probability density function for Z^N is the product of the individual functions (22)

$$p_{Y^N}(y_1, \dots, y_N) = \prod_{k=1}^{N} p_{Y_k}(y_k).$$

A parameterized likelihood function is obtained if the measurements Z^N are considered fixed and the true (but unknown) transfer functions G_0, H_0, and variance λ_0 are exchanged for the parameterized transfer functions G_θ, H_θ, and variance λ. By taking the negative logarithm of the parameterized likelihood function and removing terms that do not depend on the parameters (θ and λ), we obtain [8], [12]

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

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 *ML estimate*,

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

To reduce the number of variables, we can, for each θ, analytically derive the optimal value of λ:

$$\hat{\lambda}_N(\theta) = \frac{1}{N}\sum_{k=1}^{N}\frac{|Y_k - G_{\theta,\omega_k}U_k|^2}{|H_{\theta,\omega_k}|^2}. \tag{25}$$

The resulting concentrated estimator is obtained by inserting (25) into (23) and removing the constant term

$$\begin{aligned}\hat{\theta}_N &= \arg\min_{\theta} W_N(\theta)\\ W_N(\theta) &= \log\left(\frac{1}{N}\sum_{k=1}^{N}\frac{|Y_k - G_{\theta,\omega_k}U_k|^2}{|H_{\theta,\omega_k}|^2}\right) + \frac{1}{N}\sum_{k=1}^{N}\log|H_{\theta,\omega_k}|^2\end{aligned} \tag{26}$$

and $\hat{\lambda}_N = \hat{\lambda}_N(\hat{\theta}_N)$ defined by (25).

In general, the minimization of (23) or (26) cannot be performed analytically, and an iterative optimization strategy needs to be employed. Often Newton-type methods [4] perform well for this class of problems if started not too far away from the optimum.

3.2.1. *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

$$W_N(\omega) = \frac{|\{k \mid \omega_k < \omega, \omega_k \in \Omega_N\}|}{N}, \tag{27}$$

where $|S|$ denotes the cardinality of the set S. In probability theory, $W_N(\omega)$ corresponds to a distribution function. We assume that the sequence of sample frequencies is such that as $N \to \infty$ the function $W_N(\omega)$ converges to a function $W(\omega)$ in all points of continuity of $W(\omega)$. By using a Stieltjes integral notation, an infinite sum can be written as an integral. In our case we have that as $N \to \infty$ (see [12] for the details),

$$\begin{aligned} V_N(\theta,\lambda) \to \bar{V}(\theta,\lambda) \triangleq \int_\Omega \Bigg[& \frac{|G_{0,\omega} - G_{\theta,\omega}|^2 \Phi_u(\omega) + \Phi_v(\omega)}{|H_{\theta,\omega}|^2 \lambda} \\ & + \log(|H_{\theta,\omega}|^2 \lambda) \Bigg] \mathrm{d}W(\omega), \end{aligned} \tag{28}$$

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 (28) is uniform in $D_{\mathcal{M}}$ with probability one. It then follows that the estimate $\hat{\theta}_N, \hat{\lambda}_N$ converges to values that minimize $\bar{V}(\theta,\lambda)$.

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

$$\int_\Omega |G_{0,\omega} - G_{\theta_*,\omega}|^2 \Phi_u(\omega)\, \mathrm{d}W(\omega) = 0 \tag{29}$$

$$\int_\Omega \Big| |H_{0,\omega}|^2 - |H_{\theta_*,\omega}|^2 \Big|^2 \mathrm{d}W(\omega) = 0. \tag{30}$$

Then it can be shown that $\bar{V}(\theta,\lambda)$ in (28) 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 (29) and (30)).

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

$$\int_\Omega \frac{|G_{0,\omega} - G_{\theta,\omega}|^2 \Phi_u(\omega)}{|H_\omega|^2} \mathrm{d}W(\omega). \tag{31}$$

If the true system G_0 is not in the model class, then an approximate model will result. In this case, the estimate will be the model that in a weighted mean-square sense best approximates the transfer function of the system. As shown in (31), 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 integral simplifies into $\int_\Omega(\cdot)w(\omega)\,d\omega$, where $w(\omega) = \frac{d}{d\omega}W(\omega)$. In this case, $w(\omega)$ acts as a standard weighting function.

3.2.2. *Asymptotic variance*

Consider the case when a fixed-noise model H_ω is used in the criterion (23) and assume that the limiting set Θ_* defined by (29) is a singleton θ_*. Furthermore assume that $G'_{\omega,\theta}$ and $G''_{\omega,\theta}$ are Lipschitz continuous, where G' and G'' denote the first- and second-order derivatives with respect to the parameters θ, respectively. Define

$$Q = \int_\Omega \frac{\Phi_v(\omega)\Phi_u(\omega)2\mathrm{Re}\,\{G'_{\omega,\theta_*}(G'_{\omega,\theta_*})^*\}}{|H_\omega|^4}\,dW(\omega) \tag{32}$$

and

$$R = \int_\Omega \frac{\Phi_u(\omega)2\mathrm{Re}\,\{G'_{\omega,\theta_*}(G'_{\omega,\theta_*})^*\}}{|H_\omega|^2}\,dW(\omega) \tag{33}$$

and assume $R > \delta I$ for some $\delta > 0$. Then the estimate given by minimizing (23) is asymptotically normally distributed [12]

$$\sqrt{N}(\hat{\theta}_N - \theta_*) \in AsN(0, P_\theta) \tag{34}$$

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

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

then the size of the covariance is minimized and is equal to $P_\theta^{\mathrm{opt}} = \lambda_0 R^{-1}$. The estimator is then asymptotically *efficient*.

Discussion. It is interesting to compare the frequency domain ML-estimator (23) with the time-domain counterpart as described in [9]. The time domain (conditional) ML estimator, also known as the prediction-error method (PEM), can be described as

$$\hat{\theta}_N = \arg\min_\theta \frac{1}{N}\sum_{t=1}^{N} |H_\theta(z)^{-1}[y(t) - G_\theta(z)u(t)]|^2. \tag{35}$$

Applying Parseval's formula to the time domain criterion function in (35) reveals that the time domain ML estimator minimizes the function

$$\int_{-\pi}^{\pi} \frac{|Y_N(\omega_k) - G_{\theta,\omega_k}U_N(\omega_k)|^2}{|H_{\theta,\omega_k}|^2}\,d\omega. \tag{36}$$

A few points are worth noting. For a fixed known noise model $H_\theta(z) = H(z)$, the frequency domain ML estimate (23) and the time domain estimate (36) are essentially the same. Whenever the noise model is estimated, the additional term

$$\frac{1}{N}\sum_{k=1}^{N}\log(|H_{\theta,\omega_k}|^2\lambda) \tag{37}$$

occurs in the criterion. In fact, (37) is the log of the determinant of the transformation which changes 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 becomes of no importance because, 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. This implies that the time-domain ML estimator and the frequency domain estimator are (asymptotically) identical if the frequency domain estimator obtains data from DFT and all frequency points are used. 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 because the noise model will try to undo the effect of the prefilter [9]. 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 (29)). It could also be expected that it suffices with a less complex noise model as only the part of the true noise transfer function at the frequency points in Ω_N needs to be accounted for.

3.3. Linear least-squares methods

Although optimal, the ML method in most cases relies on solving a nonlinear optimization problem using iterative methods. A wide range of alternative methods exists here we will mention a few that are based on solving a linear regression problem (or a sequence of them). For a broad overview of other methods, we refer to [13].

Consider the following model class:

$$G_{\theta,\omega} = \frac{B_\theta(e^{i\omega})}{A_\theta(e^{i\omega})}, \qquad H_{\theta,\omega} = \frac{1}{A_\theta(e^{i\omega})}, \tag{38}$$

where $A_\theta(z)$ and $B_\theta(z)$ are polynomials with coefficients which are an affine function of the parameters θ. By disregarding λ and the log term, the ML criterion (23) simplifies to

$$\frac{1}{N}\sum_{k=1}^{N} |A_\theta(e^{i\omega_k})Y_k - B_\theta(e^{i\omega_k})U_k|^2. \tag{39}$$

The criterion (39) is quadratic in the coefficients of the A and B polynomials, and the minimizing parameters can be found by simple linear regression. If all DFT frequencies are retained in the set Ω_N, asymptotically this estimator equals the time domain ARX method [9]. A continuous, time formulation of this estimator can be found in in [7]. If only a subset of DFT frequency points used, the estimator that minimizes (39) is *not guaranteed to be asymptotically consistent* even if the true system is of the form (38). To ensure consistency, the full criterion (26) must be minimized. The least-squares solution to (39) is often used as starting values for the iterative nonlinear optimization of the full ML criterion (23).

3.3.1. *Iterative linear least-squares method*

If, on the other hand, it is desired to minimize

$$\frac{1}{N}\sum_{k=1}^{N} \left| -\frac{B_\theta(e^{i\omega_k})}{A_\theta(e^{i\omega_k})} Y_k / U_k \right|^2, \tag{40}$$

which, from an ML point of view, implies that the frequency function of the noise is constant (i.e., $H_0(e^{i\omega}) \equiv 1$), then use of the simple criterion (39) will lead to bias due to an incorrect noise model $1/A_\theta(e^{i\omega})$. A remedy to this problem was given in [15], where an iterative procedure was suggested where at iteration m a weighted criterion was minimized

$$\frac{1}{N}\sum_{k=1}^{N} |A_\theta(e^{i\omega_k})Y_k - B_\theta(e^{i\omega_k})U_k|^2 W_k^{(m)}, \tag{41}$$

where $W_k^{(1)} = 1$, $W_k^{(m)} = |\hat{A}^{(m-1)}(e^{i\omega_k})|^{-2}$, and $\hat{A}^{(m-1)}(z)$ is the estimated A polynomial from step $m-1$. The scheme often improves the performance but is not guaranteed to converge to the global minimizer of (40), see [20].

3.4. *Method of errors-in-variables*

Up to now, we have assumed that the input signal is known without errors. If the input is measured, the input will also be subject to additive noise. The estimation problem then turns into an errors-in-variables problem. Assuming the true system is given by $G_0(z) = B(z)/A(z)$, and the input and output power spectra of the noise, denoted by $\sigma_U^2(\omega)$ and $\sigma_Y^2(\omega)$, are known, the ML estimate is given by minimizing [16]

$$\sum_{k=1}^{N} \frac{|B_\theta(e^{i\omega_k})U(\omega_k) - A_\theta(e^{i\omega_k})Y(\omega_k)|^2}{\sigma_U^2(i\omega_k)|B_\theta(e^{i\omega_k})|^2 + \sigma_Y^2(\omega_k)|A_\theta(e^{i\omega_k})|^2} \tag{42}$$

if the input and output noise sources are uncorrelated. For correlated noise, a correction is subtracted from the denominator in (42). In [17] it is shown that the true input and output noise spectra can be exchanged by nonparametric estimates derived from four repeated experiments while still providing consistent estimates. If we set $\sigma_U^2 = 0$ and define $\sigma_Y^2(\omega_k) = H_{\omega_k}$ (a fixed-noise model), we conclude that minimizing (42) is identical to minimizing (23) using a fixed-noise model.

4. Conclusions

Techniques for identification of linear dynamical systems from frequency domain data have been discussed. First the transformation from time domain data to frequency domain data via the discrete Fourier transform was presented. The stochastic properties derived imply an ML estimation formulation for joint estimation of the deterministic and the stochastic part of a linear system. The asymptotic properties of the ML estimator regarding bias and variance were discussed, and the relation with the time domain ML method was covered. If all frequency data from a DFT is retained in the the estimation set, the time domain and the frequency domain ML estimators are asymptotically equal. In the light of the ML estimator, the simpler linear least-squares and iterative weighted least-squares estimators were described.

The frequency domain approach has some special features that in some applications can favor using the frequency domain ML estimator instead of the time domain counterpart:

Partial modeling. Often it is sufficient to find a model that accurately describes the true system in a limited frequency band. A low order model could thus be sufficient rather than to fit a more complex model at all frequencies. In the frequency domain, this is simply accomplished by fitting a model only at the desired frequencies, which corresponds to using an ideal bandpass filter on the raw time domain data.

Continuous time systems. If the experimental conditions are such that a multisine input can be used, the modeling in the frequency domain is straightforward. In this case the Fourier transformed data is exactly described by the continuous-time frequency function. Systems with time delay are also easy to describe in the frequency domain by using $G_\theta(i\omega)e^{-\tau i\omega}$ as the model structure, where τ is the time delay parameter.

Periodic input. When a system is excited with a periodic input applied for a sufficiently long time prior to the actual measurement interval, the effect of the initial conditions is diminished even for a finite measurement interval. Periodic excitation also makes it possible to first estimate a nonparametric noise model which can be used as a fixed-noise model in the ML criterion, see [17].

Merging data. If data is obtained by different experiments, all frequency data can be merged into one data set. Continuous-time models that are valid for large frequency ranges can be estimated from data sets obtained from several experiments, each using a different sampling frequency.

References

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

[2] K. J. Åström and B. Wittenmark, *Computer Controlled Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[3] D. R. Brillinger, *Time Series: Data Analysis and Theory*, McGraw-Hill, New York, 1981.

[4] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[5] T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[6] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 2, 2nd edn., Griffin, London, 1967.

[7] E. C. Levy, Complex curve fitting, *IRE Trans. Automat. Control*, AC-4:37–44, May 1959.

[8] L. Ljung, Building models from frequency domain data, in *IMA Workshop on Adaptive Control and Signal Processing*, Minneapolis, MN, 1994.

[9] L. Ljung, *System Identificaion: Theory of the User*, 2nd edn., Prentice-Hall, Englewood Cliffs, NJ, 1999.

[10] T. McKelvey, Frequency domain system identification with instrumental variable based subspace algorithm, in *Proc. 16th Biennial Conference on Mechanical Vibration and Noise, DETC'97, ASME*, Sacramento, CA, pp. VIB-4252, September 1997.

[11] T. McKelvey, Frequency domain identification, in R. Smith and D. Seborg, eds., *Preprints of the 12th IFAC Symposium on System Identification*, Santa Barbara, CA, June 2000 (plenary paper).

[12] T. Mckelvey and L. Ljung, Frequency domain maximum likelihood identification, in *Proc. of the 11th IFAC Symposium on System Identification*, Fukuoka, Japan, vol. 4, pp. 1741–1746, July 1997.

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

[14] R. Pintelon, J. Schoukens, and J. Vandersteen, Frequency domain system identification using arbitrary signals, in *Proc. 35th IEEE Conference on Decision and Control*, Kobe, Japan, vol. 2, pp. 2048–2051, December 1996.

[15] C. K. Sanathanan and J. Koerner, Transfer function synthesis as a ratio of two complex polynomials, *IEEE Trans. Automat. Control*, 8:56–58, January 1963.

[16] J. Schoukens and R. Pintelon, *Identification of Linear Systems: A Practical Guideline to Accurate Modeling*, Pergamon Press, London, 1991.

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

[18] T. Söderström and P. Stoica, *System Identification*, Prentice-Hall International, Hemel Hempstead, Hertfordshire, 1989.

[19] J. K. Tugnait and C. Tontiruttananon, Identification of linear system via spectral analysis given time-domain data: Consistency, reduced-order approximations and performance analysis, *IEEE Trans. Automat. Control*, AC-43:1354–1373, 1998.

[20] A. H. Whitfield, Asymptotic behaviour of transfer function synthesis methods, *Internat. J. Control*, 45(3): 1083–1092, 1987.

[21] P. Whittle, Hypothesis testing in time series analysis, Ph.D. thesis, Uppsala University, Almqvist and Wiksell, Uppsala; Hafner, New York, 1951.