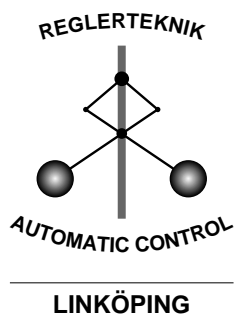


# Frequency Domain Maximum Likelihood Identification

Tomas McKelvey and Lennart Ljung

Department of Electrical Engineering  
Linköping University, S-581 83 Linköping, Sweden  
WWW: <http://www.control.isy.liu.se>  
Email: {tomas,ljung}@isy.liu.se

March 1999



Report no.: LiTH-ISY-R-2131

Proc. of the 11th IFAC Symposium on System Identification, Fukuoka, Japan, 1997,  
pp. 1741-1746

Technical reports from the Automatic Control group in Linköping are available by anonymous ftp at the address [ftp.control.isy.liu.se](ftp://control.isy.liu.se). This report is contained in the pdf-file 2131.pdf.

# FREQUENCY DOMAIN MAXIMUM LIKELIHOOD IDENTIFICATION<sup>1</sup>

Tomas McKelvey and Lennart Ljung

*Department of Electrical Engineering,  
Linköping University, S-581 83 Linköping, Sweden  
tomas@isy.liu.se, ljung@isy.liu.se*

**Abstract:** The multivariable maximum-likelihood estimate is derived for the case of frequency domain data. The relation with the time domain estimate is commented upon. The algorithm is analyzed with respect to consistency and expressions of the asymptotic variance is presented.

**Keywords:** maximum likelihood, multivariable systems, frequency domain, stochastic analysis

## 1. INTRODUCTION

The time domain prediction error method, with a quadratic criterion function, can be interpreted as the maximum-likelihood estimator if the innovation sequence has a normal distribution (Whittle 1951, Akaike 1973, Ljung 1987). Frequency domain methods which fit a rational function to the noisy data by minimizing the weighted squared sum of the model error can also be considered as maximum-likelihood (ML) estimators (Schoukens and Pintelon 1991, Ljung 1993). If the spectrum of the noise model is known the ML-estimator for the single output case is equal to a weighted least-squares fit (Schoukens and Pintelon 1991). This paper will discuss maximum-likelihood estimation using frequency data when the noise model is unknown. The single-output case has previously been discussed in (Ljung 1993, Ljung 1994) and we will here present the corresponding ML-estimator for the more general case of multi-output systems. Consistency and asymptotic variance of the ML-estimator is investigated in detail. We also highlight the similarities and point out the differences of the time and frequency domain ML-estimators. In our presentation we focus on discrete time models but the extension to continuous time modeling is straight forward (Schoukens and Pintelon 1991).

## 2. SYSTEM, MODEL AND DATA

Let us assume that the plant we are interested to obtain a model of can be described by the following linear time-invariant system

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

where  $y(t) \in \mathbb{R}^p$ ,  $u(t) \in \mathbb{R}^m$  and  $e(t) \in \mathbb{R}^p$  are the output, input and innovation signals, respectively. The operators  $G_0(q)$  and  $H_0(q)$  represent the discrete time linear systems relating the output signal to the input and noise signals respectively. For a continuous time representation, exchange the delay operator  $q$  with the differentiation operator  $p$ . The signal  $e(t)$  is assumed to be white noise with variance  $\Lambda$  and independent of the input signal  $u(t)$ .

The aim is to find a model of (1) and we construct a model set by using a parametrized model structure

$$y(t, \theta) = G(q, \theta)u(t) + H(q, \theta)e(t), \quad (2)$$

where  $G(q, \theta)$  and  $H(q, \theta)$  are the transfer function matrices of the system and noise parametrized by a vector  $\theta \in D_{\mathcal{M}} \subset \mathbb{R}^d$ .

### 2.1 Time Domain Data and the ML-Estimator

Suppose input-output data in the time domain are given

$$z^N = \{y(t), u(t) | t = 1, \dots, N\}$$

---

<sup>1</sup> This work was supported in part by the Swedish Research Council for Engineering Sciences (TFR), which is gratefully acknowledged.

From the model (2) we can define the one-step ahead predictor (in discrete time)

$$\hat{y}(t|\theta) = H(q, \theta)^{-1}G(q, \theta)u(t) + (I - H(q, \theta)^{-1})y(t)$$

The prediction errors  $\epsilon(t, \theta) = y(t) - \hat{y}(t|\theta)$  equal the innovations  $e(t)$  if the output  $y(t)$  was generated by (2) for some sequences  $u(t)$  and  $e(t)$ .

If  $e(t)$  is normal distributed zero mean random variables with covariance  $\Lambda$ , i.e. ,  $e(t) \in N(0, \Lambda)$ , the estimator

$$\hat{\theta}_N = \arg \min_{\theta} \det \frac{1}{N} \sum_{k=1}^N \epsilon(t, \theta) \epsilon(t, \theta)^T \quad (3)$$

$$\hat{\Lambda}_N = \frac{1}{N} \sum_{k=1}^N \epsilon(t, \hat{\theta}_N) \epsilon(t, \hat{\theta}_N)^T$$

is the maximum-likelihood estimator of  $\theta$  (Ljung 1987, Söderström and Stoica 1989).

## 2.2 Frequency Domain Data

For a frequency domain formulation we need a complex representation of the innovations. One possible characterization is given by the complex normal distribution.

### The Complex Normal Distribution

Let  $X \in \mathbb{C}^p$  be a complex random vector. If the real valued random vector  $\begin{bmatrix} \text{Re } X \\ \text{Im } X \end{bmatrix}$  has a normal distribution,

$$N \left( \begin{bmatrix} \text{Re } m_X \\ \text{Im } m_X \end{bmatrix}, \frac{1}{2} \begin{bmatrix} \text{Re } \Sigma_{XX} & -\text{Im } \Sigma_{XX} \\ \text{Im } \Sigma_{XX} & \text{Re } \Sigma_{XX} \end{bmatrix} \right)$$

for some  $m_X \in \mathbb{C}^p$  and some non-negative definite Hermitian matrix  $\Sigma_{XX} \in \mathbb{C}^{p \times p}$ , then  $X$  is *complex normal* (Brillinger 1981) which we denote

$$X \in N^c(m_X, \Sigma_{XX}).$$

If  $X$  is complex normal, it has a mean value  $\mathbf{E} X = m_X$ , and covariance matrix  $\mathbf{E}(X - m_X)(X - m_X)^* = \Sigma_{XX}$ . Here  $(\cdot)^*$  denotes complex conjugate transpose and  $\mathbf{E}$  mathematical expectation. We also notice that  $\mathbf{E}(X - m_X)(X - m_X)^T = 0$ . Furthermore, if  $\Sigma_{XX}$  is non-singular the probability density function of  $X$  is given by (Brillinger 1981)

$$p(\text{Re } X, \text{Im } X) = \frac{1}{\pi^p \det \Sigma_{XX}} \exp \left[ -(X - m_X)^* \Sigma_{XX}^{-1} (X - m_X) \right] \quad (4)$$

### The System Relations

Let us define the finite Fourier transform of the signal  $\{x(t)\}_{k=1}^N$  as

$$X_N(\omega) = \frac{1}{\sqrt{N}} \sum_{t=1}^N x(t) e^{-j\omega t}.$$

From the system relation (1) it is well known that (Ljung 1987, Brillinger 1981)

$$Y_N(\omega) = G_0(e^{j\omega})U_N(\omega) + V_N(\omega) + O\left(\frac{1}{\sqrt{N}}\right) \quad (5)$$

where  $V_N$  is asymptotically complex normal distributed, i.e.

$$V_N(\omega) \rightarrow N^c(0, \Phi_v(\omega)), \quad \text{as } N \rightarrow \infty$$

where  $\Phi_v(\omega) = H_0(e^{j\omega})\Lambda H_0(e^{j\omega})^*$ . Furthermore  $V_N(\omega)$  and  $V_N(\xi)$  are asymptotically independent whenever  $\omega \neq \xi$  and  $\omega, \xi \in (-\pi, \pi)$ .

Asymptotically as  $N \rightarrow \infty$  we can consider the following frequency domain system equation to hold

$$Y(\omega) = G_0(e^{j\omega})U(\omega) + H(e^{j\omega})E_0(\omega). \quad (6)$$

where  $Y, U$  are the (weak) limits of  $Y_N$  and  $U_N$  and  $E_0$  are the frequency domain innovations which are zero mean complex normal with variance  $\Lambda$ . For a continuous time formulation just exchange  $e^{j\omega}$  for  $j\omega$ . To simplify notation we will in what follows use  $G_{0,\omega} = G_0(e^{j\omega})$ ,  $G_{\omega,\theta} = G(e^{j\omega}, \theta)$  and conformally for  $H$ .

### Frequency Domain Data

In our setup we assume that we can sample the relation (6) at a sequence of frequencies  $\{\omega_k\}_{k=1}^N$  yielding the set  $Z^N = \{Y_k, U_k | k = 1, \dots, N\}$  where  $U_k = U(j\omega_k)$  and

$$Y_k = G_{0,\omega_k}U_k + H_{0,\omega_k}E_k \quad (7)$$

Here  $E_k$  is a sequence of zero mean independent complex normal variables with variance  $\Lambda$ . Note that a repeated sample at the same frequency,  $\omega_k = \omega_s$ , yields a new random variable  $E_k \neq E_s$ . Consequently,

$$Y_k \in N^c(G_{0,\omega_k}U_k, H_{0,\omega_k}\Lambda H_{0,\omega_k}^*). \quad (8)$$

Since  $\Lambda$  is the time domain covariance matrix of the noise it is real, symmetric and positive semi-definite.

### Stepped-Sine Excitation

One method to “directly” obtain frequency domain data is to use the method of stepped-sine excitation (Schoukens and Pintelon 1991). A sinusoidal excitation is applied to the plant and after the transient is sufficiently small the Fourier transform at the excitation frequency is calculated for the input and output respectively. This is done for a sequence of frequencies  $\{\omega_k, k = 1, \dots, N\}$  which yields the set  $Z^N$ . With this testing method we can guarantee that  $U_k$  is a uniformly bounded function which is needed in the consistency analysis in Section 3.

### The ML-estimator

We are now ready to derive a maximum-likelihood estimator for the frequency domain identification problem. We will base our estimation formulation on a somewhat more general setting by assuming

$\Lambda$  to be a positive definite Hermitian complex matrix.

Given samples of the Fourier transform of the input and output at  $N$  frequencies we can formulate the maximum-likelihood estimator of  $\theta$ . From (8) and (4), the scaled  $\theta$  dependent part of the negative logarithm of the likelihood function becomes

$$V_N(\theta) = \frac{1}{N} \sum_{k=1}^N \left[ \log \det (H_{\omega_k, \theta} \Lambda H_{\omega_k, \theta}^*) + \epsilon_{k, \theta}^* (H_{\omega_k, \theta} \Lambda H_{\omega_k, \theta}^*)^{-1} \epsilon_{k, \theta} \right]. \quad (9)$$

where  $\epsilon_{k, \theta} = Y_k - G_{\omega_k, \theta} U_k$ . If  $\Lambda$  is known the maximum-likelihood estimate of  $\theta$  is simply

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta). \quad (10)$$

For the case when  $\Lambda$  is unknown, we can remove  $\Lambda$  by first deriving the optimal  $\Lambda$  for each fixed value of  $\theta$ , by analytical optimization.

By introducing the notation

$$\bar{H} = \frac{1}{N} \sum_{k=1}^N \log \det (H_{\omega_k, \theta} H_{\omega_k, \theta}^*)$$

and

$$\bar{Z} = \frac{1}{N} \sum_{k=1}^N H_{\omega_k, \theta}^{-1} \epsilon_{k, \theta} \epsilon_{k, \theta}^* H_{\omega_k, \theta}^{-*}$$

we can reformulate  $V_N(\theta)$  as

$$V_N(\theta) = \log \det \Lambda + \bar{H} + \text{tr} \bar{Z} \Lambda^{-1}.$$

Let us now prove that  $V_N(\theta)$  is minimized with respect to  $\Lambda$  by the choice

$$\Lambda = \bar{Z}. \quad (11)$$

This is equivalent to the following sequence of inequalities

$$\begin{aligned} \log \det \Lambda + \text{tr} \bar{Z} \Lambda^{-1} &\geq \log \det \bar{Z} + \text{tr} I_p \\ \log \det (\Lambda \bar{Z}^{-1}) + \text{tr} \bar{Z} \Lambda^{-1} &\geq p \\ \text{tr} \bar{Z} \Lambda^{-1} - \log \det (\bar{Z} \Lambda^{-1}) &\geq p \end{aligned}$$

Assume  $\bar{Z}$  to be positive definite and define  $PP^* = \bar{Z}$  and  $X = P^* \Lambda^{-1} P$ . By construction,  $X$  is positive definite with eigenvalues  $\lambda_i > 0$ ,  $i = 1, \dots, p$ . Using the factorization of  $\bar{Z}$ , we can continue the inequalities as

$$\begin{aligned} \text{tr} PP^* \Lambda^{-1} - \log \det (PP^* \Lambda^{-1}) &\geq p \\ \text{tr} X - \log \det X &\geq p \\ \sum_{k=1}^p (\lambda_i - \log \lambda_i - 1) &\geq 0 \end{aligned}$$

The last inequality is always true since  $\lambda - \log \lambda - 1 \geq 0$  for all  $\lambda > 0$  with equality for  $\lambda = 1$ , *i.e.*, when  $\Lambda = \bar{Z}$ . This proves that  $V_N(\theta)$  is minimized with respect to  $\Lambda$  by the choice (11). If we insert (11) into (9), the maximum-likelihood estimate for unknown noise covariance emerges as

$$\begin{aligned} \hat{\theta}_N &= \arg \min_{\theta} \left[ \log \det M_N(\theta) \right. \\ &\quad \left. + \frac{1}{N} \sum_{k=1}^N \log \det (H_{\omega_k, \theta} H_{\omega_k, \theta}^*) \right] \\ M_N(\theta) &= \frac{1}{N} \sum_{k=1}^N H_{\omega_k, \theta}^{-1} \epsilon_{k, \theta} \epsilon_{k, \theta}^* H_{\omega_k, \theta}^{-*} \\ \hat{\Lambda}_N &= M_N(\hat{\theta}_N). \end{aligned} \quad (12)$$

As already mentioned, this estimator assumes  $\Lambda$  to be positive definite and Hermitian. If we would like to strengthen the estimator to only consider the case when  $\Lambda$  is a real matrix, the ML-estimator becomes more complicated.

For the single output case  $p = 1$  the ML-estimator (12) simplifies to

$$\begin{aligned} \hat{\theta}_N &= \arg \min_{\theta} \left[ \log M_N(\theta) + \sum_{k=1}^N \frac{\log |H_{\omega_k, \theta}|^2}{N} \right] \\ M_N(\theta) &= \frac{1}{N} \sum_{k=1}^N \frac{|Y_k - G_{\omega_k, \theta} U_k|^2}{|H_{\omega_k, \theta}|^2} \\ \hat{\lambda}_N &= W_N(\hat{\theta}_N). \end{aligned} \quad (13)$$

In the single output case the scalar  $\lambda$  is real.

### 2.3 Discussion

Let us briefly return to the time-domain formulation and the prediction error estimate  $\hat{\theta}_N$  given by (3) and for simplicity we consider the single output case. A straightforward application of Parseval's relation to (3) reveals that

$$\hat{\theta}_N \sim \arg \min_{\theta} \int_{-\pi}^{\pi} |E_N(\omega, \theta)|^2 d\omega,$$

where  $E_N$  is the discrete Fourier transform of  $\epsilon$  calculated on  $N$  data, *i.e.*,

$$E_N(\omega, \theta) = H_{\omega, \theta}^{-1} (Y_N(\omega) - G_{\omega, \theta} U_N(\omega)).$$

Altogether this yields the approximate expression

$$\hat{\theta}_N \sim \arg \min_{\theta} \int_{-\pi}^{\pi} \frac{|Y_N(\omega_k) - G_{\omega_k, \theta} U_N(\omega_k)|^2}{|H_{\omega_k, \theta}|^2} d\omega. \quad (14)$$

A few points are worth noticing. For a fixed known noise model  $H(q, \theta) = H(q)$ , as assumed in (Schoukens and Pintelon 1991), the frequency domain ML-estimate (13) and the time domain estimate (14) are essentially the same. Whenever the noise model also has to be estimated, the additional term

$$\frac{1}{N} \sum_{k=1}^N \log |H_{\omega_k, \theta}|^2 \quad (15)$$

occurs in the criterion. In fact (15) 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.

If the frequencies  $\omega_k$  are equidistantly distributed between 0 and  $2\pi$  this additional term again becomes less important since

$$\int_{-\pi}^{\pi} \log |H_{\omega, \theta}|^2 d\omega = 0$$

for any stable and inversely stable monic transfer function.

### 3. ASYMPTOTIC PROPERTIES

In this section we will formally prove consistency and asymptotic normality of the ML-estimate. We limit the presentation to the single-input single-output case. The extension to the multivariable case is straightforward but notationally more complex. The analysis will be based on a few assumptions.

Let the data be generated from (7) where the transfer functions, the input and noise satisfy a few assumptions.

**Assumption A1**  $U_k$  are bounded signals for all  $k$ . The admissible set of parameters  $D_{\mathcal{M}}$  is compact. The parametrized transfer functions  $G_{\omega, \theta}$ ,  $H_{\omega, \theta}$  and  $H_{\omega, \theta}^{-1}$  are continuous and uniformly bounded functions for all  $\omega \in \Omega \subset [-\pi, \pi]$  and Lipschitz continuous<sup>2</sup> in  $\theta$  in the set  $D_{\mathcal{M}} \subset \mathbb{R}^d$ . For all  $\omega \in \Omega$  and  $\theta \in D_{\mathcal{M}}$  we assume  $H_{\omega, \theta} > 0$ . ■

**Assumption A2** The noise sequence  $\{E_k\}$  are independent complex random variables which satisfy for all  $k$   $\mathbf{E} E_k = 0$ ,  $\mathbf{E} E_k E_k^* = \Lambda < \infty$ ,  $\mathbf{E} E_k E_k^T = 0$  and  $E_k$  has bounded fourth moments. The signals  $U_k$  and  $E_k$  are independent. ■

Denote by  $\Omega$  the interval of the real line to which the set of sample frequencies belongs and let  $\Omega_N$  denote the set of sample frequencies. Let us define

$$W_N(\omega) = \frac{|\{k \mid \omega_k < \omega, \omega_k \in \Omega_N\}|}{N} \quad (16)$$

where  $|\cdot|$  denotes the cardinality of a set. In probability theory  $W_N(\omega)$  corresponds to the distribution function.

**Assumption A3** The sequence of sample frequencies is such that  $W_N(\omega)$  converges to a function  $W(\omega)$  in all points of continuity of  $W(\omega)$ . ■

With the aid of the Stieltjes integral, an infinite sum can be written as an integral (Rudin 1987).

*Lemma 1.* Let  $a_{\omega, \theta}$  be a bounded Lipschitz continuous function in  $\omega \in \Omega$  and  $\theta \in D_{\mathcal{M}}$  and let  $\omega_k$  satisfy A3. Then, as  $N \rightarrow \infty$

$$\sup_{\theta \in D_{\mathcal{M}}} \left| \frac{1}{N} \sum_{k=1}^N a_{\omega_k, \theta} - \int_{\Omega} a_{\omega, \theta} dW(\omega) \right| \rightarrow 0.$$

<sup>2</sup> A function  $f$  is Lipschitz continuous in a set  $D$  if there exists a constant  $C_f$  such that  $|f(x) - f(y)| < C_f |x - y|$  for all  $x, y \in D$ .

**Proof:** First we establish pointwise convergence in  $\theta$ . By the definition of a Stieltjes integral it is immediate that

$$\frac{1}{N} \sum_{k=1}^N a_{\omega_k, \theta} = \int_{\Omega} a_{\omega, \theta} dW_N(\omega).$$

Pointwise convergence (in  $\theta$ ) then follows from assumption A3 and (Kingman and Taylor 1966, Theorem 12.1).

Consider  $f_N(\theta) = \frac{1}{N} \sum_{k=1}^N a_{\omega_k, \theta}$ . Lipschitz continuity of  $a$  now implies that

$$\begin{aligned} |f_N(\theta) - f_N(\theta_*)| &\leq \frac{1}{N} \sum_{k=1}^N |a_{\omega_k, \theta} - a_{\omega_k, \theta_*}| \\ &\leq C_a |\theta - \theta_*| \end{aligned}$$

Hence,  $f_N(\theta)$  is Lipschitz. Uniform convergence now follows from (Rudin 1987, Chapter 7 exercise 16). ■

*Lemma 2.* Let  $\{E_k\}$  be a sequence of zero mean independent random variables satisfying A2. Suppose that  $a_{k, \theta}$  is uniformly Lipschitz continuous in  $\theta$ :  $|a_{k, \theta_1} - a_{k, \theta_2}| < C_a |\theta_1 - \theta_2|$  and that  $D_{\mathcal{M}}$  is compact. Let

$$R_N(\theta) = \frac{1}{N} \sum_{k=1}^N a_{k, \theta} E_k.$$

Then

$$\sup_{\theta \in D_{\mathcal{M}}} |R_N(\theta)| \rightarrow 0, \quad \text{w.p 1 as } N \rightarrow \infty$$

**Proof:** Pointwise convergence in  $\theta$  with probability one is immediate from the standard law of large numbers, see e.g. (Chung 1974). To extend it to uniform convergence, select a dense, enumerable subset of  $D_{\mathcal{M}}$  and cover the whole set with open neighborhoods of the enumerable set. Use the Lipschitz continuity to assure that the sup over these neighborhoods is small enough. Then apply Heine-Borel's Theorem. See e.g. (Ljung 1978), pages 781-782 for the details. ■

For the case when the noise model is fixed,  $H(e^{j\omega}, \theta) = H(e^{j\omega}) = H_{\omega}$ , the ML-criterion for the single output case is

$$V_N(\theta) = \frac{1}{N} \sum_{k=1}^N \frac{|Y_k - G_{\omega_k, \theta} U_k|^2}{|H_{\omega_k}|^2}. \quad (17)$$

By regarding the generation of  $Y_k$  from (7), we get

$$V_N(\theta) = \frac{1}{N} \sum_{k=1}^N \frac{|\tilde{G}_{\omega_k, \theta} U_k + H_{0, \omega_k} E_k|^2}{|H_{\omega_k}|^2} \quad (18)$$

where  $\tilde{G}_{\omega, \theta} = G_{0, \omega} - G_{\omega, \theta}$  is the fitting error. As a limit function we define

$$\bar{V}(\theta) = \int_{\Omega} \frac{|\tilde{G}_{\omega, \theta}|^2 \Phi_u(\omega) + \Phi_v(\omega)}{|H_{\omega}|^2} dW(\omega), \quad (19)$$

where  $\Phi_u(\omega) = |U(\omega)|^2$  and  $\Phi_v(\omega) = \Lambda |H_{0, \omega}|^2$ .

*Theorem 1.* Let Assumptions A1 to A3 hold and let  $V_N(\theta)$  and  $\bar{V}(\theta)$  be given by (18) and (19). Then w.p. 1

$$\sup_{\theta \in D_{\mathcal{M}}} |V_N(\theta) - \bar{V}(\theta)| \rightarrow 0, \quad \text{as } N \rightarrow \infty \quad (20)$$

**Proof:** The sum  $V_N(\theta)$  from (18) can be decomposed into three terms

$$V_N(\theta) = \frac{1}{N} \sum_{k=1}^N |\tilde{g}_{\omega_k, \theta}|^2 + |h_{\omega_k} E_k|^2 - 2 \operatorname{Re}(a_{\omega_k, \theta} E_k) \quad (21)$$

where  $\tilde{g}_{\omega_k, \theta} = (G_{0, \omega_k} - G_{\omega_k, \theta}) U_k / H_{\omega_k}$ ,  $h_{\omega_k} = H_{0, \omega_k} / H_{\omega_k}$  and  $a_{\omega_k, \theta} = \tilde{g}_{\omega_k, \theta} h_{\omega_k}$ . By assumption A1 it follows that  $\tilde{g}$ ,  $h$  and  $a$  are Lipschitz continuous and uniformly bounded in  $\theta$  and  $\omega_k$ . By Lemma 2 we directly obtain, w.p. 1

$$\sup_{\theta \in D_{\mathcal{M}}} \left| \frac{1}{N} \sum_{k=1}^N a_{\omega_k, \theta} E_k \right| \rightarrow 0, \quad \text{as } N \rightarrow \infty$$

which implies that the third term in (21) will vanish.

Let  $v_k = E_k E_k^* - \Lambda$ . Then  $v_k$  is a sequence of zero mean random variables with bounded second moments. By Lemma 2 we get w.p.1 as  $N \rightarrow \infty$

$$\sup_{\theta \in D_{\mathcal{M}}} \left| \frac{1}{N} \sum_{k=1}^N h_{\omega_k} h_{\omega_k}^* v_k \right| \rightarrow 0$$

which implies

$$\sup_{\theta \in D_{\mathcal{M}}} \left| \frac{1}{N} \sum_{k=1}^N |h_{\omega_k} E_k|^2 - \frac{1}{N} \sum_{k=1}^N \Lambda |h_{\omega_k}|^2 \right| \rightarrow 0$$

By appealing to Lemma 1 we obtain as  $N \rightarrow \infty$

$$\sup_{\theta \in D_{\mathcal{M}}} \left| \frac{1}{N} \sum_{k=1}^N \Lambda |h_{\omega_k}|^2 - \int_{\Omega} \frac{\Phi_v(\omega)}{H_{\omega}} dW(\omega) \right| \rightarrow 0$$

By using these results on the second term we obtain w.p. 1 as  $N \rightarrow \infty$

$$\begin{aligned} & \sup_{\theta \in D_{\mathcal{M}}} \left| \frac{1}{N} \sum_{k=1}^N |h_{\omega_k} E_k|^2 - \int_{\Omega} \frac{\Phi_v(\omega)}{H_{\omega}} dW(\omega) \right| \\ & \leq \sup_{\theta \in D_{\mathcal{M}}} \left| \frac{1}{N} \sum_{k=1}^N |h_{\omega_k} E_k|^2 - \frac{1}{N} \sum_{k=1}^N \Lambda |h_{\omega_k}|^2 \right| \\ & + \sup_{\theta \in D_{\mathcal{M}}} \left| \frac{1}{N} \sum_{k=1}^N \Lambda |h_{\omega_k}|^2 - \int_{\Omega} \frac{\Phi_v(\omega)}{H_{\omega}} dW(\omega) \right| \rightarrow 0 \end{aligned}$$

By applying Lemma 1 to the first term in (21) the result follows.  $\blacksquare$

With additional notational effort the result of Theorem 1 can be extended to the MIMO case and when a parametrized noise model is used.

Let us denote by

$$D_c = \arg \min_{\theta} \bar{V}(\theta)$$

the set of limiting estimates as  $N \rightarrow \infty$ . If  $\bar{V}(\theta)$  has a unique minimum,  $D_c = \{\theta_*\}$  is a singleton.

By discarding the  $\theta$ -independent term of (19), the characterization of the limiting estimate becomes

$$D_c = \arg \min_{\theta} \int_{-\Omega}^{\Omega} \frac{|G_{0, \omega} - G_{\omega, \theta}|^2 \Phi_u(\omega)}{|H_{\omega}|^2} dW(\omega)$$

The asymptotic estimate is thus the best model in weighted mean-square sense. The weight is dependent on the input spectrum, the frequency distribution function and the assumed noise model.

If the frequency data is obtained from a time domain experiment using an  $M$  periodic input signal the input spectrum will only be non-zero for  $M$  frequency points. If each measured period delivers  $M$  points  $U_k$  and  $Y_k$  (from the FFT), the same  $M$  frequencies will be repeated over and over. In this case the distribution function  $W(\omega)$  will be a piecewise constant staircase function and the integral in the limiting criterion function (19) is just a sum of  $M$  terms.

### 3.1 Asymptotic Normality

In this section we will demonstrate that the estimated parameters are asymptotically normal distributed. We will show the result under the assumption of a fixed noise model  $H(q, \theta) = H(q)$  and that the system belongs to the model class, i.e.,  $G_0(q) = G(q, \theta_0)$  and consequently the limiting estimate  $\theta_* = \theta_0$ .

Let differentiation of a function  $f(\theta)$  w.r.t.  $\theta$  be denoted by the column vector  $f'(\theta) = \frac{d}{d\theta} f(\tilde{\theta}) \Big|_{\tilde{\theta}=\theta}$  and similarly  $f''(\theta) = \frac{d^2}{d\theta^2} f(\tilde{\theta}) \Big|_{\tilde{\theta}=\theta}$  is the corresponding matrix.

*Theorem 2.* Consider an estimate  $\hat{\theta}_N$  defined by (10) and (17). Let Assumptions A1-A3 hold. Also assume that for a unique value  $\theta_*$  interior to  $D_{\mathcal{M}}$  we have

$$\hat{\theta}_N \rightarrow \theta_*, \quad \text{w.p. 1 as } N \rightarrow \infty$$

such that  $G_{\omega, \theta_*} = G_{0, \omega}$ . Furthermore assume that  $G'_{\omega, \theta}$  and  $G''_{\omega, \theta}$  are Lipschitz continuous. Let us define

$$Q = \int_{\Omega} \frac{2\Lambda \Phi_v(\omega) \Phi_u(\omega) \operatorname{Re}\{G'_{\omega, \theta_*} (G'_{\omega, \theta_*})^*\}}{|H_{\omega}|^4} dW(\omega) \quad (22)$$

and

$$R = \int_{\Omega} \frac{2\Phi_u(\omega) \operatorname{Re}\{G'_{\omega, \theta_*} (G'_{\omega, \theta_*})^*\}}{|H_{\omega}|^2} dW(\omega) \quad (23)$$

and assume  $R > \delta I$  for some  $\delta > 0$ .

Then

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

where  $P_{\theta} = R^{-1} Q R^{-1}$ .

**Proof:** Since  $\hat{\theta}_N = \arg \min_{\theta \in D_{\mathcal{M}}} V_N(\theta)$  we have  $V'_N(\hat{\theta}_N) = 0$ . Let us expand  $V'_N(\hat{\theta}_N) = 0$  in the neighborhood of  $\theta_*$

$$0 = V'_N(\theta_*) + V''_N(\xi_N)(\hat{\theta}_N - \theta_*) \quad (24)$$

where  $\xi_N$  is between  $\hat{\theta}_N$  and  $\theta_*$ . Straightforward differentiation of  $V_N(\theta)$  gives

$$V'_N(\theta_*) = \frac{1}{N} \sum_{k=1}^N \frac{-2 \operatorname{Re}\{G'_{\omega_k, \theta_*} U_k H_{0, \omega_k}^* E_k^*\}}{|H_{\omega_k}|^2} \quad (25)$$

which is a sum of zero mean independent random variables. Let

$$X_{N,k} = \frac{-2 \operatorname{Re}\{G'_{\omega_k, \theta_*} U_k H_{0, \omega_k}^* E_k^*\}}{\sqrt{N} |H_{\omega_k}|^2}.$$

Trivially then  $\mathbf{E} X_{N,k} = 0$  and by A2 we have

$$\begin{aligned} \mathbf{E} |X_{N,k}|^2 &= \\ \frac{2\Lambda |H_{0, \omega_k}|^2 |U_k|^2 \operatorname{Re}\{G'_{\omega_k, \theta_*} (G'_{\omega_k, \theta_*})^*\}}{N |H_{\omega_k}|^4} &< \infty \end{aligned}$$

for all  $N \geq k > 0$ . Furthermore we have

$$\mathbf{E} |X_{N,k}|^3 < \frac{C}{N^{3/2}}, \quad \forall N \geq k > 0$$

for some constant  $C$  since  $E_k$  has bounded fourth order moments by Assumption A2. This implies that

$$\lim_{N \rightarrow \infty} \sum_{k=1}^N \mathbf{E} |X_{N,k}|^3 = 0.$$

Since  $\sqrt{N} V'_N(\theta_*) = \sum_{k=1}^N X_{N,k}$  we obtain by using Theorem 7.1.2 in (Chung 1974)

$$\sqrt{N} V'_N(\theta_*) \in \text{AsN}(0, Q), \quad \text{as } N \rightarrow \infty \quad (26)$$

where

$$Q = \lim_{N \rightarrow \infty} \sum_{k=1}^N \mathbf{E} |X_{N,k}|^2 \quad (27)$$

By use of Lemma 1 we obtain (22).

The Hessian of  $V_N(\theta)$  is given by

$$\begin{aligned} V''_N(\theta) &= \frac{1}{N} \sum_{k=1}^N \frac{2}{|H_{\omega_k}|^2} \operatorname{Re}\{-G''_{\omega_k, \theta} \Phi_u(\omega_k) \tilde{G}_{\omega_k, \theta}^* \\ &+ G''_{\omega_k, \theta} U_k H_{0, \omega_k}^* E_k^* + \Phi_u(\omega_k) G'_{\omega_k, \theta} (G'_{\omega_k, \theta})^*\} \end{aligned} \quad (28)$$

By A1-A2 and Lemma 2 the sum of the second term in (28) will converge to zero uniformly in  $\theta$ . Lemma 1 and A1-A3 shows that

$$\sup_{\theta \in D_{\mathcal{M}}} |V''_N(\theta) - \bar{V}''(\theta)| \rightarrow 0, \quad \text{w.p. 1 as } N \rightarrow \infty \quad (29)$$

where  $\bar{V}''(\theta) =$

$$\int_{\Omega} \frac{2\Phi_u(\omega) \operatorname{Re}\{-G''_{\omega, \theta} \tilde{G}_{\omega, \theta}^* + G'_{\omega, \theta} (G'_{\omega, \theta})^*\}}{|H_{\omega}|^2} dW(\omega)$$

By letting  $\theta = \theta_*$  we obtain  $\bar{V}''(\theta_*) = R$  where  $R$  is defined by (23).

Consider equation (24) and (29). Consequently, since  $\hat{\theta}_N \rightarrow \theta_*$  w.p. 1,

$$V''_N(\xi_N) \rightarrow \bar{V}''(\theta_*), \quad \text{w.p. 1 as } N \rightarrow \infty. \quad (30)$$

Now  $\sqrt{N}(\hat{\theta}_N - \theta_*) = [\bar{V}''(\xi_N)]^{-1} \sqrt{N} V'_N(\theta_*)$  and (30) together with (26) concludes the proof. ■

## 4. CONCLUSIONS

We have derived the multivariable ML-estimator for the case when data is given in the frequency domain. If the noise spectrum is known, the resulting estimator is equal to a weighted non-linear least squares estimator. However, if the noise spectrum is unknown, an additional sum of the logarithm of the noise spectrum appears in the criterion. The asymptotic properties of the estimator were discussed, and it was pointed out that the ML-estimator is consistent under rather general noise assumptions. Expressions for the asymptotic variance were also derived.

## 5. REFERENCES

- Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models.. *Biometrika* **20**, 255–265.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. McGraw-Hill Inc., New York.
- Chung, K. L. (1974). *A Course in Probability Theory*. Academic Press, San Diego, CA.
- Kingman, J. F. C. and S.J. Taylor (1966). *Introduction to Measure and Probability*. Cambridge at the University Press. Cambridge, UK.
- Ljung, L. (1978). Convergence analysis of parametric identification methods. *IEEE Trans. on Automatic Control* **AC-23**, 770–783.
- Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall. Englewood Cliffs, New Jersey.
- Ljung, L. (1993). Some results on identifying linear systems using frequency domain data. In: *Proc. 32nd IEEE Conference on Decision and Control, San Antonio, Texas*. pp. 3534–3538.
- Ljung, L. (1994). Building models from frequency domain data. In: *IMA Workshop on Adaptive Control and Signal Processing*. Minneapolis, Minnesota.
- Rudin, W. (1987). *Principles of Mathematical Analysis*. third ed.. McGraw-Hill.
- Schoukens, J. and R. Pintelon (1991). *Identification of Linear Systems: a Practical Guideline to Accurate Modeling*. Pergamon Press, London, UK.
- Söderström, T. and P. Stoica (1989). *System Identification*. Prentice-Hall International. Hemel Hempstead, Hertfordshire.
- Whittle, P. (1951). *Hypothesis testing in time series analysis*. Thesis. Uppsala University, Almqvist and Wiksell, Uppsala. Hafner, New York.