

Aspects on the Interpretation of Disturbances in System Identification¹

Liang-Liang Xie
Institute of Systems Science
Chinese Academy of Sciences
100080, Beijing, China

Lennart Ljung
Dept of Electrical Engineering
Linköping University
581 83, Linköping, Sweden

Abstract

The paper contains a discussion about what results about the quality of an estimated model can be achieved, if no probabilistic assumptions are introduced. Several technical results that illustrate possibilities and difficulties are also given.

1 Introduction

This contribution deals with the problem of characterizing the disturbances that act on a system. In connection with system identification applications, the role of disturbances have been discussed in various contexts. The traditional view is to regard the disturbances as stationary stochastic processes. This opens up a more or less classical stochastic framework for evaluating model quality, convergence of estimates, asymptotic covariance, as well a formal way to deal with experiment design issues, essentially aiming at minimizing variances of certain aspects of the resulting parameter estimate.

Over a number of years this basic framework has been questioned and various other approaches to describe the disturbances have been suggested. Basically they follow the idea of a "unknown but bounded" or "worst case" view of disturbances. That means that a stochastic environment is rejected and either the noise is seen as an adversary and one would like to find an identification method that guarantees certain properties even under the worse possible disturbances. The "unknown but bounded" approach is a related view where the disturbance not necessarily is an adversary but it does not necessarily possess any averaging properties, and only models that are consistent with a certain bound of the disturbances will be considered. This approach is also known as the set membership approach to identification. See, for example [4], [5], [6] for various aspects of these approaches.

In this contribution we consider what can be said about the properties of a parameter estimate when no stochas-

tic assumptions are made about the disturbance. Of particular interest are the following three aspects:

1. What are the convergence aspects of the estimate, when the disturbances are not modeled as stochastic processes?
2. Can ensemble properties, like the covariance matrix of a parameter estimate, be given an interpretation that is relevant also in a non-probabilistic disturbance framework?
3. Experiment design is usually based on minimizing the variance of a suitably chosen estimated quantity. Is it possible to show that an optimal design has some optimality properties even when applied to disturbances, that are not described by any probabilistic framework?

The paper discusses these questions in a general framework and the technical results that are offered include the following ones:

1. Convergence aspects
 - The convergence of the estimates in a general linear setting takes place as soon as the disturbances have some averaging properties, regardless whether they are described as stochastic processes or not.
 - Even if the disturbances do not have averaging properties, a characterization of the limiting estimates can be done without a probabilistic environment.
2. Variance aspects
 - It can be shown that it is not possible to obtain any ergodicity type result for the sequence of least squares prediction error estimates obtained as the number of data increases.
 - However it can be shown that if a forgetting factor algorithm is used then

¹The work of the first author was completed while visiting Linköping University as Guest Researcher. Please address all correspondence to the first author Professor L. Ljung. E-mail: ljung@isy.liu.se. The project was supported by TFR, The Swedish Research Council for Engineering Sciences

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N (\hat{\theta}_t^\lambda - \theta_0)(\hat{\theta}_t^\lambda - \theta_0)^T = P_\lambda$$

will hold with probability one, as soon as the noise disturbances has an averaging property. Here $\hat{\theta}_t^\lambda$ is the recursively estimated parameter vector at time t , using forgetting factor λ and P_λ is the ensemble-covariance matrix. This means that the covariance matrix can be given a single realization interpretation, even if no distribution is assigned to the disturbance sequence.

3. Relative qualities independent of disturbances

- Several results are shown in the special case where the input is periodic, and the model is an FIR-model of maximum degree. In this special case it can be shown that there are relationships between the estimates obtained for different input signals but the same noise sequence. These relationships are independent of the properties of that disturbance sequence. That is to say that certain design aspects will give estimates of a relative quality that does not at all depend on the disturbance sequence.

Based on these technical results, a discussion is included on whether it is possible to develop a full theory for identification, including quality measures of the traditional (ensemble-type) way of measuring the size of the model error. Can such a framework be developed for disturbances that are not described in a probabilistic setting?

2 Convergence Aspects

2.1 Disturbances with Some Averaging Properties

Consider a linear prediction model structure

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

where $G(q, \theta)$ is a stable filter in the shift operator with one delay, and $H(q, \theta)$ is an inversely stable monic filter. Observe input-output data

$$z^N = \{y(1), u(1), \dots, y(N), u(N)\} \quad (2)$$

from the system and compute the fit between (1) and the actual data:

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta); \quad (3)$$

$$V_N(\theta, z^N) = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \theta). \quad (4)$$

Determine that value of θ that gives the best fit:

$$\hat{\theta}_N = \underset{\theta}{\operatorname{argmin}} V_N(\theta, z^N). \quad (5)$$

Suppose that the observed data z^N have been generated as

$$y(t) = G_0(q)u(t) + H_0(q)e_0(t), \quad (6)$$

where the sequence $\{e_0(t)\}$ has the following properties:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N e_0(t)e_0(t-\tau) = \begin{cases} \lambda_0 & \text{if } \tau = 0; \\ 0 & \text{if } \tau \neq 0, \end{cases} \quad (7)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t)e_0(t-\tau) = 0 \text{ for all } \tau. \quad (8)$$

Then we have

$$\hat{\theta}_N \rightarrow \theta^*, \quad \text{as } N \rightarrow \infty, \quad (9)$$

where,

$$\theta^* \triangleq \underset{\theta}{\operatorname{argmin}} \int_{-\pi}^{\pi} [|G(e^{j\omega}, \theta) - G_0(e^{j\omega})|^2 \Phi_u(\omega)] \quad (10)$$

$$+ \lambda_0 |H_0(e^{j\omega})|^2 / |H(e^{j\omega}, \theta)|^2 d\omega. \quad (11)$$

The above result was proved in [1]. With this we have nailed the properties of $\hat{\theta}_N$ to those of $y(t)$ without introducing other fictitious experiments.

If $\{e_0(t)\}$ is white noise, independent of $\{u(t)\}$, then it will have the properties (7)-(8) with probability one (w.p.1), so that (9)-(10) also holds w.p.1. Notice though that the quoted result says more. It tells us that whenever (7)-(8) hold then (9)-(10) will hold (no exception on null sets). It is thus more than an ergodicity result.

We may also note that the whole probabilistic setting can then be reintroduced by just using the most elementary law of large numbers for (7)-(8). The need for more sophisticated "mixing" conditions in the convergence analysis is thus sidestepped.

2.2 Bounded Disturbances

Now, suppose that the only condition on the noise part in (6) is that it is bounded. That is, we suppose that the observed data z^N have been generated as

$$y(t) = G_0(q)u(t) + e(t), \quad (12)$$

where $|e(t)| \leq \delta$. Now the linear prediction model (1) becomes

$$\hat{y}(t|\theta) = G(q, \theta)u(t). \quad (13)$$

Then what can we say about the convergence behavior of the estimate $\hat{\theta}_N$ as defined in (5)?

First by the minimizing criterion (3)-(5), we have

$$\frac{1}{N} \sum_{t=1}^N [\tilde{G}_N(q)u(t) + e(t)]^2 \leq \frac{1}{N} \sum_{t=1}^N e^2(t),$$

where, $\tilde{G}_N(q) \triangleq G_0(q) - G(q, \hat{\theta}_N)$. Rearranging the terms and using the Cauchy-Schwartz inequality, we have, assuming that the true system G_0 is contained in

the model parameterization:

$$\begin{aligned} \frac{1}{N} \sum_{t=1}^N [\tilde{G}_N(q)u(t)]^2 &\leq -2 \frac{1}{N} \sum_{t=1}^N e(t) [\tilde{G}_N(q)u(t)] \\ &\leq 2 \left(\frac{1}{N} \sum_{t=1}^N e^2(t) \right)^{\frac{1}{2}} \left(\frac{1}{N} \sum_{t=1}^N [\tilde{G}_N(q)u(t)]^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Hence, we have for any N ,

$$\left(\frac{1}{N} \sum_{t=1}^N [\tilde{G}_N(q)u(t)]^2 \right)^{\frac{1}{2}} \leq 2 \left(\frac{1}{N} \sum_{t=1}^N e^2(t) \right)^{\frac{1}{2}} \leq 2\delta,$$

that is,

$$\frac{1}{N} \sum_{t=1}^N [\tilde{G}_N(q)u(t)]^2 \leq 4\delta^2.$$

Letting $N \rightarrow \infty$ and using Parseval's relationship, we have the following frequency function estimate error bound:

$$\limsup_{N \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} |\tilde{G}_N(e^{j\omega})|^2 \Phi_u(\omega) d\omega \leq 4\delta^2, \quad (14)$$

where $\Phi_u(\omega)$ is the spectrum of $\{u(t)\}$.

So we arrive at the following result.

Theorem 2.1 *Suppose the observed data have been generated by (12) with $|e(t)| \leq \delta$. Then using the prediction model (13) and the criterion (3)-(5), we have the asymptotic upper bound of the frequency function estimate error as in (14), provided the true system can be described within the model set.*

Instead of the average bound type result (14), one may be tempted to prove a frequency by frequency result like:

$$|\tilde{G}_N(e^{j\omega})|^2 \Phi_u(\omega) \leq B(\delta) \quad \text{for all } \omega. \quad (15)$$

But such kind of result is generally impossible if only based on the boundedness assumption of the noise. The following example demonstrates this.

Example 2.1 Suppose $u(t) = \cos(\omega t) + \cos(2\omega t)$ and $e(t) = -2G_0(q)u^2(t)$ in (12). It is easy to check that

$$y(t) = -G_0(q)[2 + \cos(\omega t) + 2\cos(3\omega t) + \cos(4\omega t)],$$

where there is no frequency 2ω component in $y(t)$. Hence the frequency function estimate error at 2ω , i.e., $\tilde{G}_N(e^{2\omega j})$ can be arbitrarily large. On the other hand, $\Phi_u(2\omega) > 0$ since $u(t)$ contains frequency 2ω component. Therefore, the left side of (15) does not have bound. Apparently the cause is that $e(t)$ introduced nonlinear dynamics.

3 Variance Aspects

3.1 Least Squares

It is well known (see e.g., [3]) that with a probabilistic setting, we can get the following convergence rate expression:

$$\sqrt{N}(\hat{\theta}_N - \theta^*) \in \text{AsN}(0, P), \quad (16)$$

which tells us that the distribution of the parameter estimate will be asymptotically normal. Also the variance of $\hat{\theta}_N$ will thus behave like $\frac{1}{N}P$ asymptotically. Now, both these statements are inherently tied to a probabilistic framework. If you make an experiment design to minimize P , you are therefore guaranteeing that your result will be good "on the average". But suppose that you are primarily concerned with the actual experiment and the quality of your actual estimate $\hat{\theta}_N$ (a very reasonable concern). We would then ask the question whether P actually tells us anything about the convergence rate of $\hat{\theta}_N$ to θ^* for the data sequence in question. The first conjecture could be that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N t \cdot (\hat{\theta}_t - \theta^*)(\hat{\theta}_t - \theta^*)^T = P \quad (17)$$

for the realization in question.

Ideally, to rid ourselves from the probabilistic framework, we should aim at proving that if $\{e_0(t)\}$ is such that (7)-(8) hold (plus possibly some other relations of the same nature) then (17) will hold. However, this we will not be able to prove and it is certainly not true as can be demonstrated by a simple example (see [2]).

3.2 Least Squares with Forgetting Factors

While impossible for the ordinary LS methods, it is possible for WLS with a forgetting factor to get an ergodicity type result.

Consider WLS with forgetting factor λ . ($0 < \lambda < 1$)

Suppose our model is linear regression model:

$$y_t = \theta^T \phi_t + e_t, \quad t \geq 1. \quad (18)$$

The minimum criterion is

$$V_N = \frac{1}{N} \sum_{t=1}^N \lambda^{N-t} (y_t - \theta^T \phi_t)^2.$$

Then the estimate error:

$$\tilde{\theta}_N = B_N \sum_{t=1}^N \lambda^{N-t} e_t \phi_t,$$

where,

$$B_N \triangleq \left(\sum_{t=1}^N \lambda^{N-t} \phi_t \phi_t^T \right)^{-1}. \quad (19)$$

Hence, the error size:

$$\begin{aligned}
\frac{1}{N} \sum_{t=1}^N \tilde{\theta}_t \tilde{\theta}_t^T &= \frac{1}{N} \sum_{t=1}^N B_t \sum_{i=1}^t \lambda^{t-i} e_i \phi_i \sum_{j=1}^t \lambda^{t-j} e_j \phi_j^T B_t^T \\
&= \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^t \sum_{j=1}^t \lambda^{2t-i-j} e_i e_j B_t \phi_i \phi_j^T B_t^T \\
&= \frac{1}{N} \sum_{\tau=0}^{N-1} \sum_{s=1}^{N-\tau} \sum_{t=s+\tau}^N \lambda^{2(t-s-\tau)+\tau} e_s e_{s+\tau} B_t \Phi(s, \tau) B_t^T \\
&= \sum_{\tau=0}^{N-1} \lambda^\tau \frac{1}{N} \sum_{s=1}^{N-\tau} e_s e_{s+\tau} \sum_{t=s+\tau}^N \lambda^{2(t-s-\tau)} B_t \Phi(s, \tau) B_t^T \\
&= \sum_{\tau=0}^{N-1} \lambda^\tau \frac{1}{N} \sum_{s=1}^{N-\tau} e_s e_{s+\tau} A(s, \tau, N), \tag{20}
\end{aligned}$$

where, $A(s, \tau, N) \triangleq \sum_{t=s+\tau}^N \lambda^{2(t-s-\tau)} B_t \Phi(s, \tau) B_t^T$, and

$$\Phi(s, \tau) \triangleq \begin{cases} \phi_s \phi_{s+\tau}^T + \phi_{s+\tau} \phi_s^T, & \tau \neq 0; \\ \phi_s \phi_s^T, & \tau = 0. \end{cases} \tag{21}$$

By (20), obviously a sufficient condition for

$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \tilde{\theta}_t \tilde{\theta}_t^T$ to exist is that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{s=1}^{N-\tau} e_s e_{s+\tau} A(s, \tau, N) \text{ exists for any } \tau \geq 0. \tag{22}$$

For bounded $\{\phi_t\}$, it is easy to see that

$$A(s, \tau) \triangleq \lim_{N \rightarrow \infty} A(s, \tau, N) \text{ exists and bounded,}$$

and also when $\phi_t = \{u_{t-1}, \dots, u_{t-n}\}^T$ and $\{u_t\}$ are periodic with period T , we have

$$\bar{A}_{p, \tau} \triangleq \lim_{k \rightarrow \infty} A(kT + p, \tau)$$

exists and bounded for any $p = 0, \dots, T-1$, which means that $A(s, \tau)$ is asymptotically periodic in s .

Hence if $\{e_t\}$ are bounded and

$$\sigma_{p, \tau} \triangleq \lim_{N_1 \rightarrow \infty} \frac{1}{N_1} \sum_{k=1}^{N_1} e_{(kT+p)} e_{(kT+p+\tau)} \text{ exists and bounded} \tag{23}$$

for any $p = 0, \dots, T-1$, and $\tau \geq 0$,

we have

$$\begin{aligned}
\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{s=1}^{N-\tau} e_s e_{s+\tau} A(s, \tau, N) &= \\
\sum_{p=0}^{T-1} \lim_{N_1 \rightarrow \infty} \frac{1}{T N_1} \sum_{k=1}^{N_1} e_{(kT+p)} e_{(kT+p+\tau)} \bar{A}_{p, \tau} &= \\
= \frac{1}{T} \sum_{p=0}^{T-1} \sigma_{p, \tau} \bar{A}_{p, \tau}.
\end{aligned}$$

Consequently, by (20)

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \tilde{\theta}_t \tilde{\theta}_t^T = \sum_{\tau=0}^{\infty} \lambda^\tau \frac{1}{T} \sum_{p=0}^{T-1} \sigma_{p, \tau} \bar{A}_{p, \tau}, \tag{24}$$

where,

$$\begin{aligned}
\bar{A}_{p, \tau} &= \lim_{k \rightarrow \infty} \lim_{N \rightarrow \infty} \sum_{t=kT+p+\tau}^N \lambda^{2(t-kT-p-\tau)} B_t \\
\Phi(kT + p, \tau) B_t &= \frac{1}{1 - \lambda^2} B_\infty \Phi(p, \tau) B_\infty
\end{aligned} \tag{25}$$

with B_t and $\Phi(s, \tau)$ defined in (19) and (21), and $B_\infty \triangleq \lim_{t \rightarrow \infty} B_t$.

Now we arrive at the following result.

Theorem 3.1 Consider a FIR model with periodic inputs with period T . Assume that the disturbance sequence $\{e_t\}$ is bounded and satisfies (23). If the parameters are estimated using WLS with forgetting factor $0 < \lambda < 1$, then we can interpret the error size asymptotically as in (24)-(25) with the averaged noise correlations in (23).

It is obvious from probabilistic point of view, (22) should hold for any quasi-stationary inputs independent of the noise. Hence the conclusion of Theorem 3.1 can be extended. The extension is stated in next theorem, where a probabilistic framework has to be used.

Theorem 3.2 Consider a FIR model with quasi-stationary inputs. Assume that the noise $\{e_t\}$ are quasi-stationary, independent of the inputs and have the following averaging property:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{s=1}^N e_s e_{s+\tau} = \sigma_\tau, \quad \text{w.p.1. for any } \tau \geq 0.$$

Then for the parameters estimated using WLS with forgetting factor $0 < \lambda < 1$, the error size is given by

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \tilde{\theta}_t \tilde{\theta}_t^T = \sum_{\tau=0}^{\infty} \lambda^\tau \sigma_\tau A(\tau), \quad \text{w.p.1.}$$

where

$$A(\tau) \triangleq \frac{1}{1 - \lambda^2} B_\infty \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{s=1}^N \Phi(s, \tau) B_\infty,$$

with $B_\infty \triangleq \lim_{N \rightarrow \infty} B_N$, B_N and $\Phi(s, \tau)$ defined in (19) and (21).

4 Relative Qualities Independent of Disturbances

Consider the LS estimation method to estimate θ in the linear regression model (18). The estimate error is given by

$$\tilde{\theta}_N = \left(\sum_{t=1}^N \phi_t \phi_t^T \right)^{-1} \sum_{t=1}^N \phi_t e_t. \quad (26)$$

Suppose for another $\{\phi'_t\}$, there exists some nonsingular matrix P such that

$$[\phi_1 \phi_2 \cdots \phi_N] = P[\phi'_1 \phi'_2 \cdots \phi'_N], \quad (27)$$

then it is obvious that for the same noise sequence $\{e_t, 1 \leq t \leq N\}$, there exists a linear transformation between $\tilde{\theta}_N$ and $\tilde{\theta}'_N$ —the estimation error using $\{\phi'_t\}$:

$$\tilde{\theta}_N = (P^T)^{-1} \tilde{\theta}'_N. \quad (28)$$

Now we start to search under what kind of conditions, such a P in (27) exists. Let's consider the FIR model with order n , i.e.,

$$\phi_t = [u_{t-1}, u_{t-2}, \cdots, u_{t-n}]^T. \quad (29)$$

(i) Periodic inputs.

Suppose $\{u_t\}$, $\{u'_t\}$ are both periodic with the same period τ . Then $\{\phi_t\}$, $\{\phi'_t\}$ are also periodic with period τ . It is obvious that $\tau \geq n$ is the necessary condition for $\left(\sum_{t=1}^N \phi_t \phi_t^T \right)$ and $\left(\sum_{t=1}^N \phi'_t \phi'^T_t \right)$ to be non-singular.

It is also easy to see that there exists a matrix P such that

$$[\phi_1 \phi_2 \cdots \phi_\tau] = P[\phi'_1 \phi'_2 \cdots \phi'_\tau] \quad (30)$$

is the necessary and sufficient condition for (27) to hold.

(ii) Sinusoidal inputs.

Consider two sequences of sinusoids inputs with the same frequencies but different amplitudes and phases:

$$u_t = \sum_{i=1}^m a_i \sin(\omega_i t + \varphi_i), \quad u'_t = \sum_{i=1}^m a'_i \sin(\omega_i t + \varphi'_i) \quad (31)$$

$\omega_i, 1 \leq i \leq m$ are different.

Assume that $\omega_i \notin \{k\pi \mid k \in N\}$ and that $2m = n$. Then m is the minimum number of frequencies to make n parameters identifiable.

We have the following result.

Theorem 4.1 Suppose the input-output dynamics can be described by an FIR model with order n . Then for two different sinusoids inputs as in (31) with $2m = n$, but the same disturbance sequence $\{e_t\}$, we have the following relationship between the frequency function estimate errors:

$$\tilde{G}_N(\omega_i) e^{j\varphi_i} = \frac{a'_i}{a_i} \tilde{G}'_N(\omega_i) e^{j\varphi'_i} \quad (32)$$

for any $1 \leq i \leq m$, $N \geq n$.

Remark 4.1 Theorem 4.1 means that the estimation error of the frequency function is inversely proportional to the amplitude of the sinusoid at the frequency in question and has nothing to do with other frequencies. This is intuitively appealing since the number of parameters $n = 2m$ is the maximum identifiable number of parameters, which gives the identified model enough flexibility to decouple the effects of different frequencies. This result holds for any disturbance $\{e_t\}$ and at any time $N \geq n$ as long as the disturbance does not depend on the inputs.

Proof. It is easy to prove that

$$\text{rank}[\phi_1 \phi_2 \cdots \phi_n] = n,$$

with ϕ_t defined in (29). Hence,

$$\text{rank} \left(\sum_{t=1}^N \phi_t \phi_t^T \right) = n, \quad \text{for } N \geq n.$$

Furthermore, we can prove that there exist c_1, c_2, \cdots, c_n (dependent only on $\omega_i, 1 \leq i \leq m$), such that for any $t > n$,

$$\phi_t = c_1 \phi_{t-1} + c_2 \phi_{t-2} + \cdots + c_n \phi_{t-n}. \quad (33)$$

Therefore,

$$[\phi_1 \phi_2 \cdots \phi_n]^{-1} [\phi_1 \phi_2 \cdots \phi_N] = [\alpha_1 \alpha_2 \cdots \alpha_N], \quad (34)$$

where, $\alpha_t = e_t, 1 \leq t \leq n$, $\alpha_t = c_1 \alpha_{t-1} + c_2 \alpha_{t-2} + \cdots + c_n \alpha_{t-n}, t > n$.

Similarly for u'_t and the corresponding ϕ'_t , we also have for any $t > n$,

$$\phi'_t = c_1 \phi'_{t-1} + c_2 \phi'_{t-2} + \cdots + c_n \phi'_{t-n}. \quad (35)$$

Thus,

$$[\phi'_1 \phi'_2 \cdots \phi'_n]^{-1} [\phi'_1 \phi'_2 \cdots \phi'_N] = [\alpha_1 \alpha_2 \cdots \alpha_N]. \quad (36)$$

Let $P \triangleq [\phi_1 \phi_2 \cdots \phi_n][\phi'_1 \phi'_2 \cdots \phi'_n]^{-1}$. Then by (34) and (36), P satisfies (27). Hence,

$$\begin{aligned} \tilde{\theta}_N &= (P^T)^{-1} \tilde{\theta}'_N \\ &= ([\phi_1 \phi_2 \cdots \phi_n]^T)^{-1} [\phi'_1 \phi'_2 \cdots \phi'_n]^T \tilde{\theta}'_N. \end{aligned} \quad (37)$$

Now let us introduce some notations. Define the matrix “row rearrangement” operator “ \rightarrow ” and “column rearrangement” operator “ \downarrow ” as:

$$A^{\rightarrow} \triangleq \begin{bmatrix} a_{m,1} & a_{m,2} & \cdots & a_{m,n} \\ a_{m-1,1} & a_{m-1,2} & \cdots & a_{m-1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,1} & a_{1,2} & \cdots & a_{1,n} \end{bmatrix};$$

$$A^{\downarrow} \triangleq \begin{bmatrix} a_{1,n} & a_{1,n-1} & \cdots & a_{1,1} \\ a_{2,n} & a_{2,n-1} & \cdots & a_{2,1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,n} & a_{m,n-1} & \cdots & a_{m,1} \end{bmatrix},$$

for $A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$.

It is easy to check the following properties of these operators:

$$C^{\rightarrow} = A^{\rightarrow} B; \quad C^{\downarrow} = AB^{\downarrow};$$

$$(C^{\rightarrow})^{\downarrow} = A^{\rightarrow} B^{\downarrow}; \quad C = A^{\downarrow} B^{\rightarrow}, \quad \text{for } C = AB.$$

The reason for introducing these operators is that $[\phi_1 \phi_2 \cdots \phi_n]^{\rightarrow}$ is a symmetric matrix and that for any $1 \leq i \leq m$, it is easy to see that there exists elementary transformation U_i (dependent only on ω_i) such that

$$U_i[\phi_1 \phi_2 \cdots \phi_n]^{\rightarrow} U_i^T = \begin{bmatrix} A_i & \mathbf{0}_{2 \times (n-2)} \\ \mathbf{0}_{(n-2) \times 2} & B_i \end{bmatrix} \quad (38)$$

$$U_i[\phi'_1 \phi'_2 \cdots \phi'_n]^{\rightarrow} U_i^T = \begin{bmatrix} A'_i & \mathbf{0}_{2 \times (n-2)} \\ \mathbf{0}_{(n-2) \times 2} & B'_i \end{bmatrix} \quad (39)$$

$$U_i \begin{bmatrix} e^{j\omega_i} \\ e^{2j\omega_i} \\ \vdots \\ e^{nj\omega_i} \end{bmatrix} = \begin{bmatrix} e^{j\omega_i} \\ e^{2j\omega_i} \\ \mathbf{0}_{(n-2) \times 1} \end{bmatrix} \quad (40)$$

where, $A_i \triangleq \begin{bmatrix} a_i \sin(\omega_i + \varphi_i) & a_i \sin(2\omega_i + \varphi_i) \\ a_i \sin(2\omega_i + \varphi_i) & a_i \sin(3\omega_i + \varphi_i) \end{bmatrix}$;

A'_i similarly defined with a'_i and φ'_i ; B_i and B'_i are $(n-2) \times (n-2)$ matrices containing no ω_i components.

Hence, in the case that $\varphi_i = \varphi'_i$, noticing that $A'_i A_i^{-1} = \frac{a'_i}{a_i}$, we have the following relationship for the estimation error of the frequency function:

$$\begin{aligned} \tilde{G}_N(\omega_i) &\triangleq \tilde{\theta}_N^T [e^{-j\omega_i} e^{-2j\omega_i} \cdots e^{-nj\omega_i}]^T \\ &= \tilde{\theta}_N^T [\phi'_1 \phi'_2 \cdots \phi'_n] [\phi_1 \phi_2 \cdots \phi_n]^{-1} \\ &\times [e^{nj\omega_i} e^{(n-1)j\omega_i} \cdots e^{j\omega_i}]^T \cdot e^{-(n+1)j\omega_i} \end{aligned} \quad (41)$$

$$= \tilde{\theta}_N^T (U_i^{-1})^{\rightarrow} \begin{bmatrix} A'_i & \mathbf{0}_{2 \times (n-2)} \\ \mathbf{0}_{(n-2) \times 2} & B'_i \end{bmatrix} (U_i^T)^{-1} \quad (42)$$

$$\cdot U_i^T \begin{bmatrix} A_i^{-1} & \mathbf{0}_{2 \times (n-2)} \\ \mathbf{0}_{(n-2) \times 2} & B_i^{-1} \end{bmatrix}$$

$$\times ((U_i^{-1})^{\rightarrow})^{-1} \begin{bmatrix} e^{j\omega_i} \\ e^{2j\omega_i} \\ \vdots \\ e^{nj\omega_i} \end{bmatrix}^{\rightarrow} \cdot e^{-(n+1)j\omega_i}$$

$$= \tilde{\theta}_N^T (U_i^{-1})^{\rightarrow} \begin{bmatrix} A'_i A_i^{-1} & \mathbf{0}_{2 \times (n-2)} \\ \mathbf{0}_{(n-2) \times 2} & B'_i B_i^{-1} \end{bmatrix} U_i^{\downarrow} \quad (43)$$

$$\times \begin{bmatrix} e^{j\omega_i} \\ e^{2j\omega_i} \\ \vdots \\ e^{nj\omega_i} \end{bmatrix}^{\rightarrow} \cdot e^{-(n+1)j\omega_i} \quad (44)$$

$$= \tilde{\theta}_N^T (U_i^{-1})^{\rightarrow} \begin{bmatrix} A'_i A_i^{-1} & \mathbf{0}_{2 \times (n-2)} \\ \mathbf{0}_{(n-2) \times 2} & B'_i B_i^{-1} \end{bmatrix} \quad (45)$$

$$\times \begin{bmatrix} e^{j\omega_i} \\ e^{2j\omega_i} \\ \mathbf{0}_{(n-2) \times 1} \end{bmatrix} \cdot e^{-(n+1)j\omega_i} \quad (46)$$

$$= \frac{a'_i}{a_i} \tilde{G}'_N(\omega_i). \quad (47)$$

In the case that $\varphi_i \neq \varphi'_i$, the calculation is somewhat more complicated. First, by explicit calculation, we can get the following equation:

$$\frac{a_i}{a'_i} A'_i A_i^{-1} \begin{bmatrix} e^{j(\omega_i + \varphi_i)} \\ e^{j(2\omega_i + \varphi_i)} \end{bmatrix} = \begin{bmatrix} e^{j(\omega_i + \varphi'_i)} \\ e^{j(2\omega_i + \varphi'_i)} \end{bmatrix},$$

which together with (45) leads to that

$$\begin{aligned} &\tilde{\theta}_N^T [e^{-j\omega_i} e^{-2j\omega_i} \cdots e^{-nj\omega_i}]^T \cdot e^{j\varphi_i} \\ &= \tilde{\theta}_N^T (U_i^{-1})^{\rightarrow} \begin{bmatrix} A'_i A_i^{-1} & \mathbf{0}_{2 \times (n-2)} \\ \mathbf{0}_{(n-2) \times 2} & B'_i B_i^{-1} \end{bmatrix} \begin{bmatrix} e^{j(\omega_i + \varphi_i)} \\ e^{j(2\omega_i + \varphi_i)} \\ \mathbf{0}_{(n-2) \times 1} \end{bmatrix} \\ &e^{-(n+1)j\omega_i} \\ &= \frac{a'_i}{a_i} \tilde{\theta}_N^T (U_i^{-1})^{\rightarrow} \begin{bmatrix} e^{j(\omega_i + \varphi'_i)} \\ e^{j(2\omega_i + \varphi'_i)} \\ \mathbf{0}_{(n-2) \times 1} \end{bmatrix} \cdot e^{-(n+1)j\omega_i} \\ &= \frac{a'_i}{a_i} \tilde{\theta}_N^T [e^{-j\omega_i} e^{-2j\omega_i} \cdots e^{-nj\omega_i}]^T \cdot e^{j\varphi'_i}, \end{aligned} \quad (48)$$

which means that in the general case, (32) holds. \diamond

References

- [1] L. Ljung. A nonprobabilistic framework for signal spectra. In *Proc. 24th IEEE Conf. on decision and Control*, pages 1056–1060, Fort Lauderdale, Fla, 1985.
- [2] L. Ljung. *The System Identification Toolbox: The Manual*. The MathWorks Inc. 1st edition 1986, 4th edition 1995, Natick, MA, 1995.
- [3] L. Ljung. *System Identification - Theory for the User*. Prentice-Hall, Upper Saddle River, N.J., 2nd edition, 1999.
- [4] P. M. Mäkilä. Worst-case input-output identification. *Int. J. Control*, 56:673–689, 1992.
- [5] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27:997–1009, 1991.
- [6] E. Walter and H. Piet-Lahanier. Estimation of parameter bounds from bounded error data: a survey. *Math. Comput Simul.*, 32:449–468, 1990.