9th IFAC Symp. on Identification, Budapest, July 1991.

C55

BIAS AND VARIANCE DISTRIBUTION IN TRANSFER FUNCTION ESTIMATION

Graham C. Goodwin
Electrical & Computer Engineering, University of Newcastle, NSW 2308, Australia

Michel Gevers
Louvain University, Bâtiment Maxwell, B1338 Louvain la Neuve, Belgium.

David Q. Mayne
Electrical Engineering, University of California, Davis, USA.

Abstract This paper gives qualitative and quantitive information on the bias and variance distribution of estimated transfer functions. The formulae show how the bias and variance distribution can be affected by the choices of design parameters : input spectrum, filters and model structure.

Keywords Error analysis, frequency response, identification, parameter estimation,

1. Introduction

Most of the early literature on the approximation of linear systems by parameterized models focused on the role played by noise (Goodwin and Payne (1977), Ljung (1987), Soderstrom and Stoica (1989)). In recent years attention has shifted to the equally significant problem of bias resulting from undermodelling (Ljung (1976), Caines (1978), Anderson Moore and Hawkes (1978), Kabaila and Goodwin (1980), La Maire et. al (1987) and Kosut (1988)).

The seminal works of Ljung (1987), Ljung and Yuan (1985) and Wahlberg and Ljung (1986) have produced a number of important qualitative insights on the distribution of bias and variance in estimated transfer functions. Here we go beyond these works in two respects. Firstly, we show the important role played by model structure in bias/variance distribution.This is achieved by use of an exact expression for the noise error rather than the asymptotic in model order result used in Ljung (1987). Secondly, we develop our results in continuous time (or equivalently the Delta operator for the discrete case). We believe this sheds more light on the role of physical parameters such as observation period, input energy and noise spectral density than is obtained in the tradition discrete time shift operator presentation.

The bias and variance formulae that we derive are based on the use of least squares for the estimation of the parameters of the nominal model. Our focus will be on formulae for the bias and variance in transfer functions; the parameters are only a vehicle for obtaining estimates of the transfer function.

Because we derive quantitative formulae for the bias and variance, we can then also make a quantitative analysis of the role of various design parameters on these two quantities.

One of our contributions of our paper is to show that the choice of a parameterization for the nominal model is related to an implicit definition of basis functions. The nominal model can be thought of as a combination of these basis functions. We examine the role of these basis functions on bias and variance as well as their interplay with the input energy distribution, the data filter frequency response and the noise distribution in some detail.

2. System Description

We denote the unknown true system frequency response by $G_T(j\omega)$. We assume that, over a given frequency range $[\omega_{\min}, \omega_{\max}]$ of interest, $G_T(j\omega)$ can be written as the sum of a parameterized but unknown nominal model, $G(\theta_0, j\omega)$ plus a residual $G_\Delta(j\omega)$:

$$G_T(j\omega) = G(\theta_0, j\omega) + G_\Delta(j\omega) \tag{2.1}$$

We assume our input-output description to be as follows:

$$y = g_T * u + n \tag{2.2}$$

where g_T is the impulse response corresponding to $G_T(j\omega)$; u, y, n denote the input, output and noise respectively; and $*$ denotes convolution in the time domain. We further assume that in the time domain, n is a wide sense stationary process (independent of G_Δ) with power spectral density $S_n(\omega)$, i.e.

$$E\{n(t)n(t-\tau)\} = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} e^{j\omega\tau} S_n(\omega)d\omega \tag{2.3}$$

It is also convenient to describe the model (2.2) in the frequency domain as

$$Y(j\omega) = G(\theta_0, j\omega)U(j\omega) + G_\Delta(j\omega)U(j\omega) + N(j\omega) \tag{2.4}$$

We acknowledge that the Fourier Transform model (2.4) is strictly only valid where the signals in the time domain have an infinite length. When only a finite data record is available, an additional error term should be added. However, for simplicity of exposition, we shall not explicitely include these terms. In the infinite data case, the Fourier transform of a time-domain stationary noise signal $n(t)$ has the property (Cox and Miller (1965)) that $N(j\omega)$ is uncorrelated in the frequency domain:

$$E\left\{N(j\omega_1)N(j\omega_2)^*\right\} = S_n(\omega_1)\delta(\omega_1 - \omega_2) \tag{2.5}$$

where $\delta(\cdot)$ is a dirac delta function.

We also introduce the notation $\phi_u(\omega)$ to denote $|U(j\omega)|^2$. This represents the input energy spectral density at frequency ω. The total energy in the signal $u(t)$ can then be decomposed into its frequency components as

$$\int_0^T u(t)^2 dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_u(\omega)d\omega \tag{2.6}$$

If $u(t)$ is a stationary signal, then one can define its power spectral density $S_u(\omega)$. Assuming ergodicity, the following relationship exists between $S_u(\omega)$ and $\phi_u(\omega)$:

$$\lim_{T\to\infty} \frac{\phi_u(\omega)}{T} = \frac{1}{2\pi} S_u(\omega) \tag{2.7}$$

Motivated by this relationship, we shall sometimes approximate $\phi_u(\omega)$ by $\frac{T}{2\pi}S_u(\omega)$ later in this paper.

3. Parameterization of the Nominal Model

We will assume that the nominal model can be parameterized as follows:

$$G(\theta,j\omega) = V(j\omega)\theta \tag{3.1}$$

Where $V(j\omega)$ is a p-row vector of known basis functions and $\theta \in I\!R^p$

One situation when the model (3.1) is immediately obtained is when one fixes the denominator in a rational transfer function model. Thus we write

$$G(\theta,j\omega) = \frac{N(\theta,j\omega)}{D(j\omega)} \tag{3.2}$$

where

$$N(\theta,j\omega) = \sum_{k=0}^{p-1}(j\omega)^k\theta_k \tag{3.3}$$

and $D(j\omega)$ is a given stable p[th] order polynominal. In this case, we have

$$V(j\omega) = \frac{1}{D(j\omega)}\left[1, j\omega, \; \ldots\ldots \; ,(j\omega)^{p-1}\right] \tag{3.4}$$

The model (3.1) also holds (at least approximately) in the following situation : suppose the nominal model is

$$G(\beta_0,j\omega) = \frac{B(\beta_0,j\omega)}{A(\beta_0,j\omega)} \tag{3.5}$$

where $B(\beta_0 j\omega), A(\beta_0,j\omega)$ are polynomials of order $(n-1)$ and n respectively and suppose that an approximation $\overline{\beta}$ of β_0 is known, then we can expand $G(\beta_0,j\omega)$ as

$$G(\beta_0,j\omega) \simeq \overline{G}(j\omega) + W(j\omega)\beta_0 \tag{3.6}$$

$$\overline{G}(j\omega) = \frac{B(\overline{\beta},j\omega)}{A(\overline{\beta},j\omega)} - W(j\omega)\overline{\beta} \tag{3.7}$$

and

$$W(j\omega) = \frac{1}{A(\overline{\beta},j\omega)^2}\Big[-B(\overline{\beta},j\omega),\ldots,-(j\omega)^{n-1}B(\overline{\beta},j\omega);$$

$$A(\overline{\beta},j\omega),\ldots,(j\omega)^{n-1}A(\overline{\beta},j\omega)\Big] \tag{3.8}$$

$$= \frac{1}{A(\overline{\beta},j\omega)^2}\left[1, j\omega, \;\ldots\;,(j\omega)^{2n-1}\right]\overline{S} \tag{3.9}$$

where $\overline{S}$ is the transpose of the Sylvester matrix for the polynomial pair $B(\overline{\beta},j\omega), A(\overline{\beta},j\omega)$ which we assume to be relatively prime at $\overline{\beta}$. Thus

$$\overline{S} = \begin{bmatrix} -b_0 & \cdots & -b_{n-1} & & & \\ 0\;-b_0 & \cdots & & -b_{n-1} & & \\ & \ddots & & & \ddots & \\ & & -b_0 & \cdots & & -b_{n-1} \\ a_0 & \cdots & a_{n-1}\,1 & & & \\ & \ddots & & \ddots & & \\ & & a_0 & & a_{n-1}\,1 & \end{bmatrix}^T \tag{3.10}$$

where a_i, b_i are the coefficients of the polynomials $A(\overline{\beta},j\omega)$ and $B(\overline{\beta},j\omega)$.

We note from (3.6), (3.9) that by defining $\overline{A}(j\omega) \triangleq A(\overline{\beta},j\omega)$:

$$V(j\omega) \triangleq \frac{1}{\overline{A}(j\omega)^2}\left[1, j\omega, \;\ldots\;,(j\omega)^{2n-1}\right] \tag{3.11}$$

$$\theta \triangleq \overline{S}\beta \tag{3.12}$$

and

$$y' \triangleq y - \overline{g} * u \tag{3.13}$$

Then we have

$$Y'(j\omega) \simeq U(j\omega)V(j\omega)\theta_0 + U(j\omega)G_\Delta(j\omega) + N(j\omega) \tag{3.14}$$

We note that is the same model as obtained in (3.4) where $D(j\omega)$ takes the value $\overline{A}(j\omega)^2$.

4. Least Squares Estimation

The least squares criterion for estimating θ can be conveniently expressed in the frequency domain using Parseval's theorem (Ljung (1987)) as

$$J = \int \| FY - FUV\theta \|^2 \tag{4.1}$$

where $F(j\omega)$ represents prefiltering of the input and output data. In this, and subsequent expressions, we have omitted the argument $d\omega$ for simplicity of notation, as well as the interval of integration. Unless otherwise stated, this interval will always be $[\omega_{\min}, \omega_{\max}]$.

The value, $\hat{\theta}$, of θ minimizing J in (4.1) is readily shown by differentation to satisfy

$$\left[\int |F|^2\phi_u V^* V\right]\hat{\theta} = \left[\int |F|^2 U^* V^* Y\right] \tag{4.2}$$

The corresponding estimate of the parameterized model is $G(\hat{\theta},j\omega) = V(j\omega)\hat{\theta}$. We shall denote the matrix in the left hand side of (4.2) as M. Equation (4.2) has a unique solution provided the spectrum $|F|^2\phi_u$ of the filtered input is such that it makes M nonsingular. We shall discuss the interelationships between the choice of filtered input spectrum, the choice of basis functions and the nonsingularity of this matrix in Sections 7 and 8.

We next turn to the study of the properties of $\hat{\theta}$. Substituting (2.4) into (4.2) yields

$$\left[\int |F|^2\phi_u V^* V\right]\hat{\theta} =$$

$$\left[\int |F|^2\left[\phi_u V^* V\theta_0 + \phi_u V^* G_\Delta + U^* V^* N\right]\right] \tag{4.3}$$

Defining $\tilde{\theta} = \hat{\theta} - \theta_0$, it then follows that

$$\left[\int |F|^2\phi_u V^* V\right]\tilde{\theta} = \int |F|^2\left[\phi_u V^* G_\Delta + U^* V^* N\right] \tag{4.4}$$

This expression shows the respective contribution in the error $\tilde{\theta}$ of the unmodelled dynamics G_Δ and of the noise N. We thus partition $\tilde{\theta}$ into

$$\tilde{\theta} = \tilde{\theta}_b + \tilde{\theta}_N \tag{4.5}$$

where $\tilde{\theta}_b$ and $\tilde{\theta}_N$ are defined by

$$\left[\int |F|^2\phi_u V^* V\right]\tilde{\theta}_b = \int |F|^2\phi_u V^* G_\Delta \quad (4.6)$$

$$\left[\int |F|^2\phi_u V^* V\right]\tilde{\theta}_N = \int |F|^2 U^* V^* N \quad (4.7)$$

We then immediately have the following result

Lemma 4.1

The error in the transfer function estimate $G_T(j\omega) - G(\hat{\theta},j\omega)$ can be decomposed as follows:

$$G_T(j\omega) - G(\hat{\theta},j\omega) = G_b(j\omega) + G_N(j\omega) \quad (4.8)$$

where

$$G_b(j\omega) = G_\Delta(j\omega) - V(j\omega)\tilde{\theta}_b \quad (4.9)$$

$$G_N(j\omega) = -V(j\omega)\tilde{\theta}_N \quad (4.10)$$

Proof:

$$\begin{aligned} G_T(j\omega) - G(\hat{\theta},j\omega) &= G_T(j\omega) - G(\theta_o,j\omega) \\ &\quad + G(\theta_o,j\omega) - G(\hat{\theta},j\omega) \\ &= G_\Delta(j\omega) - V(j\omega)\tilde{\theta} \\ &= G_\Delta(j\omega) - V(j\omega)\tilde{\theta}_b - V(j\omega)\tilde{\theta}_N \quad (4.11) \end{aligned}$$

We will call the transfer functions G_b and G_N the 'undermodelling' and 'noise' contribution of the transfer function error respectively. (In the literature, these are sometimes called bias and variance errors – Ljung (1987)).

We next analyse some possible choices and properties for the vectors $V(j\omega)$.

5. Discussion of basis functions

It is clear from expression (4.2) that the vector $V(j\omega)$ plays an important role in the estimation of the transfer function. We note that $G(\hat{\theta},j\omega)$ is simply a linear combination of the elements of $V(j\omega)$, that is

$$G(\hat{\theta},j\omega) = \sum_{i=1}^{p} \hat{\theta}_i V_i(j\omega) \quad (5.1)$$

Thus the $V_i(j\omega)$ play the role of elementary frequency responses from which $G(\hat{\theta},j\omega)$ is constructed by superposition. We shall therefore call $V_i(j\omega)$; $i = 1, \ldots, p$ the basis functions for our nominal model.

In the next two sections we shall study the role of these basis functions, not only how they effect the description of $G(\hat{\theta},j\omega)$ but also how they effect the estimated $\hat{\theta}$.

To gain some insight to the format of $V_i(j\omega)$ we will next analyze several representative choices for these basis functions.

5.1 Orthogonal Frequency Domain Pulse Functions

One possible set of basis functions is a set of non-overlapping bandbass filters. Consider the vector constructed as follows:

$$V_{2i-1}(j\omega) = B_i(j\omega) \ ; \ i = 1, \ldots, \frac{p}{2} \quad (5.2)\text{a}$$

$$V_{2i}(j\omega) = j(\text{sign}(\omega))B_i(j\omega) \ ; \ i = 1, \ldots, \frac{p}{2} \quad (5.2)\text{b}$$

where $\{B_i\}$ represents a set of non-overlapping ideal bandpass filters with bandwidths $BW_1, \ldots BW_{\frac{p}{2}}$.

A property of these basis functions is that

$$\int |F|^2\phi_u V_i^* V_k = 0 \ ; \ i \neq k \quad (5.3)$$

These basis functions are implicitely used in "so called" non-parametric frequency response estimation. It is usual to take the bandwidths $BW_1 \ldots BW_{\frac{p}{2}}$ as equal. However, this is not necessary. Indeed, it may be desirable to subdivide the frequency response more finely in some intervals if it is known *a-priori* that the frequency response is likely to exhibit more variability in these intervals.

5.2 Rational Transfer Function Basis Functions

We have argued in section 3 that a generic set of basis functions for rational transfer function estimation has the form

$$V(j\omega) = \frac{1}{D(j\omega)}\left[1, j\omega, \ldots, (j\omega)^{p-1}\right] \quad (5.4)$$

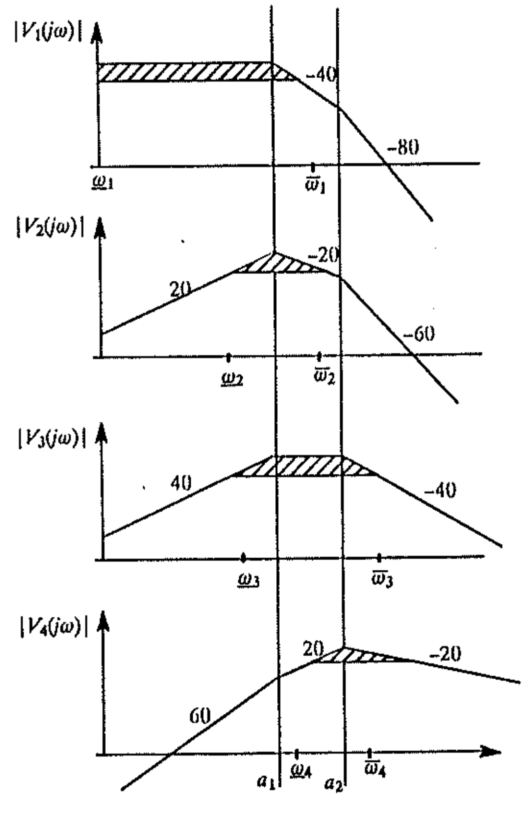
For this set of basis functions, equation (5.3) does not hold in general. However, what is true, is that basis functions whose indices differ by an odd number are orthogonal, that is (5.3) holds for $(i-k)$ odd

To reveal more of the structure of these basis functions say we have a second order rational transfer function with four parameters with basis functions chosen as in (3.11). Further, assume that $\overline{A}(s) = (s + a_1)(s + a_2)$ with a_1 and a_2 real and distinct. Then the magnitude of the frequency response of the four basis functions is as shown in Figure 5.1.

In the figure we also show the frequency support for each of the basis functions. These have been defined as the frequency interval over which the magnitude of the basis function does not drop by more than 3 dB from its maximum values. The support for $V_i(j\omega)$ is denoted by $[\underline{\omega}_i, \overline{\omega}_i]$.

Although thefunctions are not strictly orthogonal in the sense of (5.3), on noting that the result holds for $(i-k)$ odd, we see from Figure 5.1 that the functions do exhibit a high degree of orthogonality. This is apparent from the fact that the interval $[\underline{\omega}_1, \overline{\omega}_1]$ has small overlap with $[\underline{\omega}_3, \overline{\omega}_3]$ for example.

The near orthogonality of these basis functions is borne out in simulation studies (Goodwin et.al 1991).



5.3 Orthogonal Basis Functions

Given any set $\{V_i(j\omega)\}$ of basis functions, then it is always possible to find a new set $\{V'_i(j\omega)\}$ of basis functions with the same span but which satisfy (5.3) for a particular filter F and input spectrum $\phi_u(\omega)$. This can be achieved via the standard Gram Schmidt orthogonalization procedure (Goodwin et.al 1991). This amounts to taking a linear combination of the basis functions as follows:

$$\begin{bmatrix} V'_1(j\omega) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ V'_p(j\omega) \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ \gamma_1^1 & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ \gamma_{p-1}^{p-1} & & & & 1 \end{bmatrix} \begin{bmatrix} V_1(j\omega) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ V_p(j\omega) \end{bmatrix} \qquad (5.5)$$

Thus, we may assume (5.3) holds without loss of generality.

6. Undermodelling Error

We note that the nominal model can only be formed as a linear combination of the basis functions. Thus, for example, if we are given the basis functions whose amplitude response is shown in Figure 5.1, then it is clear that the true frequency response can, at best, be averaged over the frequency support of each basis function. This tells us that to get a good model at all frequencies will require us to choose basis functions consistent with the expected smoothness of the true frequency respnse.

To be more specific about the nature of the bias errors we need to say more about the actual nature of input signal and the smoothness of the true frequency response.

We first prove a preliminary result for a specific choice of input energy distribution. Roughly speaking the result says that if the filtered input spectrum has less than or exactly p points of support (where p is the number of unknown parameters), then the undermodelling error $G_b(j\omega)$ is identically zero at those points of support. Further, if there is no additive noise, this result means that at those frequencies, $G(\hat{\theta},j\omega)$ is exactly equal to $G_T(j\omega)$!

Lemma 6.1

Assume that the filtered input signal is a linear combination of l sinusoids where $l \le \frac{p}{2}$ and p is the number of basis functions in the nominal model. Then at the corresponding frequencies $\omega_1 \dots \omega_l$ in the input spectrum we have

$$G_b(j\omega_i) = 0 \; ; \; i = 1, \dots, l \qquad (6.1)$$

for all possible realizations of $G_\Delta(j\omega)$.

Proof:

By our input assumptions, we can write

$$|F|^2\phi_u = \sum_1^l \beta_k[\delta(\omega-\omega_k) + \delta(\omega+\omega_k)] \qquad (6.2)$$

Therefore (4.6) can be written as

$$\left[\sum_1^l \beta_k V^*(j\omega_k)V(j\omega_k)\right]\bar{\theta}_b = \sum_1^l \beta_k V^*(j\omega_k)G_\Delta(j\omega_k) \qquad (6.3)$$

Equivalently ; we have

$$\sum_1^l \beta_k V^*(j\omega_k)\left[V(j\omega_k)\bar{\theta}_b - G_\Delta(j\omega_k)\right] = 0 \qquad (6.4)$$

For all non trivial choices of basis functions, the vectors $V^*(j\omega_1), \dots, V^*(j\omega_l)$ will be linearly independent. Hence from (6.4) we have that

$$V(j\omega_k)\bar{\theta}_b - G_\Delta(j\omega_k) = 0 \text{ for } k = 1, \dots, l \qquad (6.5)$$

The result then follows from (4.9)

∇∇∇

The fact that in the noise free case, the estimated transfer function $\hat{G}(j\omega)$ can be made to coincide with the true transfer function, $G_T(j\omega)$, when pure sinusoidal inputs are applied is well known when $\hat{G}(j\omega)$ is estimated by non-parametric methods (Ljung (1987)). That this result also holds when a parametric model and a least squares estimation method is used, is less obvious and we have been unable to locate a proof elsewhere in the literature.

It follows immediately from lemma 6.1, that if one wants to minimize the bias at a particular frequency, ω_0, one need only apply a filtered input spectrum that has less points of support then the model has degrees of freedom provided one of the input frequencies is at ω_0. Indeed, this is a special case of the following more general result.

Theorem 6.1

Suppose we want to optimize the following weighted function of undermodelling errors:

$$J_b = \int |G_b(j\omega)|^2 X(\omega)d\omega \qquad (6.6)$$

by choice of the filtered input spectrum $|F(j\omega)|^2\phi_u(\omega)$, where $X(\omega)$ is a given non-negative definite even function of ω. Then the optimal solution is obtained using

$$|F(j\omega)|^2\phi_u(\omega) = aX(\omega) \qquad (6.7)$$

for all possible realizations of $G_\Delta(j\omega)$ and where a is any real positive scalar.

Proof:

Recall that $G_b(j\omega)$ is described by (4.9) where $\bar{\theta}_b$ is a solution of (4.6). $\bar{\theta}_b$ can alternatively be described as the minimizing solution of the following least squares minimization problem:

$$\bar{\theta}_b = \arg\min_\theta \int |G_\Delta(j\omega) - V(j\omega)\theta|^2\phi_u(\omega)|F(j\omega)|^2 \qquad (6.8)$$

Therefore the problem of optimizing J_b is equivalent to solving

$$\min_{F,\phi_u} J_b \; ; \; J_b = \int |G_\Delta(j\omega) - V(j\omega)\bar{\theta}_b|^2 X(\omega)d\omega \qquad (6.9)$$

Subject to $\bar{\theta}_b$ being described by (6.8). It then follows from Lemma 13.1 in Ljung (1987) that the optimal solution is

$$\phi_u(\omega)|F(j\omega)|^2 = aX(\omega), \qquad (6.10)$$

Where a is any positive scalar.

∇∇∇

Comments

1. It is clear from (4.6) (4.9) that the value of a has no effect on the bias error G_b. However, it will appear later, that in order to decrease the noise error G_N, we will use the largest possible value of a subject to input constraints.
2. Note that the result specializes to the single frequency case if $X(\omega)$ is $\delta(\omega-\omega_0) + \delta(\omega+\omega_0)$.
3. In the case where $X(\omega)$ is constant, the optimal solution is to use the filter F to equalize the actual input spectrum, i.e. $|F(j\omega)|^2\phi_u(\omega)$ = constant. This is an intuitively appealing result.
4. It is interesting to observe that the result holds independent of the choice of basis functions.

7. Noise Error

We next examine the component of error due to noise. We will compute the expected value of the square magnitude of the noise error G_N in the estimated transfer function.

To simplify the analysis we will require that the least squares problem has a unique solution. We argued in section 4 that equation (4.2) has a unique solution if the matrix M is nonsingular. We note that this matrix depends on the filtered input energy distribution. M can only be nonsingular if the corresponding matrix with unfiltered input is nonsingular. With this in mind, we introduce the following definition:

Definition 7.1

The input signal $u(t)$ will be called sufficiently rich for the vector $V(j\omega)$ of basis functions if there exists $a > 0$ such that

$$M_u = \int_{-\infty}^{\infty} \phi_u(\omega)V^*(j\omega)V(j\omega)d\omega \geq aI \qquad (7.1)$$

Theorem 7.1

Provided the input is sufficiently rich for the choice of basis functions, i.e. M is nonsingular, the magnitude of the noise error satisfies:

$$E\{|G_N(j\omega)|^2\} = V(j\omega)M^{-1}\left[\int |F|^2S_n|F|^2\phi_uV^*V\right]M^{-1}V^*(j\omega) \qquad (7.2)$$

Proof:

This expression follows immediately from (4.10), (4.7) and (2.5).

∇∇∇

We also have the following result

Theorem 7.2

The expected value of the square magnitude of the noise error is minimized by filtering the inputs and outputs with a filter $F(j\omega)$ such that $|F(j\omega)|^2S_n(\omega) = \text{constant}$. The optimal value of the expected square of $G_N(j\omega)$ is then given by

$$E\{|G_N(j\omega)|^2\} = V(j\omega)\left[\int \frac{\phi_u}{S_n}V^*V\right]^{-1}V^*(j\omega) \qquad (7.3)$$

Proof:

Follows immediately from the result in Appendix A.

∇∇∇

Comments

1. Since $\phi_u(\omega)$ is an energy density it grows proportionally to time T, while $S_n(\omega)$ is independent of this variable (being a power spectral density). It follows from (7.2) that the variance of $|G_N(j\omega)|$ is proportional to $\frac{1}{T}$. However, expression (7.2) also shows that the variance of $G(\hat{\theta}, j\omega)$ is not just proportional to the noise to signal ratio $\frac{S_n(\omega)}{T\phi_u(\omega)}$ at frequency ω as is suggested in Ljung (1987), where an asymptotic (in the model order) theory was used. The expression shows that the distribution of $|F|^2\phi_u$ over frequency plays a role. Equation 7.2 also shows the exact role played by the basis functions in determining the distribution of noise errors..

2. For orthogonal basis functions obtained as in section 5.3 we can compute a more insightful estimate for the noise error. In such case, we first define the following quantities:

   Peak magnitude of the ith orthogonal basis function:

$$\bar{V}_i \triangleq \sup_{\omega}|V'_i(j\omega)| \qquad (7.4)$$

   Effective support of the ith orthogonal basis function:

$$l_i(\omega) \triangleq 1 \text{ if } |V'_i(j\omega)|^2 \geq \frac{1}{2}\bar{V}_i$$
$$= 0 \quad \text{otherwise} \qquad (7.5)$$

   Effective bandwidth of ith orthogonal basis function:

$$(BW)_i \triangleq \int l_i(\omega) \qquad (7.6)$$

   Maximum Noise Power Density in effective bandwidth of ith basis function:

$$S^i_{\max} \triangleq \sup_{\omega} l_i(\omega)S_n(\omega) \qquad (7.7)$$

   Average input Power Density in effective bandwidth of ith basis function:

$$S^i_u \triangleq \frac{1}{(BW)_i}\int S_u(\omega)l_i(\omega) \qquad (7.8)$$

We then have

Corollary 7.1 Let the filter $F(j\omega)$ be chosen optimally as in Theorem 7.2 and let $\{V'_i(j\omega)\}$ be orthogonal set of basis functions with the same span as $\{V_i(j\omega)\}$. Then the noise error $G_N(j\omega)$ satisfies

$$E\{|G_N(j\omega)|^2\} \approx 8\sum_{i=1}^{p}\frac{|V'_i(j\omega)|^2S^i_{\max}}{T|\bar{V}_i|^2S^i_u(BW)_i} \qquad (7.9)$$

Further, if the set $\Gamma(\omega) = \{k : l_k(\omega) = 1\}$ is non empty, then

$$\{|G_N(j\omega)|^2\} \approx \sum_{k\in\Gamma(\omega)}\frac{8S^i_{\max}}{TS^i_u(BW)_i} \qquad (7.10)$$

Proof

By using the orthogonality of the basis function in (7.3) and by approximating the integral by values where $|V'_i(j\bar{\omega})|$ is above 0.7 of its maximum value.

∇∇∇

We note that the variance at a frequency ω decreases with: the observation length, the power spectral density of the input over the support of the corresponding basis function and the frequency bandwidth of that support. Assuming that the total frequency bandwidth of interest is fixed, and that the number of basis functions is equal to the number of parameters, this expression also shows that increasing the width of the support of one basis function reduces the variance over that range of frequencies at the expense of an increase of variance over other frequency ranges. Moreover, if the basis functions contain resonant structures, then the bandwidth of the corresponding basis function will necessarily be narrow and the variance of the frequency response estimate in the region of the resonance is thus likely to be large. Another observation is that, contrary to the impression given in Ljung (1987), increasing the number of parameters in the model does not necessarily increase the mean square noise error. This will not occur, for example, if the effective support of additional basis functions lies outside the frequency range of interest. On the other hand, there are at least two mechanisms whereby the mean square noise error can increase when additional basis functions are added. Firstly, if the new basis functions are distinguished via their amplitude response, then adding additional basis functions will reduce the bandwidth available for any one basis functions, thus increasing the term $\left(\frac{1}{TS_u^i(BW)_i}\right)$ in (7.10). Secondly, if the new basis functions are distinguished via phase (as in the case of orthogonal all pass networks with a white noise input), then the input energy $\left(TS_u^i(BW)_i\right)$ going into each basis function will remain roughly constant but more terms will appear in the set $\Gamma(\omega)$ in (7.10) for each frequency of interest thus increasing the mean square noise error.

8. Design of Input Signals and Data Filters

We have seen in section 6 that if one wishes to minimize the integral square undermodelling error weighted by $X(\omega)$ then one needs to choose the input and filters so that $\left[|F|^2\phi_u/X(\omega)\right]$ is constant. On the other hand, to minimize the mean square noise error one needs to choose the filter so that $|F|^2S_n$ = constant. Note that when both $F(j\omega)$ and $\phi_u(\omega)$ can be adjusted, then both of these conditions can be simultaneously satisfied. On the other hand if $\phi_u(\omega)$ is given (i.e. the experiment has already been performed) then one has to make a compromise between minimizing bias errors and minimizing variance errors when choosing the filter. Another observation from (7.10) is that it is clear that it is important that non-negligible input energy go into each basis function whose effective support overlaps the frequencies of interest. This is a more intuitive result than the usual statement that it suffices to have $\frac{p}{2}$ sinusoids (of arbitrary frequencies) to estimate p parameters in a model.

9. Conclusions

This paper has addressed the issue of the distribution of errors in the estimation of the frequency response of a linear system. Particular emphasis has been given to the role of the basis functions in the nominal model and their connection with errors arising from undermodelling and noise respectively. Explicit expressions have been obtained for these errors and the expressions have been used to determine the optimal design of input spectra, and data filters to minimize the errors. The results are believed to offer deeper insights into the problem of estimating linear models than have been hitherto available. Potential application areas include off-line estimation and adaptive control.

Appendix A

Lemma A.1

Let $g(\omega)$ be a real nonnegative even function of ω and let $V(j\omega)$ be a complex row-vector in $\mathbf{C}^n$. Then for all real nonnegative even functions $f(\omega)$ we have:

$$P(f(\cdot)) \triangleq \left[\int_{S_g} f(\omega)V^*(j\omega)V(j\omega)\right]^{-1}\left[\int_{S_g} f^2(\omega)g(\omega)V^*(j\omega)V(j\omega)\right]$$

$$\left[\int_{S_g} f(\omega)V^*(j\omega)V(j\omega)\right]^{-1} \geq P\left(g^{-1}(\cdot)\right)$$

where S_g denotes the support of $g(\omega)$. In other words, the minimum (in a matrix sense) of $P(f(\cdot))$ is obtained for $f(\omega) = g^{-1}(\omega)$.

References

B.D.O. Anderson, J.B. Moore and R.M. Hawkes (1978), "Model Approximation via Prediction Error Identification", *Automatic*, Vol.14, pp.615-622.

P.E. Caines (1978), "Stationary Linear and Nonlinear System Identification and Predictor set Completeness", *IEEE Trans. Auto. Control*, Vol.Ac-23, pp.583-594.

D.R. Cox and H.D. Miller (1965), "The Theory of Stochastic Processes", Chapman andHall Ltd., London.

G.C. Goodwin, B. Ninness, V. Poor (1991), "Choice of Basis Functions for Continuous and Discrete System Modelling" IFAC Symposium on Identification and System Parameter Estimation, Hungary.

G.C. Goodwin and R.L. Payne (1977), "Dynamic System Identification : Experiment Design and Data Analysis", Academic Press.

G.C. Goodwin and M. Salgado (1989), "A Stochastic Embedding Approach for Quantifying Uncertainty in the Estimation of Restricted Complexity Models", *Int. J. Adaptive Control and Signal Processing*, Vol. 3, No.4, pp.333-357.

P.V. Kabaila and G.C. Goodwin (1980), "On the Estimation of the Parameters of an Optimal Interpolation when the class of Interpolators is Restricted", *SIAM J. Control and Optimization*, Vol.18, No.2, pp.121-144.

R.L. Kosut (1988), "Adaptive Robust Control via Transfer Function Uncertainty Estimation", *Proc. 1988 ACC*, Atlanta, Georgia.

La Maire R.O., L. Valavani, M. Athans and G. Stein (1987), "A Frequency Domain Estimator for use in Adaptive Control Systems", *Proc. ACC*, Minneapolis, MN, pp.238-244.

L. Ljung (1976), "On Consistency and Identifiability", Mathematical Programming Study No.5, North-Holland, Amsterdam, pp.169-190.

L. Ljung and Z.D. Yuan (1985), "Asymptotic Properties of Black-Box Identification of Transfer Functions", *IEEE Trans. Auto. Control*, Vol. AC-30, pp.514-530.

L. Ljung (1987), "System Identification : Theory for the user", Prentice Hall.

T. Söderstrom and P. Stoica (1989), "System Identification", Prentice Hall.

B. Wahlberg and L. Ljung (1986), "Design Variables for Bias Distribution on Transfer Function Estimation", *IEEE Trans. Auto. Control*, Vol. AC-31, pp.137-144.