REFERENCES

[1] S. Lippman, "Semi-Markov decision processes with unbounded rewards," *Man. Sci.*, vol. 19, no. 7, pp. 717-731, 1973.
[2] R. L. Larsen, "Control of multiple exponential servers with application to computer systems," Ph.D. dissertation, Univ. Maryland, 1981.
[3] W. Lin and P. R. Kumar, "Optimal control of a queueing system with two heterogeneous servers," *IEEE Trans. Automat. Contr.*, vol. AC-29, no. 8, pp. 696-703, 1984.
[4] J. Walrand, "A note on 'Optimal control of a queueing system with two heterogeneous servers'," *Syst. Contr. Lett.*, vol. 4, pp. 131-134, 1984.
[5] B. Hajek, "Optimal control of two interacting service stations," *IEEE Trans. Automat. Contr.*, vol. AC-29, no. 6, pp. 421-429, 1984.
[6] C. E. Bell, "Optimal operation of a *M/M/2* queue with removable servers," *Oper. Res.*, vol. 28, no. 5, pp. 1189-1204, 1980.
[7] Y. C. Chow and W. H. Kohler, "Models for dynamic load balancing in a heterogeneous multiple processor system," *IEEE Trans. Comput.*, vol. C-28, pp. 354-361, 1979.
[8] E. L. Hahne, "Dynamic routing in an unreliable manufacturing network with limited storage," Mass. Inst. Technol., Cambridge, MA, Rep. LIDS-TH-1063, 1981.
[9] J. N. Tsitsiklis, "Convexity and characterization of optimal policies in a dynamic routing problem," Mass. Inst. Technol., Cambridge, MA, Rep. LIDS-R-1178, 1982.
[10] A. Seidmann and P. J. Schweitzer, "Part selection policy for a flexible manufacturing cell feeding several production lines," *IIE Trans.*, vol. 16, no. 4, pp. 355-362, 1984.
[11] A. K. Agrawala *et al.*, "A stochastic optimization algorithm minimizing expected flow times on uniform processors," *IEEE Trans. Comput.*, vol. C-33, pp. 351-356, 1984.
[12] I. Viniotis, "Optimal control of a multiserver queueing system with heterogeneous servers," M.Sc. thesis, Dep. Elec. Eng., Univ. Maryland, 1985.
[13] W. Whitt, "Comparing counting processes and queues," *Adv. Appl. Prob.*, vol. 13, pp. 207-220, 1981.
[14] S. Lippman, "Applying a new device in the optimization of exponential queueing systems," *Oper. Res.*, vol. 23, pp. 687-710, 1975.
[15] I. Viniotis and A. Ephremides, "An optimization problem in a multiserver queue," in *Proc. 23rd Annual Allerton Conf. on Control, Commun. and Comp.*, Urbana-Champaign, IL, pp. 555-563.
[16] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models.* New York: Wiley, 1983.
[17] J. Keilson, *Markov Chain Models—Rarity and Exponentiality.* New York: Springer-Verlag, 1979.

Adaptive Parameter Estimation for a Constrained Low-Pass System

ARYE NEHORAI AND PETRE STOICA

Abstract—**A new algorithm is presented for adaptive parameter estimation for a constrained low-pass Butterworth system model. Potential applications include filter design and adaptive decision on Nyquist rate for systems.**

I. INTRODUCTION

Output error (OE) models used in system identification schemes are known for their physical appeal and robustness with respect to the spectral properties of the measurement noise (see, e.g., [1]-[4]). Most existing OE schemes use unconstrained rational transfer function models.

In this note we consider a system identification algorithm for on-line estimation of the parameters of a special constrained OE model. This model is constrained to be a digital low-pass Butterworth filter obtained by a bilinear transformation of a continuous low-pass Butterworth filter (see, e.g., [5, pp. 227-228], [6, pp. 211-218]).

The proposed algorithm will estimate the system cut-off frequency and gain on-line. When it is known that the true system has a low-pass Butterworth structure or some transfer function similar to it and its true order is used, the algorithm will lead to a substantial saving in computations and more accurate results (by the parsimony principle [2]) than unconstrained algorithms. Potential applications include adaptive decision on Nyquist rate for systems and filter design.

Manuscript received January 5, 1987; revised March 13, 1987. This work was supported in part by the National Science Foundation under Grant DCI-8604351.

A. Nehorai is with the Department of Electrical Engineering, Yale University, New Haven, CT 06520.

P. Stoica is with the Department of Automatic Control, Polytechnic Institute of Bucharest, Bucharest, Romania.

IEEE Log Number 8716169.

II. THE MODEL AND THE GRADIENT OF ITS OUTPUT WITH RESPECT TO THE MODEL PARAMETERS

A. The Model

The OE system model is characterized by

$$y(t)=G(q^{-1})u(t)+\epsilon(t) \tag{1}$$

where $y(t)$ and $u(t)$ are the measurable system output and input, respectively, $\epsilon(t)$ denotes the model residuals, and q^{-1} the unit delay operator. In our special case $G(q^{-1})$ is assumed to be

$$G(q^{-1})=g\,\frac{B^0(q^{-1})}{A(q^{-1})} \tag{2a}$$

where q is an unknown gain, and

$$B^0(q^{-1})=(1+q^{-1})^n \tag{2b}$$

$$A(q^{-1})=\prod_{i=1}^{n}(1-\lambda_i q^{-1}) \tag{2c}$$

$$\lambda_i=\frac{1+\frac{\omega_c}{2}e^{j\alpha_i}}{1-\frac{\omega_c}{2}e^{j\alpha_i}} \qquad \alpha_i=\left(1+\frac{2i-1}{n}\right)\frac{\pi}{2} \tag{2d}$$

where ω_c denotes the normalized cut-off frequency of $G(q^{-1})$. The superscript "0" in $B^0(q^{-1})$ emphasizes that this polynomial is known. Observe from (2d) that $\lambda_i = \lambda^*_{n-i+1}$ since $\alpha_i + \alpha_{n-i+1} = 2\pi$. ("*" denotes complex conjugate.) Hence, the coefficients of $A(q^{-1})$ are real. Also $|\lambda_i| < 1$ since $\cos\alpha_i < 0$. Thus, the model $G(q^{-1})$ is guaranteed to be stable, which is an important feature since it implies that, unlike unconstrained models, stability monitoring will not be needed in the following OE algorithm.

The unknown parameter vector to be estimated is

$$\theta=[g \;\; \omega_c]^T. \tag{3}$$

The estimate $\hat{\theta}$ of θ is defined within the OE methodology by

$$\hat{\theta}=\arg\min_{\theta} E[y(t)-\hat{y}(t)]^2 \tag{4}$$

where E denotes the expectation operator, and $\hat{y}(t)$ is the model output

$$\hat{y}(t)=g\,\frac{B^0(q^{-1})}{A(q^{-1})}\,u(t). \tag{5}$$

In the course of estimation, we must constrain ω_c to be positive (which is equivalent to the stability of our model, see above). However, this is a simple constraint that can be handled easily.

Next, let us assume that the true data generating mechanism is

$$y(t)=G^*(q^{-1})u(t)+e(t) \tag{6}$$

where $e(t)$ is a disturbance that is uncorrelated with $u(s)$ for all t and s, and where the transfer function $G^*(q^{-1})$ is not necessarily of the form (2).

0018-9286/88/0100-0109$01.00

It follows from (4)–(6) that the OE model satisfies

$$\int_{-\pi}^{\pi} \left| G^*(e^{i\omega}) - g\frac{B^0(e^{i\omega})}{A(e^{i\omega})}\right|^2 \phi_u(\omega)\, d\omega = \min \tag{7}$$

where $\phi_u(\omega)$ denotes the spectral density of the input signal. Thus, the OE method provides the Butterworth model, which is the best approximation of the system's frequency characteristic in the sense of minimizing the weighted norm in (7). Note that the weighting function in (7) is $\phi_u(\omega)$. Hence, by selecting the input, one can control the distribution of approximation errors in the frequency domain. If a direct access to $u(t)$ is not available, one may prefilter the input and output data by a filter, say $H(q^{-1})$, which has the same effect as replacing $\phi_u(\omega)$ with $|H(e^{i\omega})|^2\phi_u(\omega)$. We may choose $H(q^{-1})$ for emphasizing (or deemphasizing) various frequency bands in the approximation (see also [3] and [4] for the prefiltering approach).

It is important to observe that interpretation (7) of the OE models does not require that the true system be of the Butterworth type [see (2)]. Also, the disturbance $e(t)$ may be a colored process. If $e(t)$ is white, then $\hat{y}(t)$ of (5) becomes the best (in mean-square sense) one-step prediction of $y(t)$, and the OE estimate $\hat{\theta}$ of (4) in turn becomes a prediction error estimate.

In this note we consider the integer-valued parameter n of our model as given. Note that the larger the value of n, the larger the slope of the low-pass transfer function will be near ω_c. When the order n is unknown (corresponding to unknown slope of the low-pass filter) it is possible to apply information criteria to estimate the order (e.g., [7]; also see [1]–[3]). Another approach is to replace the model (2) with a weighted sum of two models of different orders. Specifically, use

$$G(q^{-1}) = \alpha G_1(q^{-1}) + (1-\alpha)G_2(q^{-1}) \tag{8}$$

where $0 \le \alpha \le 1$, $G_1(q^{-1})$ and $G_2(q^{-1})$ are constrained models, as in (2), but of different orders. The orders of $G_1(q^{-1})$ and $G_2(q^{-1})$ correspond to the expected lower and upper bounds on the true order, based on some *a priori* information. The variable α then has to be estimated as a part of θ. However, we will not follow this idea since our main concern in this note is estimation of the parameters ω_c and g.

In Section III we present a recursive Gauss–Newton algorithm for performing the minimization in (4). Next we discuss the computation of $\hat{y}(t)$ and of the derivative $\partial\hat{y}(t)/\partial\theta$ required for this algorithm

B. Calculation of $\hat{y}(t)$ and $\partial\hat{y}(t)/\partial\theta$

The model output is given by

$$\hat{y}(t) = g\frac{B^0(q^{-1})}{A(q^{-1})}u(t) = -\sum_{i=1}^{n} a_i\hat{y}(t-i) + g\sum_{i=0}^{n} b_i^0 u(t-i) \tag{9}$$

where a_i and b_i^0 are the coefficients of the polynomials $A(q^{-1})$ and $B^0(q^{-1})$, respectively. The gradient of $\hat{y}(t)$ with respect to the parameter vector θ is found as follows. Let

$$\psi(t) = \left[\frac{\partial\hat{y}(t)}{\partial\omega_c} \;\; \frac{\partial\hat{y}(t)}{\partial g}\right]^T. \tag{10}$$

The first component of $\psi(t)$ can be found using the chain rule

$$\psi_1(t) \triangleq \frac{\partial\hat{y}(t)}{\partial\omega_c} = \sum_{i=1}^{n}\frac{\partial\hat{y}(t)}{\partial a_i}\frac{\partial a_i}{\partial\omega_c} = \sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\partial\hat{y}(t)}{\partial a_i}\frac{\partial a_i}{\partial\lambda_k}\frac{\partial\lambda_k}{\partial\omega_c}. \tag{11}$$

Here

$$\frac{\partial\hat{y}(t)}{\partial a_i} = -g\frac{B^0(q^{-1})}{A^2(q^{-1})}u(t-i) = -\frac{1}{A(q^{-1})}\hat{y}(t-i) \triangleq -\hat{y}_F(t-i). \tag{12}$$

An expression for the derivative of polynomial coefficients with respect to the polynomial zeros was derived in [8, eq. (21)] (see also [9]). Assuming the polynomial under consideration has distinct zeros, the result is

$$\frac{\partial a_i}{\partial\lambda_k} = -\sum_{l=0}^{i-1} a_l\lambda_k^{i-l-1}. \tag{13}$$

For completeness we present in the Appendix a proof of (13) which is simpler than that of [8] and provides an interesting interpretation for the Jacobian matrix $\partial a/\partial\lambda$. From (13) we get

$$\nabla_i \triangleq \frac{\partial a_i}{\partial\omega_c} = \sum_{k=1}^{n}\frac{\partial a_i}{\partial\lambda_k}\frac{\partial\lambda_k}{\partial\omega_c} = \sum_{l=0}^{i-1} a_l x_{i-l} \tag{14a}$$

$$x_i = -\sum_{k=1}^{n}\lambda_k^{i-1}\frac{\partial\lambda_k}{\partial\omega_c}. \tag{14b}$$

Since $\{\lambda_k\}$ occur in complex conjugate pairs ($\lambda_k = \lambda_{n-k+1}^*$, see above), we can write

$$x_i = \begin{cases} -2\sum_{k=1}^{n/2}\operatorname{Re}\left\{\lambda_k^{i-1}\frac{\partial\lambda_k}{\partial\omega_c}\right\} & n \text{ even} \\ -2\sum_{k=1}^{(n-1)/2}\operatorname{Re}\left\{\lambda_k^{i-1}\frac{\partial\lambda_k}{\partial\omega_c}\right\} - \lambda_{(n+1)/2}^{i-1}\frac{\partial\lambda_{(n+1)/2}}{\partial\omega_c} & n \text{ odd}\end{cases} \tag{15}$$

where Re $\{x\}$ denotes the real part of x. In our special case, from (2d)

$$\lambda_k^{i-1}\frac{\partial\lambda_k}{\partial\omega_c} = e^{j\alpha_k}\frac{\left(1+\frac{\omega_c}{2}e^{j\alpha_k}\right)^{i-1}}{\left(1-\frac{\omega_c}{2}e^{j\alpha_k}\right)^{i+1}}. \tag{16a}$$

When n is odd we have

$$\lambda_{(n+1)/2}^{i-1}\frac{\partial\lambda_{(n+1)/2}}{\partial\omega_c} = -\frac{\left(1-\frac{\omega_c}{2}\right)^{i-1}}{\left(1+\frac{\omega_c}{2}\right)^{i+1}}. \tag{16b}$$

Thus, $\psi_1(t)$ can be computed by first evaluating the sequence $\{x_i\}$ in (15), (16), then computing ∇_i by (14a), and finally from (11), (12)

$$\psi_1(t) = \sum_{i=1}^{n}\hat{y}_F(t-i)\nabla_i. \tag{17}$$

The second component of $\psi(t)$ is found as follows. From (5)

$$\psi_2(t) \triangleq \frac{\partial\hat{y}(t)}{\partial g} = \frac{B^0(q^{-1})}{A(q^{-1})}u(t) = \sum_{i=0}^{n} b_i^0 u_F(t-i) \tag{18}$$

where $u_F(t) = [1/A(q^{-1})]u(t)$.

III. The Recursive Algorithm for Estimating the Cut-Off Frequency and Gain

Based on the above results we now present the recursive Gauss–Newton OE algorithm for estimating the cut-off frequency ω_c and the gain g. For more details on OE methods see, e.g., [1], [2], and [10].

A. Summary of the OE Algorithm

Initialization: Set $\lambda(1)$, λ_o, $\hat{\theta}(0)$, $P(0)$, $\psi(1)$; $u(t)$, $u_F(t)$, $\hat{y}_F(t)$, $\hat{y}(t)$, for $t < 1$. Recommended values: $\lambda(1) = 0.95$, $\lambda_o = 0.99$, $\psi(1) = 0$,

$P(0)$: application dependent, e.g., $\approx 0.01 \cdot I$, $\hat{\theta}(0) =$ small positive values. $u(t) = u_F(t) = \hat{y}_F(t) = \hat{y}(t) = 0$ for $t < 1$.

Main Loop: Do for $t = 1, 2, \cdots$

$$\hat{y}(t) = -\sum_{i=1}^{n} \hat{a}_i(t-1)\hat{y}(t-i) + \hat{g}(t-1)\sum_{i=0}^{n} b_i^0 u(t-i)$$

$$\epsilon(t) = y(t) - \hat{y}(t)$$

$$K(t) = P(t-1)\psi(t)/[\lambda(t) + \psi^T(t)P(t-1)\psi(t)]$$

$$P(t) = [P(t-1) - K(t)\psi^T(t)P(t-1)]/\lambda(t)$$

$$\hat{\theta}(t) = [\hat{\theta}(t-1) + K(t)\epsilon(t)]_D = [\hat{\omega}_c(t) \;\; \hat{g}(t)]_D^T.$$

Compute the Butterworth filter coefficients $\{\hat{a}_i(t)\}$ as a function of n and $\hat{\omega}_c(t)$ using (2b)–(2d).

$$\hat{y}(t) = -\sum_{i=1}^{n} \hat{a}_i(t)\hat{y}(t-i) + \hat{g}(t)\sum_{i=0}^{n} b_i^0 u(t-i)$$

$$\hat{y}_F(t) = \hat{y}(t) - \sum_{i=1}^{n} \hat{a}_i(t)\hat{y}_F(t-i)$$

$$u_F(t) = u(t) - \sum_{i=1}^{n} \hat{a}_i(t)u_F(t-i)$$

$$x_i(t) = -2\sum_{k=1}^{\lfloor n/2 \rfloor} \mathrm{Re}\left\{ e^{j\alpha_k} \frac{\left[1 + \frac{\hat{\omega}_c(t)}{2} e^{j\alpha_k}\right]^{i-1}}{\left[1 - \frac{\hat{\omega}_c(t)}{2} e^{j\alpha_k}\right]^{i+1}} \right\} + \frac{\left(1 - \frac{\hat{\omega}_c(t)}{2}\right)^{i-1}}{\left(1 + \frac{\hat{\omega}_c(t)}{2}\right)^{i+1}} \quad (\leftarrow \text{for } n \text{ odd only})$$

$$\nabla_i(t+1) = \sum_{l=0}^{i-1} \hat{a}_l(t)x_{i-l}(t)$$

$$\psi_1(t+1) = -\sum_{i=1}^{n} \nabla_i(t+1)\hat{y}_F(t+1-i)$$

$$\psi_2(t+1) = \sum_{i=0}^{n} b_i^0 u_F(t+1-i)$$

$$\lambda(t+1) = \lambda_o\lambda(t) + (1-\lambda_o).$$

Comments: The notation $[\hat{\theta}]_D$ means that a projection feature is used to prevent the algorithm from getting outside the model set D. In our case, we constrain $\hat{\omega}_c(t)$ to be positive, which is equivalent to the stability of our model; see Section II. Thus, if $\hat{\omega}_c(t) < 0$ we choose, e.g., to set $\hat{\omega}_c(t)$ equals a small positive value. The notation "$\lfloor x \rfloor$" denotes the largest integer smaller than x.

Note that the algorithm above is guaranteed to converge to a local minimum of the OE loss function. Of some concern, as in any other minimization algorithm, is the possibility of the existence of local minima in the OE loss function. This issue is not addressed here as it is beyond the scope of this note.

B. A Simulated Example

The following example was simulated on a DEC VAX with a single precision arithmetic to illustrate the convergence of the algorithm. The input $u(t)$ and the noise $e(t)$ were independent zero-mean white sequences of unit variance. The system order was $n = 6$, the normalized cut-off frequency was $f_c \triangleq \omega_c/2\pi = 0.2$ and the signal-to-noise ratio SNR = 6 dB. The gain was $g = 2.12 \times 10^{-2}$. The algorithm was applied with the true order and design variables as recommended above. The initial values of the cut-off frequency and gain were $\hat{f}_c(0) = 0.05$ and $\hat{g}(0) = 0.01$. The convergence of the cut-off frequency and gain estimates is illustrated in Fig. 1(a) and (b), respectively, as a function of time.

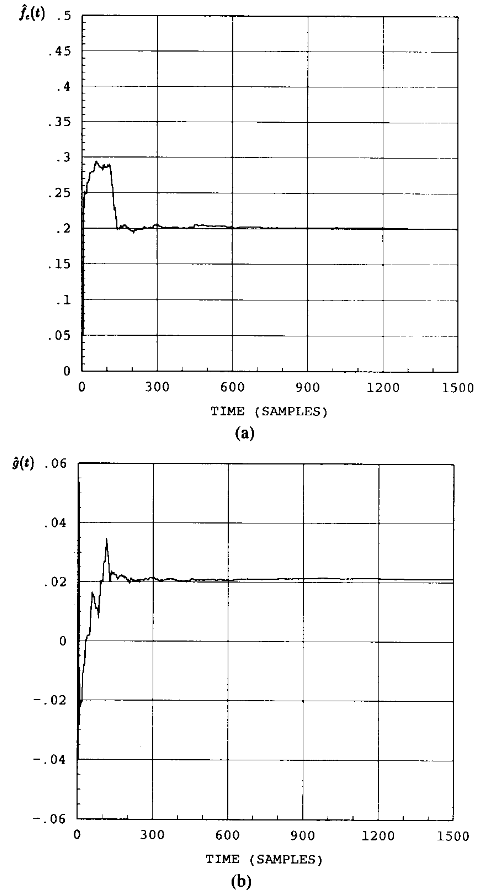

Fig. 1. Parameter estimates for Example III.2. (a) Cut-off frequency estimate (true value $f_c = 0.2$. (b) Gain estimate (true value $g = 2.12 \cdot 10^{-2}$).

C. Potential Applications

A potential application is filter design. Let S be a continuous-time filter with a low-pass frequency characteristic for which we seek a discrete-time approximation. To solve this problem we have two possibilities: 1) Use the bilinear z transform (if S is known); 2) Measure input/output data from S [using a wide-band input $u(t)$] and apply the proposed method to fit a low-pass Butterworth to these data. As explained in Section II, this is equivalent to fitting the frequency characteristic of a Butterworth filter to that of S.

For systems that can be approximated well by the low-pass Butterworth filter, the above algorithm can be applied for on-line decision on their Nyquist rate.

IV. Concluding Remarks

We have presented in this note a new algorithm for on-line estimation of the cut-off frequency and gain in a special constrained output error model. This model is of low-pass Butterworth type. When the system satisfies this structure or is similar to it and its true order is used, it is beneficial to apply the proposed algorithm to obtain more accurate results (by the parsimony principle [2]) and more efficient computations than is possible with the usual unconstrained models. In the white noise case, the algorithm becomes a recursive prediction error method, and as such its covariance attains the Cramér–Rao lower bound asymptotically when the true order is used (see, e.g., [3]). In the more general nonwhite noise case, the algorithm's covariance can be evaluated using methods described, e.g., in [2, ch. 7]. This algorithm can be used for filter design and adaptive Nyquist rate estimation. The basic method used here can be applied for deriving system identification algorithms for other constrained transfer functions, such as band-pass and band-stop. Extension to adaptive parameter estimation of constrained ARMA signals with unknown inputs in the presence of noise is presented in [12].

Appendix

The Derivative of Polynomial Coefficients with Respect to Polynomial Zeros

In this Appendix we provide a simple proof of the formula (13). Let C denote the following companion matrix:

$$C=\begin{bmatrix} -a_1 & 1 & & 0 \\ \vdots & & \ddots & \\ -a_{n-1} & 0 & & 1 \\ -a_n & 0 & \cdots & 0 \end{bmatrix}$$

associated with the polynomial $z^nA(z^{-1}) = z^n + a_1z^{n-1} + \cdots + a_n$. It is well known that the zeros $\{\lambda_k\}$ of $z^nA(z^{-1})$ are equal to the eigenvalues of C (see, e.g., [11]). Let

$$u_k=[u_{1,k}\ \cdots\ u_{n,k}]^T\neq 0$$

denote a (nonzero) eigenvector corresponding to λ_k. Thus,

$$Cu_k=\lambda_k u_k \tag{A.1}$$

or, in a more detailed form

$$\begin{cases} u_{2,k}=a_1u_{1,k}+\lambda_k u_{1,k} \\ \vdots \qquad \vdots \\ u_{n,k}=a_{n-1}u_{1,k}+\lambda_k u_{n-1,k} \\ 0=a_nu_{1,k}+\lambda_k u_{n,k}. \end{cases} \tag{A.2}$$

It readily follows by contradiction from (A.2) that $u_k \neq 0$ implies $u_{1,k} \neq 0$. Thus, we can set $u_{1,k} = 1$. In the following we assume that the eigenvector u_k has been normalized such that $u_{1,k} = 1$.

Using forward substitution, we find from (A.2) that

$$u_k=Hv_k \tag{A.3}$$

where H is the Hankel matrix

$$H=\begin{bmatrix} & & & 1 \\ & & \iddots & a_1 \\ & \iddots & & \vdots \\ 1 & a_1 & \cdots & a_{n-1} \end{bmatrix} \tag{A.4}$$

and

$$v_k=[\lambda_k^{n-1}\ \cdots\ \lambda_k\ 1]^T. \tag{A.5}$$

Next, it can easily be verified that

$$v_k^TC=\lambda_k v_k^T \tag{A.6}$$

which means that v_k is a left eigenvector of C. Left and right eigenvectors associated with different eigenvalues must be orthogonal

$$v_i^Tu_k=0 \qquad \text{for } i\neq k \tag{A.7}$$

which can be seen as follows. From (A.1) and (A.6), we get

$$\lambda_i v_i^Tu_k-\lambda_k v_i^Tu_k=v_i^TCu_k-v_i^TCu_k=0$$

which implies (A.7) since $\lambda_i \neq \lambda_k$. Writing out (A.7) for $i = 1, \cdots, n$, $i \neq k$, we obtain

$$\begin{bmatrix} \lambda_1^{n-1} \\ \vdots \\ \lambda_{k-1}^{n-1} \\ \lambda_{k+1}^{n-1} \\ \vdots \\ \lambda_n^{n-1} \end{bmatrix} + \begin{bmatrix} \lambda_1^{n-2} & \cdots & \lambda_1 & 1 \\ \vdots & & \vdots & \vdots \\ \lambda_{k-1}^{n-2} & \cdots & \lambda_{k-1} & 1 \\ \lambda_{k+1}^{n-2} & \cdots & \lambda_{k+1} & 1 \\ \vdots & & \vdots & \vdots \\ \lambda_n^{n-2} & \cdots & \lambda_n & 1 \end{bmatrix} \begin{bmatrix} u_{2,k} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ u_{n,k} \end{bmatrix} = 0. \tag{A.8}$$

Since $\{\lambda_k\}$ are distinct by assumption, the vander Monde matrix appearing in (A.5) is nonsingular. Therefore, u_k is uniquely determined by $\{\lambda_1, \cdots, \lambda_{k-1}, \lambda_{k+1}, \cdots, \lambda_n\}$ and, in particular, does not depend on λ_k. Using this property we get by differentiating (A.2) with respect to λ_k

$$\partial a/\partial\lambda_k=-u_k. \tag{A.9}$$

From (A.3) and (A.9) the Jacobian matrix $\partial a/\partial\lambda$ is equal to

$$\frac{\partial a}{\partial\lambda}=-[u_1\ \cdots\ u_n]=-HV \tag{A.10}$$

where H was defined in (A.4) and V is the vander Monde matrix $V = [v_1 \cdots v_n]$. Expression (15) is now proven immediately from the i, k entry of the matrices in (A.10). ■

References

[1] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: M.I.T. Press, 1983.

[2] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, to be published.

[3] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[4] B. Wahlberg and L. Ljung, "Design variables for bias distribution in transfer function estimation," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 134–144, Feb. 1986.

[5] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[6] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[7] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.

[8] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1124–1138, Oct. 1986.

[9] A. Nehorai and D. Starer, "An adaptive SSB carrier estimator," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Dallas, TX, Apr. 1987, pp. 2109–2112.

[10] T. Söderström and P. Stoica, "Some properties of the output error method," *Automatica*, vol. 18, pp. 93–99, Jan. 1982.

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

[12] A. Nehorai and P. Stoica, "Adaptive constrained ARMA signals in the presence of noise," in *Proc. ICASSP*, Dallas, TX, Apr. 1987, pp. 1003–1006.

Persistency of Excitation Results for Structured Nonminimal Models

MICHAEL HEYMANN

Abstract—**It is frequently convenient to employ specially structured nonminimal models for parameter estimation such as in the case of direct**

Manuscript received February 3, 1987. This work was supported in part by the Technion Fund for promotion of research.

The author is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel.

IEEE Log Number 8716170.

0018-9286/88/0100-0112$01.00