論 文
37～3～6

ARMAX모델의 매개변수 추정을 위한 최적입력 신호의 설계

Design of the Optimal Input Singals for Parameter Estimation in the ARMAX Model

李 錫 元* · 梁 興 錫**
(Suk-Won Lee · Heung-Suk Yang)

요 약

본 논문에서는 계통 전달함수와 잡음 전달 함수가 공통 분모 다항식을 갖는 ARMAX모델에서의 매개변수 추정을 위한 최적 입력의 설계 문제를 다루었다.

가정한 모델구조에 대한 정보 행렬을 상세히 유도하고 여파된 입력의 자기 상관 함수를 설계 변수로 사용함으로써 D-optimal 입력신호가 autoregressive moving average process로 실현될 수 있음을 보였다.

최적입력을 이용한 매개변수 추정에서 추정치의 표준편차가 감소함을 보이기 위해 컴퓨터 시뮬레이션을 행하였다.

Abstract- This paper considers the problem of the optimal input design for parameter estimtion in the ARMAX model in which the system and noise transfer function have the common denominator polynomial. Deriving the information matrix, in detail, for the assumed model structure and using the autocorrelation functin of the filtered input as design variables, it is shown that D-optimal input signal can be realized as an autoregressive moving average process. Computer simulations are carried out to show the standard-deviation reduction in the parameter estimtes using the optimal input signal.

1. Introduction

The problem of designing the optimal inputs for parameter estimation in dynamic systems has been extensively studied for certain classes of models.

The optimal input means that maximum information about the system can be extracted from the measured input-output data.

For the special case of the moving average model with input power constraint, Levin[1] has derived the optimal input condition which is independent of system parameters but the optimal input conditions will, in general, depend on the system parameters which are unknown.

To overcome this difficulty we have to do a prelimiary experiment to get the nominal values of the parameters. They will be considered as true values for computing the optimal input. Using this optimal input, a new improved model can be estimated. Much of the early works was surveyed and contained in[2] [3].

As the counterpart of Levin's result, a closed-loop input signal was derived analytically by use of a minimum variance feedback control law together with a white pertubation signal for an autoregressive model with an

*正 會 員：서울大 大學院 電氣工學科 博士課程
**正 會 員：서울大 工大 電氣工學科 敎授・工博.
接授日字：1987年 11月 24日
1次修正：1988年 2月 12日

output power constraint [4) 5)].

Using a Chebyshev system approach, Zarrop[6)] showed that under a certain condition, D-optimal design could be achieved with finite number of sinusoidal input frequencies without feedback.

Stoica and Söderström[7) 8)] proposed an useful input parameterization for the SISO transter function model with parametrically disjoint system and noise transter functions.

Ng, Goodwin and Söderström[9)] has shown that the minimum variance control strategy gives a D-optimality for a general linear system with output variance constraint by reparameterizing independently the system and noise transter function.

In this paper the input design problem is considered for the linear system model in which the system and noise transfer function have common parameters. Deriving the information matrix, in detail, it is shown that D-optimal open-loop input signal can be realized as an autoregressive moving average process that is easily implemented.

2. The input design problem

Consider the linear time-invariant discrete-time system model described by

$$A(z^{-1})y_k = B(z^{-1})u_k + C(z^{-1})e_k \tag{1}$$

or

$$y_k = \frac{B(z^{-1})}{A(z^{-1})}u_k + \frac{C(z^{-1})}{A(z^{-1})}e_k \tag{1'}$$

Where $\{y_k\}$ is a sequence of observations, $\{u_k\}$ is a sequence of inputs and $\{e_k\}$ is a white Gaussian noise sequence with variance σ^2 and z^{-1} is the unit backward shift operator.

Note that the system transfer function $B(z^{-1})/A(z^{-1})$ and the noise transfer function $C(z^{-1})/A(z^{-1})$ are interrelated, since they have common denomenator polynomial $A(z^{-1})$.

The polynomial $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ are given as follows.

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \cdots\cdots + a_n z^{-n} \\ B(z^{-1}) &= b_1 z^{-1} + \cdots\cdots + b_m z^{-m} \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \cdots\cdots + c_r z^{-r} \end{aligned} \tag{2}$$

It is asummed that the polynomial A(z), B(z) and C(z) are relatively prime and the polynomial A(z) and C(z) has all its zeros outside the unit circle.

The input signal $\{u_k\}$ is uncorrelated with $\{e_s\}$ for any k and s (open-loop signal).

θ is the vector of unknown parameters to be estimated more accurately.

$$\theta = [a_1 \cdots\cdots a_n \; b_1 \cdots b_m \; c_1 \cdots\cdots c_r]^T$$

A general measure of the estimation accuracy is given by the covariance matrix of the parameter estimates. If the estimator is asymptotically efficient (e.g. maximum likelihood estimator) the asymptotic covariance matrix is equal to the inverse of Fisher information matrix M defined by

$$M = E_{y|\theta}\,(\partial L/\partial\theta)^T(\partial L/\partial\theta) \tag{3}$$

Where L is the log-likelihood function, i.e., log p(Y| θ) and $(\partial L/\partial\theta)$ denotes a row vector with i-th component of $\partial L/\partial\theta_i$, $\theta_i$being the i-th component of θ.

In general, it is not possible to optimize the whole matrix. Therefore we have to select a suitable scalar function of the information matrix M to be optimized. The determinant of M is used in this paper.

Any optimal input design must also take account of the constraints on input signals. Otherwise the optimal input will clearly be an infinite power signal.

Now we can state the optimal input design as the problem of finding an input sequence $\{u_k\}$ that optimizes the suitable scalar accuracy function subject to the given constraints.

3. Information metrix structure and the design variables

A measure of efficiency in an identification experiment can be expressed as a scalar function of the information matrix which is defined by eq.(3). An expression for this information matrix is developed in detail.

For Gaussian data, the likelihood function can be written as

$$P(Y|\theta;U) = (2\pi\sigma^2)^{-\frac{N}{2}}\exp\{-(1/2\sigma^2)\sum_{k=1}^{N} w_k^2\}$$

$$\text{Where}\; Y = [y_1 \cdots y_N]^T \qquad U = [u_1 \cdots u_N]^T \tag{4}$$

$\{w_k\}$ is the residual sequence defined by

$$w_k = \{A(z^{-1})/C(z^{-1})\}[y_k - \{B(z^{-1})/A(z^{-1})\}u_k] \quad (5)$$

The log likelihood function L is easily given by

$$L = -(N/2)\log 2\pi - (N/2)\log\sigma^2 - (1/2\sigma^2)\sum_{k=1}^{N} w_k^2 \quad (6)$$

An expression for $(\partial L/\partial\theta)$ can be obtained from eq.(6).

$$\partial L/\partial\theta = -(1/\sigma^2)\sum_{k=1}^{N} w_k(\partial w_k/\partial\theta) \quad (7)$$

In general, the information will grow without bound as N increases. It is therefore reasonable to consider the average information matrix per sample defined by

$$\overline{M} = \lim_{N\to\infty}\frac{1}{N}M \quad (8)$$

It is assumed that $\sigma^2=1$ for convenience. Substituting eq.(7) into eq.(3), (8) yields

$$\overline{M} = E(\partial w_k/\partial\theta)^T(\partial w_k/\partial\theta) \quad (9)$$

An expression for $(\partial W_k/\partial\theta)$ can be obtained by differentiating eq.(5) with respect to the relevant parameters.

$$\partial w_k/\partial a_i = (B/CA)z^{-i}u_k + (1/A)z^{-i}w_k \quad (i=1,\cdots\cdots,n) \quad (10)$$

$$\partial w_k/\partial b_i = -(1/C)z^{-i}u_k \quad (i=1,\cdots\cdots,m) \quad (11)$$

$$\partial w_k/\partial c_i = -(1/C)z^{-i}w_k \quad (i=1,\cdots\cdots,r) \quad (12)$$

where, for convenience here and subsequently, we omit the argument and let $A=A(z^{-1})$, $B=B(z^{-1})$ and $C=C(z^{-1})$.

Note that $\{\partial w_k/\partial c_i\}$ do not depend on the input sequence $\{u_k\}$ and the second term in eq.(10) is resulted from the common denomenator polynomial between the system and noise transfer function.

Substituting eq.(10), (11) and (12) into eq.(9) gives the following expression for $\overline{M}$.

$$\overline{M} = \begin{pmatrix} M_{\alpha\alpha} & M_{\alpha\beta} \\ M_{\alpha\beta}^T & M_{\sigma\beta} \end{pmatrix} \quad (13)$$

Where the partition of $\overline{M}$ corresponds to the partition of θ between α and β, i.e.,

$$\theta^T = [\alpha^T \,\vdots\, \beta^T]$$
$$a^T = [a_1\cdots a_n\, b_1\cdots b_m]$$
$$\beta^T = [c_1\cdots\cdots c_r] \quad (14)$$

As an optimal criterion J, we shall use the determinant of the information matrix which is commonly used for input design. The inputs optimizing this function are usually called D-optimal inputs.

An important advantage of the determinant criterion is that it is invariant with respect to parameter transformations with nonsingular Jacobians[2) 3)].

$$J = \det\overline{M}$$
$$= \det(M_{\beta\beta})\det(M_{\alpha\alpha} - M_{\alpha\beta}M_{\beta\beta}^{-1}M_{\alpha\beta}^T) \quad (15)$$

$M_{\alpha\beta}$ and $M_{\beta\beta}$ are constant matrices independent of the input sequence $\{u_k\}$. If the system and noise transfer function have no common parameters, $M_{\alpha\beta}$ is shown to be null matrix. Only the $M_{\alpha\alpha}$ is dependent upon the input sequence $\{u_k\}$.

Therefore, in the following, only the input-dependent part of the whole information matrix, i.e., $M_{\alpha\alpha}$ will be considered in detail.

In view of eq.(9) and (13), $M_{\alpha\alpha}$ can be written as follows.

$$M_{\alpha\alpha} = E(\partial w_k/\partial\alpha)^T(\partial w_k/\partial\alpha) \quad (16)$$

Substituting eq.(10) and (11) into eq.(16), the above $M_{\alpha\alpha}$ can be expressed as the sum of two terms.

$$M_{\alpha\alpha} = M_u + M_c \quad (17)$$

Where M_u depends upon the input sequence and M_c is a constant matrix which has the diagonal elements $m_c(i,i) = E\{(1/A)w_{k-i}(1/A)w_{k-i}\}$ for $i=1,\cdots,n$ and the others are all zero. This nonzero term results from the common parameters between the system and noise transfer functions.

Substituting eq.(10) and (11) into eq.(16), the expression of M_u is given by

$$M_u = E\{(1/CA)\phi_k(1/CA)\phi_k^T\} \quad (18)$$

Where

$$\phi_k = [Bu_{k-1} \cdots Bu_{k-n}, -Au_{k-1} \cdots -Au_{k-m}]^T$$

Using the following Sylvester matrix.

$$S(B, -A) = \begin{bmatrix} 0 & b_1 \cdots\cdots b_m & 0 \cdots\cdots & 0 \\ 0 & 0 \;\; b_1 \cdots\cdots b_m & 0 \cdots\cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 \cdots\cdots & 0 \;\; b_1 \cdots\cdots & & b_m \\ -1 & -a_1 \cdots -a_n & 0 \cdots\cdots & 0 \\ 0 & -1 \;\; -a_1 \cdots\cdots -J_n & 0 \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 \cdots\cdots & 0 \;\; -a_1 \cdots & & -a_n \end{bmatrix} \quad (19)$$

M_u can be rewritten as the following expression.

$$M_u = E\{(1/CA)\,S(B,-A)\,U\,(1/CA)\,U^T S(B,-A)^T\}$$
$$= \sigma_u^2 S(B,-A)\,E\bar{U}\bar{U}^T S^T(B,-A) \quad (20)$$

Where $U = [u_{k-1} \cdots\cdots u_{k-n-m}]^T$

$$\bar{u}_k = (1/\sigma_{\bar{u}} AC)\,u_k, \quad \sigma_{\bar{u}}^2 = E\{(1/AC)\,u_k\}^2 \quad (21)$$

The above result states that M_u is completely determined by σ_u and the (n+m-1) following autocorrelations defined by

$$\rho_i = E\{\bar{u}_k \bar{u}_{k+i}\} \quad i=1, 2 \cdots\cdots, n+m-1 \quad (22)$$

Since $E\{\bar{u}_k^2\}=1$, the sequence $\{\rho_i\}$ can be viewed as the autocorrelation function of $\bar{u}_k$.

If the constraint is on the input variance, the allowable set D_u is

$$D_u = [\{u_k\} \mid E\{u_k\}^2 = \sigma_u^2]$$
$$= [\{u_k\} \mid \sigma_{\bar{u}}^2 E\{AC\bar{u}_k\}^2 = \sigma_u^2] \quad (23)$$

Since eq.(23) clearly depends on the first (n+r) autocorrelations of $\bar{u}_k$, the criterion J in eq.(15) can be optimized by choosing the σ_u^2 and the autocorrelation function ρ of $\bar{u}_k$.

$$\rho = [\rho_1 \;\; \rho_2 \cdots\cdots \rho_p]^T$$
$$\text{with } p = n + \max(m-1, r) \quad (24)$$

Since σ_u^2 is expressed as a function of $\{\rho_i\}$ from the constraint, $\{\rho_i\}$ are the only design variables for the optimization problem.

4. Optimal stochastic input realization

A sufficient condition for consistent estimates of parameters is that the input signal should be persistently exciting of appropriate order. This requirement makes it necessary that R_{n+m-1} be positive definite. Otherwise the system is not locally identifiable.

The matrix R_k is defined as

$$R_k = \begin{bmatrix} 1 & \rho_1 & \cdots\cdots\cdots & \rho_k \\ \rho_1 & 1 & \cdots\cdots\cdots & \rho_{k-1} \\ \rho_{k-1} & \cdots\cdots\cdots & 1 & \rho_1 \\ \rho_k & \rho_{k-1} \cdots\cdots & \rho_1 & 1 \end{bmatrix} \quad (25)$$

This additional constraint can be efficiently tested by using the partial autocorrelation function Φ_k properties described in the following Lemma[11].

Lemma

The following statements are equivalent:

i) $|\Phi_k| < 1 \quad k=1, \cdots, p$ and $\rho_0 > 0$

ii) R_p is positive definite

It is convenient to consider the set of the sequence $\{\rho_i\}$.

$$\mathbf{R} = \mathbf{B}(\mathbf{R}) + \mathbf{I}(\mathbf{R}) \quad (26)$$

where $\mathbf{I}(\mathbf{R}) = \{\rho \mid R_P > 0 \text{ -positive definite}\}$

and $\mathbf{B}(\mathbf{R}) = \{\rho \mid R_P > 0,\ R_{k-1} > 0,\ \det(R_k) = 0\}$

for some integer $k\ (m+n-1 < k < P)$

In case of $\rho \in I(R)$, the sequence $\{\Phi_k\}$ can be determined sequentially by the following Levinson-Durbin algorithm[10].

The relevant equations are :

$$\Phi_{k+1} = -a_{k+1,\,k+1}$$
$$= (\rho_{k+1} + a_{k,1}\,\rho_k + \cdots + a_{k,k}\rho_1)/\lambda_k^2 \quad (27.1)$$

$$a_{k+1,\,i} = a_{k,\,i} - \Phi_{k+1} a_{k,\,k+1-i}$$
$$(i=1, \cdots\cdots, k) \quad (27.2)$$

$$\lambda_{k+1}^2 = \lambda_k^2 (1 - \Phi_{k+1}^2) \quad (27.3)$$

with starting values

$$\lambda_1 = -a_{1,1} = \rho_1 \quad \lambda_1^2 = 1 - \Phi_1^2 \quad (27.4)$$

If we constrain ρ to belong to the set I(R) arbitrarily, the above recursion can be iterated for k=1, ⋯, p-1. Then the recursion eq.(27) will give, as a byproduct, the following autoregression.

$$(1+a_{p,1}-1+\cdots\cdots+a_{p,p}z^{-p})\bar{u}_k=\varepsilon_k \quad (28)$$

Where ε_k is a white noise with $E\{\cdot\varepsilon_k^2\}=\lambda_p^2$ which exactly matches the given autocorrelations $\{\rho_i\}$[11].

In such a case the following polynomial has all its zeros strictly outside the unit disc[8].

$$A_p(Z^{-1})+1+a_{p,1}z^{-1}+\cdots\cdots+a_{p,p}z^{-a} \quad (29)$$

Combining eq.(21) and (28), the optimal input u_k^* can be easily realized as an autoregressive moving average process.

$$A_p(z^{-1})u_k^*=\sigma\bar{u}^*A(z^{-1})C(z^{-1})\varepsilon_k$$
$$\text{with } E\{\epsilon_k^2\}=\lambda_p^2 \quad (30)$$

The coefficients of the polynomial A and C are given from the preliminary non-optimal input experiment.

If the constraint is on the output variance, the allowable set D_y can be also described by the first (m+r) autocorrelations of $\bar{u}_k$.

$$D_y=\{\{u_k\}|E\{(B/A)u_k\}^2=\sigma^2\}$$
$$=\{\{u_k\}|\sigma_u^2 E\{B(z^{-1})C(z^{-1})\bar{u}_k\}^2=\sigma_y^2\} \quad (31)$$

Thus the similar development as in the case of input variance constraint results in an open loop autoregressive moving average input signal.

5. Example with the computer simulations

The following example is used to demonstrate the availability of the optimal input signal.

$$y_k+a_1y_{k-1}=b_1u_{k-1}+e_k+c_1e_{k-1}$$

Where the true values of parameters are $a_1=-0.8$ $b_1=1.0$ and $c_1=0.7$. The noise variance is 0.5 and the input variance is constrained to be 1.

Maximum likelihood Estimator is used with 500 data points and the following forgetting factor[12] satisfying.

$$\lambda_t=0.99\lambda_{t-1}+0.01, \quad \lambda_o=0.95$$

Monte Carlo runs are carried out for 50 different realizations of the noise process and the white input signal is used to compare with the optimal input signal.

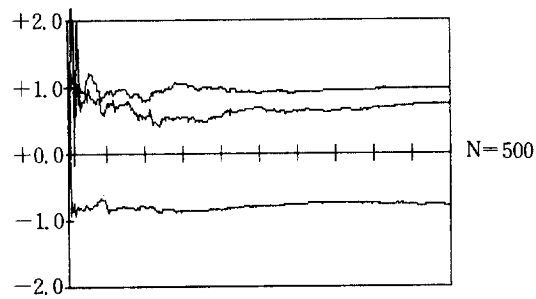

Fig. 1 Parmeter estimates using the optimal input

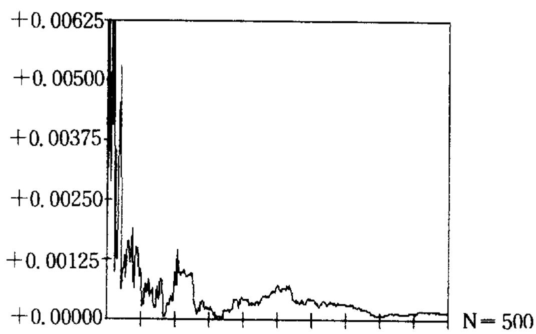

Fig. 2 MNE(k) using the white input

Table.1. mean±standard deviations computed from 50 reailzations

parameter		a_1	b_1	c_1
true value		−0.8	1.0	0.7
(a)	nominal value	−.7887987	.9956418	.6374495
	white input	−.80383±0.02797	1.00301±0.02979	.69552±0.04921
	optimal input	−.80381±0.01630	1.00086±0.01107	.70303±0.04815
(b)	nominal value	−.7676073	.9733296	.6472477
	white input	−.80190±0.02537	1.00221±0.03400	.69386±0.04935
	optimal input	−.80186±0.01736	1.00138±0.01542	.70160±0.04453

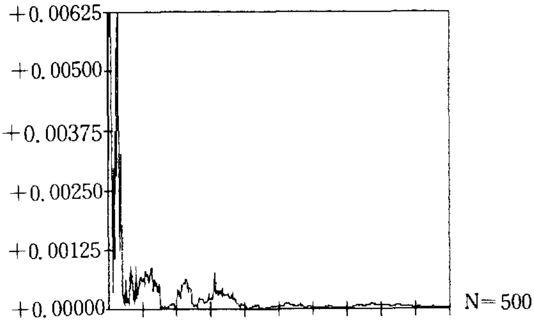

Fig. 3 MNE(k) using the optimal input

The simulation results are summarized in Table 1.

Fig. 1 shows a representative parameter estimates using the optimal input and Fig.2, 3 shows the value of the ensemble average of the mean normalized error which is defined by

$$\mathrm{MNE}(k) = \frac{1}{\mathrm{NoM}} \sum_{i=1}^{\mathrm{NOM}} \mathrm{mne}_i(k), \quad \mathrm{mne}(k) = \frac{\|\theta_t - \theta\|}{\|\theta_t\|}$$

where NoM is the Number of Monte Carlo runs, $\|\theta_t\|$ means the inner product of the true parameter vector and θ is the estimated parameter vector.

Note that the optimal input results in the smaller standard deviations of the parameter estimates and smaller MNE(k) value than the white input.

6. Conclusions

The optimal input design problem for the linear system which have the common parameters between the system and noise transfer functions, is considered.

Exploiting the assumed model structure, the information matrix is derived in detail. And choosing the autocorrelation function of the filtered input as the design variables, it is shown that the D-optimal open-loop input can be realized as an ARMA process under the input or output variance constrints.

Computer simulations are carried out to show the avilablity of the optimal input. The optimal inputs give rise to better estimation accuracy but futher simulations are necessary.

REFERENCES

1) M. J. Levin, "Optimal estimation of impulse response in the presence of noise.", IRE Trans. Circuit Theory, Vol. CT-7, pp. 50-56, 1960.
2) R. K. Mehra, "Optimal input signals for parameter estimation in dynamic systems – A survey and new results", IEEE Trans. Automat. Contr., Vol. AC-19, pp. 753-568, Dec. 1974.
3) G. C. Goodwin and R. L. Payne, Dynamic system Identification – Experiment design and data analysis. New York : Academic, 1977.
4) T. S. Ng, G.C.Goodwin and R. L. Payne, "On maximal accuracy estimation with output power constraints", IEEE Trans, Automat. Contr., Vol. AC-22, pp. 133-134, 1977.
5) T. S. Ng, E. H. Qureshi and Y. C. Cheah, "Optimal input design for an AR model with output constraints", Automatica, Vol. 20, pp. 359-363, 1984.
6) M. B. Zarrop, "A Chebyshev system approach to optimal input design", IEEE Trans. Automat. Contr., Vol. AC-24, pp. 687-698, Oct. 1979.
7) P. Stoica and T. Soderstrom, "A useful input parametrization for optimal experiment design", IEEE Trans, Automat. Contr., Vol. AC-27, pp. 986-989, Aug. 1982.
8) T. Soderstrom and P. G. Stoica, Instrumental variable methods for system identification. New York : Springer-Verlag, 1983.
9) T. S. Ng, G. C. Goodwin and T. Soderstrom, "Optimal expressiment design for linear systems with input-output constraints", Automatica, Vol. 12, pp. 571-577, 1977.
10) R. L. Ramsey, "Characterization of the partial autocorrelation function", Ann. Statist., Vol. 2, pp. 1296-1301, 1974.
11) C. T. Mullis and R. A. Roberts, "The use of second-order information in the approximation of discrete-time linear systems", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-24, pp. 226-238, 1976.
12) L. Ljung and T. Soderstrom, Theory and Practice of recursive identification, London : MIT press, 1983.