'87 韓國自動制御學術會議論文集 1987. 10.16~17

선형계통의 파라미터 추정을 위한 최적 확률 입력신호의 설계

양 홍 석 서울대학교
°이 석 원 대유공업전문대학

DESIGN OF THE OPTIMAL STOCHASTIC INPUTS FOR LINEAR SYSTEM PARAMETER ESTIMATION

H.S.Yang Dept. of Elec. Eng., Seoul National Univ.
S.W.Lee Dae Yeu Tech. Junior College

1. Introduction

It is well known that the parameter estimation accuracy is dependent upon the choice of input signal.

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

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 an optimal input conditions will, in general, depend on the system parameters which are unknown.

To overcome this difficulty we have to do a preliminary experiment to get 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 work was surveyed and contained in [3,4].

As the counterpart of Levin's result, a closed-loop input signal is derived analytically by use of a minimum variance feedback control law together with a white pertubation signal for an autoregressive model with an output power constraint [5].

Using a Chebyshev system approach Zarrop [7] showed that under a certain condition, D-optimal design could be achieved with finite number of sinusoidal input frequencies without feedback. This is clearly the strongest result we can hope to obtain in this approach.

Stoica and Söderström [8,9] proposed an useful input parameterization for the SISO transter function model with parametrically disjoint system and noise transter functions.

Ng, Goodwin and Söderström [10] 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 transfer 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. Exploiting the information matrix structure, it is shown that D-optimal open-loop input signal can be realized as an autoregressive moving average process.

2. The input design problem

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

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

or

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

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 defined as follows.

$$A(z^{-1}) = 1+a_1z^{-1}+.....+a_nz^{-n}$$
$$B(z^{-1}) = b_1z^{-1}+.....+b_mz^{-m} \qquad (2)$$
$$C(z^{-1}) = 1+c_1z^{-1}+.....+c_rz^{-r}$$

It is asummed that the polynomial A,B and C are relatively prime and the polynomial A and C 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.

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

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) the asymptotic covariance matrix is equal to the inverse of Fisher information matrix M defined by

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

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

In general, it is not possible to optimize the whole matrix. We then have to select a suitable scalar function of M to be optimized. 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 matrix structure and the input parameterization

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 matrix is developed in detail.

For Gaussian data, the likelihood function can be written as

$$P(Y|\theta;U) = (2\pi\sigma^2)\exp\{-(1/2\sigma^2)\sum_{K=1}^{N} w_K^2\} \quad (4)$$

Where $Y = [y_1 \ldots y_N]^T$

$U = [u_1 \ldots u_N]^T$

$\{w_K\}$ is the residual sequence given 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 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 easily 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

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

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

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

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

$$\partial w_K/\partial a_i = (B/CA)z^{-i}u_K + (1/A)z^{-i}w_K \quad (10)$$
$(i=1,\ldots,n)$

$$\partial w_K/\partial b_i = -(1/C)z^{-i}u_K \quad (11)$$
$(i=1,\ldots,m)$

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

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 $\{u_K\}$.

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

$$\bar{M} = \begin{bmatrix} M_{\alpha\alpha} & \vdots & M_{\alpha\beta} \\ \cdots & \vdots & \cdots \\ M_{\alpha\beta}^T & \vdots & M_{\beta\beta} \end{bmatrix} \quad (13)$$

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

$$\theta^t = [\alpha^t \mid \beta^T]$$
$$\alpha^T = [a_1 \ldots a_n\ b_1 \ldots b_m] \quad (14)$$
$$\beta^T = [c_1 \ldots c_r]$$

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 [3,8]

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

$M_{\alpha\beta}$ 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 information matrix $M_{\alpha\alpha}$ will be considered in detail.

$$M_{\alpha\alpha} = E\,(\partial w_K/\partial\alpha)^T(\partial w_K/\partial\alpha) \qquad (16)$$

$M_{\alpha\alpha}$ can be also expressed as the sum of two terms.

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

Where M_u depends upon the input sequence and M_c is a constant matrix which has the elements $m_c(i,j) = E\{(1/A)w_{K-i}(1/A)w_{K-j}\}$ for $i,j=1,\ldots,n$ and the others are all zeros. This term is resulted from the common parameters in system and noise transfer functions.

The expression of M_u is given by

$$M_u = E\,\{(1/CA)\varphi_K\,(1/CA)\varphi_K^T\} \qquad (18)$$

Where

$$\varphi_K = [Bu_{K-1}\ \ldots\ Bu_{K-n},\ -Au_{K-1}\ \ldots\ Au_{K-m}]$$

Using the following Sylvester matrix,

$$S(B,-A) = \begin{bmatrix} 0 & b_1 & \ldots\ldots & b_m & 0\ldots\ldots 0 \\ 0 & 0 & b_1 & \ldots\ldots & b_m\ 0\ldots 0 \\ & & & & \\ 0\ldots\ldots\ldots 0 & & b_1 & \ldots\ldots & b_m \\ -1 & -a_1 & \ldots\ldots & -a_n & 0\ldots 0 \\ 0 & -1 & -a_1\ldots\ldots & -a_n & 0\ldots 0 \\ & & & & \\ 0\ldots\ldots 0 & & -a_1 & \ldots\ldots & -a_n \end{bmatrix} \qquad (19)$$

$$M_u = E\{(1/CA)S(B,-A)U(1/CA)U^TS(B,-A)^T\}$$
$$= \sigma_{\bar{u}}^2 S(B,-A)E\bar{U}\bar{U}^TS^T(B,-A) \qquad (20)$$

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

$\bar{U} = [\bar{u}_{K-1}\ \ldots\ldots\ \bar{u}_{K-n-m}]^T$

$$\bar{u}_K = (1/\sigma_{\bar{u}}AC)u_K\ ,\quad \sigma_{\bar{u}}^2 = E\{(1/AC)u_K\} \qquad (21)$$

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

$$\rho_i = E\{u_K u_{K+i}\} \quad i=1,2\ldots,n+m-1 \qquad (22)$$

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

$$D_u = [\{u_K\}\,|\,E\{u_K\}^2=\sigma_u^2]$$
$$= [\{u_K\}\,|\,\sigma_{\bar{u}}^2E\{AC\bar{u}_K\}^2=\sigma_u^2] \qquad (23)$$

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 $\sigma_{\bar{u}}^2$ and the autocorrelation function ρ of $\bar{u}_K$.

$$\rho = [\rho_1\ \rho_2\ \ldots\ldots\ \rho_p]^T \qquad (24)$$

with $p = n+\max(m-1,r)$

Since $\sigma_{\bar{u}}^2$ is expressed as a function of $\{\rho_i\}$ from the constraint, $\{\rho_i\}$ are independent 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

$$R_K = \begin{bmatrix} 1 & \rho_1 & \ldots\ldots\ldots\ldots & \rho_K \\ \rho_1 & 1 & \ldots\ldots\ldots & \rho_{K-1} \\ & & & \\ & & 1 & \rho_1 \\ \rho_K & \rho_{K-1} & \ldots\ldots\ \rho_1 & 1 \end{bmatrix} \qquad (25)$$

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) $|\phi_K| < 1 \quad K=1,\dots,p$ and $\rho_0 > 0$

ii) R_p is positive definite

Let **R** be the set of the sequence $\{\rho_i\}$

It is convenient to consider

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

where $\mathbf{I(R)} = \{\rho | R_p > 0\}$

and $\mathbf{B(R)} = \{\rho | R_p \geq 0, R_{K-1} > 0, |R_K| = 0\}$

for some integer K $(m+n-1 \leq K \leq p)$

If the polynomial's order relation is $r > m-1$, $\rho \in \mathbf{B(R)}$ or $\rho \in \mathbf{I(R)}$.

If $r \leq m-1$, then $\rho \in \mathbf{I(R)}$.

In case of $\rho \in \mathbf{I(R)}$, the sequence $\{\phi_K\}$ can be determined sequentially by the following Levinson-Durbin algorithm [11].

The relevant equations are :

$$\phi_{K+1} = -a_{K+1,K+1} = (\rho_{K+1} + a_{K,1}\rho_K + \dots + a_{K,K}\rho_1)/\lambda_K^2 \tag{27.1}$$

$$a_{K+1,i} = a_{K,i} - \phi_{K+1} a_{K,K+1-i} \quad (i=1,\dots,k) \tag{27.2}$$

$$\lambda_{K+1}^2 = \lambda_K^2 (1 - \phi_{K+1}^2) \tag{27.3}$$

with starting values

$$\lambda_1 = -a_{1,1} = \rho_1 \qquad \lambda_1^2 = 1 - \phi_1^2 \tag{27.4}$$

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

$$(1 + a_{p,1}{}^{-1} + \dots + a_{p,p} z^{-p}) \bar{u}_K = \epsilon_K \tag{28}$$

Where ϵ_K is white noise with $E\{\epsilon_K^2\} = \lambda_p^2$ which exactly matches the given autocorrelations $\{\rho_i\}$ [12].

In such a case the polynomial

$$A_p(z^{-1}) = 1 + a_{p,1} z^{-1} + \dots + a_{p,p} z^{-p} \tag{29}$$

has all its zeros strictly outside the unit disc [9].

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

$$A_p(z^{-1}) u_K^* = \sigma_{\bar{u}}^* A(z^{-1}) C(z^{-1}) \epsilon_K \quad \text{with } E\{\epsilon_K^2\} = \lambda_p^2 \tag{30}$$

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 $\bar{u}_K$.

$$D_y = \{u_K | E\{(B/A)u_K\}^2 = \sigma_y^2\}$$
$$= \{u_K | \sigma_{\bar{u}}^2 E\{B(z^{-1}) C(z^{-1}) \bar{u}_K\}^2 = \sigma_y^2\} \tag{31}$$

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

5. Conclusions

The optimal input design problem for linear system which have the common parameters in the system and noise transfer functions.

Exploiting the assumed model structure and deriving the information matrix structure in detail, D-optimal open-loop stochastic input can be realized as an ARMA process under the input or output variance constraints.

In spite of the reduced order, it is necessary to develop an efficient algorithms for the optimation with respect to the ρ.

REFERENCES

[1] M.J.Levin, "Optimal estimation of impulse response in the presence of noise.", IRE Trans. Circuit Theory,Vol.CT-7, pp50-56,1960.

[2] Aoki and Staley, "On input signal synthesis in parameter identification", Automatica, Vol.6,pp431-440, 1970.

[3] 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 753-768,Dec.1974.

[4] G.C.Goodwin and R.L.Payne, Dynamic system Identification - Experiment design and data analysis. New York : Academic, 1977.

[5] 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, pp133-134, 1977.

[6] T.S.Ng, E.H.Qureshi and Y.C.Cheah, "Optimal input design for an AR model with output constraints", Automatica, Vol. 20, pp359-363, 1984.

[7] M.B.Zarrop, "A Chebyshev system approach to optimal input design", IEEE Trans. Automat. Contr., Vol. AC-24, pp687-698, Oct. 1979.

[8] P.Stoica and T.Söderström, "A useful input parametrization for optimal experiment design", IEEE Trans. Automat. Contr.,Vol. AC-27, pp986-989, Aug. 1982.

[9] T.Söderström and P.G.Stoica, Instrumental variable methods for system identification. New York :Springer-Verlag, 1983.

[10] T.S. Ng, G.C.Goodwin and T. Siderstrom, "Optimal experiment design for linear systems with input-output constraints", Automatica, Vol13,pp571-577,1977.

[11] R.L.Ramsey, "Characterization of the partial autocorrelation function", Ann.Statist.,Vol2,pp1296-1301,1974.

[12] 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, pp226-238,1976.