ON METHODS OF OPTIMUM INPUT SIGNAL SYNTHESIS

By

L. KEVICZKY

Department of Automation, Technical University, Budapest

(Received September 13, 1973)

Presented by Prof. Dr. F. CSÁKI

Introduction

In the practical parameter estimation problems the question of suitability or optimality of input signal for identification purposes are very often occurring. The input signal synthesis has been investigated by many authors and they pointed out that besides the persistently exciting condition other properties of input can also be demanded in order to fulfil some optimality criteria regarding the goodness of the parameter estimation. The relation between the input synthesis and the sensitivity analysis was examined by RAULT [11] and INOUE [5]. Many authors have dealt with this problem, approaching it from statistical aspects. LEVADI [9] investigated continuous system: approximating the output by linear series-expansion he set the minimization of the covariance matrix trace as an aim. NAHI [10] dealt with the maximization of information matrix trace, likewise for continuous systems. These works led to difficult computation methods (Fredholm equation, variation technique, and so on) which cannot be easily performed. AOKI [1, 2] called attention to the fact that the linear discrete-time systems can easily be rewritten into the form having linearity in parameters and it is more advisable to use the optimum input synthesis there. He elaborated a method for the maximization of trace of the information matrix when the equation error is white noise. This procedure minimizes asymptotically the Cramer-Rao lower bound, but its realization is getting ever more difficult concerning the computation technique for many samplings. On the basis of detailed analysis of experimental design methods we have pointed out that the methods of optimum input synthesis are different in case of static and dynamic identification and suggested the determinant of information matrix as a criterion [6]. In this paper procedures elaborated on the basis of the mentioned principles are presented.

In our investigations linear discrete-time system is used which can be described by stochastic difference equation

$$A(z^{-1})y(t) = B(z^{-1})u(t) + \lambda C(z^{-1})e(t) \tag{1}$$

where $\{y(t), u(t), e(t), t = 1, 2, \ldots, N\}$ are the measured output signal; the applied input signal and disturbance sequence (with normal distribution, zero mean, variance one and it is independent of the input), respectively. Here t means the discrete-time of the system (natural number) taking sampling time unit. Furthermore in equation (1)

$$\left.\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \ldots + a_n z^{-n} \\ B(z^{-1}) &= b_0 + b_1 z^{-1} + \ldots + b_m z^{-m}; \qquad m \leq n \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \ldots + c_k z^{-k}; \qquad k \leq n \end{aligned}\right\} \tag{2}$$

where z^{-1} is the backward shift operator [3]. Using (2) the system equation (1) can be written in the following form, as well:

$$y(t) = \mathbf{f}^T(u, y, t)\mathbf{p} + \lambda C(z^{-1})e(t), \tag{3}$$

where T means the transposition and

$$\mathbf{f}(u, y, t) = [u(t), u(t-1), \ldots, u(t-m), -y(t-1), \ldots, -y(t-n)]^T = \\ = [u(t), \mathbf{g}^T(t-1)]^T \tag{4}$$

and

$$\mathbf{p} = [b_0, b_1, \ldots, b_m, a_1, \ldots, a_n]^T. \tag{5}$$

Assuming $e(t)$ has a normal distribution it can be deduced that the information matrix regarding the estimation of parameter vector $\mathbf{p}$ is, for N samplings [7, 8]:

$$\mathbf{J}_N = \frac{1}{\lambda^2} \sum_{t=1}^{N} \mathbf{f}(u^F, x^F, t)\mathbf{f}^T(u^F, x^F, t) \tag{6}$$

where

$$\mathbf{f}(u^F, x^F, t) = [u^F(t), \ldots, u^F(t), -x^F(t-1), \ldots, -x^F(t-n)]^T. \tag{7}$$

Here

$$C(z^{-1})u^F(t) = u(t); \qquad C(z^{-1})x^F(t) = x(t) \tag{8}$$

and

$$A(z^{-1})x(t) = B(z^{-1})u(t) \tag{9}$$

(i.e. $x(t)$ is the output without noise).

Let the covariance matrix of the parameter estimate $\hat{\mathbf{p}}_N$, obtained from N samplings, be:

$$\mathbf{K}_N = E\{(\hat{\mathbf{p}}_N - \mathbf{p})(\hat{\mathbf{p}}_N - \mathbf{p})^T\} \tag{10}$$

where $E\{\ldots\}$ and $\wedge$ mean the expected and estimated value, respectively. The Cramer-Rao lower bound gives a limit for $\mathbf{K}_N$ according to which $\mathbf{K}_N \geqq \geqq \mathbf{J}_N^{-1}$ where the inequality sign indicates that the difference matrix is a non-negative definite. It is obvious that the strength of this equality is characterized in the same way by the determinant as a trace of these two matrices. Moreover, since the maximization of trace of $\mathbf{J}_N$ is also used for the minimization of trace of $\mathbf{J}_N^{-1}$ which is only asymptotically efficient, it is more worth-while to maximize the determinant of $\mathbf{J}_N$, since this directly minimizes the determinant of $\mathbf{J}_N^{-1}$, too. In the following the minimization of the determinant of the covariance matrix, for white noise equation error, and of inverse of the information matrix for general noise, is presented. The input signal $u(t)$ is assumed to be an amplitude constrained signal:

$$u_{\min} \leqq u(t) \leqq u_{\max} . \tag{11}$$

Case of least-squares structure

First, let us consider the case of least-squares (LS) structure when the equation error — the term $\lambda C(z^{-1})e(t)$ in (3) — is white noise, i.e. $C(z^{-1}) = 1$. In this case the well-known LS estimation coincides with the maximum likelihood (ML) and gives an unbiased estimation. The well-known recursive version of the method is

$$\hat{\mathbf{p}}_{N+1} = \hat{\mathbf{p}}_N + \mathbf{R}_{N+1}[y(N+1) - \mathbf{f}^T(u, y, N+1)\hat{\mathbf{p}}_N]\mathbf{f}(u, y, N+1) \tag{12}$$

where

$$\mathbf{R}_{N+1} = \mathbf{R}_N - \frac{\mathbf{R}_N\mathbf{f}(u,y,N+1)\mathbf{f}^T(u,y,N+1)\mathbf{R}_N}{1 + \mathbf{f}^T(u,y,N+1)\mathbf{R}_N\mathbf{f}(u,y,N+1)} = \frac{1}{\lambda^2}\mathbf{K}_{N+1} \tag{13}$$

and the covariance matrix

$$\mathbf{K}_{N+1} = \left[\sum_{t=1}^{N+1} \mathbf{f}(u,y,t)\mathbf{f}^T(u,y,t)\right]^{-1}\lambda^2 . \tag{14}$$

It can be deduced that the increasing rate of determinant of $\mathbf{K}_N^{-1}$ is

$$\frac{|\mathbf{K}_{N+1}^{-1}|}{|\mathbf{K}_N^{-1}|} = 1 + \mathbf{f}^T(u,y,N+1)\mathbf{R}_N\mathbf{f}(u,y,N+1) \tag{15}$$

if the sampling number N is changed to $N+1$ [8]. Here $|\ldots|$ means the determinant of a matrix. A locally optimum strategy can be formed for the maximization of $|\mathbf{K}^{-1}|$ (which is equivalent to the minimization of $|\mathbf{K}|$)

if the quadratic form on the right side of (15) is maximized by $u(N+1)$ in every step. (This is a so-called locally optimum strategy because only the next step is optimized each time.) Partitionate $\mathbf{R}_N$ according to (4):

$$\mathbf{R}_N = \begin{bmatrix} r_N & \mathbf{d}_N^T \\ \mathbf{d}_N & \mathbf{Q}_N \end{bmatrix} = \frac{1}{\lambda^2}\mathbf{K}_N \ . \tag{16}$$

By this designation we obtain for (15):

$$\frac{|\mathbf{K}_{N+1}^{-1}|}{|\mathbf{K}_N^{-1}|} = u^2(N+1)r_N + 2u(N+1)\mathbf{g}^T(N)\mathbf{d}_N + \mathbf{g}^T(N)\mathbf{Q}_N\mathbf{g}(N) + 1 \tag{17}$$

i.e. this means a parabola as a function of $u(N+1)$ having its vertex downwards (since r_N is absolutely positive). It can easily be seen that the following expression gives the optimum value of $u^0(N+1)$ — on a constrained region given by (11) — ensuring the global maximum of (17):

$$u^0(N+1) = \begin{cases} u_{\max}\,, & \text{if } \dfrac{u_{\min}+u_{\max}}{2} + \dfrac{\mathbf{g}^T(N)\mathbf{d}_N}{r_N} \geqq 0 \\ u_{\min}\,, & \text{if } \dfrac{u_{\min}+u_{\max}}{2} + \dfrac{\mathbf{g}^T(N)\mathbf{d}_N}{r_N} < 0. \end{cases} \tag{18}$$

Determination of the global maximum can be seen in Fig. 1. Here $u^*(N) = -(\mathbf{g}^T(N)\mathbf{d}_N)/r_N$ is the vertex of the parabola. In this way the optimal input signal can be generated by the on-line connection with the process for identification purposes. $u^0(t)$ depends only on $y(t-1), \ldots, y(t-n)$, consequently,

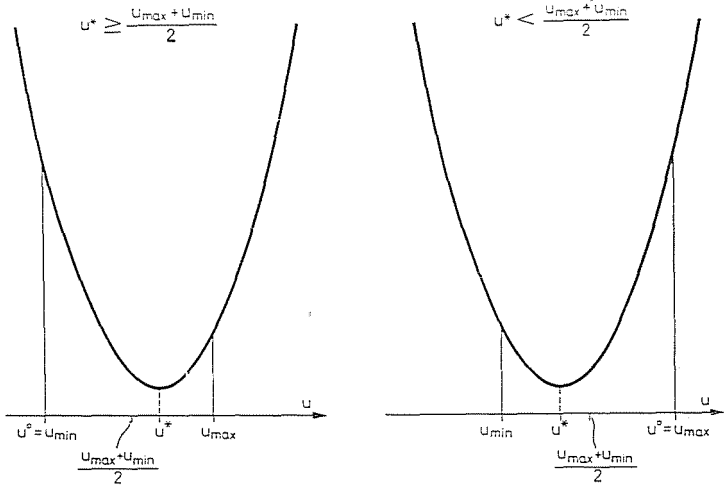

Fig. 1

on values $e(t-1), \ldots, e(t-n)$ and the independence of input signal and the measurement noise is valid in this case, too. The algorithm (18) can easily be realized because only the data applied so far are needed to generate the new $u^0(t)$. Scheme of algorithm generating optimum input signal is shown on Fig. 2.

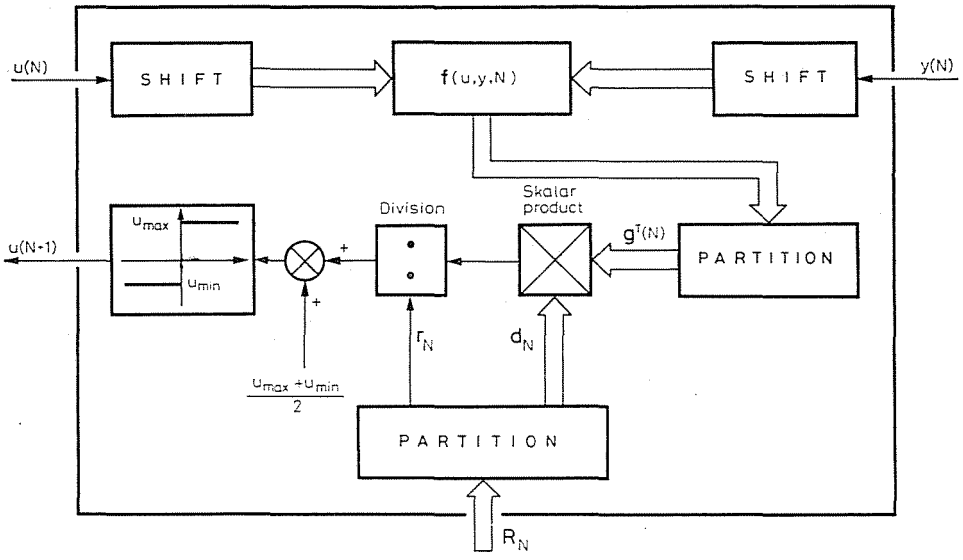

Fig. 2

Instead of the local minimization of the determinant of the covariance matrix, the local minimization of its trace can also be chosen, since

$$tr(\mathbf{R}_{N+1}) = tr(\mathbf{R}_N) - \frac{\mathbf{f}^T(u,y,N+1)\mathbf{R}_N\mathbf{R}_N\mathbf{f}(u,y,N+1)}{1+\mathbf{f}^T(u,y,N+1)\mathbf{R}_N\mathbf{f}(u,y,N+1)} \tag{19}$$

and the second term of (19) right side is a second order rational fractional function of $u^0(N+1)$. (Here $tr(\ldots)$ means the trace of a matrix.) Thus, to generate optimum input (i.e. to minimize $tr(\mathbf{R})$ or $tr(\mathbf{K})$, namely, $tr(\mathbf{K})$ is proportional to $tr(\mathbf{R})$) the global maximum of a far more difficult function than (17) has to be determined, therefore, it is reasonable to use (18).

It can be established from the comparison of equations (6) and (14) that the same algorithm can be used for the minimization of the determinant of $\mathbf{J}^{-1}$ (i.e. for the maximization of $|\mathbf{J}|$) as what was used in (18) for the minimization of $|\mathbf{K}|$. Since $C(z^{-1}) \equiv 1$, $u^F(t) = u(t)$ and $x^F(t) = x(t)$. So $\mathbf{f}(u,y,t)$ must be replaced by $\mathbf{f}(u,x,t)$. Here $x(t)$ is unknown and can be produced by prediction:

$$\hat{x}(t) = \sum_{i=0}^{m} \hat{b}_i u(t-i) - \sum_{i=1}^{n} \hat{a}_i \hat{x}(t-i) \ . \tag{20}$$

Obviously, now $\mathbf{g}(u, \hat{x}, N)$ is in (18) instead of $\mathbf{g}(N) = \mathbf{g}(u, y, N)$. The local maximization of the determinant of the information matrix does not need the knowledge of output but needs $\hat{x}(t)$, i.e. the parameter estimates $\hat{a}_i$, $\hat{b}_i$. Thus, this strategy can be performed by an on-line way (simultaneously with the identification) but it can be done in off-line way, too, in the apriori knowledge of parameter estimates. This means the optimum input sequences (so-called D-optimum) can be generated in advance to the identification. Unfortunately, we have to know far more for the synthesis (the parameter estimates themselves) and this strategy can only be realized by the successive application of the off-line identification methods.

The case of maximum likelihood structure

Such an algorithm which locally minimizes the determinant of the covariance matrix for the general form of system equation (3) cannot be constructed similarly to (18) but the algorithm suggested for the local maximization of the determinant of information matrix in LS structure can be generalized for this case. Comparing (6) and (14) it can be seen that $\mathbf{f}(u^F, x^F, t)$ corresponds to $\mathbf{f}(u, y, t)$, formally. Since $x(t)$ is unknown $\mathbf{f}(u^F, \hat{x}^F, t)$ can be determined from the predicted value $\hat{x}(t)$. By the filtering equations (8):

$$\mathbf{f}(u^F, \hat{x}^F, N+1) = [u(N+1) - q,\ \mathbf{g}^T(u^F, \hat{x}^F, N)]^T \tag{21}$$

where

$$q = \sum_{i=1}^{k} c_i u^F(t-i)\,. \tag{22}$$

Now the optimum $u^0(N+1)$ is computed according to

$$u^0(N+1) = \begin{cases} u_{\max}\,, & if\ \dfrac{u_{\min} + u_{\max}}{2} + \dfrac{\mathbf{g}^T(N)\mathbf{d}_N}{r_N} \geqq q \\ u_{\min}\,, & if\ \dfrac{u_{\min} + u_{\max}}{2} + \dfrac{\mathbf{g}^T(N)\mathbf{d}_N}{r_N} < q\,. \end{cases} \tag{23}$$

(It is to be mentioned here that now r_N and $\mathbf{d}_N$ issue from the partitionation of $\mathbf{R}_N = (\lambda^2 \mathbf{J}_N)^{-1}$.) Since there is no good on-line method to estimate the coefficients of $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ the off-line input signal synthesis — suggested for the LS structure, as well — should be applied.

In a special case, when $C(z^{-1}) = 1/H(z^{-1})$, HASTINGS' on-line method can be used to estimate the parameters of $H(z^{-1}) = 1 + h_1 z^{-1} + \ldots + h_s z^{-s}$ [4]. The formula (23) is also valid taking into account that

$$q = -\sum_{i=1}^{s} h_i u(t-i)\,. \tag{24}$$

Simulation results

The effectiveness of the elaborated methods has been proved by several simulation examples [7, 8]. Here only some of them are presented. The process was simulated for several structures according to equation (1). In these examples results of identification obtained by pseudo-random-binary-sequences (PRBS) and D-optimal sequences as inputs are compared. In both cases the input signal was an amplitude constrained signal: $-1 \leq u(t) \leq 1$. The following measure was used to compare the identification results obtained by PRBS and D-optimum input:

$$w(N) = w(t) = \frac{(\hat{\mathbf{p}}_N - \mathbf{p})^T(\hat{\mathbf{p}}_N - \mathbf{p})}{(\hat{\mathbf{p}}_0 - \mathbf{p})^T(\hat{\mathbf{p}}_0 - \mathbf{p})} .$$

Let us consider the following numerical examples:

1. *Example*

The equation of simulated process is

$$(1 - 0.8z^{-1})y(t) = (0.6 + 0.2z^{-1})u(t) + \lambda e(t)$$

and $\lambda = 0.2$.

In Fig. 3 the values $w(t)$ are shown for the cases of PRBS input and D-optimum input synthetised on-line by equation (18) when the identification was performed by on-line LS method.

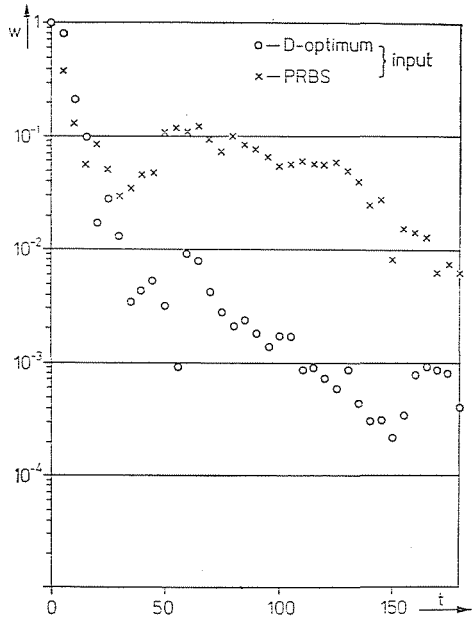

Fig. 3

2. *Example*

The equation of process is

$$(1 - 0.7z^{-1} + 0.1z^{-2})y(t) = (0.6 + 0.2z^{-1})u(t) + \lambda e(t)$$

and $\lambda = 0.2$. The values $w(t)$ can be seen in Fig. 4.

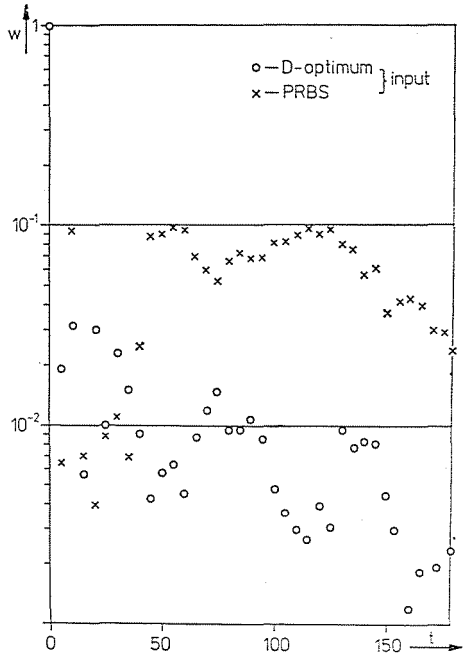

Fig. 4

3. *Example*

The simulated equation is

$$(1 - 1.5z^{-1} + 0.66z^{-2} - 0.08z^{-3})y(t) = (1 + z^{-1} - 0.5z^{-2})u(t) + \lambda e(t)$$

and $\lambda = 0.3$. The values $w(t)$ are presented in Fig. 5 for both input signals.

4. *Example*

In order to identify the parameters of system equation

$$(1 - 1.5z^{-1} + 0.7z^{-2})y(t) = (1.0z^{-1} + 0.5z^{-2})u(t) + \lambda(1 - 1.8z^{-1} + 0.9z^{-2})e(t)$$

an off-line ML estimation was performed. The following table contains the determinant of covariance matrix of parameter estimates and the variances of

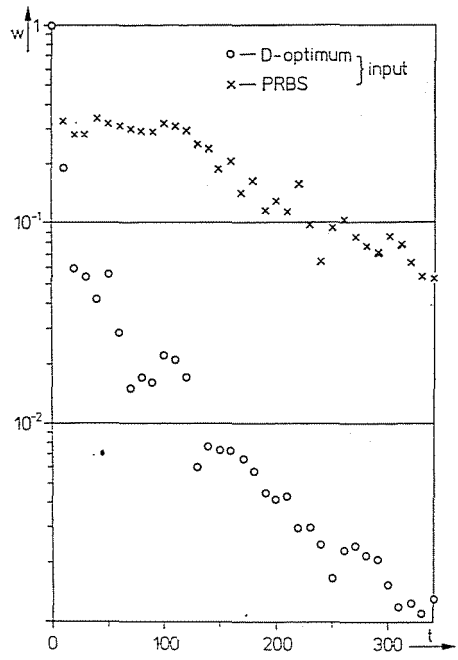

Fig. 5

the estimated values in case of PRBS input and D-optimum input synthetised by (23).

	PRBS	D-opt.
var $(\hat{a}_1)$	0.0058	0.0029
var $(\hat{a}_2)$	0.0042	0.0021
var $(\hat{b}_1)$	0.0369	0.0193
var $(\hat{b}_2)$	0.0516	0.0271
$\lvert\mathbf{K}_N\rvert$	$2.57 \cdot 10^{-18}$	$4.35 \cdot 10^{-21}$

$N = 500$
$\lambda = 1.0$

Conclusions

In this paper a locally optimum algorithm is suggested for the input signal synthesis by means of which computationally very simple method can be given to minimize the determinant of covariance matrix if the equation error is white noise, and in another case to maximize the determinant of information matrix. The local optimality is actually an analogous concept of the "one-stage

control" of stochastic control theory. Realization of the global optimality which corresponds to the "N-stage control" means computationally a far more difficult problem. (But it can be solved on the basis of this paper by dynamic programming.)

The suggested algorithms can be employed profitably in the off-line input synthesis to improve the result of identification step by step.

Summary

In this paper it was investigated how to generate optimal input signal series for the identification of linear discrete-time system in order to improve the accuracy of estimate. The determinant of the covariance matrix or the inverse of information matrix are considered as a measure of the error in the parameter estimate. We suggest very simple methods for the minimization of these criteria in case of amplitude constrained input signal.

References

1. AOKI, M.—STALEY, R. M.: On Approximate Input Signal Synthesis in Plant Parameter. Identification. First Hawaii Int. Conf. on Syst. Sciences, University of Hawaii, 1968. 363.
2. AOKI, M.—STALEY, R. M.: On Input Signal Synthesis in Parameter Identification. Automatica, **6**, 1970. 431.
3. ASTRÖM, K. J.—EYKHOFF, P.: System Identification — A Survey, IFAC, Prague, 1970.
4. HASTINGS, R. J.—SAGE, M. W.: Recursive Generalized Least-Squares Procedure for on-line Identification of Process Parameters. Proc. IEE., **116**, 1969, 2057.
5. INOUE, K.—OGINO, K.—SAWARAGI, Y.: Sensitivity Synthesis of Optimal Input for Parameter Identification. IFAC, Prague, 1970.
6. KEVICZKY, L.—BÁNYÁSZ, Cs.: Optimal Identification by Simulation of the Information Obtained from the Processes. Summer Computer Simulation Conference, San Diego, 1972.
7. KEVICZKY, L.: On Some Questions of Input Signal Synthesis. Report 7226 (B), Lund Institute of Technology, Division of Automatic Control.
8. KEVICZKY, L.: Regressziós kísérletek tervezése. Kandidátusi értekezés, BME Automatizálási Tanszék, 1972.
9. LEVADI, V. S.: Design of Input Signals for Parameter Estimation. IEEE Trans. on Aut. Control, *AE*—**11**, 205.
10. NAHI, N. E.—WALLIS, D. E.: Optimal Inputs for Parameter Estimation in Dynamic Systems with White Observation Noise. Proc. JACC, 1969, 506.
11. RAULT, A.—POULIQUEN, R.—RICHALET, J.: Sensitivity and Identification. IFAC, Dubrovnik, 1968.

Dr. László KEVICZKY, H-1521 Budapest