KYBERNETIKA — VOLUME *19* (1983), NUMBER 4

A SIMPLE RECURSIVE ESTIMATOR WITH ON-LINE ORTHOGONALIZATION OF THE INPUT SEQUENCE

JAROSLAV MARKL

The optimal one-step identification algorithm for noise free linear systems is generalized by means of on-line orthogonalization of the input sequence. The suggested generalization employs a fixed small amount of old data and thus radically accelerates convergence of estimation. An additional minor modification can guarantee numerical stability of the proposed algorithm.

1. INTRODUCTION

The optimal one-step identification algorithm ("learning identification", Kaczmarz's algorithm) [1, 6] is very attractive for various applications thanks to its numerical simplicity. However, its applicability is often restricted due to the slow convergence. An extremely slow convergence rate occurs in the following cases:

1) All input vectors have almost the same direction. This occurs when the variations of coordinates of the input vectors are relatively small in comparison with the magnitudes of their mean values.
2) The changes of direction of input vectors are relatively small. This occurs when the input sequence is significantly correlated. This is the case of identification of dynamical systems.

There are some ways how to prevent this phenomenon, however. A simple remedy to improve the slow convergence due to the former cause is to work with deviations instead of the original input quantities. A simple method precluding the slowness due to the latter cause is described in [1, 4].

This paper presents a method (inspired by [7]) that prevents the retardation due to the both causes mentioned above. It also substantially improves the rate of convergence even in the best case when the input sequence is both centred and uncorrelated. Approaches, similar to that of [7], have been recently presented in [2, 3].

The structure of the paper is as follows. First a generalization of the optimal

one-step algorithm is formulared (Section 2). In Section 3, the orthogonalization of the input sequence is investigated. In Section 4, convergence and some other properties of generalized algorithm are proved. In Section 5, the theoretical statements are illustrated by experimental results. Section 6 contains some recommendations for practical use of the algorithm.

2. THE ALGORITHM

The following notation will be used:

- t discrete time, $t = 0, 1, 2, \ldots$
- y_t scalar output
- $\mathbf{z}_t$ n-dimensional input vector
- $\mathbf{b}$ n-dimensional vector of unknown parameters
- $\mathbf{b}_t^{\wedge}$ estimate of vector $\mathbf{b}$
- $\mathbf{d}_t$ error of estimate, $\mathbf{d}_t = \mathbf{b} - \mathbf{b}_t^{\wedge}$

The optimal one-step algorithm for parameter estimation of the linear system

$$y_t = \mathbf{z}_t^{\mathrm{T}}\mathbf{b} \tag{2.1}$$

has the form

$$\mathbf{b}_t^{\wedge} = \mathbf{b}_{t-1}^{\wedge} + \frac{y_t - \mathbf{z}_t^{\mathrm{T}}\mathbf{b}_{t-1}^{\wedge}}{\mathbf{z}_t^{\mathrm{T}}\mathbf{z}_t}\,\mathbf{z}_t \tag{2.2}$$

This algorithm can be easily derived by minimizing the criterion

$$I(\mathbf{b}_t^{\wedge}) = (\mathbf{b}_t^{\wedge} - \mathbf{b}_{t-1}^{\wedge})^{\mathrm{T}}(\mathbf{b}_t^{\wedge} - \mathbf{b}_{t-1}^{\wedge}) \tag{2.3}$$

subject to the condition

$$y_t = \mathbf{z}_t^{\mathrm{T}}\mathbf{b}_t^{\wedge} \tag{2.4}$$

Thus, the new estimate $\mathbf{b}_t^{\wedge}$ complies with the new data $(y_t, \mathbf{z}_t)$ and at the same time it has the minimal distance from the previous estimate $\mathbf{b}_{t-1}^{\wedge}$.

The object of the theoretical studies in Sections 3 and 4 is the following generalization of the recursive procedure (2.2).

$$\zeta_t = \mathbf{z}_t - \sum_{j=1}^{p_t} \frac{\mathbf{z}_t^{\mathrm{T}}\zeta_{t-j}}{\zeta_{t-j}^{\mathrm{T}}\zeta_{t-j}}\,\zeta_{t-j} \tag{2.5}$$

$$\mathbf{b}_t^{\wedge} = \mathbf{b}_{t-1}^{\wedge} + \frac{y_t - \mathbf{z}_t^{\mathrm{T}}\mathbf{b}_{t-1}^{\wedge}}{\zeta_t^{\mathrm{T}}\zeta_t}\,\zeta_t \tag{2.6}$$

where p_t, $t = 1, 2, \ldots$ are adjustable parameters satisfying the following conditions:

$$p_t \in \{0, 1, 2, \ldots, n-1\} \tag{2.7}$$

$$p_t \leqq t - 1 \tag{2.8}$$

$$p_t \leqq p_{t-1} + 1 . \tag{2.9}$$

The quantity p_t will be referred to as the depth of orthogonalization. Note that the choice $p_t \equiv 0$ in (2.5), (2.6) leads back to the original algorithm (2.2). In Section 6 the algorithm (2.5), (2.6) will be slightly further generalized to guarantee its applicability under practical circumstances.

3. ORTHOGONALIZATION OF THE INPUT SEQUENCE

The formula (2.5) which transforms the sequence $\{\mathbf{z}_t\}$ into the sequence $\{\zeta_t\}$, together with constraints (2.7)–(2.9), represents some generalization of the well-known Gram-Schmidt procedure. Choosing $p_t = t - 1$, $t = 1, 2, \ldots, n$ we get the original Gram-Schmidt orthogonalization of n n-dimensional vectors. In the general case the sequence $\{p_t\}$ may be an arbitrary sequence fulfilling the conditions (2.7)–(2.9), namely an infinite one. Characteristic features of the generalized orthogonalization process are stated in the following theorem:

Theorem 3.1. Let the sequence $\{\zeta_t\}$ be generated by (2.5) in accordance with (2.7)–(2.9). Let further

$$\zeta_{t-p_t}, \zeta_{t-p_t+1}, \ldots, \zeta_{t-1} \tag{3.1}$$

be an orthogonal system of non-zero vectors. Then

1) If vector $\mathbf{z}_t$ is linearly independent of the system (3.1) then also the extended system

$$\zeta_{t-p_t}, \zeta_{t-p_t+1}, \ldots, \zeta_{t-1}, \zeta_t$$

is an orthogonal system of non-zero vectors.

2) If vector $\mathbf{z}_t$ is linearly dependent on the system (3.1) then $\zeta_t = 0$.

Proof. Simple and straightforward.

Somewhat stronger assertion can be proved for the very beginning of the orthogonalization process:

Theorem 3.2. Let

$$\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{\tau-1} \quad \tau \leqq n \tag{3.2}$$

be a system of linearly independent vectors and the vectors ζ_t, $t = 1, 2, \ldots, \tau$ be generated by (2.5) with the depth $p_t = t - 1$. Then it holds:

1) If the vector $\mathbf{z}_\tau$ is linearly independent of the system (3.2) then the system

$$\zeta_1, \zeta_2, \ldots, \zeta_{\tau-1}, \zeta_\tau$$

is an orthogonal system of non-zero vectors.

2) If the vector $\mathbf{z}_t$ is linearly dependent on the system (3.2) then $\zeta_\tau = 0$.

Proof. By induction on τ.

The following theorem and its corollaries enable us to evaluate the numerical stability of the orthogonalization process.

Theorem 3.3. Under assumptions of theorem 3.1 it holds

$$\zeta_t^{\mathrm{T}}\zeta_t = \mathbf{z}_t^{\mathrm{T}}\mathbf{z}_t - \sum_{j=1}^{p_t} \frac{(\mathbf{z}_t^{\mathrm{T}}\zeta_{t-j})^2}{\zeta_{t-j}^{\mathrm{T}}\zeta_{t-j}} \tag{3.3}$$

Proof. The proof follows immediately from (2.5).

Corollary 3.1. Let the assumptions of Theorem 3.1 hold. Then:

1) The inequality

$$\zeta_t^{\mathrm{T}}\zeta_t \leqq \mathbf{z}_t^{\mathrm{T}}\mathbf{z}_t \tag{3.4}$$

is valid and the equality in (3.4) holds if and only if the vector $\mathbf{z}_t$ is orthogonal to the linear subspace spanned by the vectors in (3.1).

2) Let $(\zeta_t^{\mathrm{T}}\zeta_t)^{(k)}$ denote the value of expression $\zeta_t^{\mathrm{T}}\zeta_t$ provided the depth of orthogonalization is $p_t = k$. The inequality

$$(\zeta_t^{\mathrm{T}}\zeta_t)^{(k)} \leqq (\zeta_t^{\mathrm{T}}\zeta_t)^{(k-1)} \tag{3.5}$$

is valid and the equality in (3.5) holds if and only if the vector $\mathbf{z}_t$ is orthogonal to ζ_{t-k}.

3) For $t \geqq n$ the relations (3.4), (3.5) can be generalized as follows:

$$0 \leqq (\zeta_t^{\mathrm{T}}\zeta_t)^{(n-1)} \leqq \ldots \leqq (\zeta_t^{\mathrm{T}}\zeta_t)^{(2)} \leqq (\zeta_t^{\mathrm{T}}\zeta_t)^{(1)} \leqq (\zeta_t^{\mathrm{T}}\zeta_t)^{(0)} = \mathbf{z}_t^{\mathrm{T}}\mathbf{z}_t \tag{3.6}$$

If the vector $\mathbf{z}_t$ is not orthogonal to any of the vectors $\zeta_{t-1}, \zeta_{t-2}, \ldots, \zeta_{t-n+1}$ then all inequalities in (3.6) are strict.

Proof. The proof follows directly from Theorem 3.3.

The assertions given above justify to consider the expression $\zeta_t^{\mathrm{T}}\zeta_t$ as a "measure of orthogonality" or as a "measure of deviation" between the vector $\mathbf{z}_t$ and the linear subspace spanned by the vectors (3.1) (in the general case when the function p_t is an arbitrary function satisfying conditions (2.7)–(2.9)) or spanned by the vectors (3.2) (in the special initial case when $p_t = t - 1$, $t \leqq n$). If the vector $\mathbf{z}_t$ is orthogonal to the subspace (maximal deviation) then $\zeta_t^{\mathrm{T}}\zeta_t = \mathbf{z}_t^{\mathrm{T}}\mathbf{z}_t$; if the vector $\mathbf{z}_t$ lies inside the subspace (minimal zero deviation) then $\zeta_t^{\mathrm{T}}\zeta_t = 0$. In all other cases $\zeta_t^{\mathrm{T}}\zeta_t \in (0; \mathbf{z}_t^{\mathrm{T}}\mathbf{z}_t)$.

Trying to eliminate (in the "measure of deviation") the dependence upon the magnitude of the vector $\mathbf{z}_t$, we can use the expression $\zeta_t^T\zeta_t/\mathbf{z}_t^T\mathbf{z}_t$ instead of $\zeta_t^T\zeta_t$. According to (3.3) we get

$$\frac{\zeta_t^T\zeta_t}{\mathbf{z}_t^T\mathbf{z}_t} = 1 - \sum_{j=1}^{p_t} \frac{(\mathbf{z}_t^T\zeta_{t-j})^2}{(\mathbf{z}_t^T\mathbf{z}_t)(\zeta_{t-j}^T\zeta_{t-j})} = 1 - \sum_{j=1}^{p_t} \cos^2 \gamma_j \tag{3.7}$$

where γ_j is the angle between vectors $\mathbf{z}_t$ and ζ_{t-j}. The measure of deviation defined by (3.7) takes the values from interval $\langle 0; 1\rangle$; the zero value corresponds to the minimal zero deviation and the value 1 to the maximal orthogonal deviation. Values close to zero indicate that the latest p_t input vectors are "almost linearly dependent" (we also say that the latest input data are "ill conditioned" or "ill exciting").

Realistic input sequences are such that the equalities in (3.6) occur quite exceptionally. From relations (3.6) it follows subsequently: the larger is the depth of orthogonalization, the greater is our expectation that the value of $\zeta_t^T\zeta_t$ will be close to zero. Since the expressions $\zeta_{t-j}^T\zeta_{t-j}$ are in denominators of (2.5) the danger of numerical instability of orthogonalization process increases with the depth of orthogonalization. This danger is further multiplied if the input sequence $\{\mathbf{z}_t\}$ is significantly correlated. Then the angle between the vector $\mathbf{z}_t$ and the linear subspace of (3.1) (general case) or of (3.2) (special initial case) is small and the value of $\zeta_t^T\zeta_t$ can be inadmissible close to zero even if the depth of orthogonalization is relatively low.

A reliable protection against the danger of numerical instability is proposed in Section 6.

4. CONVERGENCE OF THE GENERALIZED ALGORITHM

It is well known that the original algorithm (2.2) converges ($\mathbf{b}_t^\wedge \to \mathbf{b}$ for $t \to \infty$) if the input sequence is general enough. Remember that the sequence $\{\mathbf{z}_t\}$ of n-dimensional vectors is said to be general enough if there exists an infinite sequence of integers $\{k_i\}$, where $k_{i+1} < k_i + n$, such that each subset $\{\mathbf{z}_{k_i}, \mathbf{z}_{k_i+1}, \ldots, \mathbf{z}_{k_{i+1}}\}$ spans the entire n-dimensional space (see e.g. [8]).

In this section we shall prove that the generalized algorithm (2.5), (2.6) converges for each input sequence for which the original algorithm (2.2) converges. We shall also see that the rate of convergence increases with the depth of orthogonalization.

For the sake of convenience we rewrite the relation (2.6) for recursive estimation in terms of the estimation error

$$\mathbf{d}_t = \mathbf{d}_{t-1} - e_t \frac{\zeta_t}{\zeta_t^T\zeta_t} \tag{4.1}$$

where e_t is the error of the output prediction

$$e_t = y_t - \mathbf{z}_t^T\mathbf{b}_{t-1}^\wedge . \tag{4.2}$$

We denote

$$L_t = \boldsymbol{d}_t^{\mathrm{T}} \boldsymbol{d}_t = (\boldsymbol{b} - \hat{\boldsymbol{b}}_t)^{\mathrm{T}} (\boldsymbol{b} - \hat{\boldsymbol{b}}_t) . \tag{4.3}$$

The rate of convergence at time t will be measured by means of a difference of the function L_t

$$\Delta L_t = L_t - L_{t-1} \tag{4.4}$$

The function $L_t = L(\boldsymbol{d}_t)$ is the Lyapunov function for the iterative process (4.1): $L(\boldsymbol{d}_t)$ is positive definite and equal to zero for $\boldsymbol{d}_t = 0$ and, as it will be shown – see (4.8), the difference $\Delta L(\boldsymbol{d}_t)$ is negative semidefinite.

Lemma 4.1. Let the sequence of estimates $\{\hat{\boldsymbol{b}}_t\}$ be generated by algorithm (2.5), (2.6), or, equivalently, the sequence of estimation errors be generated by algorithm (2.5), (4.1). Then it holds for $t = 1, 2, \ldots$.

$$\boldsymbol{d}_{t-1}^{\mathrm{T}} \zeta_t = e_t \tag{4.5}$$

$$\boldsymbol{d}_t^{\mathrm{T}} \zeta_{t-r} = 0 \quad \text{for} \quad r = 0, 1, 2, \ldots, p_t \tag{4.6}$$

Proof. The proof can be easily performed by induction.

Theorem 4.1. Let $\Delta L_t^{(k)}$ denote the value of ΔL_t provided the depth of orthogonalization is $p_t = k$ $(k \in \{0, 1, \ldots, n-1\}, k \leqq t-1)$.
Then

$$-\infty < \Delta L_t^{(n-1)} \leqq \Delta L_t^{(n-2)} \leqq \ldots \leqq \Delta L_t^{(1)} \leqq \Delta L_t^{(0)} \leqq 0 . \tag{4.7}$$

Proof. From (4.1) we get

$$\boldsymbol{d}_t^{\mathrm{T}} \boldsymbol{d}_t = \boldsymbol{d}_{t-1}^{\mathrm{T}} \boldsymbol{d}_{t-1} - 2 \frac{e_t}{\zeta_t^{\mathrm{T}} \zeta_t} \boldsymbol{d}_{t-1}^{\mathrm{T}} \zeta_t + \frac{e_t^2}{\zeta_t^{\mathrm{T}} \zeta_t}$$

and hence using (4.5) and according to notation (4.3), (4.4)

$$\Delta L_t = - \frac{e_t^2}{\zeta_t^{\mathrm{T}} \zeta_t} \tag{4.8}$$

or, if we express explicitly the depth of orthogonalization

$$\Delta L_t^{(k)} = - \frac{e_t^2}{(\zeta_t^{\mathrm{T}} \zeta_t)^{(k)}} \tag{4.9}$$

The inequalities (4.7) follow immediately from (4.9) and (3.6). □

In reality the equalities in (4.7) occur only exceptionally (see the end of the previous section) and that is why the rate of convergence of the generalized algorithm is growing up with the depth of orthogonalization.

The identification with maximal admissible depth of orthogonalization is described by the following theorem:

Theorem 4.2. Let $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$ be linearly independent. Then the estimation algorithm (2.5), (2.6) with depth $p_t = t - 1$ reaches the exact estimate at most in n steps, i.e. $\hat{\mathbf{b}}_n = \mathbf{b}$.

Proof. If vectors $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$ are linearly independent then the vector $\zeta_1, \zeta_2, \ldots, \zeta_n$ are linearly independent too (Theorem 3.2) and therefore the system of equations (4.6) (Lemma 4.1) with $p_t = n - 1$ has only trivial solution $\mathbf{d}_n = \mathbf{b} - \hat{\mathbf{b}}_n = 0$. □

5. EXPERIMENTAL VERIFICATION

Estimator (2.5), (2.6) was experimentally studied for two types of the orthogonalization processes:

1) Continuous orthogonalization of depth p, $p \in \{0, 1, \ldots, n-1\}$ In this case the depth is constant, naturally except the initial phase. The function p_t is given by

$$p_t = \begin{cases} t - 1 & \text{for } t = 1, 2, \ldots, p + 1 \\ p & t = p + 2, \; p + 3, \ldots \end{cases} \tag{5.1}$$

i.e.

$$\{p_t\} = \{0, 1, 2, \ldots, p - 1, p, p, p, \ldots\} .$$

2) Repeated orthogonalization of depth p, $p \in \{0, 1, \ldots, n-1\}$. In this case the orthogonalization depth is a periodic function which permanently repeats the initial phase of (5.1). The function p_t is given by

$$p_t = t - 1 \bmod (p + 1) \tag{5.2}$$

or

$$\{p_t\} = \{0, 1, \ldots, p, 0, 1, \ldots, p, 0, 1, \ldots, p, \ldots\}$$

Note that both functions (5.1) and (5.2) satisfy the conditions (2.7)–(2.9).

We shall compare how the original and the generalized algorithm works in two unfavourable situations mentioned in the introduction, namely, when the input sequence is a) not centred, b) correlated.

The system to be identified has the form of (2.1) with $n = 5$, $\mathbf{b} = (1, -2, 3, -4, 5)$. The initial estimate $\hat{\mathbf{b}}_0$ was chosen to be the zero vector. The accuracy of identification is measured in the time interval $\langle t_1, t_2 \rangle$ by a logarithm of averaged squared distance between the true parameter vector and its estimate, i.e. by the quantity

$$\log_{10} \left(\frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} \mathbf{d}_t^{\mathrm{T}} \mathbf{d}_t \right) \tag{5.3}$$

The results of comparison for the case of non-centred inputs are summarized in Table 1. The input sequence was gaussian, uncorrelated, with the mean $\mathsf{E}\{\mathbf{z}_t\} = (200, 100, 50, 100, 200)$ and with the variance $\mathsf{D}\{\mathbf{z}_t\} = (25, 16, 9, 4, 1)$. The columns of the table correspond to various variants of generalized algorithm (2.5), (2.6). In particular, the first column corresponds to original algorithm (2.2). In the

table the values of (5.3) are tabulated. The symbol "—" used there denotes the numerical destruction of computation due to the finite word-length of the computer. In all experiments the same realization of random sequence was used.

The Table 2 compares the runs of identification for the case of correlated inputs.

Table 1. The run of identification in the case of non-centred inputs.

$\langle t_1; t_2 \rangle$	Original alg. (2.2)	Generalized algorithm (2.5), (2.6)			
		Continuous orthog. (5.1)		Repeated orthog. (5.2)	
	$p = 0$	$p = 1$	$p = 4$	$p = 1$	$p = 4$
1— 10	1·696	1·47	1·05	1·39	1·05
11— 20	1·695	1·07	1·32	0·36	—6·90
21— 30	1·695	0·37	2·35	—0·50	—6·50
31— 40	1·694	0·05	4·43	—1·35	—7·32
41— 50	1·693	—0·06	6·17	—2·65	—6·05
51— 60	1·693	—0·16	—	—3·81	—5·63
61— 70	1·692	—0·21	—	—4·27	—6·19
71— 80	1·692	—0·38	—	—4·93	—8·05
81— 90	1·691	—0·54	—	—6·45	—7·55
91—100	1·690	—0·63	—	—6·95	—7·09
141—150	1·688	—3·75	—	—7·85	—7·18
191—200	1·686	—6·01	—	—8·41	—6·89

Table 2. The run of identification in the case of correlated inputs.

$\langle t_1; t_2 \rangle$	Original alg. (2.2)	Generalized algorithm (2.5), (2.6)			
		Continuous orthog. (5.1)		Repeated orthog. (5.2)	
	$p = 0$	$p = 1$	$p = 4$	$p = 1$	$p = 4$
1— 10	1·56	1·53	1·13	1·45	1·13
11— 20	1·49	1·20	—4·71	0·42	— 7·96
21— 30	1·37	0·92	—3·59	— 1·05	—10·91
31— 40	1·23	0·44	—3·19	— 2·22	— 8·54
41— 50	1·13	0·25	—2·06	— 2·99	— 9·21
51— 60	1·05	0·07	—1·91	— 3·50	—10·77
61— 70	0·99	—0·02	—1·11	— 4·21	—10·04
71— 80	0·77	—0·44	0·50	— 4·70	— 9·96
81— 90	0·53	—1·19	1·90	— 5·02	—10·20
91—100	0·24	—2·06	3·20	— 5·60	—10·46
141—150	—0·73	—5·25	—	—10·02	— 9·10
191—200	—2·24	—8·60	—	—10·39	— 8·77

The coordinates of the input vector were independent, gaussian, zero-mean and unit-variance random processes with the common correlation function $\mathsf{R}(\tau_1, \tau_2) = q^{|\tau_1 - \tau_2|}$ where τ_1, τ_2 are instants of discrete time and $q (0 \leq q < 1)$ is the parameter of correlation. The value of this parameter was chosen $q = 0{,}9$.

Fom the reported simulations (and many other unreported ones) the following conclusions can be derived:

1) Orthogonalization of input sequence radically accelerates the convergence of estimation.

2) With orthogonalization depth both rate of convergence and numerical instability are growing.

3) When the continuous orthogonalization is applied then the consequences of numerical noise can be catastrophic: the sequence of estimates can start to diverge after some time.

4) If the repeated orthogonalization is used then the numerical noise can cause fluctuation and a decrease of estimation accuracy.

5) With the same depth the repeated orthogonalization gives better results than the continuous one (more exact and faster estimation).

The experimental conclusions 1) and 2) are in a full agreement with the theoretical conclusions of Section 3 and 4. The observations 3) – 5) can be theoretically explained in the following manner. If the continuous orthogonalization is used then the numerical errors can cumulate within unlimited interval of time in contrast with the case of repeated orthogonalization when these errors are cumulated only in a limited time interval (after p steps a new orthogonalization process starts and it is not burdened by numerical errors of preceding one). In the case of continuous orthogonalization the present course of orthogonalization process depends on its entire past history and probably it implies rather worse excitation of the proper estimation algorithm (2.6) in comparison with the case of the repeated orthogonalization.

6. USE OF THE ALGORITHM IN REAL CONDITIONS

When real systems are identified we must take into account that the input-output relation is always to a certain extent a random one and that the parameters of the identified object need not be strictly constant. For these reasons it is necessary to replace equation (2.1), describing the object to be identified, by the more general equation

$$y_t = \mathbf{z}_t^{\mathrm{T}} \mathbf{b}_t + \eta_t \tag{6.1}$$

The quantity η_t in (6.1) is a random component of the input-output mapping and it involves the measurement errors as well as possible structural inadequacies of the mathematical model. The variability in time of system parameters can be caused

by actual non-stationarity of the object or by only local validity of the model structure (then $\boldsymbol{b}_t$ depends to some extent on $\mathbf{z}_t$).

For practical use in real conditions, i.e. to estimate the parameters of system (6.1), we shall generalize estimator (2.5), (2.6) as follows

$$\zeta_t = \mathbf{z}_t - \sum_{j=1}^{p_t} \frac{\mathbf{z}_t^{\mathrm{T}} \zeta_{t-j}}{\zeta_{t-j}^{\mathrm{T}} \zeta_{t-j}} \zeta_{t-j} \tag{6.2}$$

$$\zeta_t^{\mathrm{T}} \zeta_t < h_t \Rightarrow \text{data } (\mathbf{z}_t, y_t) \text{ are dropped out from the processing (as if they never came)} \tag{6.3}$$

$$\hat{\boldsymbol{b}}_t = \hat{\boldsymbol{b}}_{t-1} + \varkappa_t \frac{y_t - \mathbf{z}^{\mathrm{T}} \hat{\boldsymbol{b}}_{t-1}}{\zeta_t^{\mathrm{T}} \zeta_t} \zeta_t \tag{6.4}$$

where p_t, h_t, $\varkappa_t$ are the control parameters (control functions) of the algorithm. The quantity p_t is the orthogonalization depth restricted by (2.7)–(2.9). The quantity h_t is a threshold for the admissible magnitude of $\zeta_t^{\mathrm{T}} \zeta_t$ restricted by $h_t \geqq 0$. The quantity $\varkappa_t$ is the gain of proper estimation algorithm limited by the inequalities $0 < \varkappa_t \leqq 1$. Let us note that particular choice $h_t \equiv 0$, $\varkappa_t \equiv 1$ leads back to algorithm (2.5), (2.6) discussed in the previous sections.

The main reason for non-zero choice of the threshold h_t is to guarantee the numerical security of computation according to formulas (6.2) and (6.4). If we keep in mind only this reason then it is sufficient to choose h_t as a relatively small positive constant h. The magnitude of h depends on the accuracy with which the numbers are represented in the computer and is selected in such a way that the division by $\zeta_t^{\mathrm{T}} \zeta_t$ in (6.2), (6.4) could not cause significant computional errors. Another reason for a nonzero choice of h_t is the interruption of identification if the excitation of the system is bad. The choice of appropriately great threshold causes the identification algorithm to work only if the identified system is well excited.

The reason for $\varkappa_t < 1$ is a decrease of the estimate variance when the identified system is noisy. If the identified system is deterministic, then the optimal choice of control function $\varkappa_t$ is $\varkappa_t \equiv 1$.

Although we do not treat the identification of stochastic systems theoretically we nevertheless give some "rules of thumb" how to choose the control parameters in this important case. In Sections 3 and 5 there was theoretically and experimentally demonstrated how the orthogonalization and consequently the whole identification process is sensitive to numerical noise. It can be shown that the same effect has also the noise expressed by the quantity η_t in (6.1). The sensitivity to the noise of both kind is growing up with the depth of orthogonalization and is greater for the continuous orthogonalization than for the repeated one. When the system to be identified is significantly noisy it is therefore reasonable to use only repeated orthogonalization and at most of the depth $p = 1$. Deeper orthogonalization is admissible only at the very beginning of the identification process when we expect a considerable distance between the parameter vector $\boldsymbol{b}_t$ and its estimate $\hat{\boldsymbol{b}}_t$ and the identified system can

be approximately considered as almost deterministic one. Providing the system is noisy, time-invariant and exactly of the form (6.1), the exact estimates can be reached if we put $p_t = 0$ from certain time instant t_0 and if the gain $\varkappa_t$ satisfies the well-know conditions of stochastic approximation.

In general, the influence of control parameters on the algorithm functioning can be characterised as follows. A greater depth of orthogonalization p_t or a lower threshold h_t or a greater gain $\varkappa_t$ yields increasing activity of the algorithm: the algorithm reacts stronger both to excitation and to noise. This consequently implies an increasing rate of convergence on the one hand and an increasing variance of the estimates on the other hand. Conversely, decreasing p_t or increasing h_t or decreasing $\varkappa_t$ has the inverse impact: slower convergence and higher finite accuracy of estimation.

As a matter of fact, formulas (6.2)–(6.4) represent a whole class of identification algorithms. When we specify the control functions p_t, h_t, $\varkappa_t$ then the element of this class is determined. The problem of an appropriate (or even optimal in some sense) choice of functions p_t, h_t, $\varkappa_t$ for a given identification situation is a very complex and difficult one. Probably only in some special cases the problem of optimal control of sequential identification can be exactly stated and solved. For example, in [5] this was done for algorithm (6.2)–(6.4) with the special choice $p_t \equiv 0$, $h_t \equiv 0$ and for the simple situation when all processes which take part in the identification (excitation noise, non-stationarity) are random. uncorrelated and mutually independant.

In most practical cases we choose the control function h_t, $\varkappa_t$ as constants ($h_t \equiv h$, $\varkappa_t \equiv \varkappa$) and p_t according to (5.2) with constant p. When the external conditions of identification change and we have reliable information about these changes then there is a possibility of an appropriate on-line modification of the control parameters p, h, $\varkappa$.

7. CONCLUSION

The generalized algorithm (6.2)–(6.4) can be widely used for system identification. In contrast with the original algorithm (2.2) the algorithm (6.2)–(6.4) can effectively work even under difficult identification conditions (providing the control parameters are chosen appropriately). Its main characteristic features are:

- numerical simplicity (for the recommended depth of orthogonalization $p = 1$, the number of arithmetic operations per iteration depends only linearly on the number of estimated parameters)
- on-line indication of ill excitation and the possibility to keep numerical stability at a determined level
- the possibility of on-line (or only one-shot) adaptation of estimation algorithm to the external conditions of the identification process.

(Received May 20, 1982.)

REFERENCES

[1] E. D. Avedjan and Ja. Z. Cypkin: A generalized Kacmarz's algorithm (in Russian). Avtomatika i telemechanika (1979), 1, 72—78.

[2] P. A. Janakiraman and S. Renganathan: A new on-line identification method for discrete multivariable linear systems. Automatica *17* (1981), 4, 653—655.

[3] C. G. Kim and V. Gourishankar: A model reference adaptive technique of identification for discrete systems. Automatica *17* (1981), 4, 637—640.

[4] J. Markl: Accelerating the Kaczmarz algorithm convergence in the case of input process time correlation (in Russian). Avtomatika i telemechanika (1980), 8, 70—73.

[5] J. Markl: On-line identification as an optimal controlled process (in Czech). Automatizace *24* (1981), 10, 252—256.

[6] J. Nagumo and A. Noda: A learning method for system identification. IEEE Trans. Automat. Control *AC-12* (1967), 3, 282—287.

[7] L. C. Westphal: An improved identifier for discrete multivariable linear systems. IEEE Trans. Automat. Control *AC-23* (1978), 5, 860—865.

[8] D. Graupe and E. Fogel: A unified sequential identification structure based on convergence cosiderations. Automatica *12* (1976), 1, 53—59.

RNDr. Jaroslav Markl, Řídicí systémy — Kancelářské stroje, k.ú.o., Ostrava (Control Systems Department of "Kancelářské stroje" Ostrava), Starobělská 13, 704 00 Ostrava 4. Czechoslovakia.