Appl. Math. and Comp. Sci., 1997, vol.7, No.1, 125–154

TWO-STAGE IDENTIFICATION OF INTERCONNECTED STEADY-STATE SYSTEMS WITH CASCADE STRUCTURE: A PARAMETRIC APPROACH. PART 2: EMPIRICAL IDENTIFICATION ALGORITHM

ZYGMUNT HASIEWICZ*

The paper deals with an empirical counterpart of the two-stage parametric identification algorithm for serially structured systems, introduced and motivated in (Hasiewicz, 1996). An empirical two-stage identification scheme is formulated and the convergence (with probability one) of the corresponding model parameter estimates to the desired reference values, when the number of measurement data records becomes large, is shown. The asymptotic rate of convergence of the estimates (in probability) is established. Some illustrative numerical examples are also included.

1. Introduction

This paper constitutes a continuation of the work (Hasiewicz, 1996), where the problem of parametric identification of a static serially structured composite system was formulated in a probabilistic framework, and a two-stage procedure for a system structure preserving model selection from a parametric class of models was introduced and motivated theoretically, while assuming full probabilistic knowledge of the system. Here, we shall consider the corresponding empirical identification algorithm, exploiting only measurement data recorded in the system in the noisy environment, i.e. adapted to realistic conditions, and check whether the algorithm is capable of achieving the 'theoretical' approximate cascade system model, derived and validated in Part 1.

To recall the problem, let us consider the identification task of an interconnected static system with cascade structure being a collection of n components (subsystems) $S_1, S_2, \ldots, S_n$ and described by the equations:

$$y_1 = F_1^*(c_1, z_1), \quad y_i = F_i^*(c_i, u_i, z_i), \quad u_i = y_{i-1}, \quad i = 2, 3, \ldots, n \tag{1}$$

where $y_i \in Y_i \subset \mathbb{R}^{l_i}$, $u_i \in U_i \subset \mathbb{R}^{m_i}$ and $c_i \in C_i \subset \mathbb{R}^{s_i}$ denote respectively outputs, interaction (non-manipulable) inputs and external excitations of the element S_i, and $z_i \in Z_i \subset \mathbb{R}^{d_i}$ is a stochastic noise disturbing the subsystem. It is assumed that the true characteristics of the system components $F_i^* : C_i \times U_i \times Z_i \to \mathbb{R}^{l_i}$ are completely

* Institute of Engineering Cybernetics, Technical University of Wrocław, ul. Janiszewskiego 11/17, 50–372 Wrocław, Poland, e-mail: zhas@ict.pwr.wroc.pl.

unknown and approximate models of individual subsystems are to be selected within certain, arbitrarily chosen, sets of models, specified by parameter sets $A_i \subset \mathbb{R}^{p_i}$ and some pre-selected mappings $\Phi_i : \mathbb{R}^{s_i} \times \mathbb{R}^{m_i} \times A_i \to \mathbb{R}^{l_i}$. The interconnected complex model of a series system is thus defined parametrically by the set of equations

$$y_{1M} = \Phi_1(c_1, a_1), \quad y_{iM} = \Phi_i(c_i, u_{iM}, a_i), \quad u_{iM} = y_{i-1,M}, \quad i = 2, 3, \ldots, n \tag{2}$$

where $a_i \in A_i$, and $y_{iM} \in \mathbb{R}^{l_i}$ and $u_{iM} \in \mathbb{R}^{m_i}$ are respectively the outputs and interaction inputs of the i-th submodel within the complex model. There is no reason to suppose that the model structure (2) includes the perfect description of the true system (1) (for $z_i = 0$). This assumption is fundamental throughout the paper (cf. the pertinent remarks in Section 2 of Part 1).

The aim of system identification is to determine in the set of models (2) the best approximate system model in the sense of minimizing the global mean-squared model output error

$$Q(a_1, a_2, \ldots, a_n) = Q_1(a_1) + \sum_{i=2}^{n} Q_i(a_i \mid a_1, a_2, \ldots, a_{i-1}) \tag{3}$$

where

$$\begin{cases} Q_1(a_1) = E\left\|y_1 - \Phi_1(c_1, a_1)\right\|^2 \\ Q_i(a_i \mid a_1, a_2, \ldots, a_{i-1}) = E\left\|y_i - \Phi_i(c_i, u_{iM}, a_i)\right\|^2 \end{cases} \tag{4}$$

stand respectively for measures of accuracy of the first and i-th submodel in the cascade complex model ($\|\cdot\|$ denotes the Euclidean vector norm in $\mathbb{R}^{l_i}$ and E is the expectation with respect to $(c_1, c_2, \ldots, c_i, y_i)$).

As was explained in Part 1, even if the expectations (4) could be determined explicitly, the exact solution to the problem (i.e. the strict minimization of (3) on $A = A_1 \times A_2 \times \ldots \times A_n$ subject to (2)) would not be a good decision because of computational complexity of the task caused by the structural constraints (couplings) introduced by the interconnected complex model. Therefore, the following, easy-to-use but only suboptimal in general, two-stage procedure for system identification was proposed as an alternative when assuming full probabilistic knowledge of the system:

Stage 1. Compute b_i^* minimizing the indices

$$q_i(b_i) = E\left\|y_i - K_{iM}(\bar{c}_i, b_i)\right\|^2 \tag{5}$$

on the sets B_i, independently for $i = 1, 2, \ldots, n$.

Stage 2. Using b_i^* from Stage 1, find a minimum point

$$\hat{a}_i = \arg\min_{a_i \in A_i} \left\|b_i^* - \varphi_i(b_{i-1}^*, a_i)\right\| \tag{6}$$

for $i = 2, 3, \ldots, n$, taking $\hat{a}_1 = b_1^*$ for $i = 1$, and accept $\hat{a}_i$'s as the complex parameters of the model.

In the above, for $i = 1, 2, \ldots, n$, $K_{iM}(\bar{c}_i, b_i)$ stands for a model of the aggregate $AG_i = \{S_1, S_2, \ldots, S_i\}$ comprising the subsystems $S_1, S_2, \ldots, S_i$ (cf. Fig. 1 in Part 1), composed of (2) as follows:

$$\begin{cases} y_{iM} = K_{iM}(\bar{c}_i, b_i) = \Phi_i\Big(c_i, K_{i-1,M}(\bar{c}_{i-1}, b_{i-1}), a_i\Big) \\ \bar{c}_i = (\bar{c}_{i-1}, c_i) = (c_1, c_2, \ldots, c_{i-1}, c_i) \quad \text{for} \quad i \geq 2 \end{cases} \tag{7}$$

and $K_{1M}(\bar{c}_1, b_1) = \Phi_1(c_1, a_1)$, $\bar{c}_1 = c_1$ for $i = 1$. In turn, the set B_i contains all admissible parameters b_i of the aggregate model (7), while the function $\varphi_i(b_{i-1}, a_i)$ describes the manner of stacking the parameters b_{i-1} and a_i in the model (7):

$$\begin{cases} b_1 = a_1 \\ b_i = \varphi_i(b_{i-1}, a_i), \quad i \geq 2 \end{cases} \tag{8}$$

Obviously, $B_i = \varphi_i(B_{i-1} \times A_i)$ for $i \geq 2$ and $B_1 = A_1$. Due to assumptions (c), (d) and (e) in Section 2 of Part 1, the functions $K_{iM}(\bar{c}_i, b_i)$ and $\varphi_i(b_{i-1}, a_i)$ are well-defined, continuous in both arguments, and the sets A_i and B_i are compact.

One can easily recognize that, by a subsequent aggregation and dis-aggregation of consecutive subsystem models (their evolving complexes (7)), the routine reduces the problem of an interconnected cascade system identification to a set of independent standard identification tasks of single-element systems (aggregates AG_i at Stage 1) and some auxiliary computations (solving the problems (6) in Stage 2). Hence, essential savings in complex model computation may rightly be expected. This benefit, however, is achieved at the expense of the loss in quality (accuracy) of the resulting model in comparison with the 'desired' optimum model, as follows from the analysis in Section 4 of Part 1.

This part is devoted to an empirical version of the algorithm (5)–(6). We assume that prior knowledge of the system (i.e. of probability distributions of $(\bar{c}_i, y_i)$'s for the system at hand) is insufficient, and excludes computation of the expectations used in Sections 3–5 in Part 1 for the derivation and a theoretical analysis of the algorithm. Instead, in addition to assumptions (a) and (b) of Part 1 which state that $c = (c_1, c_2, \ldots, c_n)$ and the system noise $z = (z_1, z_2, \ldots, z_n)$ are mutually independent random vectors with finite variances, with zero mean and independent components in the latter case, we shall assume that:

(α) vectors c and $y = (y_1, y_2, \ldots, y_n)$ can be measured, which results in a set of data points $\{(c^k, v^k)\}_{k=1}^N$ (the measurement (c^k, v^k) at instant k), where

$$v^k = \begin{cases} y^k & \text{for a perfect output measurement} \\ y^k + \eta^k & \text{for a noisy output measurement} \end{cases} \tag{9}$$

and η (η^k at instant k), is a random measurement noise with zero mean and finite variance;

(β) $\{c^k\}$, $\{z^k\}$, $\{\eta^k\}$ are mutually independent stationary (with respect to k) white-noise sequences (i.i.d. random variables).

The last assumption, which neglects any correlation between system excitations, is made in the paper for the following reasons:

i. to diminish the influence of the complexity of the system structure on the problem under consideration; actually, the complexity of the system itself, and not of the correlation structure of system excitations, is the true focus of our considerations;

ii. to simplify and clarify the analysis;

iii. to solve the problem in the theoretically simplest form.

This contribution should thus be considered rather as a starting point to develop more involved complex system identification problems, better suited for 'practical' situations e.g. by taking into account possible dependences between system inputs. However, the 'ideal' situation considered in the paper can also occur in practice. For instance, this will be the case if in our example from Section 2 of Part 1 (ore concentration process) the controls $\{c^k\}$ vary around a fixed set-point (nominal control value) due to random inaccuracies of the performing devices (feeders) of follow-up controllers. Such accidental fluctuations can be, of course, independent events.

In Section 2, the empirical counterpart of the two-stage algorithm (5)–(6), based on the measurements $\{(c^k, v^k)\}$, is formulated and a particular form in the case of linear subsystem models is discussed as an example. Then, in Section 3, asymptotic properties of the empirical two-stage identification scheme are analyzed and the convergence (with probability one) of the resulting model parameter estimates to the 'theoretical' outcomes b_i^* and $\hat{a}_i$ of Stages 1 and 2 of the algorithm (cf. (5) and (6)) for a large number of observations N is shown. Next, in Section 4, a study of the asymptotic rate of convergence is made. Finally, in Section 5, finite-sample properties of the algorithm are examined through computer simulations and illustrative numerical examples are given.

2. Empirical Identification Algorithm

Assume that a set of measurements $\{(c^k, v^k)\}_{k=1}^N$ is given. Replacing the theoretical expectations in (5) with the corresponding sample means and next applying the same methodology of cascade system identification as in (5)–(6) of Section 1 yields the following empirical variant of the two-stage identification algorithm:

Stage 1. Compose the models $K_{iM}(\bar{c}_i, b_i)$ of the aggregates $AG_i = \{S_1, S_2, \ldots, S_i\}$ by stacking the models (2) of subsystems according to (7). Then compute b_{iN} which minimize the sample means

$$q_{iN}(b_i) = \frac{1}{N} \sum_{k=1}^{N} \left\| v_i^k - K_{iM}(\bar{c}_i^k, b_i) \right\|^2 \tag{10}$$

on the sets B_i, independently for $i = 1, 2, \ldots, n$.

Stage 2. Using b_{iN} obtained in Stage 1, determine the minimum points

$$a_{iN} = \arg \min_{a_i \in A_i} \left\| b_{iN} - \varphi_i(b_{i-1,N}, a_i) \right\| \tag{11}$$

for $i = 2, 3, \ldots, n$, assuming $a_{1N} = b_{1N}$ for $i = 1$, and take a_{iN}'s as the parameters of the cascade complex model (2).

The above algorithm, suited for the application in real conditions, is easy to implement and for a practical implementation only readily accessible, standard, numerical optimization procedures are needed. The 'raw' measurements are used directly only in Stage 1 of the routine, to compute b_{iN}, while in Stage 2 the role of the data is taken over by the corresponding b_{iN}'s. Thus, after pre-processing in Stage 1, the original data $\{(c^k, v^k)\}_{k=1}^N$ may be wiped out from computer memory (next merely a relatively small number n of b_{iN}'s, generally much less than the whole number N of measurement data points, is sufficient to be stored for further computations). Hence, a gain in memory load can additionally be achieved in the method, compared with the situation where a full set of measurements $\{(c^k, v^k)\}_{k=1}^N$ must be memorized. These savings grow up with the number N of sample points.

The algorithm takes a particularly simple form for a special case of linear models in (2) (linearization of system characteristics), as shown in the following example.

Example 1. Assume linear models for subsystems

$$y_{1M} = B_1 c_1, \quad y_{iM} = A_i u_{iM} + B_i c_i, \quad i = 2, 3, \ldots, n \tag{12}$$

as in Example 1 of Part 1. Then the aggregate models $K_{iM}(\bar{c}_i, b_i)$ are linear (cf. (7))

$$y_{iM} = K_{iM}(\bar{c}_i, b_i) = \bar{K}_{iM} \bar{c}_i \tag{13}$$

and for the matrix parameters $\bar{K}_{iM}$ $(= b_i)$ we have (cf. (8))

$$\bar{K}_{iM} = \left[A_i \bar{K}_{i-1,M}, B_i \right] \tag{14}$$

for $i = 2, 3, \ldots, n$ and $\bar{K}_{1M} = B_1$ for $i = 1$. Actually, $\bar{K}_{iM,N}$ $(= b_{iN})$ minimizing (10) in Stage 1 is as follows:

$$\bar{K}_{iM,N} = V_{iN} \bar{C}_{iN}^T (\bar{C}_{iN} \bar{C}_{iN}^T)^{-1} \tag{15}$$

provided that the matrix inverse exists, where

$$\bar{C}_{iN} = \left[\bar{c}_i^1, \bar{c}_i^2, \ldots, \bar{c}_i^N \right], \quad V_{iN} = \left[v_i^1, v_i^2, \ldots, v_i^N \right]$$

are arrays of measurements. Accordingly, (A_{iN}, B_{iN}) $(= a_{iN})$ minimizing (11) in Stage 2 are of the form

$$\begin{cases} A_{iN} = \bar{K}_{iM,N}^{(i)} \bar{K}_{i-1,M,N}^T \left[(\bar{K}_{i-1,M,N} \bar{K}_{i-1,M,N}^T)^+ \right]^T \\ B_{iN} = \bar{K}_{iM,N}^i \end{cases} \tag{16}$$

where, together with the empirical counterpart of (14), we make use of the dependence $\bar{K}_{iM,N} = [\bar{K}_{iM,N}^{(i)}, \bar{K}_{iM,N}^i]$, $\bar{K}_{iM,N}^i$ being the segment of $\bar{K}_{iM,N}$ related to c_i, and

$(\bar{K}_{i-1,M,N}\bar{K}^T_{i-1,M,N})^+$ is a pseudo-inverse of $\bar{K}_{i-1,M,N}\bar{K}^T_{i-1,M,N}$. Both (15) and (16) can easily be computed by applying standard linear least-squares algorithms, separately for each $i = 1, 2, \ldots, n$.

Remark 1. If in Example 1 the cascade system generating the data $\{(\bar{c}_i^k, v_i^k)\}$ is truly linear and disturbed with additive noise (i.e. the set of models (12) contains the correct description of the system (1)), then the algorithm (15)–(16) converts into the two-stage parameter estimation algorithm of linear composite systems, worked out earlier in (Hasiewicz, 1988).

An important advantage for practical applications is that the quality of the model resulting from the approach (the actual value of the suboptimality index Δ^*; see (20) in Part 1 of the paper) may easily be estimated empirically from the set of measurements $\{(c^k, v^k)\}$ at hand. In fact, for the model obtained we can readily compute the empirical value Δ_N of Δ^* as follows (see (20) and (9) in Part 1):

$$\Delta_N = \frac{\sum_{i=1}^{n} \left[q_{iN}(\hat{b}_{iN}) - q_{iN}(b_{iN}) \right]}{\sum_{i=1}^{n} q_{iN}(b_{iN})} \tag{17}$$

where $q_{iN}(b_i)$ are given by (10), and $\hat{b}_{iN}$ are determined according to the recursive formula (see (17) in Part 1)

$$\hat{b}_{1N} = b_{1N}, \quad \hat{b}_{iN} = \varphi_i(\hat{b}_{i-1,N}, a_{iN}), \quad i = 2, 3, \ldots, n$$

b_{iN} and a_{iN} being the outcomes of Stage 1 and Stage 2 of the empirical identification algorithm. As a result, the degree of suboptimality of the model, and hence also the efficiency of the method for a given system, may be simply measured, by means of the index Δ_N, in each particular case. Let us note that the convergence of the empirical mismatch (cf. (23) in Part 1 and (11))

$$\varepsilon_{iN} = \left\| b_{iN} - \varphi_i(b_{i-1,N}, a_{iN}) \right\|$$

to zero for each $i = 2, 3, \ldots, n$ as $N \to \infty$ means asymptotic optimality of the complex model with parameters a_{iN} (in the sense of (3); see Remark 2 in Part 1). Therefore, the empirical equation errors ε_{iN} (or the cumulative equation error $\sum_{i=2}^{n} \varepsilon_{iN}$) can be used as the complementary test quantities for checking optimality (the range of suboptimality) of the empirical complex model provided by the two-stage approach.

The main theoretical problem is now to examine whether b_{iN} and a_{iN} determined by the empirical two-stage identification algorithm converge, as a sample size $N \to \infty$, respectively to the particular values b_i^* and $\hat{a}_i$ of the composite model parameters following from the theoretical origin (5)–(6) of the method, introduced and validated in Part 1 of the paper. This problem is considered in the next section.

3. Convergence Analysis

We begin examining the asymptotic behaviour of b_{iN} and a_{iN} with the following basic lemma:

Lemma 1. *In the case specified in Section 1 (assumptions* (α)*-*(β) *and (a)-(e) of Part 1), we have*

$$\sup_{b_i \in B_i} \left| q_{iN}(b_i) - q_i(b_i) \right| \to 0 \quad w.p.\ 1 \quad as \quad N \to \infty \tag{18}$$

for each $i = 1, 2, \ldots, n$, *where* $q_{iN}(b_i)$ *and* $q_i(b_i)$ *are as in (10) and (5), respectively.*

Proof. See Appendix A. ■

The lemma enables us to relate the minimum points b_{iN} of the empirical indices $q_{iN}(b_i)$ to the minimum points b_i^* of the indices $q_i(b_i)$ in Stage 1 of the algorithm (cf. Sections 1 and 2). Namely, from the uniformity in b_i of the convergence of the objective functions $q_{iN}(b_i)$ to $q_i(b_i)$ in (18) it follows (cf. e.g. Ljung, 1976; 1978; Vapnik, 1982) that for the corresponding minimum points b_{iN} and b_i^* the first required convergence does hold, i.e.

$$b_{iN} \to b_i^* \quad \text{as} \quad N \to \infty \tag{19}$$

with probability one for $i = 1, 2, \ldots, n$. Let us denote in turn the discrepancy in Stage 2 by

$$d_i(a_i) = \left\| b_i^* - \varphi_i(b_{i-1}^*, a_i) \right\|$$

and

$$d_{iN}(a_i) = \left\| b_{iN} - \varphi_i(b_{i-1,N}, a_i) \right\|$$

for the empirical version of the algorithm, and recall that a_{iN} and $\hat{a}_i$ are exactly the minimum points of $d_{iN}(a_i)$ and $d_i(a_i)$, respectively (see (11) and (6)):

$$a_{iN} = \arg \min_{a_i \in A_i} d_{iN}(a_i)$$

and

$$\hat{a}_i = \arg \min_{a_i \in A_i} d_i(a_i)$$

(with this notation $d_i(\hat{a}_i) = \hat{\epsilon}_i$, the equation error of (8) for $b_i = b_i^*$, $b_{i-1} = b_{i-1}^*$ and $a_i = \hat{a}_i$; cf. (23) in Part 1). Thus the second desired convergence:

$$a_{iN} \to \hat{a}_i \quad \text{as} \quad N \to \infty \tag{20}$$

with probability one for $i = 1, 2, \ldots, n$ could be deduced in a similar simple way as the convergence (19) above if the uniform in a_i (sup-norm) convergence, with probability one, of $d_{iN}(a_i)$ to $d_i(a_i)$ were validated as $N \to \infty$. Indeed, the following result may be proved.

Lemma 2. *Under the assumptions of Lemma 1, we have the convergence*

$$\sup_{a_i \in A_i} \left| d_{iN}(a_i) - d_i(a_i) \right| \to 0 \ \ w.p.\ 1 \ \ as \ \ N \to \infty \tag{21}$$

for $i = 1, 2, \ldots, n$.

Proof. See Appendix B. ■

The uniform almost-sure convergence (21) of the objective functions in Stage 2 implies as an immediate consequence the almost-sure convergence (20) of the corresponding parameter (minimum points) values. Combining (19) and (20), we are thus led to the following theorem:

Theorem 1. *Under the conditions stated in Section 1, for the two-stage identification algorithm, we have*

(i) $b_{iN} \to b_i^*$

(ii) $a_{iN} \to \hat{a}_i$

with probability one as $N \to \infty$, *for* $i = 1, 2, \ldots, n$, *where* b_i^*, $\hat{a}_i$ *and* b_{iN}, a_{iN} *are the outcomes of Stage 1 and Stage 2 of the algorithm in the theoretical (Section 1) and the empirical (Section 2) version, respectively.*

Therefore, in the case considered, essentially all properties hoped for convergence hold. It should be emphasized that the uniform convergence of $q_{iN}(b_i)$ to $q_i(b_i)$ and of $d_{iN}(a_i)$ to $d_i(a_i)$ from Lemma 1 and Lemma 2 is the central property related to the conclusions (i) and (ii) in the theorem. By the model parameter convergence (ii) and continuity of models Φ_i (assumption (d) in Part 1), we can easily ascertain that also the following convergence takes place:

$$\Phi_{iN}(c_i, u_{iM}) \to \hat{\Phi}_i(c_i, u_{iM}) \ \text{ w.p. 1 as } \ N \to \infty$$

where $\Phi_{iN}(c_i, u_{iM}) = \Phi_i(c_i, u_{iM}, a_{iN})$ and $\hat{\Phi}_i(c_i, u_{iM}) = \Phi_i(c_i, u_{iM}, \hat{a}_i)$, at each (continuity) point (c_i, u_{iM}) attainable in the cascade complex model (2) with parameters $a_i = \hat{a}_i$. This means that for each $i = 1, 2, \ldots, n$ the empirical models Φ_{iN} are pointwise strongly consistent estimates of the theoretical counterparts (reference models) $\hat{\Phi}_i$, which would be obtained by the two-stage identification algorithm in the theoretical version, i.e. under full probabilistic knowledge of the system. In the special case of Lipschitz continuous models Φ_i, with respect to the parameters a_i uniformly distributed on $C_i \times \hat{U}_{iM}$ (the sets of inputs accessible in the complex model), the convergence of Φ_{iN} to $\hat{\Phi}_i$ is uniform, i.e.

$$\sup_{(c_i, u_{iM}) \in C_i \times \hat{U}_{iM}} \left\| \Phi_{iN}(c_i, u_{iM}) - \hat{\Phi}_i(c_i, u_{iM}) \right\| \to 0 \ \text{ w.p. 1 as } \ N \to \infty$$

Generally, the convergence in (ii) of empirical model parameters implies asymptotic equivalence (with probability one) of the empirical model outputs to the 'desired' reference model outputs which would result in the complex model with parameters $a_i = \hat{a}_i$.

Remark 2. If the minimum points of $q_{iN}(b_i)$, $q_i(b_i)$, $d_{iN}(a_i)$ or $d_i(a_i)$ are not unique on B_i and A_i, respectively, then the convergence in (i) and (ii) of Theorem 1 is meant as

(i) $\operatorname{dist}(b_{iN_k}, B_i^*) \to 0$ w.p. 1 as $N_k \to \infty$

(ii) $\operatorname{dist}(a_{iN_m}, \hat{A}_i) \to 0$ w.p. 1 as $N_m \to \infty$

for every convergent subsequence $\{b_{iN_k}\}$ of $\{b_{iN}\}$ and $\{a_{iN_m}\}$ of $\{a_{iN}\}$, where B_i^* and $\hat{A}_i$ are the sets of minimum points of $q_i(b_i)$ and $d_i(a_i)$, respectively (Ljung, 1978).

Questions involving the rates of convergence in (19) and (20) are investigated below.

4. Rate of Convergence

The speed of convergence in (19) and (20) will be studied under some additional 'regularity' requirements imposed on the problem. Namely, we shall assume that the parameter sets A_i and B_i, and the sets of external excitations C_i are compact and that the objective functions $q_i(b_i)$ and $d_i(a_i)$ in Stage 1 and Stage 2 of the theoretical version of the algorithm are twice continuously differentiable and strictly convex, at least in some convex neighbourhoods $N(b_i^*) \subset B_i$ and $N(\hat{a}_i) \subset A_i$ around the minimum points b_i^* and $\hat{a}_i$, respectively. Moreover, similarly as in Section 4 of Part 1, we assume that the functions $K_{iM}(\bar{c}_i, b_i)$ and $\varphi_i(b_{i-1}, a_i)$ in Stage 1 and Stage 2 are Lipschitz continuous with respect to b_i and b_{i-1}, now however uniformly distributed on the respective sets $\bar{C}_i = C_1 \times C_2 \times \ldots \times C_i$ and A_i, i.e. that for some positive constants we have

$$\left\|K_{iM}(\bar{c}_i, b_i^1) - K_{iM}(\bar{c}_i, b_i^2)\right\| \leq \text{const}\left\|b_i^1 - b_i^2\right\| \qquad \forall\, b_i^1, b_i^2 \in B_i \tag{22}$$

and

$$\left\|\varphi_i(b_{i-1}^1, a_i) - \varphi_i(b_{i-1}^2, a_i)\right\| \leq \text{const}\left\|b_{i-1}^1 - b_{i-1}^2\right\| \quad \forall\, b_{i-1}^1, b_{i-1}^2 \in B_{i-1} \tag{23}$$

One can easily ascertain that the first of the last two requirements, by the mean-value theorem, coincides in fact with the assumption of differentiability of $q_i(b_i)$ and compactness of B_i and C_i, while the second is exactly the same as (27) and (28) in Part 1. Such requirements are satisfied e.g. for the linear models (12) of Example 1.

Also, like in Section 3 (see the proof of Lemma 1 in Appendix A), we shall confine ourselves to the case of accurate output measurements (i.e. to $v_i^k = y_i^k$) and we shall assume that the model output errors $\{\|y_i^k - K_{iM}(\bar{c}_i^k, b_i)\|\}_{k=1}^N$ in (10) are uniformly bounded, i.e. for each k

$$\left\|y_i^k - K_{iM}(\bar{c}_i^k, b_i)\right\| \leq L_i < \infty \quad \forall\, b_i \in B_i \tag{24}$$

($L_i > 0$ is some constant) with probability one (relative to the probability measure of $(\bar{c}_i^k, y_i^k)$'s produced by the system). This means that for each $\bar{c}_i$ (from a compact set of inputs $\bar{C}_i$) the system outputs y_i are (uniformly) bounded, $\|y_i\| \leq L_i$ (almost surely), i.e. implicitly, for additive noise z_i, that the system noise is bounded (cf. Fig. 1 in Part 1).

In this case, in the neighbourhood $N(b_i^*)$, we obtain by Taylor's theorem

$$q_i(b_i) = q_i(b_i^*) + \frac{1}{2}(b_i - b_i^*)^T \nabla^2 q_i(b_i^*)(b_i - b_i^*) + o\Big(\|b_i - b_i^*\|^2\Big)$$

where $o(\|b_i - b_i^*\|^2)/\|b_i - b_i^*\|^2 \to 0$ as $b_i \to b_i^*$ and the Hessian $\nabla^2 q_i(b_i^*)$ is a symmetric positive-definite matrix. Thus for $b_i = b_{iN}$, by the convergence (19), we get that asymptotically (for a sufficiently large N)

$$(0 \leq) \quad q_i(b_{iN}) - q_i(b_i^*) = \frac{1}{2}(b_{iN} - b_i^*)^T \nabla^2 q_i(b_i^*)(b_{iN} - b_i^*)$$

with probability one and hence, by positive definiteness of $\nabla^2 q_i(b_i^*)$ and the equivalence of norms in $\mathbb{R}^{h_i}$ ($h_i = \dim b_i$), we conclude that for N growing large

$$\|b_{iN} - b_i^*\|^2 \leq \text{const}\, |q_i(b_{iN}) - q_i(b_i^*)| \tag{25}$$

Similarly, replacing $q_i(b_i)$ with $d_i(a_i)$ and repeating the arguments, we obtain that asymptotically

$$\|a_{iN} - \hat{a}_i\|^2 \leq \text{const}\, |d_i(a_{iN}) - d_i(\hat{a}_i)| \tag{26}$$

However, from the relation

$$|d_i(a_{iN}) - d_i(\hat{a}_i)| \leq 2 \sup_{a_i \in A_i} |d_{iN}(a_i) - d_i(a_i)| \tag{27}$$

and the inequality (cf. the proof of Lemma 2 in Appendix B)

$$\sup_{a_i \in A_i} |d_{iN}(a_i) - d_i(a_i)| \leq \|b_{iN} - b_i^*\| + \sup_{a_i \in A_i} |\rho_{iN}(a_i)|$$

where

$$\rho_{iN}(a_i) = \|\varphi_i(b_{i-1,N}, a_i) - \varphi_i(b_{i-1}^*, a_i)\|$$

together with (23), it follows that

$$|d_i(a_{iN}) - d_i(\hat{a}_i)| \leq \text{const}\, \max\Big\{\|b_{iN} - b_i^*\|, \|b_{i-1,N} - b_{i-1}^*\|\Big\} \tag{28}$$

Consequently, combining (25) and (26) with (28) yields the following chain of inequalities:

$$\begin{aligned} &\|b_{iN} - b_i^*\|^2 \leq \text{const}\, |q_i(b_{iN}) - q_i(b_i^*)| && \text{(for Stage 1)} \\ &\|a_{iN} - \hat{a}_i\|^2 \leq \text{const}\, \max\Big\{\|b_{iN} - b_i^*\|, \|b_{i-1,N} - b_{i-1}^*\|\Big\} && \text{(for Stage 2)} \end{aligned} \tag{29}$$

each with probability one, where the constants above were readjusted whenever necessary.

Therefore we see that, in fact, it suffices to show the rate of convergence to zero of the error function $|q_i(b_{iN}) - q_i(b_i^*)|$ in Stage 1 of the algorithm as N tends to infinity, for $i = 1, 2, \ldots, n$. To this end, we shall investigate thereafter how fast, for a

given constant $\alpha > 0$, the probability of the error $P\{|q_i(b_{iN}) - q_i(b_i^*)| > \alpha\}$ decreases to zero as N increases.

Lemma 3. *Under the above assumptions, for a sufficiently large* N

$$P\Big\{|q_i(b_{iN}) - q_i(b_i^*)| > \alpha\Big\} = O\Big(\exp(-N\alpha^2\omega_i)\Big) \tag{30}$$

where $\omega_i = 1/(32L_i^4)$ *(see (24)), for* $i = 1, 2, \ldots, n$.

Proof. See Appendix C. ■

Remark 3. To prove (30) in Appendix C, we need in fact only the Lipschitz continuity of $K_{iM}(\bar{c}_i, b_i)$ (assumption (22)) and the boundedness requirement of the model output errors in Stage 1 (assumption (24)). Notice that the faster the asymptotic rate of convergence of $P\{|q_i(b_{iN}) - q_i(b_i^*)| > \alpha\}$ is, the more restrictive demands are imposed in the bound (24) (the smaller L_i) and it rapidly decreases with L_i growing large. However, for each finite L_i the probability $P\{|q_i(b_{iN}) - q_i(b_i^*)| > \epsilon/N^p\}$ with $0 < p < 1/2$, $\epsilon > 0$, still tends to zero at an exponential rate as $N \to \infty$.

Remark 4. It is noteworthy that the rate of convergence in (30) does not depend on the dimension h_i of the parameter vector b_i^* being estimated. Therefore the important advantage of dimensionality invariance of the convergence rate, celebrated in the parametric inference in the classical estimation theory, is observed.

The following result is a fairly straightforward consequence of Lemma 3, where for a sequence of random variables $\{\xi_N\}$ by writing $\xi_N = O(r_N)$ in probability we mean that $\gamma_N \xi_N / r_N \to 0$ in probability as $N \to \infty$ for any number sequence $\{\gamma_N\}$ convergent to zero (Greblicki, 1994; Greblicki and Pawlak, 1994).

Corollary 1. *Let all the assumptions of Lemma 3 be satisfied. Then for a sufficiently large* N

$$|q_i(b_{iN}) - q_i(b_i^*)| = O(N^{-1/2}) \quad \textit{in probability} \tag{31}$$

Proof. The proof follows immediately from (30) when substituting $\epsilon|r_N| \,/\, |\gamma_N|$, $\epsilon > 0$, for the constant α. ■

Thus the in-probability convergence of $q_i(b_{iN})$ to $q_i(b_i^*)$ (i.e. of b_{iN} to b_i^* in the sense of the objective function values) attains asymptotically the order $O(N^{-1/2})$, i.e. the optimal rate of convergence in probability for parametric estimators (see e.g. (Bickel and Doksum, 1977; Chapter 4.4)).

Remark 5. By (A1) in Appendix A and (C1) and (C14) in Appendix C, the order $O(\exp(-N\alpha^2\omega_i))$ in (30) and $O(N^{-1/2})$ in (31) characterizes also, for a large N, the rate of uniform convergence in (18) of Lemma 1. In fact, we obtain immediately

$$P\Big\{\sup_{b_i \in B_i} |q_{iN}(b_i) - q_i(b_i)| > \alpha\Big\} = O\Big(\exp(-N\alpha^2\bar{\omega}_i)\Big)$$

where $\bar{\omega}_i = 4\omega_i$, and

$$\sup_{b_i \in B_i} |q_{iN}(b_i) - q_i(b_i)| = O(N^{-1/2}) \quad \text{in probability.}$$

Lemma 3, Corollary 1 and the relations (29) provide the basis for the following rate-of-convergence theorem for the estimates b_{iN} and a_{iN} obtained in Stages 1 and 2 of the two-stage empirical identification algorithm.

Theorem 2. *Let all the requirements of this section hold. Then asymptotically (for a sufficiently large N)*

$$P\Big\{\|b_{iN} - b_i^*\| > \alpha\Big\} = O\Big(\exp(-N\alpha^4\sigma_{i1})\Big)$$

$$P\Big\{\|a_{iN} - \hat{a}_i\| > \alpha\Big\} = O\Big(\exp(-N\alpha^8\sigma_{i2})\Big)$$

where σ_{i1} and σ_{i2} are some positive constants, and respectively

$$\|b_{iN} - b_i^*\| = O(N^{-1/4}) \quad \textit{in probability}$$

$$\|a_{iN} - \hat{a}_i\| = O(N^{-1/8}) \quad \textit{in probability}$$

Proof. The conclusion is obtained by combining (29) with (30) and (31) and straightforward calculation. ■

Remark 6. If the models Φ_i preselected in (2) are Lipschitz with respect to the parameters a_i, uniformly on the sets $C_i \times \hat{U}_{iM}$ (cf. Section 3), then for the models Φ_{iN} and $\hat{\Phi}_i$ as in Section 3 we conclude easily that, for large N,

$$P\Big\{\sup_{(c_i,u_{iM})\in C_i\times\hat{U}_{iM}} \|\Phi_{iN}(c_i,u_{iM}) - \hat{\Phi}_i(c_i,u_{iM})\| > \alpha\Big\} = O\Big(\exp(-N\alpha^8\rho_{i2})\Big)$$

where $\rho_{i2} > 0$ is some constant, and respectively

$$\sup_{(c_i,u_{iM})\in C_i\times\hat{U}_{iM}} \|\Phi_{iN}(c_i,u_{iM}) - \hat{\Phi}_i(c_i,u_{iM})\| = O(N^{-1/8}) \quad \text{in probability}$$

which gives asymptotic rates of uniform convergence of the empirical models Φ_{iN} to the desired 'theoretical' counterparts $\hat{\Phi}_i$ on the sets of inputs attainable in the complex model, in this special case.

Theorem 2 says that the probability of occurence of errors $\|b_{iN} - b_i^*\|$ and $\|a_{iN} - \hat{a}_i\|$ exceeding α in estimation of the minimum points b_i^* in Stage 1 and $\hat{a}_i$ in Stage 2, respectively, tends, for each $\alpha > 0$, to zero as $N \to \infty$ at an exponential rate. However, the speed of convergence is at each stage different and in Stage 2 is twice smaller, in the meaning of the order, than in Stage 1. Generally, the established (theoretical) rates of convergence (in probability) of the estimates b_{iN} and a_{iN} to the reference points b_i^* and $\hat{a}_i$ are rather slow. This should not be however discouraging when considering practical implementation of the method. There are three reasons

for that. First, our results are asymptotic in nature, i.e. they characterize only large-sample properties of the estimates. Second, they provide merely upper bounds on the convergence speed (i.e. give guaranteed rates) and, third, they refer to all, also unfavourable conditions, where the objective functions $q_i(b_i)$ and $d_i(a_i)$ in Stages 1 and 2 are arbitrarily flat (therefore e.g. the guaranteed rate of convergence of $q_i(b_{iN})$ to $q_i(b_i^*)$ in Corollary 1 is essentially higher than the rate of convergence of 'pure' minimum points b_{iN} to b_i^* in Theorem 2). In each particular implementation the speed of convergence of the estimation error, in both stages, to a fair accuracy may be thus much faster (see e.g. Remark 1 and the paper cited therein), even if not all of the conditions imposed in this section hold (e.g. the system noise z_i is not bounded). In the next section, we present some results of computer simulation justifying this belief experimentally, for a moderate number of data records.

5. Simulation Study

Computer simulation was performed for the cascade system consisting of $n = 3$ interconnected static subsystems with the following general description:

Subsystem S_1:

$$y_{11} = p^*_{111}c_1 + p^*_{112}c_1^2 + p^*_{113}c_1^3 + z_{11}$$

Subsystem S_2:

$$y_{21} = p^*_{211}c_2 + p^*_{212}c_2^2 + p^*_{213}y_{11} + p^*_{214}c_2y_{11} + z_{21}$$

$$y_{22} = p^*_{221}c_2 + p^*_{222}c_2^2 + p^*_{223}y_{11} + p^*_{224}c_2y_{11} + z_{22}$$

Subsystem S_3:

$$y_{31} = p^*_{311}c_3 + p^*_{312}c_3^2 + p^*_{313}y_{21} + p^*_{314}y_{22} + p^*_{315}c_3y_{21} + p^*_{316}c_3y_{22} + z_{31}$$

$$y_{32} = p^*_{321}c_3 + p^*_{322}c_3^2 + p^*_{323}y_{21} + p^*_{324}y_{22} + p^*_{325}c_3y_{21} + p^*_{326}c_3y_{22} + z_{32}$$

$$y_{33} = p^*_{331}c_3 + p^*_{332}c_3^2 + p^*_{333}y_{21} + p^*_{334}y_{22} + p^*_{335}c_3y_{21} + p^*_{336}c_3y_{22} + z_{33}$$

being linear in the interaction inputs (outputs of the preceding subsystems; cf. examples in Section 5 of Part 1), and the parameters p^*_{ijk} freely chosen by the experimenter. For subsequent subsystems linear models were assumed, as in Example 1 of Section 2 (system linearization):

Model M_1:

$$y_{11,M} = a_{111}c_1$$

Model M_2:

$$y_{21,M} = a_{211}y_{11,M} + a_{212}c_2$$

$$y_{22,M} = a_{221}y_{11,M} + a_{222}c_2$$

Model M_3:

$$y_{31,M} = a_{311}y_{21,M} + a_{312}y_{22,M} + a_{313}c_3$$

$$y_{32,M} = a_{321}y_{21,M} + a_{322}y_{22,M} + a_{323}c_3$$

$$y_{33,M} = a_{331}y_{21,M} + a_{332}y_{22,M} + a_{333}c_3$$

It is also assumed that each system noise z_{ij} is Gaussian $N(0, \sigma_{z,ij})$ and the external excitations c_i of consecutive subsystems are generated with uniform distribution in the range $|c_i| \leq c_{i,\max}$ (then $Ec_i = 0$ and $\sigma_{c,i} = c_{i,\max}/3^{1/2}$) and that they are stochastically independent. Similarly to the analysis in Sections 3 and 4, we assume that there is no measurement noise η_i. The estimates $\bar{K}_{iM,N}$ (cf. (15)) of the optimum aggregate model parameters $\bar{K}^*_{iM}$ in Stage 1 (cf. (13) and Example 1 in Part 1) were computed iteratively (for a growing number N of data points in the experiment) by using the recursive linear least-squares algorithm and the corresponding estimates $a_{ijk,N}$ (resp. (A_{iN}, B_{iN}) for the empirical model (12)) of the subsystem model parameters $\hat{a}_{ijk}$ (i.e. of the least-squares solutions to the equations (14); cf. (6)) were computed in Stage 2 according to (16). With reference to the above assumptions and the conclusion of Example 1 from Part 1, we remark that in the case under consideration the two-stage approach provides the optimum linear system model (in the sense of (3) and (4)), i.e. we get $\hat{a}_{ijk} = a^*_{ijk}$, say.

The aim of the experiment was to examine the effect of (i) the structure of the system, (ii) the strength of couplings (interactions) among system elements, and (iii) the intensity of noise on the efficiency of the two-stage identification method. To this end, three evolving structures of the cascade system have been assumed: linear, weakly non-linear and non-linear. The effect of interactions within individual system elements was varied by changing the ratio between the parameters joint with interaction inputs and that associated with external inputs c_i, and assuming respectively 0.1 for weak couplings, 0.5 for moderate couplings and 1 (or the same order of parameter values) for strong couplings. Finally, three levels of noise (defined by the ratio of the corresponding noise dispersion to the sum of the modulus of the average (nominal) value of the adequate output in the system and dispersion of this output without noise) were assumed, namely small noise (1%), medium noise (5%), and large noise (10%).

For brevity, the efficiency of the algorithm for a given sample size N is further characterized by the cumulative index, referring to the whole set of complex model parameters (average relative estimation error) and defined as follows:

$$aree(N) = \frac{1}{TP}\sum_i\sum_j\sum_k \frac{|a_{ijk,N} - \hat{a}_{ijk}|}{|\hat{a}_{ijk}|}100\,\% \tag{32}$$

where TP is the total number of parameters in the cascade model and the summands are relative inaccuracies of the corresponding parameter estimates $a_{ijk,N}$ with reference to the desired (here: optimum) parameter values $\hat{a}_{ijk}$, following from the two-stage approach in the theoretical version.

Example 2. First, we consider a linear cascade system and for the case of strong couplings assume the following parameter values in the general system description:

- for subsystem S_1:

$$p^*_{111} = 1, \quad p^*_{112} = 0, \quad p^*_{113} = 0$$

- for subsystem S_2:

$$p^*_{211} = 2, \quad p^*_{212} = 0, \quad p^*_{213} = 3, \quad p^*_{214} = 0$$

$$p^*_{221} = 3, \quad p^*_{222} = 0, \quad p^*_{223} = 2, \quad p^*_{224} = 0$$

- for subsystem S_3:

$$p^*_{311} = 4, \quad p^*_{312} = 0, \quad p^*_{313} = 4, \quad p^*_{314} = 4, \quad p^*_{315} = 0, \quad p^*_{316} = 0$$

$$p^*_{321} = 5, \quad p^*_{322} = 0, \quad p^*_{323} = 5, \quad p^*_{324} = 5, \quad p^*_{325} = 0, \quad p^*_{326} = 0$$

$$p^*_{331} = 6, \quad p^*_{332} = 0, \quad p^*_{333} = 6, \quad p^*_{334} = 6, \quad p^*_{335} = 0, \quad p^*_{336} = 0$$

It is also assumed that each of the external excitations c_i varies in the range $|c_i| \leq 1$. In the case considered (the linear system and a linear model to be found), the identification problem to be solved is reduced, in fact, to the system parameter estimation task and the two-stage identification procedure turns into the two-stage parameter estimation algorithm, as indicated in Remark 1 of Section 2. Thus the parameters of the best model (the reference values in the index (32)) are now obviously as follows:

- for model M_1:

$$\hat{a}_{111} = 1$$

- for model M_2:

$$\hat{a}_{211} = 3, \quad \hat{a}_{212} = 2, \quad \hat{a}_{221} = 2, \quad \hat{a}_{222} = 3$$

- for model M_3:

$$\hat{a}_{311} = \hat{a}_{312} = \hat{a}_{313} = 4, \quad \hat{a}_{321} = \hat{a}_{322} = \hat{a}_{323} = 5, \quad \hat{a}_{331} = \hat{a}_{332} = \hat{a}_{333} = 6$$

and this is simultaneously the theoretical outcome of the two-stage identification method. The estimation error $aree(N)$, versus data set size N, arising in the recovery of these parameters due to the two-stage algorithm for various noise levels is depicted in Figs. 1, 2 and 3 for strong, moderate and weak couplings in the system, respectively. Exemplary final model parameters obtained by the algorithm under

strong interactions and medium noise in the system for $N = 500$ observations are as follows:

$$a_{111,500} = 1.002$$

$$a_{211,500} = 3.000, \quad a_{212,500} = 1.992$$

$$a_{221,500} = 2.008, \quad a_{222,500} = 3.008$$

$$a_{311,500} = 4.011, \quad a_{312,500} = 3.983, \quad a_{313,500} = 4.068$$

$$a_{321,500} = 4.965, \quad a_{322,500} = 5.097, \quad a_{323,500} = 5.113$$

$$a_{331,500} = 5.956, \quad a_{332,500} = 6.026, \quad a_{333,500} = 6.076$$

Based on the values of $aree(N)$ we deduce that the parameter estimates behave satisfactorily for the strong and moderate couplings in the system and that the efficiency of the algorithm significantly deteriorates for weak couplings. This feature is also apparent in further examples and can be explained by the particular set-up of the two-stage identification algorithm, where the existence of interconnections among subsystems is crucial for the very idea of the method. For instance, for $N = 500$ and strong couplings we obtained in an experiment $aree(500) = 0\%$ under small noise, $aree(500) = 0.6\%$ under medium noise, and $aree(500) = 1\%$ under large noise. Similarly, for moderate couplings, we then have $aree(500) = 0\%$ under small noise,

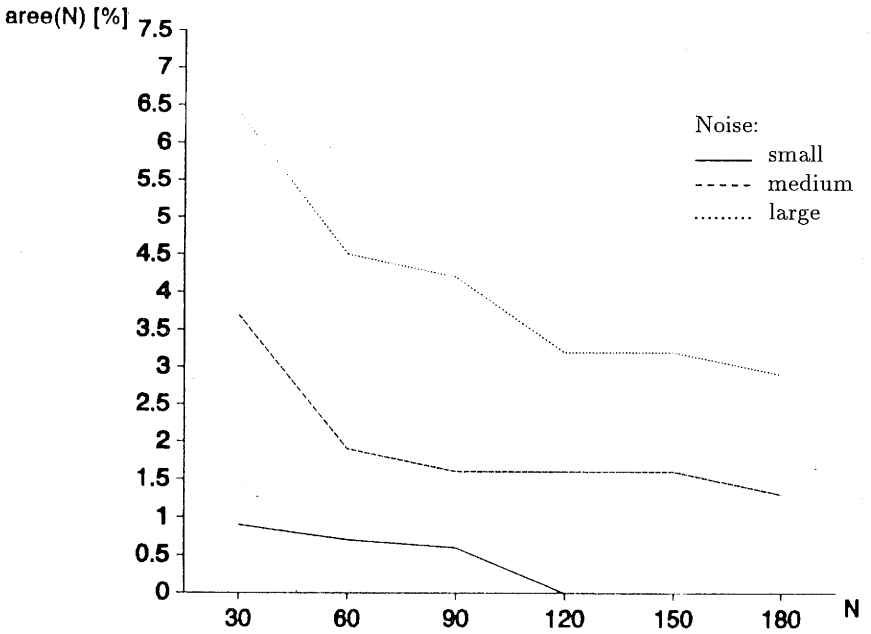

Fig. 1. Average relative estimation error versus data set size: linear system, strong couplings.

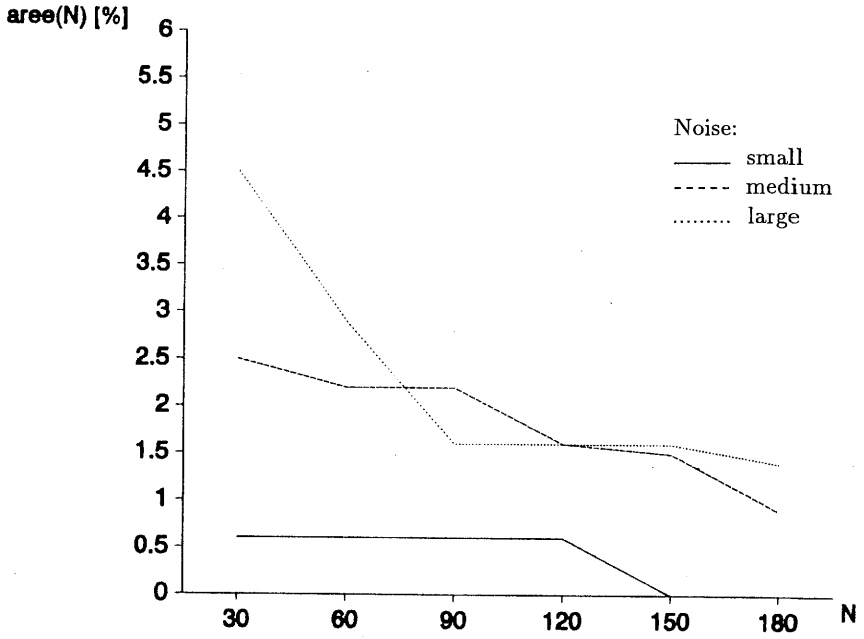

Fig. 2. Average relative estimation error versus data set size: linear system, moderate couplings.

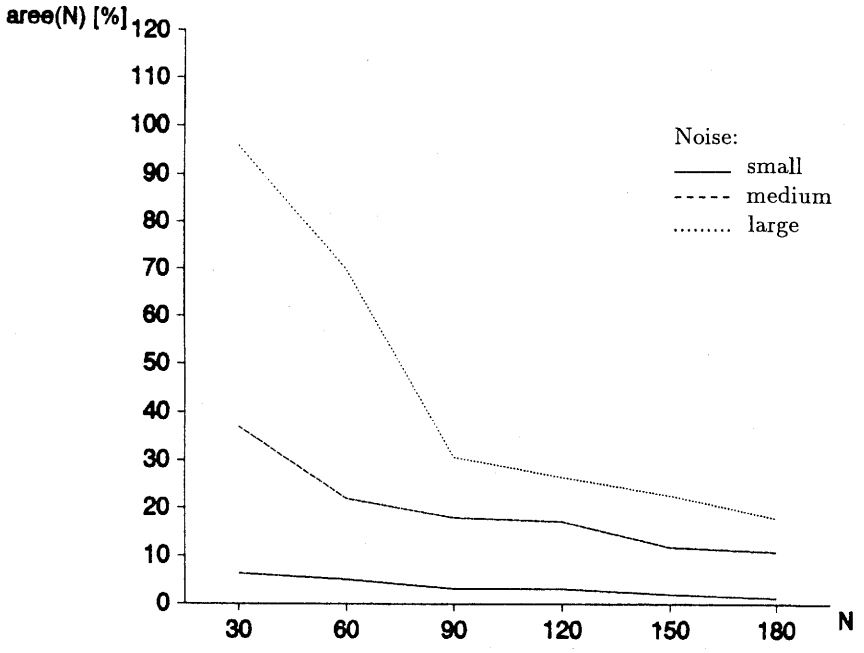

Fig. 3. Average relative estimation error versus data set size: linear system, weak couplings.

$aree(500) = 0.6\%$ under medium noise, and $aree(500) = 0.9\%$ under large noise, while for weak couplings it was then merely achieved, for the corresponding noise levels, $aree(500) = 1.2\%$, $aree(500) = 6\%$, and $aree(500) = 7.4\%$. The efficiency of the algorithm depends visibly on the intensity of noise and decreases with the growing noise intensity level, which is evidenced by the above-quoted error values and Figs. 1–3. For example, the same estimation accuracy of 1% (in the sense of $aree(N)$) is obtained, for the strong couplings, for $N = 20$ if the noise is small, $N = 200$ if the noise is medium, and $N = 500$ if the noise is large. A similar accuracy for moderate couplings and the same noise levels is obtained for $N = 20$, $N = 280$, and $N = 300$. The same tendency is observed for weak couplings.

Example 3. As a weakly non-linear cascade system we consider the general system description from the beginning of this section, setting for the strong interactions the following parameters:

- for subsystem S_1:

$$p^*_{111} = 1, \quad p^*_{112} = 0.1, \quad p^*_{113} = 0.1$$

- for subsystem S_2:

$$p^*_{211} = 2, \quad p^*_{212} = 0.1, \quad p^*_{213} = 3, \quad p^*_{214} = 0.1$$
$$p^*_{221} = 3, \quad p^*_{222} = 0.2, \quad p^*_{223} = 2, \quad p^*_{224} = 0.1$$

- for subsystem S_3:

$$p^*_{311} = 4, \quad p^*_{312} = 0.1, \quad p^*_{313} = 4, \quad p^*_{314} = 4, \quad p^*_{315} = p^*_{316} = 0.1$$
$$p^*_{321} = 5, \quad p^*_{322} = 0.2, \quad p^*_{323} = 5, \quad p^*_{324} = 5, \quad p^*_{325} = p^*_{326} = 0.1$$
$$p^*_{331} = 6, \quad p^*_{332} = 0.3, \quad p^*_{333} = 6, \quad p^*_{334} = 6, \quad p^*_{335} = p^*_{336} = 0.1$$

Actually, for $|c_i| \leq 1$, $i = 1, 2, 3$, the reference parameters of the linear complex model, resulting from the two-stage approach in the case of full probabilistic knowledge of the system, are as follows:

- for model M_1:

$$\hat{a}_{111} = 1.06$$

- for model M_2:

$$\hat{a}_{211} = 3, \quad \hat{a}_{212} = 2.003, \quad \hat{a}_{221} = 2, \quad \hat{a}_{222} = 3.003$$

- for model M_3:

$$\hat{a}_{311} = \hat{a}_{312} = 4, \quad \hat{a}_{313} = 4.027$$

$$\hat{a}_{321} = \hat{a}_{322} = 5, \quad \hat{a}_{323} = 5.027$$

$$\hat{a}_{331} = \hat{a}_{332} = 6, \quad \hat{a}_{333} = 6.027$$

and the estimation error $aree(N)$ for evolving sample size N and various noise levels is presented in Fig. 4 for strong couplings, Fig. 5 for moderate couplings, and Fig. 6 for weak couplings. The behaviour of $aree(N)$ and the influence of particular experiment factors is similar as in Example 2, but its rate of convergence to zero is now generally smaller, i.e. the comparable estimation accuracy is obtained for a larger number of data records than in the previous example. For comparison, $aree(500)$ is now as follows: for the strong couplings and small noise 0.7%, for medium noise 1.2%, and for large noise 2.5%; for the moderate couplings we have respectively 0.1%, 0.6% and 0.93%, and for the weak couplings 3.2%, 5.7%, and 8.2%. In turn, the 1% accuracy is achieved in the case of the strong couplings and small noise in the system for $N \geq 250$, under medium noise for $N \geq 500$, and under large noise for $N > 850$. In the case of moderate couplings $N \geq 250$ is then needed for small and medium noise, and $N \geq 650$ for large noise.

Example 4. Finally, we consider a non-linear series system which has, for the strong couplings, the following parameter values in the general system description:

- for subsystem S_1:

$$p^*_{111} = 1, \quad p^*_{112} = 1, \quad p^*_{113} = 1$$

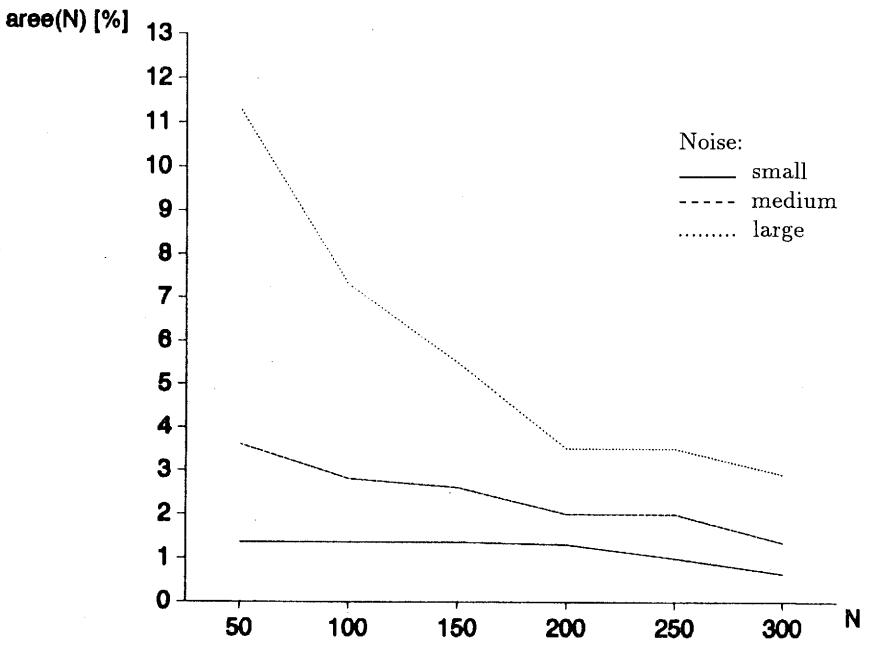

Fig. 4. Average relative estimation error versus data set size: weakly non-linear system, strong couplings.

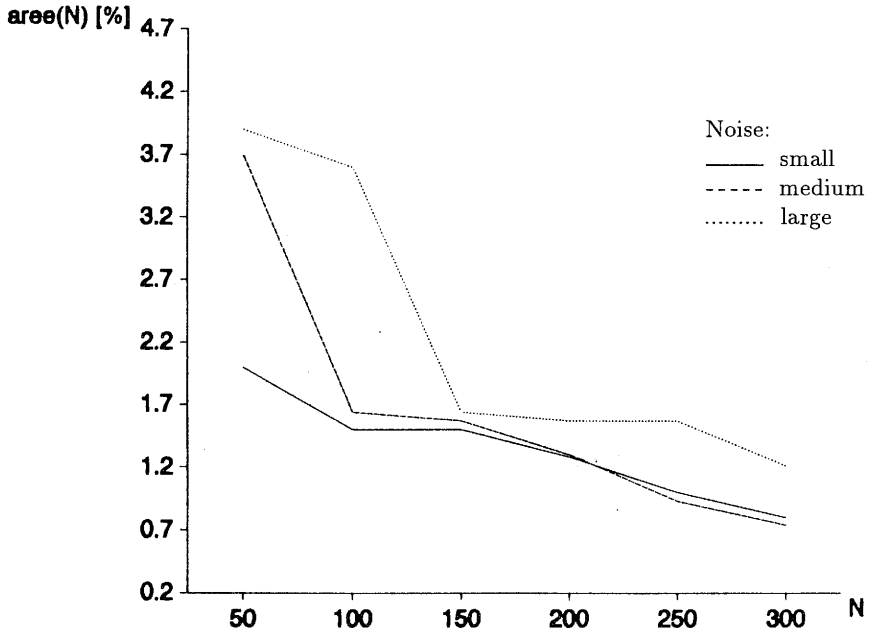

Fig. 5. Average relative estimation error versus data set size: weakly non-linear system, moderate couplings.

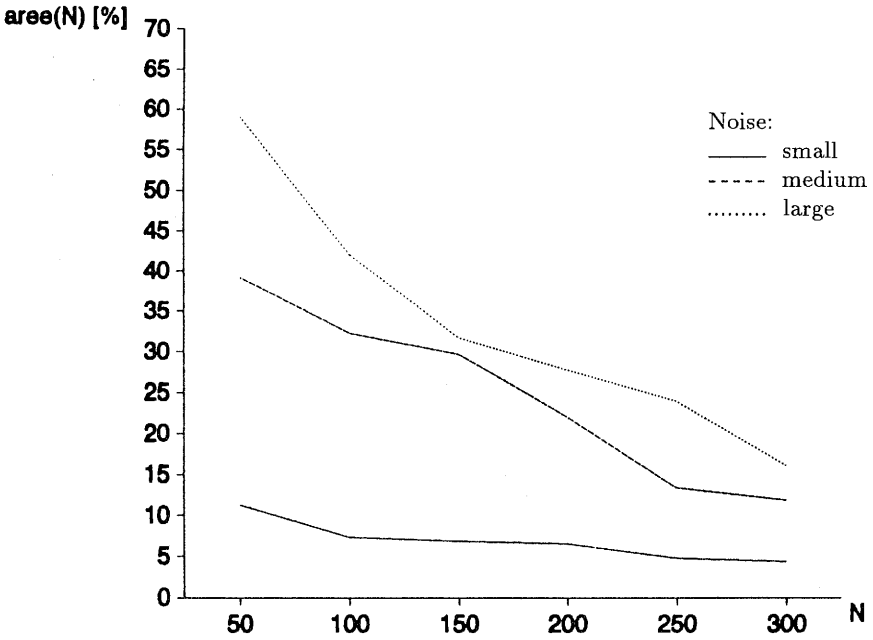

Fig. 6. Average relative estimation error versus data set size: weakly non-linear system, weak couplings.

- for subsystem S_2:

$$p^*_{211} = 2, \quad p^*_{212} = 1, \quad p^*_{213} = 3, \quad p^*_{214} = 0.1$$

$$p^*_{221} = 3, \quad p^*_{222} = 2, \quad p^*_{223} = 2, \quad p^*_{224} = 0.1$$

- for subsystem S_3:

$$p^*_{311} = 4, \quad p^*_{312} = 1, \quad p^*_{313} = 4, \quad p^*_{314} = 4, \quad p^*_{315} = p^*_{316} = 0.1$$

$$p^*_{321} = 5, \quad p^*_{322} = 2, \quad p^*_{323} = 5, \quad p^*_{324} = 5, \quad p^*_{325} = p^*_{326} = 0.1$$

$$p^*_{331} = 6, \quad p^*_{332} = 3, \quad p^*_{333} = 6, \quad p^*_{334} = 6, \quad p^*_{335} = p^*_{336} = 0.1$$

In this case, for uniformly distributed c_i and $|c_i| \leq 1$, the corresponding reference linear model, resulting from the two-stage approach under full probabilistic knowledge of the system, has the parameters

- for model M_1:

$$\hat{a}_{111} = 1.6$$

- for model M_2:

$$\hat{a}_{211} = 3, \quad \hat{a}_{212} = 2.033, \quad \hat{a}_{221} = 2, \quad \hat{a}_{222} = 3.033$$

- for model M_3:

$$\hat{a}_{311} = \hat{a}_{312} = 4, \quad \hat{a}_{313} = 4.267$$

$$\hat{a}_{321} = \hat{a}_{322} = 5, \quad \hat{a}_{323} = 5.267$$

$$\hat{a}_{331} = \hat{a}_{332} = 6, \quad \hat{a}_{333} = 6.267$$

The behaviour of $aree(N)$ for an increasing number N of data records is illustrated in Figs. 7, 8 and 9 for strong, moderate and weak couplings in the system, and various noise levels. As can be seen, the character of the plots and the influence of particular factors on the efficiency of the scheme remains, in principle, the same, but (in comparison with the preceding two examples) the rate of convergence of the error $aree(N)$ to zero decreases, i.e. a similar estimation accuracy is obtained for a larger number of data records, as shown by the diagrams in Figs. 7–9 when compared with the plots in Figs. 1–3 and Figs. 4–6. The test value $aree(500)$ is now, for small, medium and large noise, respectively, as follows: for the strong couplings 3.4%, 3.6% and 5.5%; for the moderate couplings 1.35%, 2% and 4.6%; and for the weak couplings 4%, 6.7%, 9.2%. Next, the 1% accuracy needs now in any case $N \geq 850$ observations, i.e. a significantly larger number of data records.

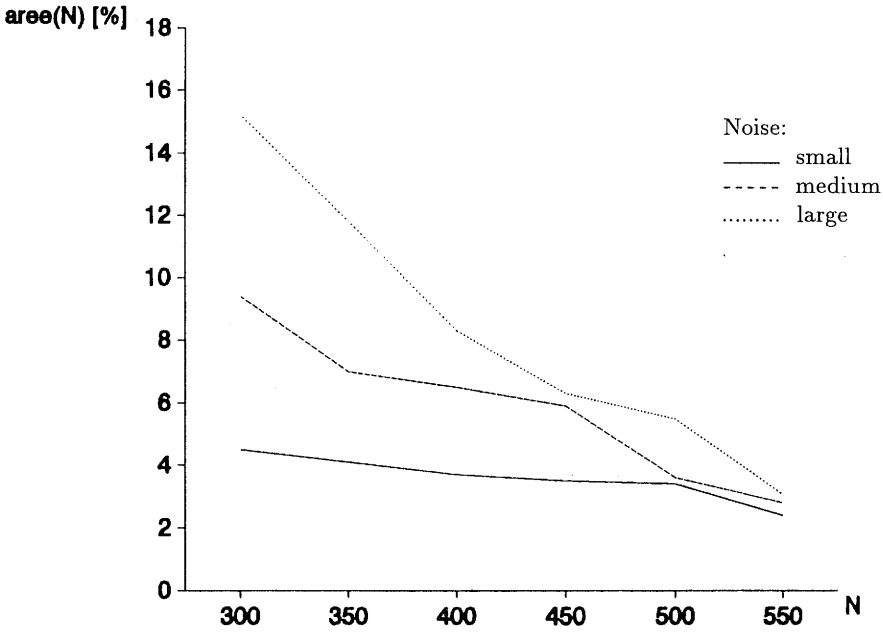

Fig. 7. Average relative estimation error versus data set size: non-linear system, strong couplings.

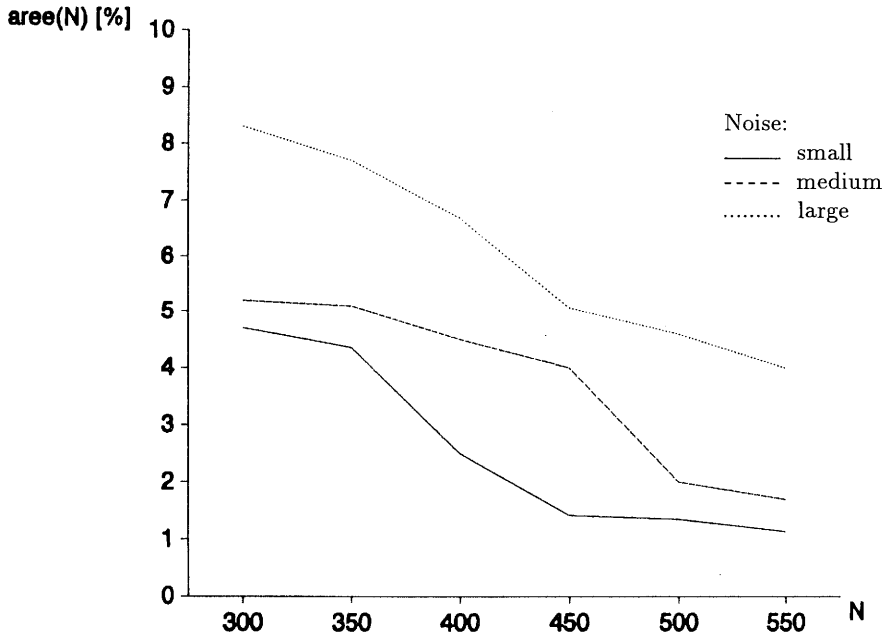

Fig. 8. Average relative estimation error versus data set size: non-linear system, moderate couplings.

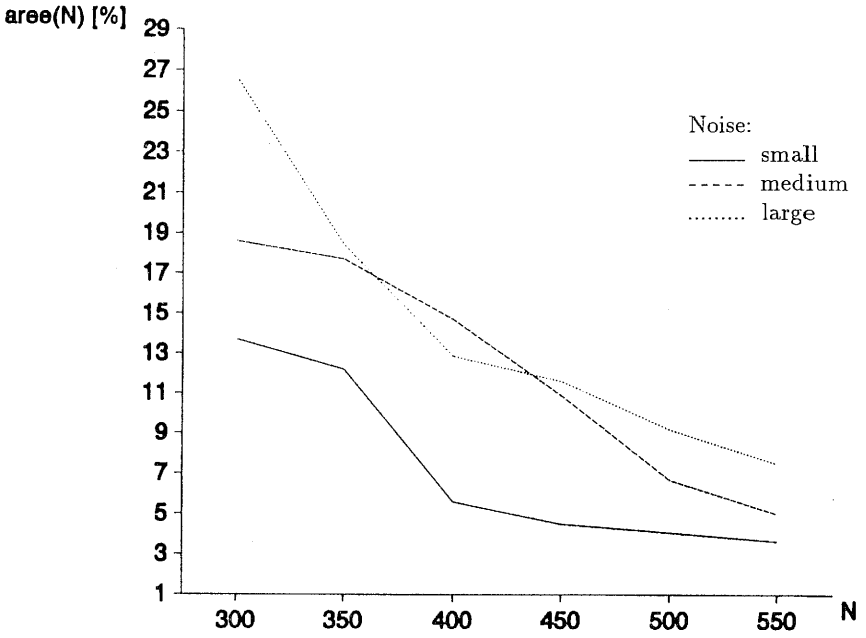

Fig. 9. Average relative estimation error versus data set size: non-linear system, weak couplings.

6. Conclusions

The empirical version of the two-stage identification algorithm from Part 1, presented in this paper, offers a computationally simple way of practical model construction for complex static systems with cascade structure. The algorithm takes advantage of the particular series structure of the system and reduces the problem of selecting the structural complex model from a given parametric collection of models (optimal in the global sense (3) and (4)) to a set of standard identification tasks of single-element systems (aggregates) solved independently at Stage 1, and a set of complementary, local, optimization (disaggregation) tasks solved in a decentralized manner at Stage 2. Thus the algorithm is easily programmable on a computer and easy to implement. The price which is paid for the simplicity of the algorithm is, in a general case, asymptotic suboptimality of the resulting empirical complex model instead of asymptotic optimality (Section 3 and Part 1 of the paper). As shown by the examples in Section 5, the efficiency of the method depends essentially on (i) the structure of the system, (ii) the strength of couplings among system elements, and (iii) the intensity of noise, and it deteriorates under extreme conditions, when disturbances grow in the system or interactions within subsystems diminish. The efficiency is also smaller for increasing non-linearities in the system (compare Examples 3 and 4 with the effect gained in the linear system identification of Example 2, where in fact, with the linear models assumed, the system parameter estimation problem was solved by the approach). However, as follows from the numerical experience, the algorithm

works rather well under moderate disturbances, interactions and non-linearities in the system, i.e. in practically relevant cases. The scheme can fail for less emphatic interconnections among system elements, particularly for large noises. It should be emphasized, however, that if interconnections in the system are really weak, and interaction inputs influence insignificantly the system outputs, then the complex model of a series system can in fact be obtained in a completely decentralized way by applying standard identification methods to individual subsystems as to a set of independent autonomous objects. Then any non-standard identification method is in fact needless (see the remarks in Section 2 of Part 1 of the paper).

Acknowledgments

The author would like to express his sincere thanks to the editors and the anonymous reviewers for a careful reading of the manuscript and helpful comments that improved the readability of the paper. I am also grateful for suggestions concerning the directions of further investigations. This work was supported in part by the State Committee for Scientific Research.

Appendix A

Proof of Lemma 1. For brevity, we shall prove the lemma for the case of accurate output measurements, i.e. for $v^k = y^k$ in (9). The general case (with $\eta^k \neq 0$) is only a bit different and can be treated analogously.

Write

$$\delta_i(\bar{c}_i, y_i; b_i) = \left\| y_i - K_{iM}(\bar{c}_i, b_i) \right\|^2 - q_i(b_i)$$

and notice that

$$\sup_{b_i \in B_i} \left| q_{iN}(b_i) - q_i(b_i) \right| = \sup_{b_i \in B_i} \left| \delta_{iN}(b_i) \right| \tag{A1}$$

where

$$\delta_{iN}(b_i) = \frac{1}{N} \sum_{k=1}^{N} \delta_i(\bar{c}_i^k, y_i^k; b_i)$$

We can easily recognize that in the case considered (cf. assumptions in Section 1)

$$\delta_{iN}(b_i) \to 0 \quad \text{w.p. 1 as } N \to \infty$$

for each $b_i \in B_i$, where we have used the strong law of large numbers and the fact that $E\delta_i(\bar{c}_i, y_i; b_i) = 0$ (cf. (5) in Section 1). By compactness of B_i and continuity of $\delta_{iN}(b_i)$ (uniform on B_i, in fact), for any $\gamma_N > 0$ there exists a positive r_N (we drop here the additional index i for simplicity) and a finite collection of points

$b_i^j \in B_i$, $j = 1, 2, \ldots, M(r_N)$, $M(r_N) < \infty$ (a finite r_N-net of B_i from the Hausdorff theorem), such that

$$B_i \subseteq \bigcup_{j=1}^{M(r_N)} Bl(b_i^j, r_N)$$

and (see e.g. (Gikhman and Skorokhod, 1974; Chapter III, Section 5))

$$\sup_{b_i \in Bl(b_i^j, r_N)} \left|\delta_{iN}(b_i) - \delta_{iN}(b_i^j)\right| \leq \gamma_N \quad \text{w.p. 1}$$

for every $j = 1, 2, \ldots, M(r_N)$, where $Bl(b_i^j, r_N)$ is the open ball of radius r_N with centre b_i^j. Hence it follows without difficulty that

$$\sup_{b_i \in B_i} \left|\delta_{iN}(b_i)\right| \leq \gamma_N + \max_{1 \leq j \leq M(r_N)} \left|\delta_{iN}(b_i^j)\right|$$

with probability one. Letting now $\gamma_N \to 0$ as $N \to \infty$ in the above bound, using (A1) and including the convergence (A2) yields the required (18). ■

Appendix B

Proof of Lemma 2. Write

$$\rho_{iN}(a_i) = \left\|\varphi_i(b_{i-1,N}, a_i) - \varphi_i(b_{i-1}^*, a_i)\right\|$$

By the inequality

$$\sup_{a_i \in A_i} \left|d_{iN}(a_i) - d_i(a_i)\right| \leq \left\|b_{iN} - b_i^*\right\| + \sup_{a_i \in A_i} \left|\rho_{iN}(a_i)\right|$$

and the convergence (19) in Section 3, it suffices to prove that

$$\sup_{a_i \in A_i} \left|\rho_{iN}(a_i)\right| \to 0 \quad \text{w.p. 1 as } N \to \infty \tag{B1}$$

The proof of (B1) may proceed in exactly the same way as that for $\delta_{iN}(b_i)$ in Appendix A (cf. (A1) and further steps), by using compactness of A_i and continuity (uniform) of $\rho_{iN}(a_i)$ on A_i as an argument, and noticing that

$$\rho_{iN}(a_i) \to 0 \quad \text{w.p. 1 as } N \to \infty$$

for each $a_i \in A_i$; the latter again by the convergence (19) and continuity of $\varphi_i(b_{i-1}, a_i)$ with respect to the parameters b_{i-1} (cf. Section 1). ■

Appendix C

Proof of Lemma 3. Applying the inequality (27) of Section 4 to the criterion function $q_i(b_i)$ and using (A1) in Appendix A, we obtain

$$\left|q_i(b_{iN}) - q_i(b_i^*)\right| \leq 2 \sup_{b_i \in B_i} \left|\delta_{iN}(b_i)\right|$$

Hence

$$P\Big\{\left|q_i(b_{iN}) - q_i(b_i^*)\right| > \alpha\Big\} \leq P\Big\{\sup_{b_i \in B_i} \left|\delta_{iN}(b_i)\right| > \frac{\alpha}{2}\Big\} \triangleq UB_{iN} \tag{C1}$$

Obviously (see the proof of Lemma 1 in Appendix A)

$$\sup_{b_i \in B_i} \left|\delta_{iN}(b_i)\right| \leq \max_{1 \leq j \leq M(r_N)} \Big\{ \sup_{b_i \in Bl(b_i^j, r_N)} \left|\delta_{iN}(b_i)\right| \Big\} \tag{C2}$$

Now, let us rename for convenience $B_i = B_{i,0}$, $\delta_{iN}(b_i) = \delta_{iN,0}(b_i)$, $r_N = r_{N,0}$, $M(r_N) = M_0(r_{N,0})$ and let

$$\begin{aligned} \delta_{iN,1}^j(b_i) &= \delta_{iN,0}(b_i) - \delta_{iN,0}(b_i^j) \\ \delta_{iN,2}^{j,\tau}(b_i) &= \delta_{iN,1}^j(b_i) - \delta_{iN,1}^j(b_i^{j,\tau}) \\ &\vdots \end{aligned} \tag{C3}$$

where $b_i^{j,\tau}$ is an element of the closure of the ball $Bl(b_i^j, r_{N,0})$ (denoted further by $B_{i,1}^j$). Exploiting (C2) and (C3), we see that

$$\Big(\sup_{b_i \in B_i} \left|\delta_{iN}(b_i)\right| = \Big) \sup_{b_i \in B_{i,0}} \left|\delta_{iN,0}(b_i)\right| \leq \max_{1 \leq j \leq M_0(r_{N,0})} \Big\{\left|\delta_{iN,0}(b_i^j)\right|\Big\}$$

$$+ \max_{1 \leq j \leq M_0(r_{N,0})} \Big\{ \sup_{b_i \in Bl(b_i^j, r_{N,0})} \left|\delta_{iN,1}^j(b_i)\right| \Big\}$$

and hence (cf. e.g. Lemma 1 in Section 4.4 of (Tucker, 1967))

$$\begin{aligned} UB_{iN} &= P\Big\{\sup_{b_i \in B_{i,0}} \left|\delta_{iN,0}(b_i)\right| > \frac{\alpha}{2}\Big\} \leq P\Big\{\max_{1 \leq j \leq M_0(r_{N,0})} \Big\{\left|\delta_{iN,0}(b_i^j)\right|\Big\} > \frac{\alpha}{4}\Big\} \\ &\quad + P\Big\{\max_{1 \leq j \leq M_0(r_{N,0})} \Big\{ \sup_{b_i \in Bl(b_i^j, r_{N,0})} \left|\delta_{iN,1}^j(b_i)\right| \Big\} > \frac{\alpha}{4}\Big\} \\ &\leq \sum_{j=1}^{M_0(r_{N,0})} P\Big\{\left|\delta_{iN,0}(b_i^j)\right| > \frac{\alpha}{4}\Big\} \\ &\quad + \sum_{j=1}^{M_0(r_{N,0})} P\Big\{ \sup_{b_i \in Bl(b_i^j, r_{N,0})} \left|\delta_{iN,1}^j(b_i)\right| > \frac{\alpha}{4}\Big\} \\ &= G_{iN,0} + H_{iN,1} \end{aligned} \tag{C4}$$

Recalling the definition of $\delta_{iN,0}(b_i)$ ($\delta_{iN}(b_i)$ in Appendix A), the fact that $E\delta_i(\bar{c}_i, y_i; b_i) = 0$ and the assumption (24) of Section 4, ensuring that for each j

$$|\delta_i(\bar{c}_i^k, y_i^k; b_i^j)| \leq \kappa_{i,0} \qquad \forall k \in \overline{1,N}$$

($\kappa_{i,0} = L_i^2$), and next applying Hoeffding's inequality to the components of the first summand, $G_{iN,0}$, in the above bound, see e.g. (Hoeffding, 1963; Karlin and Studden, 1976; Chapter XIV, Section 1), we get

$$P\Big\{|\delta_{iN,0}(b_i^j)| > \frac{\alpha}{4}\Big\} \leq 2\exp\Big(-\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\Big) \tag{C5}$$

($\beta_{i,0} = 1/32$) and consequently

$$G_{iN,0} \leq 2\sum_{j=1}^{M_0(r_{N,0})} \exp\Big(-\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\Big) \tag{C6}$$

In turn, for the components of the second summand, $H_{iN,1}$, we have

$$P\Big\{\sup_{b_i \in Bl(b_i^j, r_{N,0})} |\delta_{iN,1}^j(b_i)| > \frac{\alpha}{4}\Big\} \leq P\Big\{\sup_{b_i \in B_{i,1}^j} |\delta_{iN,1}^j(b_i)| > \frac{\alpha}{4}\Big\} \tag{C7}$$

where $B_{i,1}^j$ is the closure of the open ball $Bl(b_i^j, r_{N,0})$. Since $B_{i,1}^j$ is a compact set, $\delta_{iN,1}^j(b_i)$ is, for each j, of the same kind as $\delta_{iN,0}(b_i)$ (see (C3)) thus, by the same type of argument as in (C4), we obtain that for every j

$$P\Big\{\sup_{b_i \in B_{i,1}^j} |\delta_{iN,1}^j(b_i)| > \frac{\alpha}{4}\Big\} \leq G_{iN,1}^j + H_{iN,2}^j \tag{C8}$$

where

$$\begin{aligned} G_{iN,1}^j &= \sum_{\tau=1}^{M_1^j(r_{N,1})} P\Big\{|\delta_{iN,1}^j(b_i^{j,\tau})| > \frac{\alpha}{8}\Big\} \\ H_{iN,2}^j &= \sum_{\tau=1}^{M_1^j(r_{N,1})} P\Big\{\sup_{b_i \in Bl(b_i^{j,\tau}, r_{N,1})} |\delta_{iN,2}^{j,\tau}(b_i)| > \frac{\alpha}{8}\Big\} \end{aligned} \tag{C9}$$

and $b_i^{j,\tau}, \tau = 1, 2, \ldots, M_1^j(r_{N,1})$, is a finite $r_{N,1}$-net of the set $B_{i,1}^j(r_{N,1} < r_{N,0})$. Hence

$$H_{iN,1} \leq \sum_{j=1}^{M_0(r_{N.0})} (G_{iN,1}^j + H_{iN,2}^j) \tag{C10}$$

Since, by the assumptions (22) and (24) in Section 4, for each j and each τ

$$|\delta_i(\bar{c}_i^k, y_i^k; b_i^{j,\tau}) - \delta_i(\bar{c}_i^k, y_i^k; b_i^j)| \leq \text{ const } r_{N,0} \qquad \forall k \in \overline{1,N}$$

we can again use Hoeffding's inequality, now with respect to the components of $G^j_{iN,1}$, yielding

$$G^j_{iN,1} \leq 2 \sum_{\tau=1}^{M_1^j(r_{N,1})} \exp\Big(-\frac{N\alpha^2}{r^2_{N,0}}\beta_{i,1}\Big) \tag{C11}$$

($\beta_{i,1} > 0$ is some constant). For the components of $H^j_{iN,2}$ we can then proceed as for the components of $H_{iN,1}$ above. The scheme is continued in this fashion and leads to the following relation (see (C4), (C10) and further by analogy):

$$UB_{iN} \leq G_{iN,0} + \sum_{j=1}^{M_0(r_{N,0})} \Big(G^j_{iN,1} + \sum_{\tau=1}^{M_1^j(r_{N,1})} (G^{j,\tau}_{iN,2} + \cdots)\Big)$$

which, including (C6), (C11) and further analogous bounds, gives

$$UB_{iN} \leq 2 \sum_{j=1}^{M_0(r_{N,0})} \Big(\exp\big(-\frac{N\alpha^2}{\kappa^2_{i,0}}\beta_{i,0}\big) + \sum_{\tau=1}^{M_1^j(r_{N,1})} \Big(\exp\big(-\frac{N\alpha^2}{r^2_{N,0}}\beta_{i,1}\big)$$

$$+ \sum_{\omega=1}^{M_2^{j,\tau}(r_{N,2})} \Big(\exp\big(-\frac{N\alpha^2}{r^2_{N,1}}\beta_{i,2}\big) + \cdots\Big)\Big)\Big)$$

or, eventually, the following sequence of nested sums:

$$UB_{iN} \leq 2M_0(r_{N,0})\Big(\exp(-\frac{N\alpha^2}{\kappa^2_{i,0}}\beta_{i,0}) + M_1(r_{N,1})\Big(\exp(-\frac{N\alpha^2}{r^2_{N,0}}\beta_{i,1})$$

$$+ M_2(r_{N,2})\Big(\exp\big(-\frac{N\alpha^2}{r^2_{N,1}}\beta_{i,2}\big) + \cdots\Big)\Big)\Big) \tag{C12}$$

where

$$M_1(r_{N,1}) = \max_j M_1^j(r_{N,1}), \quad M_2(r_{N,2}) = \max_{j,\tau} M_2^{j,\tau}(r_{N,2}), \ \ldots$$

and $r_{N,0} > r_{N,1} > r_{N,2} > \cdots$. Setting now

$$r_{N,0} = \Big(\frac{1}{N}\Big)^{p_0}, \ r_{N,1} = \Big(\frac{1}{N}\Big)^{p_1}, \ r_{N,2} = \Big(\frac{1}{N}\Big)^{p_2}, \ldots, \ 0 < p_0 < p_1 < p_2 < \cdots$$

which can be done since the above reasoning holds for any arbitrary choice of $r_{N,0}$, $r_{N,1}$, $r_{N,2}, \ldots$, and noticing that with this choice ($h_i = \dim b_i$)

$$M_0(r_{N,0}) \sim N^{p_0 h_i}, \ M_1(r_{N,1}) \sim N^{(p_1-p_0)h_i}, \ M_2(r_{N,2}) \sim N^{(p_2-p_1)h_i}, \ \ldots$$

where $\mu_N \sim \nu_N$ means that $\mu_N/\nu_N = \text{const} + o(1)$, const $\neq 0$, as $N \to \infty$, we can easily recognize that for $N \to \infty$ the following is true:

$$\begin{aligned} M_1(r_{N,1}) \exp\big(-\frac{N\alpha^2}{r_{N,0}^2}\beta_{i,1}\big) &= o\Big(\exp\big(-\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\big)\Big) \\ M_2(r_{N,2}) \exp\big(-\frac{N\alpha^2}{r_{N,1}^2}\beta_{i,2}\big) &= o\Big(\exp\big(-\frac{N\alpha^2}{r_{N,0}^2}\beta_{i,1}\big)\Big) \\ &\vdots \end{aligned} \tag{C13}$$

the latter by the fact that $x^p e^{-qx} \to 0$ as $x \to \infty$ for every $p, q > 0$. Consequently, including (C12) and (C13), for a sufficiently large N we obtain

$$UB_{iN} \leq 2M_0(r_{N,0}) \exp\big(-\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\big)$$

Since, obviously, $M_0(r_{N,0}) = O(N^{p_0 h_i})$ as $N \to \infty$, we have

$$UB_{iN} = O\Big(N^{p_0 h_i} \exp\big(-\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\big)\Big)$$

for each $p_0 > 0$. However, for $p_0 h_i \ll 1$ we realize from the Taylor series expansion that

$$N^{-p_0 h_i} \exp\Big(\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\Big) = \exp\Big(\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\Big) + N^{-p_0 h_i} - 1$$

which gives asymptotically (for $N \to \infty$) on the right-hand side

$$\exp\big(\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\big)$$

In consequence, since in the above consideration $p_0 > 0$ may be arbitrary, we conclude that for a sufficiently large N

$$UB_{iN} = O\Big(\exp\big(-\frac{N\alpha^2}{\kappa_{i,0}^2}\beta_{i,0}\big)\Big) \tag{C14}$$

as was to be proved, taking account of (30) in Section 4, (C1) and the constants $\kappa_{i,0} = L_i^2$ and $\beta_{i,0} = 1/32$. ■

References

Bickel P.J. and Doksum K.A. (1977): *Mathematical Statistics. Basic Ideas and Selected Topics.* — San Francisco: Holden-Day.

Gikhman I.I. and Skorokhod A.V. (1974): *The Theory of Stochastic Processes, Vol. 1.* — New York: Springer.

Greblicki W. (1994): *Nonparametric identification of Wiener systems by orthogonal series.* — IEEE Trans. Automat. Contr., Vol.39, No.10, pp.2077–2086.

Greblicki W. and Pawlak M. (1994): *Cascade non-linear system identification by a non-parametric method.* —Int. J. Systems Sci., Vol.25, No.1, pp.129–153.

Hasiewicz Z. (1988): *Two-stage procedure for parameter estimation in large-scale interconnected linear zero-memory systems.* — Int. J. Systems Sci., Vol.19, No.3, pp.497–516.

Hasiewicz Z. (1996): *Two-stage identification of interconnected steady-state systems with cascade structure: a parametric approach; Part 1: Background.* — Appl. Math. Comp. Sci., Vol.6, No.1, pp.101–120.

Hoeffding W. (1963): *Probability inequalities for sums of bounded random variables.* — J. Amer. Statist. Assoc. Vol.58, No.1, pp.13–30.

Karlin S. and Studden W.J. (1976): *Tchebycheff Systems with Applications in Analysis and Statistics.* — New York: Wiley.

Ljung L. (1976): *On consistency and identifiability.* — Math. Programm. Study, Vol.5, No.2, pp.169–190.

Ljung L. (1978): *Convergence analysis of parametric identification methods.* — IEEE Trans. Automat. Contr., Vol.23, No.5, pp.770–783.

Tucker H.G. (1967): *A Graduate Course in Probability.* — New York: Academic Press.

Vapnik V. (1982): *Estimation of Dependences Based on Empirical Data.* — New York: Springer.

Received: August 14, 1995
Revised: March 20, 1996