Kybernetika

Jaromír Štěpán
Robust quasi-linear system identification

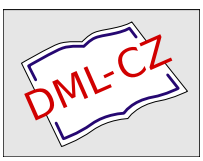
Kybernetika, Vol. 24 (1988), No. 4, 259--277

Persistent URL: http://dml.cz/dmlcz/124416

Terms of use:

© Institute of Information Theory and Automation AS CR, 1988

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://project.dml.cz

KYBERNETIKA — VOLUME 24 (1988), NUMBER 4

ROBUST QUASI-LINEAR SYSTEM IDENTIFICATION

JAROMÍR ŠTĚPÁN

The internally robust estimator, which can start from a small sample size, will be proposed. It is the nonlinear Output Error (OE) method which uses the linear Gauss estimator as an etalon for the solution of the nonlinear identification problem and so we can speak about the quasi-linear method. The close connection with the approximation method described in ([10], [12]) opens the way for deriving approximate models.

1. INTRODUCTION

The system identification is a necessary step in any application of the theory to real world systems. So it is a link between the mathematical resp. hypothetical-deductive theories and the real world. This position is the source of many difficulties especially by the design of the mathematical apparatus. We can find plenty of nice formal identification theories in the literature. Unfortunately these formal theories are mostly based on very strong assumptions and so legitimate many procedures or methods which stand in contradiction with the real world. They are mostly based on the exact system description using error free computers, exact measurements and limitless time. Let us discuss some of these assumptions.

(i) The measurement errors were imbedded in formal theories in a stochastic zero-mean component neglecting the fundamental fact that each real measurement instrument generates data corrupted by an unknown bias. Therefore we must start from the fact that the measurement instruments used in the practice have different accuracy and reliability characteristics and the measured data are mostly given with a precision lower than two decimal places. Just this limited measurement accuracy, i.e. the residual systematic errors, creates the basic part of an uncertainty band (see [2], [8], [11]).

(ii) Often one does not have the control over disturbances and the form of the input signal. Non-zero initial conditions are mostly assumed but they cannot be

exactly fulfilled. All these uncertain influences must be included in an uncertainty band and decrease more and more the usable information in measured signals.

(iii) We ask the model which is a good representation of the real system. However, the model structure is seldom identical to that of the true system. Approximate models are more robust than the exact model if measured signals are corrupted with an uncertainty band (cf. [11]).

(iv) It follows from preceding points that the accuracy of a parameter vector estimate cannot be increased arbitrarly by unlimited multiplying the number of measured data (cf. [8], [11]).

So in this paper we shall try to derive the identification procedure which is able to master the mentioned difficulties:

(a) This procedure must be able to solve simultaneously the approximation and estimation problem.

(b) Usable results can be obtained from a small number of measured data.

(c) This procedure must be sufficiently robust to the uncertainty band which corrupts measured signals.

From all these demands it follows that such a method must substantially differ from the known methods. It must be a nonlinear off line working OE method. The estimator, which fulfils these demands, is proposed in the next sections. It starts from the Damped Nonlinear Least Squares (DNLS) method which was published in the literature ([10], [12]) as the method for the approximation of signals. So we shall use for it the notation – the DNLS estimator.

The organization of the paper is as follows. The identification problem is formulated and the nonlinear regression functions for the general input signal are derived in Section 2. The theory of the nonlinear regression analysis is discussed with respect to different modifications of the Newton method in Section 3. The DNLS estimator is derived and the main propositions are given in Section 4. The relation of the DNLS estimator to the Equation Error (EE) methods is discussed in Section 5. The internal robustness of the DNLS estimator is illustrated in Section 6 by an example.

2. PROBLEM FORMULATION

Consider a SISO linear system given by

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\,\boldsymbol{x}(t) + \boldsymbol{\beta}\,u(t)\,, \quad y(t) = \boldsymbol{\gamma}\,\boldsymbol{x}(t)\,, \tag{2.1}$$

where the system triple $(\boldsymbol{A}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ has the appropriate dimension. For the zero initial conditions the corresponding transfer function can be written as

$$F(s) = \boldsymbol{\gamma}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\,\boldsymbol{\beta} = \frac{M(s)}{N(s)}\,. \tag{2.2}$$

The pertinent model can be described with the substitute transfer function $\bar{F}(s)$

(for $\bar{n} \leqq n, \bar{m} \leqq m$)

$$\bar{F}(s, \boldsymbol{b}, \boldsymbol{a}) = \frac{\bar{M}(s)}{\bar{N}(s)}, \tag{2.3}$$

where $\bar{M}(s) = 1 + \sum_{j=1}^{\bar{m}} b_j s^j$ and $\bar{N}(s) = \sum_{i=0}^{\bar{n}} a_i s^i$ are Hurwitz polynomials without common factor.

The pertinent output signal $\bar{y}(t)$ is nonlinear in coefficients a_i $(i = 0, 1, \ldots \bar{n})$ as it follows from the gradient

$$\begin{aligned} \operatorname{grad} \bar{F}(s, \boldsymbol{b}, \boldsymbol{a}) = 1/\bar{N}^2(s) \{ & s\,\bar{N}(s), s^2\,\bar{N}(s), \ldots, s^{\bar{m}}\,\bar{N}(s), \\ & - \bar{M}(s), -s\,\bar{M}(s), \ldots, -s^{\bar{n}}\,\bar{M}(s) \} . \end{aligned} \tag{2.4}$$

For the sake of simplicity we shall consider only the nonlinear part of the identification problem, i.e. $\bar{F}(s)$ with $\bar{M}(s) = 1$

$$\bar{F}(s, \boldsymbol{a}) = \frac{1}{\sum_{i=0}^{\bar{n}} a_i s^i} . \tag{2.5}$$

This case is more illustrative for the explanation and more useful for applications. In the literature [12] it is shown how the identification starting from transfer function (2.3) can be simply transformed in the considered case.

2.1. Signals description

The identification problem can start only from measured points of output and input signals, i.e. from the discrete alternative. In this paper we shall consider the discrete alternative which is based on the continuous case. The reasons for such approach are two:

(i) The real systems are continuous in nature and so we obtain rather easily physical insight in control problems.

(ii) The use of Z-transformation is connected with numerical errors (cf. [4]) and does not allow introducing the nonlinear regression functions.

Now let $\tilde{\boldsymbol{y}} = [\tilde{y}(1), \tilde{y}(2), \ldots, \tilde{y}(q)]^{\mathrm{T}}$ be for $\tilde{y}(t_k) = \tilde{y}(k)$ the measured points of the output signal $y(t)$ pertinent to the system S given by transfer function (2.2) and let $\bar{\boldsymbol{y}} = [\bar{y}(1), \bar{y}(2), \ldots, \bar{y}(q)]^{\mathrm{T}}$ be a signal pertinent to the model $\bar{S}$ given by transfer function (2.5). So we consider signals on the time interval $\Delta t = \tau_2 - \tau_1$ with the sampling interval $\Delta\tau = \Delta t/q$.

The output signal of the model can be written in the form of the nonlinear regression function (see [10], [12])

$$\bar{y}(t, \boldsymbol{a}) = \sum_{i=0}^{\bar{n}} a_i\, v^{(i)}(t) , \tag{2.6}$$

where

$$\mathscr{L}\{v^{(i)}(t)\} = \frac{s^i\, u(s)}{\bar{N}^2(s)} .$$

It follows from the pertinent Laplace transform

$$\mathscr{L}\{\bar{y}(t, \boldsymbol{a})\} = \bar{F}(s, \boldsymbol{a})\, u(s) = \mathscr{L}\{\sum_{i=0}^{\bar{n}} a_i\, v^{(i)}(t)\} = \tag{2.7}$$

$$= \frac{\sum_{i=0}^{\bar{n}} a_i s^i}{\bar{N}^2(s)}\, u(s) = \frac{u(s)}{\bar{N}(s)}\,.$$

The discrete form of this regression function is given by

$$\bar{y}(k) = \sum_{i=0}^{\bar{n}} a_i\, v^{(i)}(k) \quad (k = 1, 2, \ldots, q)\,. \tag{2.8}$$

2.2. The identification problem

The identification of the system S given by the measured points $\tilde{\boldsymbol{y}}$ can be formulated in the following way: To the signal $\tilde{\boldsymbol{y}} \in V$ find a model pertinent to the signal $\bar{\boldsymbol{y}} \in \bar{V} (\bar{V} \subset V)$ such that

$$Q(\boldsymbol{a}) = (\tilde{\boldsymbol{y}} - \bar{\boldsymbol{y}})^{\mathrm{T}} (\tilde{\boldsymbol{y}} - \bar{\boldsymbol{y}}) = \sum_{k=1}^{q} [\tilde{y}(k) - \bar{y}(k)]^2 \tag{2.9}$$

takes the minimum value. So we consider the identification problem in sense of the nonlinear least squares. We start from the fact that the classical theory of errors and, in particular, the principle of least squares is a nonprobabilistic method (cf. [3] pp. 249). If we use an approximate model then the approximation error is a bias and the system identification may be a nonprobabilistic problem (cf. [11]).

2.3. General input

The general regression function of the type (2.6) resp. (2.8) must be derived before the system identification can be solved. First let us introduce the sensitivity function $v_s^{(i)}(t)$ pertinent to the unit step. Then the regression function for an arbitrary input $u(t)$ can be written in two forms. The first alternative pertinent to relation (2.8) is based on a sequence of steps (for $u(0) = 0$)

$$\bar{y}(k) = \sum_{i=0}^{\bar{n}} \sum_{l=1}^{k} a_i\, v_s^{(i)}(l) \left[u(k - l + 1) - u(k - l)\right] = \tag{2.10}$$

$$= \sum_{i=0}^{\bar{n}} a_i\, v_s^{(i)}(k) \circ u(k) \quad (k = 1, 2, \ldots, q)\,,$$

where $v_s^{(i)}(k) \circ u(k) = \sum_{l=1}^{k} v_s^{(i)}(l) \left[u(k - l + 1) - u(k - l)\right]$.

The second alternative is given by a sequence of impulses (for $v^{(i)}(0) = 0$)

$$\bar{y}(k) = \sum_{i=0}^{\bar{n}} a_i\, u(k) \circ v_s^{(i)}(k)\,, \tag{2.11}$$

where

$$u(k) \circ v_s^{(i)}(k) = \sum_{l=1}^{k} u(l) \left[v_s^{(i)}(k - l + 1) - v_s^{(i)}(k - l) \right].$$

3. ROBUSTNESS OF NONLINEAR IDENTIFICATION METHODS

Let us sketch the derivation of the Newton method to get the insight in the nonlinear regression analysis, and especially in the problems connected with its robustness. Let us start with error function (2.9) in the form

$$Q(\boldsymbol{a}) = \boldsymbol{f}^{\mathrm{T}}(\boldsymbol{a}) \boldsymbol{f}(\boldsymbol{a}), \tag{3.1}$$

where $f(k, \boldsymbol{a}) = \tilde{y}(k) - \bar{y}(k, \boldsymbol{a})$ $(k = 1, 2, \ldots, q)$, which can be approximated with the first three members of the Taylor expansion

$$Q(\boldsymbol{a} + \Delta\boldsymbol{a}) = Q(\boldsymbol{a}) + Q'(\boldsymbol{a}) \Delta\boldsymbol{a} + \tfrac{1}{2} Q''(\boldsymbol{a}) \Delta\boldsymbol{a}^2 + \ldots . \tag{3.2}$$

The solution must be iterative and so relation (3.2) with ${}^{j+1}\boldsymbol{a} = {}^{j}\boldsymbol{a} + \Delta^{j}\boldsymbol{a}$ will be used. Superscripts on the left indicate the iteration steps. Then the solution is given by the derivative of the error function

$$Q({}^{j+1}\boldsymbol{a}) \doteq Q({}^{j}\boldsymbol{a}) + Q'({}^{j}\boldsymbol{a}) \Delta^{j}\boldsymbol{a} + \tfrac{1}{2} Q''({}^{j}\boldsymbol{a}) \Delta^{j}\boldsymbol{a}^2 \tag{3.3}$$

at $\Delta^{j}\boldsymbol{a}$, i.e. from the condition

$$Q'({}^{j}\boldsymbol{a}) + Q''({}^{j}\boldsymbol{a}) \Delta^{j}\boldsymbol{a} = 0 \tag{3.4}$$

we obtain the known result (see [10], [14])

$$\Delta^{j}\boldsymbol{a} = - \left[Q''({}^{j}\boldsymbol{a}) \right]^{-1} Q'({}^{j}\boldsymbol{a}). \tag{3.5}$$

Now let us introduce the notation for the first partial derivatives of ${}^{j}\bar{y}(t, {}^{j}\boldsymbol{a})$ by

$$\partial_i {}^{j}\bar{y}(t, {}^{j}\boldsymbol{a}) = \frac{\partial}{\partial a_i} {}^{j}\bar{y}(t, {}^{j}\boldsymbol{a}) \quad (i = 0, 1, \ldots . \bar{n})$$

and the second derivatives by

$$\partial_i \partial_e {}^{j}\bar{y}(t, {}^{j}\boldsymbol{a}) = \frac{\partial^2}{\partial a_i \partial a_e} {}^{j}\bar{y}(t, {}^{j}\boldsymbol{a}) \quad (i, e = 0, 1, \ldots, \bar{n}).$$

The elements of the matrix $Q'({}^{j}\boldsymbol{a})$ can be derived from the following relations

$$\partial_i Q({}^{j}\boldsymbol{a}) = 2 \sum_{k=1}^{q} \partial_i f(k, {}^{j}\boldsymbol{a}) f(k, {}^{j}\boldsymbol{a}) = -2 \sum_{k=1}^{q} \partial_i {}^{j}\bar{y}(k, {}^{j}\boldsymbol{a}) f(k, {}^{j}\boldsymbol{a}) \tag{3.6}$$

$$(i = 0, 1, \ldots, \bar{n}) \, (k = 1, 2, \ldots, q)$$

and the pertinent matrix form is given by

$$Q'({}^{j}\boldsymbol{a}) = -2\boldsymbol{H}^{\mathrm{T}}({}^{j}\boldsymbol{a}) \boldsymbol{f}({}^{j}\boldsymbol{a}), \tag{3.7}$$

where $\boldsymbol{H}({}^{j}\boldsymbol{a})$ is the Jacobian matrix.

Similarly we obtain the elements of the Hessian matrix

$$\partial_i \partial_e Q({}^j\boldsymbol{a}) = 2\sum_{k=1}^{q} \partial_i {}^j\bar{y}(k, {}^j\boldsymbol{a})\, \partial_e {}^j\bar{y}(k, {}^j\boldsymbol{a}) - \tag{3.8}$$

$$- 2\sum_{k=1}^{q} \partial_i \partial_e {}^j\bar{y}(k, {}^j\boldsymbol{a})\, f(k, {}^j\boldsymbol{a}) \quad (i, e = 0, 1, \ldots, \bar{n}),\ (k = 1, 2, \ldots, q).$$

The Newton method is given by relations (3.5), (3.7) and (3.8). If we neglect the second part of relation (3.8) then we get the Gauss-Newton method and relation (3.5) has the form

$$\Delta {}^j\boldsymbol{a} = [\boldsymbol{H}^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}({}^j\boldsymbol{a})]^{-1}\, \boldsymbol{H}^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{f}({}^j\boldsymbol{a}). \tag{3.9}$$

Before we analyse the robustness of the considered methods let us first define the term "robustness" because its meaning is vague in the literature. We shall use this term for the insensitivity of identification and control algorithms to errors of input data. Let us emphasize in this connection the fundamental problem of control practice how to include in the pertinent algorithms the measured data, i.e. data which are mostly given with precision lower than two decimal places. On the other side the computation on digital computers can have essentially higher precision, e.g. the double precision. Now, if we are not able to separate these different precision levels then the robustness analysis is a complicated problem which is hardly solvable (cf. [11]). So we must distinguish at least two parts of the robustness problem. The ,,internal" robustness is connected with the high precision of the computation. It must be first assured and then the analysis of the "external" robustness, i.e. the robustness to errors of measured data, makes sense. This external robustness is connected with the empirical (resp. a posteriori) identifiability of systems (cf. [7], [8]). The internal robustness is therefore the topic of numerical mathematics whereas the external robustness is related to statistics.

Let us revert to the robustness analysis of the considered methods. Here we can essentially simplify the analysis of the internal robustness and to analyze only the inversion of the information matrix, i.e. the main computation line of the considered algorithms. The elements of the information matrix pertinent to the Newton method are given by measured data too as it follows from relation (3.8). So we cannot expect that the pertinent inversion will be internally robust because we mostly lose during the computation for $\bar{n} > 2$ more than two decimal places. Here we can formulate more general statement: All methods, which are based on the inversion of matrices given by measured data, cannot be in nontrivial cases internally robust. Therefore the internal robustness of the Kalman-Bucy filter or EE estimators can be expected only near the asymptotic solution (see Section 5).

On the contrary the inversion of the matrix $(\boldsymbol{H}^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}({}^j\boldsymbol{a}))$ pertinent to the Gauss-Newton method is internally robust because the elements of this matrix can be computed with the double precision. So the loss of a few decimal places during the computation cannot influence the results. This conclusion is corroborated with

the experience given by testing different modifications of the Newton method (cf. [14] pp. 229).

Let us remark that the internal robustness of the individual iteration steps does not lead automatically to the convergence of the sequence of iterations. So we must pay for the mentioned advantage by the transition to the next step. The Gauss-Newton method can be effectively used if the following condition holds (see [10], [12])

$$(\boldsymbol{H}^{\mathrm{T}}({}^{j+1}\boldsymbol{a})\,\boldsymbol{H}({}^{j+1}\boldsymbol{a})) \approx (\boldsymbol{H}^{\mathrm{T}}({}^{j}\boldsymbol{a})\,\boldsymbol{H}({}^{j}\boldsymbol{a}))\,. \tag{3.10}$$

At the present time there is no procedure how to predict $\boldsymbol{H}({}^{j+1}\boldsymbol{a})$ from the parameters of the jth iteration step. The change of $\boldsymbol{H}({}^{j}\boldsymbol{a})$ to $\boldsymbol{H}({}^{j+1}\boldsymbol{a})$ is not under the computational control (cf. [10]). This is the main reason why the applicability region of the Gauss-Newton method is so small and why different modifications are used.

This problem was solved by deriving the DNLS method in ([10], [12]). In the next section we shall show how these results can be used by deriving the robust estimator.

4. THE DNLS ESTIMATOR

The DNLS estimator must always start from a curve of signals or better from measured points of signals. That is the main difference to the DNLS method which was derived for the signal approximation from a known original transfer function (cf. [10], [12]). On the other hand the idea of this approximation method, how to solve the problem connected with relation (3.10), can be used by deriving the DNLS estimator. As we shall show later it is necessary to introduce the linear Gauss estimator as an etalon for the solution of the nonlinear regression problem. In this way the DNLS estimator loses the character of the Newton method and belongs rather to the quasi-linear regression methods.

In the next sections we shall use if possible the same notation as in the literature ([10], [12]) where the continuous case was considered. Let us show how this continuous case is related to the discrete version pertinent to the DNLS estimator. If the integral part $\int_{\tau_2}^{\infty} {}^{j}\bar{y}^2(t)\,\mathrm{d}t \to 0$ and the discretization error is sufficiently small then the difference between these two cases is approximately given by the factor $\Delta\tau$ as it follows with respect to Paragraph 2.2 for $\tau_1 = 0$ and $\Delta\tau = \tau_2/q$ from the relation

$$\|{}^{j}\bar{y}\|_{L_2}^2 = \int_0^{\infty} {}^{j}\bar{y}^2(t)\,\mathrm{d}t \doteq \Delta\tau\|{}^{j}\bar{\boldsymbol{y}}\|^2 = \Delta\tau\sum_{k=1}^{q} {}^{j}\bar{y}^2(k)\,. \tag{4.1}$$

4.1. The linear regression function

Let us assume that the regression function (2.6) resp. the discrete case given by (2.8) is linear, i.e. the sensitivity functions $v^{(i)}(t)$ resp. $v^{(i)}(k)$ $(k = 1, 2, \ldots, q)$ are known. Then the discrete linear regression function can be written in the matrix form

$$\bar{\boldsymbol{y}}_L = \boldsymbol{H}_L\boldsymbol{a}\,. \tag{4.2}$$

Subscripts L indicate the linear case. The necessary condition that $\hat{\boldsymbol{a}}$ be a minimizer of $Q_L = (\tilde{\boldsymbol{y}} - \boldsymbol{z})^{\mathrm{T}} (\tilde{\boldsymbol{y}} - \boldsymbol{z})$ for $\tilde{\boldsymbol{y}} \in V$ and $\boldsymbol{z} \in \overline{V}(\overline{V} \subset V)$ is that

$$\left.\frac{\partial Q_L}{\partial \boldsymbol{a}}\right|_{\hat{\boldsymbol{a}}=\boldsymbol{a}} = 2\boldsymbol{H}_L^{\mathrm{T}}\boldsymbol{H}_L\hat{\boldsymbol{a}} - 2\boldsymbol{H}_L^{\mathrm{T}}\tilde{\boldsymbol{y}} = 0\,. \tag{4.3}$$

We obtain the known result

$$\hat{\boldsymbol{a}} = (\boldsymbol{H}_L^{\mathrm{T}}\boldsymbol{H}_L)^{-1}\,\boldsymbol{H}_L^{\mathrm{T}}\tilde{\boldsymbol{y}} \tag{4.4}$$

and the solution is given by the unique best regression function $\boldsymbol{z} = \boldsymbol{H}_L\hat{\boldsymbol{a}}$.

The following proposition will be later useful.

Proposition 4.1. If the system given by transfer function (2.5) is stable, i.e. $\overline{N}(s, \boldsymbol{a})$ is the Hurwitz polynomial, then the matrix $(\boldsymbol{H}_L^{\mathrm{T}}(\boldsymbol{a})\,\boldsymbol{H}_L(\boldsymbol{a}))$ is nonsingular.

Proof. Each stable signal pertinent to transfer function (2.5) can be written in the form of regression function (4.2) and therefore the backwards linear regression, i.e. the minimizer of $Q_L^{+} = (\tilde{\boldsymbol{y}}_L - \boldsymbol{z})^{\mathrm{T}} (\tilde{\boldsymbol{y}}_L - \boldsymbol{z}) = 0$, must exist. □

4.2. The gradient vector

The linear regression function from Paragraph 4.1. can be introduced in each iteration step of the procedure given by relation (3.9). In this paragraph we compare the gradient vectors pertinent to the linear case given by ${}^{j}\tilde{\boldsymbol{y}}_L = \boldsymbol{H}_L({}^{j}\boldsymbol{a})\,{}^{j}\boldsymbol{a}$ and the nonlinear case given by ${}^{j}\tilde{\boldsymbol{y}} = \boldsymbol{H}({}^{j}\boldsymbol{a})\,{}^{j}\boldsymbol{a}$. For the sake of simplicity we shall start from the continuous case to be able to use the pertinent Laplace transforms. The considered functions resp. functionals can be used if the pertinent systems are stable. Therefore let us introduce the following definition:

Definition 4.1. The vector of the coefficients ${}^{j}\boldsymbol{a}$ pertinent to the model with the transfer function ${}^{j}\overline{F}(s, {}^{j}\boldsymbol{a}) = {}^{j}\overline{N}^{-1}(s)$ is an element of the subset Ω_a of stable vectors ${}^{j}\boldsymbol{a}$ if ${}^{j}\overline{N}(s) = \sum_{i=0}^{n} {}^{j}a_i s^i$ is the Hurwitz polynomial.

In this paper we assume that the stability difference between the continuous and discrete case is negligible small with respect to the error given by an uncertainty band.

Now let us prove the following proposition:

Proposition 4.2. If

(i) ${}^{j}\boldsymbol{a} \in \Omega_a$,

(ii) ${}^{j}\tilde{y}_L, {}^{j}\tilde{y} \in \overline{V}(\overline{V} \subset V)$,

then the gradient vector of ${}^{j}\tilde{y}(t, {}^{j}\boldsymbol{a})$ at ${}^{j}\boldsymbol{a}$ is given by the negative gradient vector ${}^{j}\tilde{y}_L(t, {}^{j}\boldsymbol{a})$ at ${}^{j}\boldsymbol{a}$, i.e. it holds

$$\boldsymbol{g}(t, {}^{j}\boldsymbol{a}) = -\boldsymbol{g}_L(t, {}^{j}\boldsymbol{a})\,. \tag{4.5}$$

Proof. The gradient vector of ${}^{j}\tilde{y}_L(t, {}^{j}\boldsymbol{a})$ at ${}^{j}\boldsymbol{a}$ for the known functions ${}^{j}v^{(i)}(t)$

$(i = 0, 1, \ldots, \bar{n})$ follows directly from relations (2.6)

$$\boldsymbol{g}_L(t, {}^j\boldsymbol{a}) = [{}^jv^{(0)}(t), {}^jv^{(1)}(t), \ldots, {}^jv^{(\bar{n})}(t)]^{\mathrm{T}}. \tag{4.6}$$

The gradient vector pertinent to the nonlinear regression function ${}^j\bar{y}(t, {}^j\boldsymbol{a})$ is given by the relation

$$\begin{aligned}\partial_i\, {}^j\bar{y}(t, {}^j\boldsymbol{a}) &= {}^jv^{(i)}(t, {}^j\boldsymbol{a}) + \sum_{e=0}^{\bar{n}} a_e\, \partial_i\, {}^jv^{(e)}(t, {}^j\boldsymbol{a}) = \\ &= {}^jv^{(i)}(t, {}^j\boldsymbol{a}) + \mathscr{L}^{-1}\left\{\sum_{e=0}^{\bar{n}} a_e\, \partial_i \frac{u(s)\, s^e}{{}^j\overline{N}^2(s, {}^j\boldsymbol{a})}\right\} = \\ &= {}^jv^{(i)}(t, {}^j\boldsymbol{a}) + \mathscr{L}^{-1}\left\{\frac{-2\, {}^j\overline{N}(s, {}^j\boldsymbol{a})\, u(s)\, s^i \sum_{e=0}^{\bar{n}} {}^ja_e s^e}{{}^j\overline{N}^4(s, {}^j\boldsymbol{a})}\right\} = \\ &= {}^jv^{(i)}(t, {}^j\boldsymbol{a}) - 2\, {}^jv^{(i)}(t, {}^j\boldsymbol{a}) = -{}^jv^{(i)}(t, {}^j\boldsymbol{a}).\end{aligned} \tag{4.7}$$

Hence

$$\boldsymbol{g}(t, {}^j\boldsymbol{a}) = [-{}^jv^{(0)}(t), -{}^jv^{(1)}(t), \ldots, -{}^jv^{(\bar{n})}(t)]^{\mathrm{T}}. \tag{4.8}$$

□

From this proposition it follows with respect to relations (2.6), (2.8) and (4.2) that it must hold for the pertinent Jacobian matrices

$$\boldsymbol{H}({}^j\boldsymbol{a}) = -\boldsymbol{H}_L({}^j\boldsymbol{a}). \tag{4.9}$$

4.3. Basic relations of the DNLS estimator

Relation (4.4) can be used in the form

$${}^{j+1}\hat{\boldsymbol{a}} - {}^j\boldsymbol{a} = \Delta\, {}^j\boldsymbol{a}_L = [\boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}_L({}^j\boldsymbol{a})]^{-1}\, \boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, (\tilde{\boldsymbol{y}} - {}^j\bar{\boldsymbol{y}}) \tag{4.10}$$

with respect to the relation

$${}^j\boldsymbol{a} = [\boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}_L({}^j\boldsymbol{a})]^{-1}\, \boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, {}^j\bar{\boldsymbol{y}}.$$

Relation (4.10) differs from relation (3.9) only in the sign. Therefore the DNLS estimator is given by the following two relations

$$\begin{aligned}{}^{j+1}\boldsymbol{a} &= {}^j\boldsymbol{a} + {}^j\mu[\boldsymbol{H}^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}({}^j\boldsymbol{a})]^{-1}\, \boldsymbol{H}^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{f}({}^j\boldsymbol{a}) = \\ &= {}^j\boldsymbol{a} - {}^j\mu[\boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}_L({}^j\boldsymbol{a})]^{-1}\, \boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{f}({}^j\boldsymbol{a})\end{aligned} \tag{4.11}$$

and

$$\begin{aligned}{}^{j+1}\hat{\boldsymbol{a}} &= {}^j\boldsymbol{a} + {}^j\mu[\boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}_L({}^j\boldsymbol{a})]^{-1}\, \boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{f}({}^j\boldsymbol{a}) = \\ &= {}^j\boldsymbol{a} - {}^j\mu[\boldsymbol{H}^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}({}^j\boldsymbol{a})]^{-1}\, \boldsymbol{H}^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{f}({}^j\boldsymbol{a}),\end{aligned} \tag{4.12}$$

where ${}^j\mu \in (0, 1)$ is the damping factor.

If we multiply relations (4.11) and (4.12) containing $\boldsymbol{H}_L({}^j\boldsymbol{a})$ with the matrix $\boldsymbol{H}_L({}^j\boldsymbol{a})$ so we obtain the following functions

$${}^j\bar{\boldsymbol{y}}({}^j\boldsymbol{a}, {}^{j+1}\boldsymbol{a}, {}^j\mu) = \boldsymbol{H}_L({}^j\boldsymbol{a})\, {}^{j+1}\boldsymbol{a} = {}^j\bar{\boldsymbol{y}}({}^j\boldsymbol{a}) - {}^j\mu\, \Delta\, {}^j\bar{\boldsymbol{y}} \tag{4.13}$$

and

$$^{j}z(^{j}a, {}^{j+1}\hat{a}, {}^{j}\mu) = H_L(^{j}a)\,{}^{j+1}\hat{a} = {}^{j}\bar{y}(^{j}a) + {}^{j}\mu\,\Delta\,{}^{j}\bar{y} \tag{4.14}$$

where $^{j}\bar{y}(^{j}a) = H_L(^{j}a)\,{}^{j}a$, $\Delta\,{}^{j}\bar{y} = H_L(^{j}a)\,\Delta\,{}^{j}a_L$ and $\Delta\,{}^{j}a_L = -\Delta\,{}^{j}a = {}^{j+1}\hat{a} - {}^{j}a = {}^{j}a - {}^{j+1}a$.

The linear case given by (4.12) and (4.14) serves as an etalon for the solution of the pertinent nonlinear case given by (4.11) and (4.13).

Let us prove one important property of this estimator connected with the internal robustness.

Proposition 4.3. If $\{^{j}a\} \in \Omega_a$ then the nonsingular matrix

$$H^{\mathrm{T}}(^{j}a)\,H(^{j}a) = H_L^{\mathrm{T}}(^{j}a)\,H_L(^{j}a) \tag{4.15}$$

must always exist.

The proof follows directly from the linear part of the DNLS estimator, i.e. from Proposition 4.1 and 4.2.

4.4. Main propositions

First let us discuss the prediction of the matrix $H(^{j+1}a)$ from the parameters of the jth step, i.e. the problem connected with relation (3.10). Let us start with the following proposition:

Proposition 4.4. If

(i) $\{^{j}a\} \in \Omega_a$,

(ii) $^{j+1}\bar{y}(^{j+1}a, {}^{j}\mu) - {}^{j}\bar{y}(^{j}a, {}^{j+1}a, {}^{j}\mu) \doteq 2\,{}^{j}\mu\,\Delta\,{}^{j}\bar{y}$,

then it holds

$$^{j+1}\bar{y}(^{j+1}a, {}^{j}\mu) \doteq {}^{j}z(^{j}a, {}^{j+1}\hat{a}, {}^{j}\mu)\,. \tag{4.16}$$

Proof. Relation (4.16) resp. Hypothesis (ii) are valid for $^{j}\mu = 0$. Now Hypothesis (ii) holds according to relation (4.7) for $^{j}\mu > 0$ if we are sufficiently near to the linear case, i.e.

$$\begin{aligned} ^{j+1}\bar{y}(^{j+1}a, {}^{j}\mu) &= H_L(^{j+1}a)\,{}^{j+1}a \doteq {}^{j}\bar{y}(^{j}a, {}^{j+1}a, {}^{j}\mu) + \\ &+ 2\,{}^{j}\mu\,\Delta\,{}^{j}\bar{y} = H_L(^{j}a)\,{}^{j+1}a + 2\,{}^{j}\mu\,H_L(^{j}a)\,\Delta\,{}^{j}a_L\,. \end{aligned} \tag{4.17}$$

Insert (4.13) into (4.17) and with regard to (4.14) the result is

$$^{j+1}\bar{y}(^{j+1}a, {}^{j}\mu) \doteq {}^{j}\bar{y}(^{j}a) + {}^{j}\mu\,\Delta\,{}^{j}\bar{y} = {}^{j}z(^{j}a, {}^{j+1}\hat{a}, {}^{j}\mu)\,. \qquad \square$$

The prediction of the matrix $H(^{j+1}a)$ resp. $H^{\mathrm{T}}(^{j+1}a)\,H(^{j+1}a)$ is replaced with the prediction of the function $^{j+1}\bar{y}$. Therefore the functionals $\tilde{y}^{\mathrm{T}}\,{}^{j+1}\bar{y}$ and $^{j+1}\bar{y}^{\mathrm{T}}\,.\,{}^{j+1}\bar{y} = \|^{j+1}\bar{y}\|^2$ can be assessed with respect to the etalon ^{j}z according to the relations

$$\tilde{y}^{\mathrm{T}}\,{}^{j+1}\bar{y} = \tilde{y}^{\mathrm{T}}\,{}^{j}z + v_1(^{j}\mu) \tag{4.18}$$

and

$$\|^{j+1}\bar{y}\|^2 = \|^{j}z\|^2 + v_2(^{j}\mu)\,, \tag{4.19}$$

where the errors $v_1({}^j\mu)$ and $v_2({}^j\mu)$ can be governed by the damping factor according to two demands:

(i) The solution must be sufficiently near to the linear etalon to be able to predict the characteristic parameters in $(j+1)$th step without a laborious experimentation.

(ii) The factor ${}^j\mu$ must be selected as large as possible to get a rapid convergence.

Now let us introduce the following functionals

$$ {}^j\varphi^2 = \|{}^j\bar{\boldsymbol{y}} - {}^j\boldsymbol{z}\|^2 = \|{}^j\bar{\boldsymbol{y}} - {}^j\tilde{\boldsymbol{y}}\|^2 = \|\Delta\, {}^j\bar{\boldsymbol{y}}\|^2 \tag{4.20} $$

and

$$ {}^j\varrho = {}^j\bar{\boldsymbol{y}}^{\mathrm{T}}\, \Delta\, {}^j\bar{\boldsymbol{y}} = \tilde{\boldsymbol{y}}^{\mathrm{T}}\, {}^j\bar{\boldsymbol{y}} - \|{}^j\bar{\boldsymbol{y}}\|^2 . \tag{4.21} $$

With respect to these functionals we can write

$$ \|{}^j\boldsymbol{z}\|^2 = \|{}^j\bar{\boldsymbol{y}}\|^2 + 2\,{}^j\mu\,{}^j\varrho + {}^j\mu^2\,{}^j\varphi^2 \tag{4.22} $$

resp.

$$ \|{}^j\bar{\boldsymbol{y}}\|^2 = \|{}^j\tilde{\boldsymbol{y}}\|^2 - 2\,{}^j\mu\,{}^j\varrho + {}^j\mu^2\,{}^j\varphi^2 . \tag{4.23} $$

Further let us introduce the error of the linear solution

$${}^j\delta^2 = \|\tilde{\boldsymbol{y}} - {}^j\boldsymbol{z}\|^2 = \|\tilde{\boldsymbol{y}}\|^2 - 2\tilde{\boldsymbol{y}}^{\mathrm{T}}\,{}^j\boldsymbol{z} + \|{}^j\boldsymbol{z}\|^2$$

and with respect to relation (4.3) in the form

$$ {}^j\hat{\boldsymbol{a}}^{\mathrm{T}}\, \boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}_L({}^j\boldsymbol{a})\, {}^j\hat{\boldsymbol{a}} - {}^j\hat{\boldsymbol{a}}^{\mathrm{T}}\, \boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \tilde{\boldsymbol{y}} = \|{}^j\boldsymbol{z}\|^2 - {}^j\boldsymbol{z}^{\mathrm{T}}\tilde{\boldsymbol{y}} = 0 \tag{4.24} $$

we obtain

$$ {}^j\delta^2 = \|\tilde{\boldsymbol{y}}\|^2 - \|{}^j\boldsymbol{z}\|^2 . \tag{4.25} $$

Similarly we can derive from relation (4.3)

$$ {}^j\boldsymbol{a}^{\mathrm{T}}\, \boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \boldsymbol{H}_L({}^j\boldsymbol{a})\, {}^j\hat{\boldsymbol{a}} - {}^j\boldsymbol{a}^{\mathrm{T}}\, \boldsymbol{H}_L^{\mathrm{T}}({}^j\boldsymbol{a})\, \tilde{\boldsymbol{y}} = {}^j\bar{\boldsymbol{y}}^{\mathrm{T}}\,{}^j\boldsymbol{z} - {}^j\bar{\boldsymbol{y}}^{\mathrm{T}}\tilde{\boldsymbol{y}} = 0 . \tag{4.26} $$

The connection of the linear and nonlinear part of the DNLS estimator opens the way for deriving the very rich theory (cf. [10], [12]). Here we shall present only some relations which are necessary for deriving the next propositions.

Only the parameters of the jth step are used in the following proposition and so the damping factor ${}^j\mu = 1$ is considered.

Proposition 4.5. If $\{{}^j\boldsymbol{a}\} \in \Omega_a$ then for the functions $\tilde{\boldsymbol{y}} \in V$, ${}^j\boldsymbol{z} \in \overline{V}(\overline{V} \subset V)$, ${}^j\bar{\boldsymbol{y}} \in \overline{V}(\overline{V} \subset V)$ the following relations hold:

$$ \|\tilde{\boldsymbol{y}} - {}^j\bar{\boldsymbol{y}}\|^2 = {}^j\delta^2 + {}^j\varphi^2 , \tag{4.27} $$

$$ \tilde{\boldsymbol{y}}^{\mathrm{T}}\, \Delta\, {}^j\bar{\boldsymbol{y}} = \tilde{\boldsymbol{y}}^{\mathrm{T}}({}^j\bar{\boldsymbol{y}} - {}^j\boldsymbol{z}) = {}^j\varrho + {}^j\varphi^2 , \tag{4.28} $$

$$ \|\tilde{\boldsymbol{y}}\|^2 - \tilde{\boldsymbol{y}}^{\mathrm{T}}\,{}^j\bar{\boldsymbol{y}} = {}^j\delta^2 + {}^j\varphi^2 + {}^j\varrho . \tag{4.29} $$

Proof. Relation (4.27) can be arranged

$$ \|\tilde{\boldsymbol{y}} - {}^j\boldsymbol{z} + {}^j\boldsymbol{z} - {}^j\bar{\boldsymbol{y}}\|^2 = \|\tilde{\boldsymbol{y}} - {}^j\boldsymbol{z}\|^2 + \|{}^j\boldsymbol{z} - {}^j\bar{\boldsymbol{y}}\|^2 + $$
$$ + 2(\tilde{\boldsymbol{y}} - {}^j\boldsymbol{z})^{\mathrm{T}}\,({}^j\boldsymbol{z} - {}^j\bar{\boldsymbol{y}}) $$

and with respect to (4.20), (4.24), (4.25) and (4.26) we get (4.27). Similarly we obtain

for (4.28) resp. (4.29) with respect to (4.22), (4.24) and (4.26)

$$\tilde{\mathbf{y}}^{\mathrm{T}}({}^{j}\mathbf{z} - {}^{j}\bar{\mathbf{y}}) = \|{}^{j}\mathbf{z}\|^2 - {}^{j}\bar{\mathbf{y}}^{\mathrm{T}}\,{}^{j}\mathbf{z} = \|{}^{j}\mathbf{z}\|^2 - \|{}^{j}\bar{\mathbf{y}}\|^2 - {}^{j}\bar{\mathbf{y}}^{\mathrm{T}}\,\Delta\,{}^{j}\bar{\mathbf{y}} = \\ = \|\Delta\,{}^{j}\bar{\mathbf{y}}\|^2 + {}^{j}\bar{\mathbf{y}}^{\mathrm{T}}\,\Delta\,{}^{j}\bar{\mathbf{y}}$$

resp. with respect to (4.21), (4.26) and (4.27)

$$\|\tilde{\mathbf{y}}\|^2 - \tilde{\mathbf{y}}^{\mathrm{T}}\,{}^{j}\bar{\mathbf{y}} = \|\tilde{\mathbf{y}} - {}^{j}\bar{\mathbf{y}}\|^2 - \|{}^{j}\bar{\mathbf{y}}\|^2 + \tilde{\mathbf{y}}^{\mathrm{T}}\,{}^{j}\bar{\mathbf{y}} = {}^{j}\delta^2 + {}^{j}\varphi^2 + {}^{j}\bar{\mathbf{y}}\,\Delta\,{}^{j}\bar{\mathbf{y}} = \\ = {}^{j}\delta^2 + {}^{j}\varphi^2 + {}^{j}\varrho\,. \qquad \square$$

The main result is given by relation (4.27). The separation of the error ${}^{j}\delta$ pertinent to the linear part from the error ${}^{j}\varphi$ pertinent to the nonlinear part of the DNLS estimator is the main novelty against all other nonlinear regression methods. This separation allows deriving the simple condition for the existence of the global minimum (see Paragraph 4.5). Further we can explain why the EE estimators are linear whereas the OE estimators are nonlinear and so to analyze the failure of the EE estimators in many practical tasks (see Section 5).

Finally let us show how the important functional ${}^{j+1}\varrho$ and ${}^{j+1}\varphi$ can be predicted from the parameters of the jth step.

Proposition 4.6. If $\{{}^{j}\mathbf{a}\} \in \Omega_a$ and $\tilde{\mathbf{y}} \in V$, ${}^{j}\bar{\mathbf{y}} \in \overline{V}(\overline{V} \subset V)$, ${}^{j}\mathbf{z} \in \overline{\overline{V}}$ $(\overline{\overline{V}} \subset V)$ then the following relations are valid

$$(4.30) \qquad {}^{j+1}\varrho = {}^{j+1}\varrho_P + v_1({}^{j}\mu) - v_2({}^{j}\mu)\,,$$

$$(4.31) \qquad {}^{j+1}\varphi^2 = {}^{j+1}\varphi_P^2 + v_3({}^{j}\mu) + {}^{j}\delta^2 - {}^{j+1}\delta^2\,,$$

where ${}^{j+1}\varrho_P = (1 - {}^{j}\mu)({}^{j}\varrho + {}^{j}\mu\,{}^{j}\varphi^2)$, ${}^{j+1}\varphi_P^2 = (1 - {}^{j}\mu)^2\,{}^{j}\varphi^2$ and $v_3({}^{j}\mu) = v_2({}^{j}\mu) - 2v_1({}^{j}\mu)$.

Proof. Relation (4.30) resp. (4.31) can be derived from relation (4.21) with regard to (4.18), (4.19), (4.22), (4.28)

$${}^{j+1}\varrho = \tilde{\mathbf{y}}^{\mathrm{T}}\,{}^{j+1}\bar{\mathbf{y}} - \|{}^{j+1}\bar{\mathbf{y}}\|^2 = \tilde{\mathbf{y}}^{\mathrm{T}}\,{}^{j}\mathbf{z} + v_1({}^{j}\mu) - \|{}^{j}\mathbf{z}\|^2 - v_2({}^{j}\mu) = \\ = \tilde{\mathbf{y}}^{\mathrm{T}}\,{}^{j}\bar{\mathbf{y}} + {}^{j}\mu\tilde{\mathbf{y}}^{\mathrm{T}}\,\Delta\,{}^{j}\bar{\mathbf{y}} - \|{}^{j}\bar{\mathbf{y}}\|^2 - 2\,{}^{j}\mu\,{}^{j}\varrho - {}^{j}\mu^2\,{}^{j}\varphi^2 + v_1({}^{j}\mu) - \\ - v_2({}^{j}\mu) = (1 - {}^{j}\mu)({}^{j}\varrho + {}^{j}\mu\,{}^{j}\varphi^2) + v_1({}^{j}\mu) - v_2({}^{j}\mu)$$

resp. from relation (4.29) with regard to (4.18), (4.28), (4.30)

$${}^{j+1}\varphi^2 = \|\tilde{\mathbf{y}}\|^2 - \tilde{\mathbf{y}}^{\mathrm{T}}\,{}^{j+1}\bar{\mathbf{y}} - {}^{j+1}\delta^2 - {}^{j+1}\varrho = \|\tilde{\mathbf{y}}\|^2 - \tilde{\mathbf{y}}^{\mathrm{T}}\,{}^{j}\bar{\mathbf{y}} - \\ - {}^{j}\mu({}^{j}\varrho + {}^{j}\varphi^2) - v_1({}^{j}\mu) - {}^{j+1}\delta^2 - {}^{j+1}\varrho_P - v_1({}^{j}\mu) + \\ + v_2({}^{j}\mu) = (1 - {}^{j}\mu)({}^{j}\varrho + {}^{j}\varphi^2) - (1 - {}^{j}\mu)({}^{j}\varrho + {}^{j}\mu\,{}^{j}\varphi^2) + v_3({}^{j}\mu) + {}^{j}\delta^2 - \\ - {}^{j+1}\delta^2\,. \qquad \square$$

Relation (4.31) can be written in other forms

$$(4.32) \qquad {}^{j+1}\varphi^2 + {}^{j+1}\delta^2 = {}^{j+1}\eta^2 = {}^{j}\delta^2 + {}^{j+1}\varphi_P^2 + v_3({}^{j}\mu)$$

resp. with respect to relation (4.30)

$$^{j+1}\eta^2 = {}^{j}\delta^2 + {}^{j+1}\varphi_P^2 + {}^{j+1}\varrho_P - {}^{j+1}\varrho - v_1({}^{j}\mu) = \tag{4.33}$$
$$= {}^{j}\delta^2 + (1 - {}^{j}\mu)({}^{j}\varrho + {}^{j}\varphi^2) - {}^{j+1}\varrho - v_1({}^{j}\mu) .$$

Relations (4.30) to (4.33) are formally equivalent with the relations in the literature (cf. Paragraph 4.2 in [12]). We have reached stage by stage to the relations which allow using the procedures resp. the algorithm from the paper [12] where pertinent details connected with deriving the optimal damping factor can be found.

4.5. The convergence problem

The given propositions simplify the convergence analysis e.g. the simple condition for the existence of the global minimum can be derived. First let us introduce the notation: the asterisk on the left indicates the coefficients and functionals pertinent to the global minimum and $(\cdot)({}^{j}\mu = 1) = (\cdot)(1)$, e.g. $^{j+1}\eta^2({}^{j}\mu = 1) = {}^{j+1}\eta^2(1)$.

Proposition 4.7. If

(i) $\{{}^{j}\boldsymbol{a}\} \in \Omega_a$,

(ii) ${}^{j}\varphi^2 > v_3(1) > 0$ for $\forall_j \geqq 0$,

then the sequence $\{{}^{j}\boldsymbol{a}\}$ converges as $j \to \infty$ to the global minimum $^*\boldsymbol{a}$ given by the etalon $^*z$, i.e. it holds $^*\eta = {}^*\delta$ for $^*\varphi = 0$.

Proof. By Hypothesis (i) the matrix $\boldsymbol{H}_L^{\mathrm{T}}({}^{j}\boldsymbol{a})\,\boldsymbol{H}_L({}^{j}\boldsymbol{a})$ is nonsingular with respect to Proposition 4.1 and thus the linear solution must exist in each iteration step. Let us suppose that

$$\tilde{\boldsymbol{y}} = \boldsymbol{H}_L({}^{j}\boldsymbol{a})\,{}^{j+1}\hat{\boldsymbol{a}} . \tag{4.34}$$

Then we can write the relation

$$Q_L({}^{j}\boldsymbol{a}) = \|\tilde{\boldsymbol{y}} - \boldsymbol{H}_L({}^{j}\boldsymbol{a})\,{}^{j}\boldsymbol{a}\|^2 =$$
$$= \|\tilde{\boldsymbol{y}}\|^2 - \tilde{\boldsymbol{y}}^{\mathrm{T}}\boldsymbol{H}_L({}^{j}\boldsymbol{a})\,{}^{j}\boldsymbol{a} - {}^{j}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{H}_L^{\mathrm{T}}({}^{j}\boldsymbol{a})\,\tilde{\boldsymbol{y}} + {}^{j}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{H}_L^{\mathrm{T}}({}^{j}\boldsymbol{a})\,\boldsymbol{H}_L({}^{j}\boldsymbol{a})\,{}^{j}\boldsymbol{a}$$

in the form

$$Q_L({}^{j}\boldsymbol{a}) = \|\tilde{\boldsymbol{y}}\|^2 + \|\boldsymbol{H}_L({}^{j}\boldsymbol{a})({}^{j}\boldsymbol{a} - {}^{j+1}\hat{\boldsymbol{a}}\|^2 - \|\boldsymbol{H}_L({}^{j}\boldsymbol{a})\,{}^{j+1}\hat{\boldsymbol{a}}\|^2 . \tag{4.35}$$

Since each term in (4.35) is positive it must hold

$$Q_L({}^{j}\boldsymbol{a}) = \|\tilde{\boldsymbol{y}} - \boldsymbol{H}_L({}^{j}\boldsymbol{a})\,{}^{j}\boldsymbol{a}\|^2 \geqq \|\tilde{\boldsymbol{y}}\|^2 - \|\boldsymbol{H}_L({}^{j}\boldsymbol{a})\,{}^{j+1}\hat{\boldsymbol{a}}\|^2 . \tag{4.36}$$

Relation (4.34) is another form of relation (4.4) for the jth iterations step and so relation (4.36) proves that (4.4) yields an absolute minimum for the linear case in each iteration step.

With respect to Proposition 4.6 the discussion of the global minimum will be limited on the case with ${}^{j}\mu = 1$. Therefore we shall start from relation (4.31) in the form

$$^{j+1}\varphi^2(1) = v_3(1) + {}^{j}\delta^2 - {}^{j+1}\delta^2(1) \tag{4.37}$$

and with regard to Hypothesis (ii) we obtain

$$^{j+1}\eta^2(1) = v_3(1) + {}^j\delta^2 < {}^j\eta^2 = {}^j\varphi^2 + {}^j\delta^2 . \tag{4.38}$$

If we suppose $^j\delta = {}^*\delta$ then it must hold according to (4.38) and Hypothesis (ii) $^{j+1}\eta^2(1) \geqq {}^*\delta^2$. The sign of equality can hold only for $v_1(1) = 0$ and $v_2(1) = 0$, i.e. $v_3(1) = 0$, and we obtain according to relations (4.18), (4.19) the equality $^*\bar{\boldsymbol{y}} = {}^*\boldsymbol{z}$ and so with respect to (4.36) $^*\eta = {}^*\delta$. □

The simpler proof for the existence of the global minimum was used in the paper [10]. The end of the iteration process is given with respect to relations (4.11) and (4.12) by $\Delta\, {}^*\boldsymbol{a}_L = 0$, i.e. $\|\Delta\, {}^*\bar{\boldsymbol{y}}\| = {}^*\varphi = 0$ (cf. relations (4.13) and (4.14)), and the global minimum is then given by relation (4.36) for $^*\boldsymbol{a}$. The local minima cannot exist. It follows from the fact that the spaces L_p and l_p for $1 < p < \infty$ are strict convex.

Let us emphasize in this connection that none of the procedures so far described in the literature is really reliable as a general purpose procedure for solving the system identification from an arbitrary initial vector of coefficients. Therefore the key problem of the system identification is to find an initial vector which is sufficiently near to the global minimum and then to obtain the stable sequence $\{{}^j\boldsymbol{a}\} \in \Omega_a$ (cf. [12]). So at least three groups of problems must be distinguished.

If the initial vector of coefficients

$$^0\boldsymbol{a} \in \Phi_0 = \{{}^j\boldsymbol{a}: {}^j\varphi < {}^j\delta, {}^j\delta \doteq {}^*\delta\} , \tag{4.39}$$

i.e. the initial vector $^0\boldsymbol{a}$ is near the global minimum, then all known methods are practically equivalent (cf. [12]). The convergence of the sequence $\{{}^j\boldsymbol{a}\}$ is governed in this region by the linear case. Therefore the linear EE estimators can generate acceptable results in this region (see Section 5). But this region is small and so it is practically impossible to get $^0\boldsymbol{a} \in \Phi_0$.

The second group of problems is given by the cases in which the sufficiently good linear etalon exists. The boundary of this group is approximately given by the condition $^1\varphi^2 > {}^1\delta^2 - {}^*\delta^2$ resp. for $^1\delta \gg {}^*\delta$ by the simplified condition $^1\varphi^2 > {}^1\delta^2$ (cf. Table 3 in [12]). The effective use of the DNLS estimator is connected above all with this second group.

Problems, which belong to the third group, must start from a misleading etalon, i.e. from parameters $^1\varphi^2 < {}^j\delta^2 - {}^*\delta^2$. Here the methods with a broader convergence domain, i.e. the gradient resp. compromise-gradient techniques, should be used. Unfortunately all these methods have the lower convergence rate and so they are not suited for the solution of control problems (cf. [12], [14]). The theory of the DNLS estimator together with the apparatus of the control theory open mostly simpler and rapider ways for the solution of such cases (cf. [12]).

5. The relation to the EE estimator

The greater part of identification procedures described in the literature belongs to the linear Equation Error (EE) methods (cf. [5], [13]). So the relation of the DNLS estimator, which belongs to the Output Error (OE) methods, should be analyzed to find its position in theory and practical tasks. The main difference between the EE and OE methods lies in the fact that the EE procedures are linear in the unknown parameters, while the OE estimators are nonlinear (see [13]). So the EE approach looks simpler as it follows from the pertinent difference equation

$$\tilde{y}(k) = \boldsymbol{\theta}^{\mathrm{T}}\, \boldsymbol{r}(k) + w(k)\,, \tag{5.1}$$

where $\boldsymbol{\theta} = [\bar{a}_1, \ldots, \bar{a}_n, \bar{b}_1, \ldots, \bar{b}_m]^{\mathrm{T}}$ and

$$\boldsymbol{r}(k) = [-\tilde{y}(k-1)\,, \ldots, -\tilde{y}(k-n), u(k-1), \ldots, u(k-m)]^{\mathrm{T}}\,.$$

The parameter vector $\boldsymbol{\theta}$ is to be estimated from the measured points $\tilde{y}(k)$ and $\boldsymbol{r}(k)$. The "equation error" $w(k)$ is then given by

$$Q_E(\boldsymbol{\theta}) = \frac{1}{q} \sum_{k=1}^{q} [\tilde{y}(k) - \boldsymbol{\theta}^{\mathrm{T}}\, \boldsymbol{r}(k)]^2\,. \tag{5.2}$$

The pertinent solution has the form

$$\boldsymbol{\theta}(q) = [\sum_{k=1}^{q} \boldsymbol{r}(k)\, \boldsymbol{r}^{\mathrm{T}}(k)]^{-1} \sum_{k=1}^{q} \boldsymbol{r}(k)\, \tilde{y}(k) \tag{5.3}$$

provided the inverse exists.

The noise in (5.1) can induce an asymptotic estimation bias (see [5], [13]) and so it is necessary to introduce an additional complexity into the solution to obviate this difficulty. E.g. the Instrumental Variable (IV) method (see [5],]13]) is designed to overcome this convergence problem. The instrumental variable $\boldsymbol{\varrho}(k)$, which should be uncorrelated with $w(k)$, is introduced and we obtain

$$\boldsymbol{\theta}(q) = [\sum_{k=1}^{q} \boldsymbol{\varrho}(k)\, \boldsymbol{r}^{\mathrm{T}}(k)]^{-1} \sum_{k=1}^{q} \boldsymbol{\varrho}(k)\, \tilde{y}(k)\,. \tag{5.4}$$

A common choice of the instrumental variable is

$$\boldsymbol{\varrho}(k) = [-y_M(k-1), \ldots, -y_M(k-n), u(k-1), \ldots, u(k-m)]^{\mathrm{T}}\,, \tag{5.5}$$

where $y_M(k)$ is the output of a deterministic model driven by the actual input $u(k)$. This is the non-symmetric IV method where the instrumental variable can be derived by the recursive-iterative instrumental variable algorithm (see [13] pp. 134). This version is suited for the comparison of EE methods with the DNSL estimator.

First the EE methods cannot be internally robust. The matrices in relations (5.3) and (5.4) are given by measured data and so its inversion may be an ill-posed problem especially for a small sample size.

The second disadvantage, i.e. the pseudolinearity of difference equation (5.1), is the reason why the recursive EE methods are used for a relatively large sample

size, i.e. for $q > 100$ (cf. examples in [5], [13]). This confirms the use of iterative steps by the IV methods (cf. [13] pp. 134): The recursive-iterative solution of IV equations yields with respect to its off-line nature better results than the equivalent on-line version in the sense that, for a given finite set of sampled data, the recursive-iterative estimates have lower estimation error variance. Of course, this advantage becomes less as the data base is increased in size, simply because the additional improvement obtained via iteration itself becomes less.

Let us discuss this problem from another viewpoint. All purely theoretical analysis of the estimators is based largely on the asymptotic behaviour (see [5], [13]). So only the cases with a large sample size are covered with these theoretical results. On the other hand the small sample theory has not been developed (cf. [13], pp. 214) and here the theory of the DNLS estimator can be useful. Let us outline some points which can contribute to the solution of the problems connected with a small sample size:

(i) The DNLS estimator is internally robust (cf. Section 3) and so it can yield estimates under fairly mild, non-restrictive conditions on the nature and the size of the measured data. This is an important advantage with respect to the uncertainty band as it was discussed in the introduction.

(ii) The close connection with the linear Gauss estimator (cf. relations (4.4), (4.12) and (4.14)) allows to distinguish the region in which the identification problem is strongly nonlinear. Here the DNLS estimator can generate the solution without increasing the sample size. This may be useful by an adaptive control when the input changes rapid, e.g. by an identification from a step response (see [1], [9]).

(iii) The DNLS estimator is well suited for deriving the simplified models of real systems (cf. [11], [12]). The pertinent theory constitutes a basis for including an approximation error, i.e. a multiplicative systematic error, in identification procedures (cf. [11]).

6. EXAMPLES

To demonstrate the internal robustness of the DNLS estimator the identification of the system with the transfer function

$$F_1(s) = \frac{1}{5 + 16s + 79s^2 + 120s^3}$$

was considered. All experiments with $q = 16$ and $\Delta\tau = 2$ were realized on the digital computer. The error of measured data was simulated by a roundoff noise, i.e. points of the unit step response $y(k)$ $(k = 1, 2, \ldots, 16)$ were given on one decimal place. The sampling instants follow from the relation $t_k = 2k - 1$. This approach has one important advantage over the random noise: the experiments can be simply repeated and verified. It must be assumed in identification problems that little will be known of the model parameter values. So starting function ${}^0\bar{y}$ given by the transfer

function

$$^{0}\bar{F}_1(s) = \frac{1}{5 + 12s + 40s^2 + 40s^3},$$

which is far from the global minimum, was used. The considered example shows that the DNLS estimator can solve problems which start from signals corrupted with a large uncertainty band and given with a small sample size.

Table 1.

j	$^{j}\mu_{opt}$ $^{j}\varphi^2$	$^{j}\delta^2$ $^{j}\eta^2$	$v_1(1)$ $v_3(1)$	$\Delta\,^{j}a_2$ $\Delta\,^{j}a_3$
1	0·82 $6{\cdot}8 \cdot 10^{-2}$	$1{\cdot}8 \cdot 10^{-2}$ $8{\cdot}6 \cdot 10^{-2}$	$-3{\cdot}8 \cdot 10^{-2}$ $2{\cdot}3 \cdot 10^{-2}$	18 21·6
2	0·66 $2{\cdot}9 \cdot 10^{-2}$	$1{\cdot}1 \cdot 10^{-2}$ $4{\cdot}0 \cdot 10^{-2}$	$-8{\cdot}9 \cdot 10^{-3}$ $2{\cdot}2 \cdot 10^{-2}$	52·4 −23·2
3	0·72 $5{\cdot}2 \cdot 10^{-3}$	$1{\cdot}45 \cdot 10^{-2}$ $1{\cdot}98 \cdot 10^{-2}$	$9{\cdot}0 \cdot 10^{-3}$ $1{\cdot}6 \cdot 10^{-3}$	− 7·6 131
4	0·93 $9{\cdot}5 \cdot 10^{-4}$	$1{\cdot}36 \cdot 10^{-2}$ $1{\cdot}45 \cdot 10^{-2}$	$-1{\cdot}3 \cdot 10^{-3}$ $2{\cdot}0 \cdot 10^{-5}$	− 9·2 −18·3
5	0·84 $8{\cdot}7 \cdot 10^{-5}$	$1{\cdot}35 \cdot 10^{-2}$ $1{\cdot}35 \cdot 10^{-2}$	$-2{\cdot}5 \cdot 10^{-5}$ $1{\cdot}6 \cdot 10^{-5}$	−0·04 14·6
6	1 $3{\cdot}4 \cdot 10^{-6}$	$1{\cdot}35 \cdot 10^{-2}$ $1{\cdot}35 \cdot 10^{-2}$	$-4{\cdot}1 \cdot 10^{-6}$ $1{\cdot}1 \cdot 10^{-7}$	− 0·4 − 1·0

Table 1 contains important parameters computed with a double precision PL 1 program based upon the algorithm from the literature [12]. The first two steps illustrate the influence of the nonlinearity of the pertinent regression functions (cf. parameters $^{j}\varphi$ and $^{j}\delta$ in Table 1). On the other hand the solution for $j > 3$, i.e. for the conditions $^{j}\delta \doteq {}^{*}\delta$ and $^{j}\varphi < {}^{j}\delta$, is governed by the linear case (cf. $v_1(1)$ and $v_3(1)$ in Table 1). The convergence rate of the coefficients $^{j}a_2$ and $^{j}a_3$ to the resulting values given by the transfer function

$$^{6}\bar{F}_1(s) = \frac{1}{4{\cdot}7 + 16{\cdot}7s + 75s^2 + 131s^3}$$

follows from $\Delta\,^{j}a_2$ and $\Delta\,^{j}a_3$ in Table 1.

The signals $^{j}\bar{y}$ and pertinent sensitivity functions $^{j}v^{(i)}$ $(i = 0, 1. \ldots, \bar{n})$ were computed with the procedure based on scaling and squaring (see [6]). While the double precision can be used in the main computation line (cf. Section 3) the obtained results are sufficiently correct with respect to the low accuracy of measured data. This relatively complicated way was used to secure the numerical robustness (cf. Paragraph 2.1) and to make possible deriving simplified models.

Let us show the importance of this last problem, i.e. deriving simplified models, and consider the identification of the system with the transfer function

$$F_2(s) = \frac{1}{5 + 16s + 77s^2 + 134s^3 + 72s^4}$$

by the same conditions ($q = 16$, $\Delta\tau = 2$) as in the previous example. The reader can test how difficult is deriving the reliable result especially if the starting point is far from the global minimum. On the other hand we can simply derive the approximate model given by the transfer function

$${}^6\bar{F}_2(s) = \frac{1}{4{\cdot}8 + 18{\cdot}4s + 67{\cdot}8s^2 + 175s^3}$$

which is more reliable as the model for $\bar{n} = n$ (see [11]).

Let us add that the applicability of derived models must be tested with respect to the external robustness, i.e. according to the conditions connected with the empirical identifiability (see [7], [8], [9]). This problem is beyond the scope of this paper.

7. CONCLUSION

The DNLS estimator is an iterative procedure with maximum utility in those situations where the structure of the model is known or simplified models can be used and the parameters characterizing the model are time-invariant over the observation interval. This interval can be very short (cf. example in Table 1). The DNLS estimator is internally robust (see Section 3), i.e. it can start from signals corrupted with a broad uncertainty band. It is constructed as the method for deriving the interpolation function with the known structure from given points and the inaccuracy of these points is not implicated in the numerical procedure (cf. relations (4.11), (4.12) and Propositions 4.1, 4.3, 4.7). Its close connection to the approximation method from [12] opens the way for the use of approximate models in control theory and practice (see [11]).

(Received June 2, 1987.)

REFERENCES

[1] K. J. Aström: Theory and applications of adaptive control — a survey. Automatica *19* (1983) 471—486.

[2] M. Athans: Advances and open problems on the control of large scale systems. In: Proc. of VII IFAC Congress (Helsinki 1978), Pergamon Press, Oxford 1978, 2371—2382.

[3] T. L. Fine: Theories of Probability. Academic Press, New York 1973.

[4] G. C. Goodwin: Some observations on robust estimation and control. In: Proc. of 7th IFAC/IFORS Symp. on Identification and Parameter Estimation, York 1985, Pergamon Press, Oxford 1985.

[5] L. Ljung and T. Söderström: Theory and Practice of Recursive Identification. MIT Press, Massachusetts 1983.

[6] C. Moler and Ch. van Loan: Nineteen dubious ways to compute the exponential of a matrix. SIAM Rev. *20* (1978), 801—836.
[7] J. Štěpán: Some problems of system identification. Kybernetika *7* (1971), 2, 133—155.
[8] J. Štěpán: Messgenauigkeit und Regelungstheorie. Regelungstechnik und Prozess-Datenverarbeitung *20* (1972), 203—207.
[9] J. Štěpán: Control of regulator-configurations in large scale systems. In: Fourth Formator Symposium, Academia, Prague 1983.
[10] J. Štěpán: Nonlinear approximation in control problems. Kybernetika *19* (1983), 6, 419—504.
[11] J. Štěpán: Approximation versus identification. In: Proc. of 7th IFAC/IFORS Symp. on Identification and Parameter Estimation, York 1985, Pergamon Press, Oxford 1985.
[12] J. Štěpán: A new method for the nonlinear approximation of signals. Part I: The optimal damping factor. Kybernetika *22* (1986), 5, 425—438. Part II: The convergence problem. Kybernetika *22* (1986), 6, 503—513.
[13] P. Young: Recursive Estimation and Time-Series Analysis. Springer-Verlag, Berlin 1984.
[14] M. A. Wolfe: Numerical Methods for Unconstrained Optimization. Van Nostrand Reinhold, New York 1978.

Ing. Jaromír Štěpán, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.