Kybernetika

Jaromír Štěpán
A new method for the nonlinear approximation of signals. II. The convergence problem

A NEW METHOD FOR THE NONLINEAR APPROXIMATION OF SIGNALS

Part II: The convergence problem

JAROMÍR ŠTĚPÁN

The second part deals with the convergence of the Damped Nonlinear Least Squares (DNLS) method which was derived in the first part of this paper. The pertinent algorithm is given and the effectivity of the proposed method is illustrated with the help of some examples.

6. THE CONVERGENCE OF THE DNLS METHOD

Two problems must be distinguished by the convergence analysis. The first problem is connected with existence of a convergent sequence by a given starting point, i.e. with the convergence domain. The convergence rate is the second problem. Let us emphasize that the character of control problems, e.g. the character of an adaptive control, shows the priority of the convergence rate.

6.1. The convergence domain

The optimization procedures presented in the preceding paragraphs must be completed with the proposition which gives a sufficient condition to ensure the existence of a convergent sequence $\{^j a\}$. The asterisk on the left indicates the coefficients and functionals pertinent to the global minimum.

Proposition 6.1. If

(i) the set $L^{'}(\hat{a})$ defined by

$$L^{'}(\hat{a}) = \{ a \in \Omega_{\bar{n}} | \eta^2(a) < \eta^2(\hat{a}) \} \text{ is compact for each } \hat{a} \in \Omega_{\bar{n}} (\Omega_{\bar{n}} \subset \Omega_a),$$

(ii) $\{^j a\}$ is the DNLS sequence,

then $\{^j a\}$ converges to the global minimum $^* a$ as $j \to \infty$.

Proof. By hypothesis (ii) $^j a$ are stable, i.e. $^j a \in \Omega_a$ for $\forall j \geq 0$ and so the linear approximation according to relation (4.3) must exist in each iteration step. Therefore the nonsingular matrices $G(^j a)$ for $\forall j \geq 0$ must always exist. The sequence

of linear errors $^j\delta$ converges by hypothesis (ii) (cf. hypothesis (iii) in Definition 4.2) more rapid to the global minimum than the sequence of errors $^j\varphi$.

By hypothesis (i) and the fact that by hypothesis (ii)

$$^{j+1}\eta^2(^j\mu) < {^j\eta^2} \quad (\forall j \geqq 0)$$

$\{^j a\}$ lies in a compact set. Let us assume that this compact set is $L(^0 a)$ where $^0 a$ is an initial vector of coefficients. Now by hypothesis (i) $L(^0 a)$ is compact and $\{^j a\}$ has then a subsequence which converges to a point in $L(^0 a)$. Therefore $\{^j a\}$ has at least one limit point in $L(^0 a)$. We have proved in [9] (see Proposition 4.4) that there is exactly one limit point $^* a$ of $\{^j a\}$ given by the linear case, i.e. for $^*\eta = {^*\delta}$. Therefore $\{^j a\}$ converges to $^* a$. $\qquad\qquad\qquad\square$

Let us emphasize in connection with Proposition 6.1 that none of the procedures so far described in the literature is really reliable as a general purpose procedure for solving the signal approximation resp. the system identification (cf. [5], [12]). If an initial vector $^0 a$ lies near the stability boundary then the DNLS sequence cannot exist. Here there is only one possibility — to use some gradient technique or a compromise-gradient technique, e.g. the Marquardt or Fletcher method, and so to get $^1 a \in \Omega_a$. The basic gradient technique is notoriously slow [11] and it may be to 10^3 times slower than the Gauss-Newton method [5]. In this case increasing of the convergence domain results in decreasing the convergence rate. The point to be made here is that the DNLS method is rapid and so suited for control problems, but unstable with respect to convergence, whereas gradient techniques or compromise-gradient techniques, e. g. the Marquardt method, exhibit the opposite characteristics, good convergence and longer computation time.

6.2. The convergence rate

The convergence rate of the DNLS method is given by Proposition 4.3 for the cases which are sufficiently near to the ideal case $(^{j+1}\delta \doteq {^j\delta})$. We can write for $^j\mu_{opt} < 0.95$

$$(6.1) \qquad\qquad \Lambda_r = \frac{^r\varphi^2}{^1\varphi^2} \approx \prod_{j=1}^{r-1} (1 - {^j\mu_{opt}})^2 \,.$$

Two facts are important. From Propositions 4.3 and 6.1 it follows that only errors $^j\varphi^2$ are decisive for the assessment of the convergence rate. Secondly only the damping factors $^j\mu_{opt} < 0.95$ can be used in relation (6.1) with respect to numerical errors and with respect to the fact that the linear error converges stepwise too. It follows from relation (4.28) for the condition $0 < {^{j+1}\varphi^2} < 2\,{^{j+1}\varphi_P^2}$

$$(6.2) \quad \left| v_3(^j\mu_{opt}) + {^j\delta^2} - {^{j+1}\delta^2}(^j\mu_{opt}) \right| < {^{j+1}\varphi_P^2}(^j\mu_{opt} > 0.95) < 2.5 \cdot 10^{-3}\,{^j\varphi^2} \,.$$

The convergence analysis of the procedure from paragraph 5.1 is more complicated.

504

The convergence coefficient

$$(6.3) \qquad {}^j\lambda = \frac{{}^j\eta^2 - {}^{j+1}\eta_P^2({}^j\mu_{opt})}{{}^j\varphi^2 - 2\,{}^{j+1}\varphi_P^2({}^j\mu_{opt})}$$

can be derived with regard to the predicted error given by (5.5) for the condition $R = {}^j\varphi^2$. With a few manipulations we obtain for ${}^jZ = R/{}^j\varphi^2$ in virtue of (5.4) and (5.5)

$$(6.4) \qquad {}^j\lambda = \frac{\left(1 + {}^jZ\psi(2 - \psi)\right)\left(1 + {}^jZ\right)}{\left(1 + {}^jZ\right)^2 - 2\,{}^jZ^2(1 - \psi)^2}.$$

The cases with ${}^jZ > 1$ are important for the use of the DNLS method far from the global minimum. It follows from the relations

$$(6.5) \qquad {}^j\lambda({}^jZ = 0,\ \forall\psi,\ {}^j\mu_{opt} = 1) = 1\,,$$

$$ {}^j\lambda({}^jZ = 1,\ \forall\psi,\ {}^j\mu_{opt} = (1 + \psi)/2) = 1\,.$$

The region ${}^jZ \in (0, 1)$ is pertinent to the region ${}^j\mu_{opt} \in ((1 + \psi)/2, 1)$, i.e. to the region which is for $\psi > 0{\cdot}4$ near the global minimum.

The following proposition will be useful.

Proposition 6.2. If

(i) $\{{}^ja\}$ is the DNLS sequence,
(ii) ${}^j\mu_{opt} \in (0{\cdot}293;\ 1)$ is given by relation (5.4),
(iii) ${}^jZ > 1$ and $\psi \in (-0{\cdot}414;\ 1)$,
then ${}^j\lambda > 1$.

P r o o f. Let us start from the condition

$$(6.6) \qquad {}^j\lambda - 1 = \frac{{}^jZ({}^jZ - 1)(\psi - 1)^2}{\left(1 + {}^jZ\right)^2 - 2\,{}^jZ^2(1 - \psi)^2} > 0\,.$$

The important root of the denominator is given by the relation $1 - 2(1 - {}^j\mu_{opt})^2 = 0$ and so the pertinent critical ratio is given by ${}^jZ_m = \sqrt{0{\cdot}5}/(1 - \sqrt{(0{\cdot}5)} - \psi)$. Now we can derive from ${}^j\lambda({}^jZ = 2, \psi = 0) - 1 = 2$ that ${}^j\lambda > 1$ for $\psi \in (0{\cdot}293, 1)$ and ${}^jZ \in (1, \infty)$ resp. for $\psi \in (-0{\cdot}414;\ 0{\cdot}293)$ and ${}^jZ \in (1, {}^jZ_m)$. $\square$

The convergence rate can be analyzed using the convergence coefficient given by relation (6.3)

$$(6.7) \qquad \frac{{}^{j+1}\eta_P^2({}^j\mu_{opt})}{{}^j\eta^2} = 1 - \left(1 - 2(1 - {}^j\mu_{opt})^2\right)\frac{{}^j\varphi^2}{{}^j\eta^2}\,{}^j\lambda$$

and we obtain for ${}^{j+1}\eta^2({}^j\mu_{opt}) \doteq {}^{j+1}\eta_P^2({}^j\mu_{opt})$ and ${}^j\varphi^2\,{}^j\lambda/{}^j\eta^2 > 1$

$$(6.8) \qquad \frac{{}^{j+1}\eta^2({}^j\mu_{opt})}{{}^j\eta^2} < 2(1 - {}^j\mu_{opt})^2\,.$$

We can conclude: the convergence rate for $^j\mu_{opt} < 0.5$ is small, i.e. $^{j+1}\eta^2(^j\mu_{opt})/^j\eta^2 >$ > 0.5, and the cases with $^j\mu_{opt} < 0.3$ can be used only for deriving the better initial vector $^j a$. So these cases create the transient region to the gradient-compromise techniques.

7. THE ALGORITHM AND EXAMPLES

In this section we shall sketch one simple alternative of the DNLS algorithm and in the second part we shall illustrate the use of this algorithm with help of examples and confront the DNLS method with other techniques.

7.1. The algorithm of the DNLS method

First let us formulate the procedure for experiments with the different damping factor $^j\mu_k$ $(k = 1, 2, ..., p)$. We shall use the notation the DNLS experiment, i.e. the procedure DNLSE.

Algorithm 7.1. Procedure DNLSE

It is assumed that $(y, {}^j\bar{y})$, $\|{}^j\bar{y}\|^2$, $^j\varrho$, $^j\varphi^2$, $^j\delta^2$, $^j a$, $\Delta^j a$ and $^j\mu_k$ are given.
1. Compute $^{j+1}a = {}^j a + {}^j\mu_k\,\Delta^j a$.
2. Test the stability (Definition 4.1).
3. If $^{j+1}a \notin \Omega_a$ then go to 5.
4. Compute $^{j+1}\varrho(^j\mu_k)$, $^{j+1}\eta^2(^j\mu_k)$, $v_1(^j\mu_k)$, $v_3(^j\mu_k)$ and go to 6.
5. The vector $^{j+1}a(^j\mu_k)$ is unstable.
6. Stop.

Now we can derive the main procedure corresponding to the DNLS method.

Algorithm 7.2.

It is assumed that several initial vectors $^0_l a$ $(l = 1, 2, ..., r)$ are given.
1. Set the first initial vector $^0_1 a$, $j = 0$.
2. Compute $^0\varrho$. If $^0\varrho < 0$ then go to 16. Set $j = j + 1$.
3. Compute $(^jv^{(i)}, {}^jv^{(k)})$, $(^jv^{(i)}, {}^j\bar{y})$, $(^jv^{(i)}, y)$ for $i, k = 0, 1, ..., \bar{n}$, and solve the system of linear equations according to (3.6) and (3.7).
4. Compute $(y, {}^j\bar{y})$, $\|{}^j\bar{y}\|^2$, $^j\varrho$, $^j\varphi^2$, $^j\delta^2$, $\Delta\,^j a$, and if $^j\varphi^2 < \varepsilon_1$ then go to 17.
5. Test the DNLS sequence, compute $^j\pi$, and if $j > 3$ and $^j\pi < 1$ then go to 16.
6. Call DNLSE (1). If $v_3(1) < 0$ then use (5.11), go to 15.
7. Call DNLSE (0.5).
8. If $^{j+1}a(1) \notin \Omega_a$ and $^{j+1}a(0.5) \notin \Omega_a$ then go to 16.
9. If $^{j+1}a(1) \notin \Omega_a$ and $^{j+1}a(0.5) \in \Omega_a$ then call DNLSE (0.6). If $^{j+1}a(0.6) \in \Omega_a$ and $v_3(0.5) < 0$ then $^j\mu_{opt} = 0.6$, else $^j\mu_{opt} = 0.5$, go to 15.

10. If $v_3(1) > 0$ and $v_3(0.5) > 0$ then compute $^j\mu_{opt}$ according to (5.4) for $^j\mu_2 = 1$ and $^j\mu_1 = 0.5$, go to 14.

11. If $v_1(1) < 0$ then call DNLSE (0.6). If $v_3(0.6) > 0$ resp. $v_3(0.6) < 0$ compute $^j\mu_{opt}$ using (5.4) resp. (5.7) and (5.4) for $^j\mu_2 = 1$ and $^j\mu_1 = 0.6$, go to 14.

12. If $v_1(1) > 0$ then call DNLSE (0.7). If $^{j+1}\varrho(0.7) < 0$ then call DNLSE (0.6), compute $^j\mu_{opt}$ using (5.11), go to 15.

13. If $^{j+1}\varrho(0.7) > 0$ then compute $^j\mu_{opt}$ using (5.11), go to 15.

14. Compute $^{j+1}a = {}^ja + {}^j\mu_{opt} \Delta\, {}^ja$. Call DNLSE $(^j\mu_{opt})$. If $^{j+1}\eta^2(^j\mu_{opt}) - {}^{j+1}\eta_P^2(^j\mu_{opt}) > \varepsilon_2$ then $^j\mu_{opt} = {}^j\mu_{opt} - 0.1$.

15. Compute $^{j+1}a = {}^ja + {}^j\mu_{opt} \Delta\, {}^ja$, set $j = j + 1$ and go to 3.

16. Set the next initial vector ^0_la, $j = 0$, go to 2.

17. Stop.

Let us discuss the details of Algorithm 7.2. Suitable values of ε_i $(i = 1, 2)$ are given by $\varepsilon_1 = 10^{-7}$, $\varepsilon_2 = 0.04\, {}^j\varphi^2$.

The test of the DNLS sequence is only a sufficient condition of the convergence. Therefore we can admit that it is not fulfilled in the first three iteration steps and so to include the cases which are near to the DNLS sequences.

The key problem of all modifications of the Newton resp. Gauss-Newton method is deriving a suitable initial vector ^0a. Here we must distinguish two situations. If the signal approximation is solved from the given transfer function (2.1) (with $M(s) = 1$) then we can start from the first $(\bar{n} + 1)$ coefficients of the polynomial $N(s)$ (cf. the example in [9]) and the first two points in Algorithm 7.2 can be left out.

The other situation arises in identification problems where we expect useful applications of the DNLS method. The original transfer function (2.1) is unknown and so we must start from a curve of the function $y(t)$. Deriving a suitable initial vector ^0a is therefore a complicated problem and different ways can be used to solve it. In this paper we shall consider only one possibility which is closely connected with the DNLS method. Apparently the simplest way, how to get a "good" initial vector ^0a and simultaneously the acceptable convergence rate, consists in deriving $^0a \in \Phi_1 = \{^ja\colon {}^j\varrho \geq 0\}$, i.e. the initial vector ^0a which satisfies the condition

$$(7.1) \qquad \qquad \|^0y\|^2 \leq (y, {}^0\bar{y}) < \|y\|^2 .$$

The explanation follows from the discussion connected with relations (5.9), (5.11), (6.1) and (6.8).

This way has two advantages:

(i) The needed functionals can be simply computed for different vectors $^0_la\,(l = 1, 2, \ldots, r)$.

(ii) The region Φ_1 is sufficiently large and so the simple experimentation (e.g. with monitoring) with respect to relations (4.11), (4.18) and (7.1) can be used.

7.2. Examples

To illustrate the effectivity of the DNLS method by the use of the optimal damping factor the system from the paper [9]

$$S: F(s) = 1/(1 + 17s + 87 \cdot 24s^2 + 190 \cdot 84s^3 + 193 \cdot 04s^4 + 87 \cdot 84s^5 + 14 \cdot 4s^6)$$

was used. The response of the closed control loop to the unit step given by

$$\mathscr{L}\{y_{st}(t)\} = F_w(s) = \frac{K\,F(s)}{s(1 + K\,F(s))} = \frac{K}{s(K + N(s))}$$

was approximated with $\bar{F}_w(s) = K/(s\bar{N}_c(s))$ for $\bar{n} = 3$. Here K is a given gain coefficient $(K = 3)$. Let us assume the situation by the identification. In this case we know only the function $y(t)$. First we must find the suitable starting function $^0\bar{y}(t)$ and the pertinent transfer function $^0\bar{F}_w(s)$.

Let us test suitability of the starting function $^0\bar{F}_w(s) = 3/(3 \cdot 5 + 17s + 35s^2 + 50s^3)$ with the help of relation (4.11), i.e. $^0\varrho = 7 \cdot 36 \, . \, 10^{-2}$, resp. (7.1), i.e. $\|^0\bar{y}\|^2 = 2 \cdot 8556 < < (y, {}^0\bar{y}) = 2 \cdot 929 < \|y\|^2 = 4 \cdot 5925$. The test is positive and so Algorithm 7.2 can be used. Table 1 contains important parameters computed with a double precision PL 1 program based upon Algorithm 7.2.

Table 1.

j	$^j\mu_{opt}$ $^j\varphi^2$	$^j\delta^2$ $^j\eta^2$	$^j\pi$	$\Delta\,^ja_3(1)$ $v_1(1)$	$^j\mu_H$ $^j\mu_D$
1	0·6 9·8 . 10^{-1}	6·1 . 10^{-1} 1·59	— —	$-62 \cdot 9$ —	div. div.
2	0·5 5·8 . 10^{-1}	2·1 . 10^{-1} 7·9 . 10^{-1}	1·0	241 $-8 \cdot 8$. 10^{-1}	0·46 0·47
3	0·69 3·7 . 10^{-1}	4·0 . 10^{-2} 4·1 . 10^{-1}	1·2	90 5·2 . 10^{-1}	0·41 0·4
4	0·92 1·5 . 10^{-2}	1·318 . 10^{-2} 2·8 . 10^{-2}	13·2	-61 $-2 \cdot 9$. 10^{-3}	0·92 0·9
5	0·97 3·1 . 10^{-5}	1·418 . 10^{-2} 1·421 . 10^{-2}	14·9	5·3 . 10^{-1} $-1 \cdot 4$. 10^{-6}	0·97 0·95
6	0·95 8·0 . 10^{-8}	1·418 . 10^{-2} 1·418 . 10^{-2}	33·9	$-7 \cdot 2$. 10^{-2} $-2 \cdot 2$. 10^{-9}	0·95 0·93

Let us confront the DNLS method with other techniques. First let us compare in Table 2 the DNLS method given by Algorithm 7.2 (col. 1), the main representative of line search techniques — the Hartley method [11] given by relation (5.13) (col. 2), the combination of the first two methods given by relation (5.15) (col. 3) and two

Table 2.

j	$^{j}\mu_{opt}$ $^{j+1}\varphi^2(^{j}\mu_{opt})$	$^{j}\mu_{H}$ $^{j+1}\varphi^2(^{j}\mu_{H})$	$^{j}\mu_{D}$ $^{j+1}\varphi^2(^{j}\mu_{D})$	$^{j+1}\varphi^2(1)$	$^{j+1}\varphi^2(0\cdot95)$
1	$0\cdot6$ $5\cdot8 \cdot 10^{-1}$	div.	div.	div.	div.
2	$0\cdot5$ $3\cdot7 \cdot 10^{-1}$	$0\cdot46$ $3\cdot7 \cdot 10^{-1}$	$0\cdot47$ $3\cdot7 \cdot 10^{-1}$	div.	div.
3	$0\cdot69$ $1\cdot5 \cdot 10^{-2}$	$0\cdot39$ $1\cdot1 \cdot 10^{-1}$	$0\cdot38$ $1\cdot2 \cdot 10^{-1}$	$8\cdot7 \cdot 10^{-1}$	$5\cdot6 \cdot 10^{-1}$
4	$0\cdot92$ $3\cdot1 \cdot 10^{-5}$	$0\cdot77$ $1\cdot3 \cdot 10^{-3}$	$0\cdot81$ $1\cdot8 \cdot 10^{-3}$	$1\cdot4 \cdot 10^{-1}$	$7\cdot5 \cdot 10^{-2}$
5	$0\cdot97$ $8\cdot0 \cdot 10^{-8}$	$0\cdot94$ $5\cdot9 \cdot 10^{-6}$	$0\cdot91$ $8\cdot3 \cdot 10^{-6}$	$2\cdot2 \cdot 10^{-3}$	$8\cdot6 \cdot 10^{-4}$

alternatives of the Gauss-Newton method with $^{j}\mu = 1$ resp. $^{j}\mu = 0\cdot95$ for $\forall j \geqq 0$ (col. 4 resp. 5). The coefficient vector ^{2}a in col. 1, resp. ^{3}a in col. 3, was used as the initial vectors in the subsequent columns. The values $^{j}\mu_H$ and $^{j}\mu_D$ in Table 1 serve only for comparison in a single step and not for the computation of $\{^{j}a\}$. Both tables are sufficiently instructive.

Table 3.

	$^{1}N_{c}(s)$	$^{1}\varrho$ $^{1}\varphi^2$	$^{1}\delta^2$ $^{1}\eta^2$	$\varDelta\,^{1}a_3(1)$ $^{1}\mu_k$
1	$4 + 17s + 2s^2 + 2s^3$	$0\cdot58$ $0\cdot29$	$1\cdot9$ $2\cdot19$	$-1\cdot48$ $1\cdot35$
2	$4 + 17s + 4,36s^2 + 2s^3$	$0\cdot57$ $0\cdot33$	$1\cdot8$ $2\cdot13$	$-8\cdot56$ $0\cdot23$
3	$4 + 17s + 18\cdot7s^2 + 2s^3$	$0\cdot55$ $0\cdot88$	$1\cdot1$ $1\cdot98$	-131 $0\cdot015$
4	$4 + 17s + 57\cdot9s^2 + 2s^3$	$0\cdot63$ $0\cdot87$	$0\cdot29$ $1\cdot16$	158 —
5	$3\cdot75 + 15\cdot8s + 70\cdot6s^2 + 166s^3$	$4\cdot8 \cdot 10^{-8}$ $3\cdot7 \cdot 10^{-9}$	$1\cdot418 \cdot 10^{-2}$ $1\cdot418 \cdot 10^{-2}$	$1\cdot4 \cdot 10^{-2}$ —

Table 3 contains examples of different initial vectors for the example in Table 1 which cannot be solved by the DNLS method. It shows the given values $^{j}\varphi^2$, $^{j}\delta^2$ and above all the critical damping factor $^{j}\mu_k$ which is given by the relation $^{j}a_i +$ $+ \, ^{j}\mu_k \, \varDelta \, ^{j}a_i = 0$ for $\varDelta \, ^{j}a_i < 0$, $^{j}a_i \to 0$ and $i = 0, 1, 2, 3$. If we use in example No. 3 the damping factor $^{1}\mu > \, ^{1}\mu_k = 0\cdot015$ then we obtain $^{2}a \notin \Omega_a$ with respect to the coefficient $^{2}a_3 < 0$.

The results in the given tables show that we must distinguish at least three groups of problems. If the initial vector $^0a \in \Phi_0 = \{^ja: {}^j\varphi < {}^j\delta, {}^j\delta \doteq {}^*\delta\}$, i.e. the initial vector is near the global minimum, then all known methods are practically equivalent ($^j\mu_{opt}, {}^j\mu_D, {}^j\mu_H$ for $j > 3$ in Table 1). The cases in col. 4 resp. 5 in Table 2 complete this conclusion and simultaneously illustrate the discussion connected with relation (6.1). Let us verify relation (6.1) using the results from Table 1

$$A = \frac{^6\varphi^2}{^2\varphi^2} = 1\cdot37 \cdot 10^{-7} \approx \prod_{j=2}^{6} (1 - {}^j\mu)^2 = 3\cdot8 \cdot 10^{-7}.$$

Let us add that a lot different convergence theorems, which can be found in the literature, are connected with the region Φ_0.

The second group of problems is given by the cases in which the sufficiently good linear etalon (cf. Definition 4.2) exists. The example in Table 1 and 2 belongs to this group. Here the Gauss-Newton method is divergent ($j = 1, 2$ in Table 2). The Hartley method (col. 2) and the Hartley-DNLS combination (col. 3) give better results and are practically equivalent. The first method needs for the solution two experiments — the second only one experiment. The DNLS method is better than these two methods (col. 1 in Table 2 and $j = 3$ for $v_1(1) > 0$ in Table 1). Here let us remark that the main contribution of the DNLS method lies not only in these numerical results but above all in the derived theory which allows to demarcate the regions of the effective use for different techniques.

The third group is connected with problems which must start from a misleading linear etalon (compare $^*\delta^2 = 1\cdot418 \cdot 10^{-2}$ with $^1\delta^2$ in Table 3). Here the methods with a broader convergence domain, i.e. the gradient resp. compromise-gradient techniques and the quasi-Newton methods [3], must be used. But all these methods have the lower convergence rate (cf. the discussion in paragraph 6.1) resp. are less robust (cf. Fletcher [2], pp. 95). The reader can test how lengthy is the solution of the considered example starting from initial vector No. 1 in Table 3 using the compromise-gradient techniques, e.g. with the Marquardt method.

Let us emphasize in this connection that the theory of the DNLS method together with the apparatus of the control theory open mostly simpler and faster ways for solving such cases. The simplest way is to find another better initial vector (points 1, 2, 16 in Algorithm 7.2). Further we can extend the DNLS method and compute in each experiment parameters $^{j+1}\delta$ and $^{j+1}\varphi$ to find a better linear etalon and so a better initial vector (see $^1\delta^2$ in Table 3). Let us show one of these possibilities which starts from Proposition 6.1. The initial vector No. 1 in Table 3 is very near to the control stability boundary. Let us try to find a new initial vector which is nearer to the middle of the stability region. From the stability conditions it follows that only $\Delta {}^ja_i > 0$ ($i = 0, 1, 2, 3$) can be accepted. We compute in the considered example $\Delta {}^1a = [-1\cdot5, -3\cdot7, +2\cdot36, -1\cdot48]^T$ and so only $\Delta {}^1a_2 = 2\cdot36$ can be used to get new initial vector No. 2 in Table 3. The further initial vectors, i.e. No. 3 and 4, were derived with the same procedure and then Algorithm 7.2 was used. The result

No. 5 fulfilling the condition $^j\varphi^2 < \varepsilon_1$ was computed in ten steps. This procedure must be tested with a decrease in $^j\delta$ and an increase in the factor $^j\pi$ to find the starting point for using Algorithm 7.2. We can conclude: It is apparently more attractive to derive some classification of initial vectors based on the functional properties of the considered class of functions as to search for a method with a necessary slow convergence starting from an arbitrary initial vector.

8. THE CASE WITH TRANSFER FUNCTION (2.2)

The gradient of substitute transfer function (2.2)

(8.1)
$$\text{grad } \bar{F}(s, \boldsymbol{b}, \boldsymbol{a}) = \frac{1}{\bar{N}^2(s)} \left\{ s \, \bar{N}(s), s^2 \, \bar{N}(s), \ldots, s^{\bar{m}} \, \bar{N}(s), -\bar{M}s), -s\bar{M}(s), \ldots, -s^{\bar{n}}\bar{M}(s) \right\}$$

shows that the substitute signal $^j\bar{y}(t, \boldsymbol{b}, \boldsymbol{a})$ is linear in coefficients b_k ($k = 1, 2, \ldots, \bar{m}$). Therefore the influence of these coefficients can be simply included in the relations derived in the preceding sections. We shall use the same notation if possible. First the new vector of coefficients must be introduced

(8.2)
$$\boldsymbol{c} = \left[a_0, a_1, \ldots, a_{\bar{n}}, b_1, \ldots, b_{\bar{m}} \right]^{\mathrm{T}}.$$

The pertinent gradient vector is

(8.3)
$$^j\boldsymbol{g}(t, \boldsymbol{c}) = \left[-^jv^{(0)}(t, {}^j\boldsymbol{c}), {}^jv^{(1)}(t, {}^j\boldsymbol{c}), \ldots \right.$$
$$\left. \ldots, -^jv^{(\bar{n})}(t, {}^j\boldsymbol{c}), {}^jw^{(1)}(t, {}^j\boldsymbol{a}), \ldots, {}^jw^{(\bar{m})}(t, {}^j\boldsymbol{a}) \right]^{\mathrm{T}},$$

where

$$\mathscr{L}\left\{ {}^jv^{(i)}(t, {}^j\boldsymbol{c}) \right\} = \frac{{}^j\bar{M}(s)\, s^i}{{}^j\bar{N}^2(s)}\, u(s),$$

$$\mathscr{L}\left\{ {}^jw^{(k)}(t, {}^j\boldsymbol{a}) \right\} = \frac{s^k}{{}^j\bar{N}(s)}\, u(s).$$

Now fundamental relation (3.6) can be written in the form

(8.4)
$$^{j+1}\boldsymbol{c} = {}^j\boldsymbol{c} + {}^j\mu G^{-1}({}^j\boldsymbol{c}) \int_0^\infty \boldsymbol{g}(t, {}^j\boldsymbol{c}) \left[y(t) - {}^j\bar{y}(t, {}^j\boldsymbol{c}) \right] \mathrm{d}t = {}^j\boldsymbol{c} + {}^j\mu\, \Delta\, {}^j\boldsymbol{c},$$

where the matrix $G({}^j\boldsymbol{c}) = \int_0^\infty \boldsymbol{g}(t, {}^j\boldsymbol{c})\, \boldsymbol{g}^{\mathrm{T}}(t, {}^j\boldsymbol{c})\, \mathrm{d}t$ must be nonsingular and

$$\Delta\, {}^j\boldsymbol{c} = \left[\Delta\, {}^ja_0, \Delta\, {}^ja_1, \ldots, \Delta\, {}^ja_{\bar{n}}, \Delta\, {}^jb_1, \ldots, \Delta\, {}^jb_{\bar{m}} \right]^{\mathrm{T}}.$$

The gradient vector $\boldsymbol{g}_L(t, {}^j\boldsymbol{c})$ pertinent to the linear part of the DNLS method takes the form

(8.5)
$$\boldsymbol{g}_L(t, {}^j\boldsymbol{c}) = \left[{}^jv^{(0)}(t, {}^j\boldsymbol{c}), {}^jv^{(1)}(t, {}^j\boldsymbol{c}), \ldots, {}^jv^{(\bar{n})}(t, {}^j\boldsymbol{c}), {}^jw^{(1)}(t, {}^j\boldsymbol{a}), \ldots \right.$$
$$\left. \ldots, {}^jw^{(\bar{m})}(t, {}^j\boldsymbol{a}) \right]^{\mathrm{T}}.$$

From relations (8.3) and (8.5) it follows

$$(8.6) \qquad g_L(t, {}^j c) \doteq -g(t, {}^j c) .$$

We see better the new situation from the functions pertinent to relations (4.1) and (4.2)

$$(8.7) \qquad {}^j \bar{y}(t, {}^j c, {}^{j+1} c, {}^j \mu) = {}^j \bar{y}(t, {}^j c) - {}^j \mu \, \Delta \, {}^j \bar{y}(t, {}^j c, {}^j a) + \Delta \, {}^j \bar{y}(t, {}^j a, {}^j b) =$$

$$= {}^j \bar{y}(t, {}^j c) - {}^j \mu \sum_{i=0}^{\bar{n}} \Delta \, {}^j a_i \, {}^j v^{(i)}(t, {}^j c) + \sum_{k=1}^{\bar{m}} \Delta \, {}^j b_k \, {}^j w^{(k)}(t, {}^j a) ,$$

$$(8.8) \qquad {}^j z(t, {}^j c, {}^{j+1} \hat{c}, {}^j \mu) = {}^j \bar{y}(t, {}^j c) + {}^j \mu \, \Delta \, {}^j \bar{y}(t, {}^j c, {}^j a) + \Delta \, {}^j y(t, {}^j a, {}^j b) =$$

$$= {}^j \bar{y}(t, {}^j c) + {}^j \mu \sum_{i=0}^{\bar{n}} \Delta \, {}^j a_i \, {}^j v^{(i)}(t, {}^j c) + \sum_{k=1}^{\bar{m}} \Delta \, {}^j b_k \, {}^j w^{(k)}(t, {}^j a) .$$

The component pertinent to the vector ${}^j b = [{}^j b_1, {}^j b_2, \ldots, {}^j b_{\bar{m}}]^T$ is linear and has consequently the positive sign in both relations and so ${}^j \mu = 1$ is used. The new function

$$(8.9) \qquad {}^j \hat{y}(t, {}^j c) = {}^j \bar{y}(t, {}^j c) + \Delta \, {}^j \bar{y}(t, {}^j a, {}^j b)$$

can be introduced and then relations (8.7) and (8.8) can be written in the form of relations (4.1) and (4.2)

$$(8.10) \qquad {}^j \bar{y}(i, {}^j c, {}^{j+1} c, {}^j \mu) = {}^j \hat{y}(i, {}^j c) - {}^j \mu \, \Delta \, {}^j \bar{y}(i, {}^j c, {}^j a) ,$$

$$(8.11) \qquad {}^j z(t, {}^j c, {}^{j+1} \hat{c}, {}^j \mu) = {}^j \hat{y}(t, {}^j c) + {}^j \mu \, \Delta \, {}^j \bar{y}(t, {}^j c, {}^j a) ,$$

where

$$\mathscr{L}\{{}^j \hat{y}(t, {}^j c)\} = \frac{\sum\limits_{i=0}^{\bar{n}} {}^j a_i s^i \, {}^j \overline{M}(s) \, u(s)}{{}^j \overline{N}^2(s)} + \frac{\sum\limits_{k=1}^{\bar{m}} \Delta \, {}^j b_k s^k}{{}^j \overline{N}(s)} = \frac{{}^{j+1} \overline{M}(s) \, u(s)}{{}^j \overline{N}(s)} .$$

Therefore all relations derived in this paper resp. in the paper [9] hold for the case with substitute transfer function (2.2).

9. CONCLUSION

Let us summarize the main results. The DNLS method is based on deriving the optimal damping factor (Proposition 5.1, relations (5.9), (5.11), (5.15), Algorithm 7.2) with the help of the linear etalon ${}^j z(t)$ (relations (3.7), (4.2), Propositions 4.1 to 4.3). The sequence $\{{}^j a\}$ is governed in the iteration endsteps only by the linear case and the global minimum is linear (cf. Proposition 4.4 in [9]). So we can speak about the quasi-linear method. This quasi-linearity is the fundamental difference to all other modifications of the Newton resp. Gauss-Newton method. Thus the DNLS method is suited for the solution of all control problems and further for the solution of problems by which the linear etalon can be found. From this fact it follows that the linear control theory can be used in subsequent problems (e.g. for an optimization).

The convergence of the DNLS sequence (Definition 4.2) was proved in Proposition 6.1 and the convergence rate was discussed in paragraph 6.2 (cf. Proposition 6.2). The importance of the DNLS method lies not only in better numerical results (Table 1, 2 and 3) but above all in the pertinent theory which opens the way for the further research on the field of the control theory, e.g. for the comparison of the EE and OE identification methods, the explanation of the failure of the EE methods for a small number of measured data, deriving the robust identification procedure, etc. [10].

REFERENCES

[1] M. Athans: Advances and open problems on the control of large scale systems. Proceedings of IFAC Congress, Helsinki 1978, Pergamon Press, Oxford 1978.

[2] R. Fletcher: Practical Methods of Optimization. — Vol. 1: Unconstrained Optimization. J. Wiley, New York 1980.

[3] L. Lukšan: Quasi-Newton method without projections for unconstrained minimization. Kybernetika *18* (1982), 4, 290—306.

[4] G. Meinardus: Approximation von Funktionen und ihre numerische Behandlung. Springer-Verlag, Berlin 1984.

[5] J. Militký: Řešení úloh regrese (Solution of regression problems). Dům techniky, Ostrava 1980.

[6] J. Štěpán: Approximace funkcí v úlohách regulační techniky (Approximation of functions in control problems). Kybernetika *3* (1967), 3, 277—292.

[7] J. Štěpán: Some problems of system identification. Kybernetika *7* (1971), 2, 133—155.

[8] J. Štěpán: Decentralizované řízení složitých systémů (Decentralized control of large scale systems). Automatizace *26* (1983), 12, 301—304.

[9] J. Štěpán: Nonlinear approximation in control problems. Kybernetika *19* (1983), 6, 491—504.

[10] J. Štěpán: Approximation versus identification. Proceedings of IFAC Symposium on Identification, York 1985, Pergamon Press, Oxford 1985.

[11] M. A. Wolfe: Numerical Methods for unconstrained Optimization. Van Nostrand Reinhold, New York 1978.

[12] P. Young: Parameter estimation for continuous-time models. A survey. Proceedings of IFAC Congress, Helsinki 1978, Pergamon Press, Oxford 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.