

Miskolc Mathematical Notes

A Publication of the University of Miskolc

VOLUME 5 (2004), NUMBER 1



MISKOLC UNIVERSITY PRESS



Miskolc Mathematical Notes

VOLUME 5 (2004), NUMBER 1

CONTENTS

KATALIN BALLA AND ROSWITHA MÄRZ

Linear boundary value problems for differential algebraic equations . . . 3

IVAN CHAJDA

Congruences in transitive relational systems 19

ATTILA GÁTI

Automatic error analysis with Miller's method 25

ROBERT HAKL

On periodic-type boundary value problems for functional differential equations with a positively homogeneous operator 33

VU HOANG LINH

On the high order asymptotic solution of certain wave equations 57

ANDREI RONTÓ

On the unique solvability of linear equations determined by monotone decomposable operators 71

ALY A. SHABAN

Cooperation in patchy environment with cross-diffusion 83

SVATOSLAV STANĚK

A nonlocal singular boundary value problem for second-order differential equations 91

Z. SZABÓ

Parabola-parabola combined method 105

MIKLÓS RONTÓ

The Second International Workshop "Constructive Methods for Non-Linear Boundary Value Problems" 113



LINEAR BOUNDARY VALUE PROBLEMS FOR DIFFERENTIAL ALGEBRAIC EQUATIONS

KATALIN BALLA AND ROSWITHA MÄRZ

[Received: September 15, 2003]

ABSTRACT. By the use of the corresponding shift matrix, the paper gives a criterion for the unique solvability of linear boundary value problems posed for linear differential algebraic equations up to index 2 with well-matched leading coefficients. The solution is constructed by a proper Green function. Another characterisation of the solutions is based upon the description of arbitrary affine linear subspaces of solutions to linear differential algebraic equations in terms of solutions to the adjoint equation. When applied to boundary value problems, the result provides a constructive criterion for unique solvability and allows one to reduce the problem to initial value problems and linear algebraic equations.

Mathematics Subject Classification: 65L80, 34A09

Keywords: Differential algebraic equations, boundary value problems, adjoint equations

1. INTRODUCTION

FOR THE linear differential algebraic equations (DAEs for the sake of brevity in what follows) of the form

$$A(t) (D(t)x(t))' + B(t)x(t) = q(t) \quad (1.1)$$

with continuous, quadratic matrix-valued functions A , D and B with complex entries, the “index-1” and “index-2” notion was introduced in [2]. A theorem on the unique solvability of the properly formulated initial value problems (IVPs) for (1.1) equipped with these indices was proven. It was shown that, under the same conditions, the adjoint equation

$$-D^*(t) (A^*(t)y(t))' + B^*(t)y(t) = p(t) \quad (1.2)$$

is of the same index, and the proper IVP for (1.2) is solvable simultaneously with that for (1.1). Meanwhile, some properties of the inherent ordinary differential equation

Supported by Hungarian National Scientific Foundation Grants (OTKA) ## T029572, T031807 and by DFG Research Center “Mathematics for Key Technologies” (FZT 86) in Berlin.

(ODE) of (1.1) were investigated. The fundamental matrices for (1.1) and a specific one called the normalised fundamental matrix were also introduced.

The main goal of this paper is to study the boundary value problems (BVPs) for (1.1) up to index 2. The assertion on the simultaneous solvability of (1.1) and (1.2) turns out to be the keystone in the analysis of BVPs.

The paper is organised as follows. In Section 2, we recall the basic definitions and some propositions concerning equation (1.1). For the sake of completeness, we also define “index-0” equations. The results of [2] can be extended to “index-0” equations in an obvious way. The solvability theorem for IVPs posed for the pair (1.1) and (1.2) is cited in this section. Existence results for two-point BVPs in terms of Green’s function and shift matrix will be stated in Section 3. In Section 4, we describe affine linear subspaces of solutions to (1.1) by the help of solutions to (1.2). The transfer of boundary conditions (BCs) for BVPs both with separated and non-separated BCs and the related constructive existence theorem will be the topic of Section 5. The paper is concluded with some remarks on numerical implementation in the final Section 6.

2. PRELIMINARIES

We consider equations of the form (1.1), where A, D and B are continuous $m \times m$ matrix functions with complex entries on closed interval $\mathcal{I} = [a, b]$, and q is a continuous vector-valued function with complex components on $\mathcal{I}$. Parallel to (1.1), equation (1.2) is involved in our study, p also being a continuous vector-valued function with complex components on $\mathcal{I}$. The pair of leading terms in (1.1) is assumed to be *well-matched* in the following sense:

Condition C1 ([2]). For every $t \in \mathcal{I}$, the equality

$$\ker A(t) \oplus \operatorname{im} D(t) = \mathbb{C}^m \quad (2.1)$$

is true, and there exist continuously differentiable functions $a_1, \dots, a_{m-r}$ and $d_1, \dots, d_r$ such that

$$\ker A(t) = \operatorname{span} \{a_1(t), \dots, a_{m-r}(t)\}, \quad \operatorname{im} D(t) = \{d_1(t), \dots, d_r(t)\}, \quad t \in \mathcal{I}. \quad (2.2)$$

We proved

Lemma 1 ([2, Lemma 2.1]). *Equation (1.1) has well-matched leading coefficients A and D if and only if the leading coefficients A^* and D^* of equation (1.2) do so.*

If R is the continuously differentiable projector function realizing the decomposition (2.1), i. e., $\ker R(t) = \ker A(t)$ and $\operatorname{im} R(t) = \operatorname{im} D(t)$, $t \in \mathcal{I}$, then R^* is the projector function corresponding to the decomposition induced by A^* and D^* .

Remark 1. We are mainly interested in considering singular well-matched leading pairs $A(t)$ and $D(t)$. Assumption (2.1), however, includes the case where both matrices $A(t)$ and $D(t)$ are nonsingular over the entire interval $\mathcal{I}$. Then, $r = m$ and $R(t) \equiv I$ where I is the $m \times m$ identity matrix. The considerations of [2] can be easily extended

to the case of nonsingular well-matched leading terms. Equation (1.1) turns into a standard explicit ODE if $A(t) \equiv D(t) \equiv I$.

Definition 1 ([2, Definition 2.1]). A vector function $x : I \rightarrow \mathbb{C}^m$ is called a solution of (1.1) if $x \in C_D^1(I) := \{x \in C(I) : Dx \in C^1(I)\}$ and (1.1) is satisfied pointwise.

A solution of (1.2) is defined similarly. A kind of Lagrange identity is stated.

Lemma 2. *Let the matrix functions A and D be well-matched. Then, for every pair of solutions $x \in C_D^1$, $y \in C_{A^*}^1$ of the homogeneous equations (1.1) and (1.2), respectively, the identity*

$$y^*(t)A(t)D(t)x(t) = \text{const} \quad \text{for } t \in I. \quad (2.3)$$

holds.

A key tool in the investigation of problems carried out in [2] is a chain of matrix-valued and subspace-valued functions associated with (1.1), namely,

$$\begin{aligned} G_0 &:= AD, & B_0 &:= B; \\ \text{for } i = 0, 1, & & Q_i, P_i, W_i &\text{ are projector functions: } Q_i^2 = Q_i, \quad W_i^2 = W_i, \\ & & N_i &:= \ker G_i = \text{im } Q_i, \quad P_i = I - Q_i, \\ & & \ker W_i &= \text{im } G_i, \\ & & G_{i+1} &:= G_i + B_i Q_i, \quad B_{i+1} = B_i P_i, \\ & & S_i &:= \{z \in \mathbb{C}^m : B_i z \in \text{im } G_i\} = \ker W_i B_i. \end{aligned} \quad (2.4)$$

In the sequel, D^- denotes the reflexive generalised inverse (RGI) function of D such that $DD^- = R$ and $D^-D = P_0$; A^- is an RGI function of A such that $A^-A = R$ and $AA^- = I - W_0$; G_1^- stands for the RGI function of G_1 such that $G_1 G_1^- = I - W_1$ and $G_1^- G_1 = P_1$. We recall [5] that a matrix $T^- \in L(\mathbb{C}^k, \mathbb{C}^l)$ is an RGI of a matrix $T \in L(\mathbb{C}^l, \mathbb{C}^k)$ if it satisfies the equalities $T^- T T^- = T^-$ and $T T^- T = T$. The products $P_{RGI1} := T T^-$ and $P_{RGI2} := T^- T$ are projectors. If P_{RGI1}, P_{RGI2} are given projectors such that $\text{im } P_{RGI1} = \text{im } T$ and $\ker P_{RGI2} = \ker T$, then they define an RGI T^- uniquely.

Due to condition C1, $\dim \text{im } G_0(t) \equiv r$. Let $\dim \text{im } G_1(t) = r_1(t)$. Based on the properties of terms in the chain, an index may be assigned to some equations of the form (1.1) if, in addition to condition C1, another requirement is also fulfilled.

Condition C2 ([2]). The dimensions of $D(t)S_1(t)$ and $D(t)N_1(t)$ are constant,

$$\dim D(t)S_1(t) =: \varrho \quad \text{and} \quad \dim D(t)N_1(t) =: \nu, \quad (2.5)$$

and there exist continuously differentiable functions $s_1^D, \dots, s_\varrho^D$ and $n_1^D, \dots, n_\nu^D$ such that for all $t \in I$,

$$D(t)S_1(t) = \text{span} \{s_1^D(t), \dots, s_\varrho^D(t)\}, \quad D(t)N_1(t) = \text{span} \{n_1^D(t), \dots, n_\nu^D(t)\}.$$

Here, we extend Definition 2.2 from [2] as follows:

Definition 2. Let conditions C1 and C2 be valid. Equation (1.1) is said to be

(0) an “index-0” tractable DAE if

$$N_0(t) = \{0\}, \quad t \in \mathcal{I}, \quad (2.6)$$

(1) an “index-1” tractable DAE if

$$N_0(t) \neq \{0\}, \quad (2.7)$$

$$N_0(t) \cap S_0(t) = \{0\}, \quad t \in \mathcal{I}, \quad (2.8)$$

(2) an “index-2” tractable DAE if

$$\dim N_0(t) \cap S_0(t) = \text{const} > 0, \quad (2.9)$$

$$N_1(t) \cap S_1(t) = \{0\}, \quad t \in \mathcal{I}. \quad (2.10)$$

When $r < m$, the chain associated with an equation is not uniquely defined due to the freedom in the choices of the projectors. The index, however, does not depend on these choices. Thus, the index value, if it exists, is an inherent property of the equation. One may choose a specific projector $\hat{Q}_1$ so that $\ker \hat{Q}_1(t) = S_1(t)$. The related terms in the chain will be marked by “^” (a hat). For equations equipped with an index, the assumptions ensure $r_1(t) \equiv \text{const} =: r_1$ and $\varrho = r + r_1 - m$, $v = m - r_1$. In the “index-0” and “index-1” cases, $r_1 = m$. The function $D\hat{P}_1D^-$ is a continuously differentiable projector function: for every t , it projects onto DS_1 along $DN_1 \oplus \ker A$.

Setting $A_* = -D^*$, $D_* = A^*$, and $B_* = B^*$, one can construct a chain similar to (2.4) starting with A_* , D_* , B_* , i. e., for equation (1.2). The terms derived in this chain will be marked by an additional first subscript “*” (a star).

With the inclusion of the “index-0” equations, Theorem 5.1 of [2] reads as follows:

Theorem 1. *Equation (1.1) is of index μ , $\mu = 0, 1, 2$, if and only if equation (1.2) possesses this property.*

The main point in the proof [2] of this theorem consists in showing that

$$\begin{aligned} DS_1 &= R(A^*N_{*1})^\perp = (A^*N_{*1} \oplus \ker D^*)^\perp, \\ A^*S_{*1} &= R^*(DN_1)^\perp = (DN_1 \oplus \ker A)^\perp. \end{aligned} \quad (2.11)$$

The so-called inherent regular ODE for DAE (1.1) has the form

$$u' + DG_0^{-1}BD^{-1}u = A^{-1}q \quad (2.12)$$

in the “index-0” case, and it has the form

$$u' - R'u + DG_1^{-1}BD^-u = DG_1^{-1}q \quad (2.13)$$

in the “index-1” case. If the DAE is of index 2, then the inherent ODE is

$$u' - (D\hat{P}_1D^-)'u + D\hat{P}_1\hat{G}_2^{-1}BD^-u = N_0q, \quad (2.14)$$

where

$$N_0q := D\hat{P}_1\hat{G}_2^{-1}q + (D\hat{P}_1D^-)'D\hat{Q}_1\hat{G}_2^{-1}q, \quad N_0q = D\hat{P}_1D^-N_0q.$$

For the inherent ODEs derived from an “index-1” DAE it was shown that if $u(\tilde{t}) \in \text{im } D(\tilde{t})$ for some $\tilde{t} \in \mathcal{I}$, then $u(t) \in \text{im } D(t)$ for all $t \in \mathcal{I}$. Similarly, in the “index-2” case $u(\tilde{t}) \in \text{im } D(\tilde{t})\hat{P}_1(\tilde{t})$ involves $u(t) \in \text{im } D(t)\hat{P}_1(t)$. Equations (2.13) and (2.14) are independent of the choice of P_0 and P_0, P_1 , respectively.

Let x be a solution of equation (1.1). If (1.1) is a DAE of index 0, then Dx is a solution of (2.12). In the “index-1” case, Dx is a solution of (2.13). In the “index-2” case, function $D\hat{P}_1x$ is a solution of (2.14).

Finally, we recall the solvability statement for IVPs.

Theorem 2 ([2, Theorems 3.1 and 3.2]). *Let $t_0 \in \mathcal{I}$. Assume that one of the following conditions is satisfied:*

- (i) (1.1) is an “index-0” or “index-1” DAE, the inclusion $q \in C(\mathcal{I})$ holds, and the initial condition has the form

$$D(t_0)x(t_0) = d_0 \quad \text{with} \quad d_0 \in \text{im } D(t_0);$$

- (ii) (1.1) is an “index-2” DAE, the inclusion $q \in C^1_{DQ_1G_2^{-1}}(\mathcal{I})$ holds, and the initial condition has the form

$$D(t_0)\hat{P}_1(t_0)x(t_0) = d_0 \quad \text{with} \quad d_0 \in \text{im } D(t_0)\hat{P}_1(t_0). \quad (2.15)$$

Then there exists a unique solution x of the IVP.

Now, the assertion on simultaneous solvability of DAEs (1.1) and (1.2) with proper right-hand sides and initial conditions appears to be a direct consequence of Theorems 1 and 2.

Note that for the “index-0” equations the initial condition is equivalent simply to condition $x_0 \in \mathbb{C}^m$ and the equation may be considered formally a particular case of “index-1” equations with $Q_0 = W_0 = 0$, $r = m$. In turn, an “index-1” DAE may be considered formally a particular case of “index-2” equations with $Q_1 = W_1 = 0$; then $\varrho = r$, $D\hat{P}_1D^- = R$, $D(t)N_1(t) \equiv \{0\}$, $G_2 = G_1$. Thus, in the next sections it is sufficient to prove the statements only for the “index-2” DAEs.

3. THE APPROACH BASED ON GREEN’S FUNCTION

Let equation (1.1) be tractable with index μ , $\mu \in \{0, 1, 2\}$. Denote the maximal fundamental solution matrix normalised at $t_0 \in \mathcal{I}$ by $X(t, t_0)$, i. e., $X(t, t_0) \in L(\mathbb{C}^m)$ and $X(\cdot, t_0)$ is the matrix-valued solution of the IVP

$$A(DX)' + BX = 0, \quad D(t_0)\hat{P}_1(t_0)(X(t_0) - I) = 0. \quad (3.1)$$

We recall from [2] the following properties of the maximal fundamental solutions:

$$\text{im } X(t, t_0) = \text{im } \Pi_{\text{can}\mu}(t) \quad \ker X(t, t_0) = \ker \Pi_{\text{can}\mu}(t_0), \quad t \in \mathcal{I},$$

where $\Pi_{\text{can}\mu}$ is a projector function onto the geometric solution space of the homogeneous DAE (1.1) ($q = 0$), $S_{\text{ind}\mu}(t) = \text{im } \Pi_{\text{can}\mu}(t)$,

$$\Pi_{\text{can}\mu} := KP_0\hat{P}_1, \quad K := I - Q_0\hat{P}_1\hat{G}_2^{-1}BP_0 - Q_0\hat{Q}_1D^-(D\hat{Q}_1D^-)'D, \quad (3.2)$$

U is nonsingular.

Let the RGI $X(t, t_0)^- \in L(\mathbb{C}^m)$ of $X(t, t_0)$ be defined by the relations

$$\begin{aligned} X(t, t_0)X(t, t_0)^- &= \Pi_{\text{can}\mu}(t), \\ X(t, t_0)^-X(t, t_0) &= \Pi_{\text{can}\mu}(t_0). \end{aligned}$$

(See Section 2 for the definition of RGI). The usual group properties

$$X(t_1, t_2)X(t_2, t_3) = X(t_1, t_3), \quad X(t_1, t_2)^- = X(t_2, t_1).$$

hold. It follows from Theorem 2 that, for all $q \in C_{DQ_1G_2^{-1}}(I)$ and $x^0 \in \mathbb{C}^m$, the IVP

$$A(Dx)' + Bx = q, \quad D(t_0)\hat{P}_1(t_0)(x(t_0) - x^0) = 0, \quad (3.3)$$

is uniquely solvable. Due to the linearity, the solution can be split into two terms,

$$x(t) = X(t, t_0)x^0 + \tilde{x}(t), \quad t \in I, \quad (3.4)$$

where $\tilde{x}$ denotes the solution of the IVP

$$A(Dx)' + Bx = q, \quad D(t_0)\hat{P}_1(t_0)x(t_0) = 0. \quad (3.5)$$

In [2], it was shown that every solution of (1.1) can be represented in the form $x = \Pi_{\text{can}\mu}x + \mathcal{N}_1q$, where

$$\mathcal{N}_1q := (P_0\hat{Q}_1 + Q_0\hat{P}_1)\hat{G}_2^{-1}q + Q_0\hat{Q}_1D^-(DQ_1G_2^{-1}q)'.$$

Thus, one can obtain the relation

$$\tilde{x}(t) = \int_{t_0}^t X(t, s)(\mathcal{N}_0q)(s)ds + (\mathcal{N}_1q)(t), \quad t \in I. \quad (3.6)$$

Now we turn to the BVP for (1.1) with the boundary condition

$$K_ax(a) + K_bx(b) = d, \quad (3.7)$$

$d \in L_{BC}$, where $L_{BC} := \text{im}(K_a | K_b) \subseteq \mathbb{C}^m$ is the linear subspace associated with the boundary condition. The values $x^0 \in \mathbb{C}^m$ in (3.4) that yield solutions of the BVP (1.1), (3.7) must satisfy the linear system

$$Mx^0 = d - K_a\tilde{x}(a) - K_b\tilde{x}(b) \quad (3.8)$$

with the “shift matrix” M ,

$$M := K_aX(a, t_0) + K_bX(b, t_0). \quad (3.9)$$

Theorem 3. *Let DAE (1.1) be tractable with index μ , $\mu \in \{0, 1, 2\}$. Then, for arbitrary $d \in L_{BC}$ and $q \in C_{DQ_1G_2^{-1}}^1(I)$, the BVP (1.1), (3.7) is uniquely solvable if and only if the shift matrix M satisfies the conditions*

$$\ker M = \ker \Pi_{\text{can}\mu}(t_0), \quad (3.10)$$

$$\text{im } M = L_{BC}. \quad (3.11)$$

Proof. By construction, the relations $\ker \Pi_{\text{can}\mu}(t_0) \subseteq \ker M$, $\text{im } M \subseteq L_{BC}$ are true.

Let the BVP (1.1), (3.7) be uniquely solvable for all $d \in L_{BC}$ and $q \in C_{DQ_1G_2^{-1}}^1(I)$. Put $q = 0$. For every $d \in L_{BC}$, there is an $x^0 \in \mathbb{C}^m$ such that $Mx^0 = d$. Hence, $L_{BC} \subseteq \text{im } M$, i. e., (3.11) holds.

Moreover, since the homogeneous BVP (1.1), (3.7) with $d = 0$ and $q = 0$ has only the trivial solution, the IVP

$$A(Dx)' + Bx = 0, \quad D(t_0)\hat{P}_1(t_0)(x(t_0) - x^0) = 0, \quad x^0 \in \ker M$$

may have only the identically vanishing solution. This means that

$$\ker M \subseteq \ker D(t_0)\hat{P}_1(t_0) = \ker \Pi_{\text{can}\mu}(t_0)$$

must be true, and consequently, (3.10) holds.

Conversely, let (3.10) and (3.11) be satisfied. Then, for every $d \in L_{BC}$ and $q \in C_{DQ_1G_2^{-1}}^1(I)$, a solution of the BVP is determined by (3.4) and (3.8). The relations $d = 0$ and $q = 0$ imply $\tilde{x} = 0$ and $Mx^0 = 0$. Thus, $x^0 \in \ker M = \ker X(t, t_0)$. Now (3.4) leads us to the solution x which equal identically to zero. $\square$

Remark 2. The conditions (3.10), (3.11) ensure that $\text{rank } M = \varrho = r + r_1 - m$.

When (3.10) and (3.11) are true, we can introduce an RGI $M^- \in L(\mathbb{C}^m)$ of M such that $M^-M = \Pi_{\text{can}\mu}(t_0)$ holds.

Theorem 4. Let DAE (1.1) have tractability index μ , $\mu \in \{0, 1, 2\}$, and let conditions (3.10) and (3.11) be satisfied. Then the solution of BVP (1.1), (3.7) with $d \in L_{BC}$, $q \in C_{DQ_1G_2^{-1}}^1(I)$ is given by the formula

$$\begin{aligned} x(t) = & X(t, t_0)M^-d + \int_a^b \mathcal{G}(t, s)(N_0q)(s)ds \\ & + (N_1q)(t) - X(t, t_0)M^- \{K_a(N_1q)(a) + K_b(N_1q)(b)\}, \end{aligned} \quad (3.12)$$

where Green's function $\mathcal{G}$ is defined as follows:

$$\mathcal{G}(t, s) = \begin{cases} X(t, t_0)M^-K_aX(a, t_0)X(s, t_0)^-, & s \leq t, \\ -X(t, t_0)M^-K_bX(b, t_0)X(s, t_0)^-, & s > t. \end{cases}$$

Proof. It follows from (3.8) that $\Pi_{\text{can}\mu}(t_0)x^0 = M^-(d - K_a\tilde{x}(a) - K_b\tilde{x}(b))$, whereas from (3.4), one obtains (3.12) by standard calculations. Note that (3.12) is defined in a unique manner, while there is freedom in the choice of M^- . $\square$

Remark 3. The map $\mathcal{L} : C_D^1(I) \rightarrow C(I) \times L_{BC}$ defined by the relation

$$\mathcal{L}x := (A(Dx)' + Bx, K_ax(a) + K_bx(b)), \quad x \in C_D^1(I)$$

is linear and bounded. It acts bijectively between $C_D^1(I)$ and $C_{DQ_1G_2^{-1}}^1(I) \times L_{BC}$. Recall that, in the case where $\mu = 2$, the set $C_{DQ_1G_2^{-1}}^1(I)$ is a proper dense subset of $C(I)$.

Hence, when $\mu = 2$, $\mathcal{L}$ has a densely defined unbounded inverse $\mathcal{L}^{-1}$. However, if we equip $C_{DQ_1G_2^{-1}}^1(I)$ with a natural norm and consider $\mathcal{L}$ as a mapping $\mathcal{L} : C_D^1(I) \rightarrow C_{DQ_1G_2^{-1}}^1(I) \times L_{BC}$, in this setting, $\mathcal{L}$ has a bounded inverse.

4. AFFINE LINEAR SUBSPACES OF SOLUTIONS

In [1], for a subclass of homogeneous “index-1” DAEs (1.1) with $D = P_0$ and continuously differentiable coefficients A , P_0 and B , we described the linear subspaces of solutions in terms of the adjoint equation. In this section, we formulate and prove an analogous theorem for the affine linear subspaces of the solutions for DAE (1.1) with an arbitrary function D well-matched with A . The DAE is assumed to be of index μ , $\mu \in \{0, 1, 2\}$, and it is not necessarily homogeneous. The smoothness conditions on the coefficients A , D and B are exactly as in Section 2, i. e., they must allow for assignment of an index only. The function q is assumed to be of the class required by Theorem 2; in the “index-0” and “index-1” cases, q is only continuous, whereas $q \in C_{DQ_1G_2^{-1}}^1(I)$ in the “index-2” case.

A set of functions $\mathcal{M} \subset C_D^1(I)$ is called an affine linear subspace of functions $x \in C_D^1(I)$ if $\mathcal{M} = \tilde{x} + \mathcal{L}_{\mathcal{M}}$, where $\tilde{x} \in C_D^1(I)$ and $\mathcal{L}_{\mathcal{M}} \subset C_D^1(I)$ is a linear subspace.

Let us put

$$M(t) = \{v \in \mathbb{C}^m : v = x(t), x \in \mathcal{M}\}$$

and

$$L_M(t) = \{w \in \mathbb{C}^m : w = z(t), z \in \mathcal{L}_{\mathcal{M}}\}.$$

If $\dim L_M(t) \equiv \text{const} =: l$, then $\dim \mathcal{M} := \dim \mathcal{L} := l$.

The set $\mathcal{M}_{\text{ind}\mu}$ of all solutions of the DAE (1.1) is an affine linear subspace of dimension $\varrho = r + r_1 - m$ in $C_D^1(I)$. This fact follows immediately from the representation (3.4). The linear subspace $L_{M_{\text{ind}\mu}}(t) \in \mathbb{C}^m$ corresponding to the affine linear subspace $M_{\text{ind}\mu}(t)$ describes the geometric constraint to which every solution of the homogeneous equation is subjected. It reads as follows:

$$L_{M_{\text{ind}\mu}}(t) = S_{\text{ind}\mu}(t) = \text{im } \Pi_{\text{can}\mu}(t).$$

Lemma 3. *The set $\mathcal{M}_{\text{ind}\mu}$ admits an equivalent description in the form*

$$\left\{x \in C_D^1(I) : W_0 Bx = W_0 q, Hx = \mathcal{H}(q)\right\} \quad (4.1)$$

where the matrix function H is defined by the relation

$$H = D\hat{Q}_1 D^- [A^- B - (D\hat{Q}_1 D^-)' D] \quad (4.2)$$

and the linear map $\mathcal{H} : C_{DQ_1G_2^{-1}}^1(I) \rightarrow C(I)$ is given by the formula

$$\mathcal{H}(q) = D\hat{Q}_1 D^- [A^- q - (DQ_1 G_2^{-1} q)']. \quad (4.3)$$

Proof. Let us denote the set (4.1) by $\tilde{\mathcal{M}}$. Multiplying equation (1.1) by W_0 , we get the first required relation, while the second one, $Hx = \mathcal{H}(q)$, is the so-called hidden constraint derived in [2]. Thus, $\mathcal{M}_{\text{ind}\mu} \subset \tilde{\mathcal{M}}$. Now it is enough to study the kernels of W_0B and H (argument t is omitted). Instead of showing that $\dim(\ker W_0B \cap \ker H) = \varrho$, we check the intersection of kernels of W_0BK and HK with invertible matrix function K from (3.2), noting that the identities $W_0BK = W_0\hat{G}_2$ and

$$HK = D\hat{Q}_1D^-A^-\hat{G}_2(I - \hat{P}_1P_0)$$

can be verified by direct computation (we omit the details for the sake of brevity). If $W_0\hat{G}_2x = 0$, then there exist $y : y = P_0y, x = \hat{G}_2^{-1}ADy = \hat{P}_1P_0y = \hat{P}_1y$, i. e., $\hat{Q}_1x = 0$. If, additionally, $0 = HKx = D\hat{Q}_1D^-A^-\hat{G}_2(I - \hat{P}_1P_0)x$, then $0 = D\hat{Q}_1P_0y$, i. e., $\hat{G}_2y = ADy$. Therefore, $\hat{G}_2y = \hat{G}_2x$, i. e., $x = y$. Finally, $x = y = P_0y = P_0x = P_0\hat{P}_1x$. This yields $\ker H \cap \ker W_0B = \text{im } KP_0\hat{P}_1 = \text{im } \Pi_{\text{can}\mu}$. $\square$

Remark 4. Observe that $M_{\text{ind}0}(t)$ coincides with $\mathbb{C}^m$ because $W_0 = 0$ and $\hat{Q}_1 = 0$. For $\mu = 1$, W_0 is non-trivial while $\hat{Q}_1$ vanishes. For $\mu = 2$, both W_0 and $\hat{Q}_1$ are non-trivial, and the hidden constraint $H(t)z = \mathcal{H}(q)(t)$ is active.

For the purposes of the following assertions, we decompose equation (1.1) using the identity

$$I = Q_{*0}^* + \hat{Q}_{*1}^*P_{*0}^* + \hat{P}_{*1}^*P_{*0}^*. \quad (4.4)$$

We obtain

$$Q_{*0}^*Bx = Q_{*0}^*q, \quad (4.5)$$

$$\hat{Q}_{*1}^*A(Dx)' + \hat{Q}_{*1}^*P_{*0}^*Bx = \hat{Q}_{*1}^*P_{*0}^*q, \quad (4.6)$$

$$\hat{P}_{*1}^*A(Dx)' + \hat{P}_{*1}^*P_{*0}^*Bx = \hat{P}_{*1}^*P_{*0}^*q. \quad (4.7)$$

Equations (4.6) and (4.7) are equivalent to

$$A^{*-} \hat{Q}_{*1}^*A(Dx)' + A^{*-} \hat{Q}_{*1}^*P_{*0}^*Bx = A^{*-} \hat{Q}_{*1}^*P_{*0}^*q, \quad (4.8)$$

$$A^{*-} \hat{P}_{*1}^*A(Dx)' + A^{*-} \hat{P}_{*1}^*P_{*0}^*Bx = A^{*-} \hat{P}_{*1}^*P_{*0}^*q. \quad (4.9)$$

Since $\hat{Q}_{*1}^*Q_{*0}^*B = \hat{Q}_{*1}^*AD$, it follows from (4.5) that

$$A^{*-} \hat{Q}_{*1}^*ADx = A^{*-} \hat{Q}_{*1}^*Q_{*0}^*q. \quad (4.10)$$

Due to Condition C2 and Theorem 1, the projectors $A^*\hat{Q}_{*1}A^{*-}$ and $A^*\hat{P}_{*1}A^{*-}$ are differentiable functions, therefore so are the functions $A^{*-}\hat{Q}_{*1}^*A$ and $A^{*-}\hat{P}_{*1}^*A$. It immediately gives that $A^{*-}\hat{Q}_{*1}^*Q_{*0}^*q \in C^1$ is a necessary condition for a function x to be a solution. One can check, however, that

$$D\hat{Q}_1\hat{G}_2^{-1} = A^{*-}\hat{Q}_{*1}^*Q_{*0}^*. \quad (4.11)$$

Further, due to $AD = \hat{G}_2P_1P_0$ and $AD = -P_{*0}^*P_{*1}^*\hat{G}_{*2}^*$, the identities $D\hat{P}_1 = D\hat{G}_2^{-1}AD = -D\hat{G}_{*2}^*AD = A^{*-}\hat{P}_{*1}^*AD$ hold. Thus, in Theorem 2, we could use matrix functions

associated with equation (1.2), i. e., we could suppose

$$A^{*-*} \hat{Q}_{*1}^* Q_{*0}^* q \in C^1,$$

and replace (2.15) by

$$P_{*1}^*(t_0)A(t_0)D(t_0)x(t_0) = d, \quad d \in \text{im } P_{*1}^*(t_0)A(t_0)D(t_0).$$

If $x \in C_D^1$, then the first term in (4.8) rewrites as

$$A^{*-*} \hat{Q}_{*1}^* A[(A^{*-*} \hat{Q}_{*1}^* A D x)' - (A^{*-*} \hat{Q}_{*1}^* A)' D x].$$

Thus, combined with (4.10), i. e., with (4.5), equation (4.8) rewrites as

$$A^{*-*} \hat{Q}_{*1}^* A[(A^{*-*} \hat{Q}_{*1}^* Q_{*0}^* q)' - (A^{*-*} \hat{Q}_{*1}^* A)' D x] = A^{*-*} \hat{Q}_{*1}^* P_{*0}^* (q - Bx). \quad (4.12)$$

This is exactly the so-called hidden constraint $Hx = \mathcal{H}(q)$ in a different form. Indeed, since $\ker W_0 = \ker Q_{*0}^*$ and (4.5) hold, the multiplier $I - W_0$ may be inserted before the term $q - Bx$ in (4.12). On the other hand, the relation $\ker W_0 = \ker Q_{*0}^*$ involves that (4.5) and $W_0 Bx = W_0 q$ are equivalent. In fact, we checked the following statement:

Lemma 4. *A function $x \in C_D^1$ satisfies (4.5) and (4.6) if and only if $x \in \mathcal{M}_{\text{ind}\mu}$.*

Using (4.5) and (4.12), we can rewrite (4.9) as a regular ODE

$$\begin{aligned} u' - (A^{*-*} \hat{P}_{*1}^* A)' u - (A^{*-*} \hat{P}_{*1}^* A) A^{*-*} B G_{*2}^{-*} A u \\ = [(A^{*-*} \hat{P}_{*1}^* A)' D - A^{*-*} \hat{P}_{*1}^* P_{*0}^* B] G_{*2}^{-*} Q_{*0}^* q + A^{*-*} \hat{P}_{*1}^* P_{*0}^* q, \end{aligned} \quad (4.13)$$

for $u := A^{*-*} \hat{P}_{*1}^* A(Dx)$. Equation (4.13) is nothing else but the inherent ODE (2.14) in terms connected with (1.2). Indeed, the term by term coincidence can be verified by direct computation. The forms (4.12) and (4.13) show that both the hidden constraint and the inherent ODE are independent of the chosen projectors P_0, P_1 since so are P_{*0}, P_{*1} ; this assertion was proven in [2] in a different way.

Remark 5. Since $Q_{*0}^* G_{*2}^* = Q_{*0}^* B$, (4.5) defines the projection of the function x onto $\text{im } G_{*2}^{-*} Q_{*0}^* G_{*2}^*$:

$$G_{*2}^{-*} Q_{*0}^* G_{*2}^* x = G_{*2}^{-*} Q_{*0}^* q. \quad (4.14)$$

This is an equivalent of the first equation in formula (4.1) of Lemma 3. A combination of (4.14) with the second equation $Hx = \mathcal{H}(q)$ in (4.1) defines another projection of the function x . Namely, we can obtain the relations

$$G_{*2}^{-*} V^* G_{*2}^* x = G_{*2}^{-*} [\hat{Q}_{*1}^* (P_{*0}^* - Q_{*0}^*) q - \hat{Q}_{*1}^* A (A^{*-*} \hat{Q}_{*1}^* Q_{*0}^* q)'], \quad (4.15)$$

$$V = [P_{*0}^* + A^{*-*} (A^* \hat{P}_{*1}^* A^{*-*})' A^*] \hat{Q}_{*1}^*, \quad V^2 = V, \quad V Q_{*0}^* = Q_{*0}^* V = 0,$$

and, therefore, the system of equations in (4.1) becomes equivalent to system (4.14)–(4.15) which defines two projections of the function x . This observation will be used in Section 5.

Theorem 5. Let (1.1) be tractable with index μ , $\mu \in \{0, 1, 2\}$ and $q \in C_{DQ_1G_2^{-1}}^1(I)$. Then a set $\mathcal{K} \subset C_D^1(I)$ is a k -dimensional affine linear subspace of solutions of the DAE (1.1) if and only if, for all $t \in I$,

$$K(t) = \{w \in \mathbb{C}^m : y^*(t)A(t)D(t)w + v^*(t) = 0, \quad w \in M_{\text{ind}_\mu}(t)\} \quad (4.16)$$

where $y : I \rightarrow L(\mathbb{C}^s, \mathbb{C}^m)$, $\dim \text{im } y(t) \equiv s$, $v : I \rightarrow \mathbb{C}^s$, $s = \varrho - k$, and

$$-D^*(A^*y)' + B^*y = 0, \quad (4.17)$$

$$v' + q^*y = 0. \quad (4.18)$$

Theorem 5 states that any affine linear subspace within the whole solution set can be segregated by the help of functions that are solutions of the homogeneous adjoint DAE and solutions of an explicit ODE.

Proof. We denote the set on the right-hand side of (4.16) by $\tilde{K}(t)$ and provide the proof for $\mu = 2$.

Let $\mathcal{K} \subset C_D^1(I)$ be a k -dimensional affine linear subspace of solutions of the DAE (1.1) of index 2 and choose an arbitrary $x_a \in \mathcal{K}$. Let

$$\mathcal{L}_{\mathcal{K}} := \{\ell \in C_D^1(I) : \ell = x - x_a, \quad x \in \mathcal{K}\}$$

be the linear subspace of functions corresponding to $\mathcal{K}$ and $L_K(t) \in \mathbb{C}^m$, $t \in I$, $t \in I$, be the corresponding subspaces. Note that, for every t , we have $D(t)\hat{Q}_1(t)L_K(t) \equiv \{0\}$ and $\dim D(t)L_K(t) = \dim L_K(t)$. Fix $t_0 \in I$. Let

$$L_K^c(t_0) := (D(t_0)L_{\mathcal{K}}(t_0) \oplus DN_1(t_0) \oplus \ker A(t_0))^\perp.$$

One has

$$\dim L_{\mathcal{K}}^c(t_0) = m - [k + \nu + (m - r)] = r - k - \nu = s.$$

Thus, there exist s linearly independent vectors $z_1^0, \dots, z_s^0$ spanning $L_{\mathcal{K}}^c(t_0)$. Since

$$L_{\mathcal{K}}^c(t_0) \subset (DN_1(t_0) \oplus \ker A(t_0))^\perp = A^*(t_0)S_{*1}(t_0),$$

the IVPs for the homogeneous equation (1.2) ($p = 0$) with the initial conditions

$$A^*(t_0)\hat{P}_{*1}(t_0)y(t_0) = z_i, \quad i = 1, \dots, s,$$

have unique solutions y_i .

The above solutions $y_1, \dots, y_s$ of the homogeneous equation (1.2) are linearly independent. Indeed, assume the contrary, that is for the solution

$$\xi(t) = \sum_{i=1}^s c_i y_i(t)$$

with at least one non-zero c_i , the equality $\xi(\tilde{t}) = 0$ holds for some $\tilde{t}$. The IVP for homogeneous equation (1.2) with initial condition $A^*(\tilde{t})\hat{P}_{*1}(\tilde{t})y(\tilde{t}) = 0$ has the unique

solution $y = 0$. This is in contradiction to the fact that

$$A^*(t_0)\hat{P}_{*1}(t_0)\xi(t_0) = \sum_{i=1}^s c_i z_i \neq 0.$$

Set $v_i^0 = -z_i^{0*}D(t_0)x_a(t_0)$ and let $v_i : \mathcal{I} \rightarrow \mathbb{C}$ be the solution of the IVP for the ODE $v_i' + q^*y_i = 0$ satisfying $v_i(t_0) = v_i^0$.

Let $x \in \mathcal{K}$. Then

$$\begin{aligned} (y_i^*(t)A(t)D(t)x(t) + v_i^*(t))' = \\ \left[(y_i^*(t)A(t))'D(t) \right] x(t) + y_i^*(t) [A(t)(D(t)x(t))'] - y_i^*(t)q(t) = \\ y_i^*(t)B(t)x(t) + y_i^*(t)[q(t) - B(t)x(t)] - y_i^*(t)q(t) = 0. \end{aligned}$$

Thus,

$$y_i^*(t)A(t)D(t)x(t) + v_i^*(t) \equiv y_i^*(t_0)A(t_0)D(t_0)x(t_0) + v_i^*(t_0). \quad (4.19)$$

Note that

$$A^*(t_0)\hat{Q}_{*1}(t_0)y_i(t_0) = 0.$$

Thus, the expression in right-hand side of (4.19) can be brought to the form

$$y_i^*(t_0)\hat{P}_{*1}^*(t_0)A(t_0)D(t_0)x(t_0) - z_i^{0*}D(t_0)x_a(t_0) = z_i^*D(t_0)[x(t_0) - x_a(t_0)].$$

Since $x(t_0) - x_a(t_0) \in L_{\mathcal{K}}(t_0)$, the latter expression vanishes by construction.

Let $y : \mathcal{I} \rightarrow L(\mathbb{C}^s, \mathbb{C}^m)$ be defined “columnwise” by vector-valued functions y_i , $i = 1, \dots, s$, as $y(t) := (y_1(t), \dots, y_s(t))$. Clearly, y is a solution of DAE (4.3). Similarly, $v : \mathcal{I} \rightarrow \mathbb{C}^s$, $v(t) := (v_1(t), \dots, v_s(t))$ satisfies ODE (4.4).

Thus, we checked that for $x \in \mathcal{K} \subset \mathcal{M}_{\text{ind}\mu}$ and every fixed $t \in \mathcal{I}$, $w := x(t)$ belongs to the set $\tilde{K}(t)$.

For the second part, for each t let the set $\tilde{K}(t)$ be given. We have

$$y^*AD = y^*\Pi_{* \text{can} 2}^*AD = y^*AD\Pi_{\text{can} 2} = y^*AD\hat{P}_1.$$

On the other hand, by construction, the equality

$$s = \dim \text{im } y = \dim \text{im } A^*y$$

holds. Thus,

$$\dim(\ker y^*A \cap DS_1) = (m - s) - [v - (m - r)] = r - s - v = r - (\varrho - k) - v = k,$$

i. e., $\dim \tilde{K}(t) \equiv k$.

Fix a $\tilde{t} \in \mathcal{I}$. Due to the above considerations, there exists $w_0 \in \mathbb{C}^m$ such that $w_0 \in \tilde{K}(\tilde{t})$ and there exist k linearly independent vectors $w_1, \dots, w_k \in \mathbb{C}^m$ such that

$$y^*(\tilde{t})A(\tilde{t})D(\tilde{t})w_i = 0$$

and

$$w_i = \Pi_{\text{can} 2}w_i.$$

Let us consider the solution x_0 of (1.1) with initial value $x_0(\tilde{t}) = w_0$ and solutions $x_1, \dots, x_k$ of homogeneous equations (1.1) with initial values $x_i(\tilde{t}) = w_i$, $i = 1, \dots, k$, respectively. For every t , we have

$$x_0(t) + \text{span} \{x_1(t), \dots, x_k(t)\} \subset \tilde{K}(t).$$

A similar reasoning that we applied when showing the linear independence of the solutions of the homogeneous adjoint equation with linearly independent initial values yields that for each t , the affine linear set $x_0(t) + \text{span} \{x_1(t), \dots, x_k(t)\}$ is of dimension k . Thus,

$$\tilde{K}(t) = x_0(t) + \text{span} \{x_1(t), \dots, x_k(t)\}.$$

On the other hand, $x_0 + \text{span} \{x_1, \dots, x_k\}$ is an affine linear solution set as it was claimed. $\square$

5. TRANSFER OF BOUNDARY CONDITIONS

5.1. Separated boundary conditions. First let us consider the BVP for (1.1) with separated boundary condition (3.7), i. e., $K_a^* = (K_{a1}^* \mid 0)$, $K_b^* = (0 \mid K_{b2}^*)$, $d^* = (d_1^* \mid d_2^*)$, where $K_{a1} \in L(\mathbb{C}^m, \mathbb{C}^{m_a})$, $K_{b2} \in L(\mathbb{C}^m, \mathbb{C}^{m_b})$, $d_1 \in \mathbb{C}^{m_a}$, $d_2 \in \mathbb{C}^{m_b}$, and the symbol 0 stands for the zero matrix of the appropriate dimension.

Both sets of solutions defined by one and the other boundary conditions, i. e.,

$$\mathcal{K}_a := \{x \in \mathcal{M}_{\text{ind } 2} : K_{a1}x(a) = d_1\}$$

and

$$\mathcal{K}_b := \{x \in \mathcal{M}_{\text{ind } 2} : K_{b2}x(b) = d_2\}$$

are affine linear solution sets and so is $\mathcal{K} := \mathcal{K}_a \cap \mathcal{K}_b$, the solution set of BVP. Due to Remark 5, $\mathcal{K}_a$ and $\mathcal{K}_b$ admit the equivalent representation

$$\mathcal{K}_a := \{x \in \mathcal{M}_{\text{ind } 2} : \tilde{K}_{a1}x(a) = \tilde{d}_1\},$$

$$\mathcal{K}_b := \{x \in \mathcal{M}_{\text{ind } 2} : \tilde{K}_{b2}x(b) = \tilde{d}_2\},$$

where

$$\tilde{K}_{a1} = K_{a1}G_{*2}^{-*}(a)(I - Q_{*0}^*(a) - V^*(a))G_{*2}^*(a), \quad (5.1)$$

$$\tilde{K}_{b2} = K_{b2}G_{*2}^{-*}(b)(I - Q_{*0}^*(b) - V^*(b))G_{*2}^*(b), \quad (5.2)$$

$$\tilde{d}_1 = d_1 - G_{*2}^{-*}(a)[(P_{*1}^*(a)Q_{*0}^*(a) + Q_{*1}^*(a)P_{*0}^*(a))q(a) \quad (5.3)$$

$$- Q_{*1}^*(a)A(a)(A^{*-*}Q_{*1}^*Q_{*0}^*q)'(a)], \quad (5.4)$$

$$\tilde{d}_2 = d_2 - G_{*2}^{-*}(b)[(P_{*1}^*(b)Q_{*0}^*(b) + Q_{*1}^*(b)P_{*0}^*(b))q(b) \quad (5.5)$$

$$- Q_{*1}^*(b)A(b)(A^{*-*}Q_{*1}^*Q_{*0}^*q)'(b)].$$

Note that

$$(I - Q_{*0}^*(t) - V^*(t))G_{*2}^*(t) = -[I + A(t)(A^{*-*}Q_{*1}^*A)'(t)A^{*-*}(t)]P_{*1}^*(t)A(t)D(t).$$

We can always assume that the boundary conditions are given in their modified form and the matrices $\tilde{K}_{a1}, \tilde{K}_{a1}$ are of full rank. Let us set

$$y_{aa} := K_*(a)A^{*-}(a)D^{*-}(a)\tilde{K}_{a1}^*$$

and

$$y_{bb} := K_*(b)A^{*-}(b)D^{*-}(b)\tilde{K}_{b2}^*,$$

where K_* is the counterpart of K from (3.2), i. e.,

$$K_* := I - Q_{*0}\hat{P}_{*1}\hat{G}_{*2}^{-1}B^*P_{*0} - Q_{*0}\hat{Q}_{*1}A^{*-}(A^*\hat{Q}_{*1}A^{*-})'A^*$$

Clearly, the conditions $\tilde{K}_{a1}x(a) = \tilde{d}_1$ and $y_{aa}^*A(a)D(a)x(a) = \tilde{d}_1$ are identical, and the same is true for the pair of conditions $\tilde{K}_{b2}x(b) = \tilde{d}_2$ and $y_{bb}^*A(a)D(a)x(a)x(b) = \tilde{d}_2$.

Let y_a and y_b be the solutions of (4.17) with initial values $y_a(a) = y_{aa}$ and $y_b(b) = y_{bb}$. In parallel, let v_a and v_b be the solutions of (4.18) constructed with the corresponding y_a and y_b and initial values $v(a) = \tilde{d}_1$ and $v(b) = \tilde{d}_2$, respectively. Due to Theorem 5, a function $x \in C_D^1$ is a solution of BVP (1.1) if and only if, for every t , $x(t)$ satisfies the system

$$y_a^*(t)A(t)D(t)x(t) = -v_a^*(t) \quad (5.6)$$

$$y_b^*(t)A(t)D(t)x(t) = -v_b^*(t), \quad (5.7)$$

$$\begin{aligned} V^*(t)G_{*2}^*(t)x(t) &= \hat{Q}_{*1}^*(t)(P_{*0}^*(t) - Q_{*0}^*(t))q(t) \\ &\quad - \hat{Q}_{*1}^*(t)A(t)(A^{*-}\hat{Q}_{*1}^*Q_{*0}^*q)'(t), \end{aligned} \quad (5.8)$$

$$Q_{*0}^*(t)B(t)x(t) = Q_{*0}^*(t)q(t). \quad (5.9)$$

By construction, the first pair of equations is linearly independent of the second one. Let $\tau := \dim \operatorname{im} (y_a \mid y_b)$. Also by construction, the equalities

$$\dim \operatorname{im} D^*A^*(y_a \mid y_b) \equiv \dim \operatorname{im} (y_a \mid y_b) \equiv \dim \operatorname{im} (y_{aa} \mid y_{bb})$$

hold. We have proved above that $\dim M(t) \equiv r + r_1 - m$. The Fredholm alternative for (5.6)–(5.9) now yields the following statement.

Theorem 6. *For any $q \in C_{D\hat{Q}_1G_2^{-1}}^1$ and $\tilde{d}_1 \in \operatorname{im} \tilde{K}_{a1}$, $\tilde{d}_2 \in \operatorname{im} \tilde{K}_{b2}$, a unique solution x exists if and only if*

$$r + r_1 - m - \tau = 0.$$

5.2. Non-separated boundary conditions. By using Moszyński's trick [4], we transform the problem into an equivalent one with separated boundary conditions. For

$t \in [a, (a+b)/2]$, set

$$\begin{aligned}\bar{x}(t) &:= \begin{pmatrix} x(t) \\ x(b+a-t) \end{pmatrix}, & \bar{q}(t) &:= \begin{pmatrix} q(t) \\ q(b+a-t) \end{pmatrix}, \\ \bar{A}(t) &:= \text{diag}(A(t), A(b+a-t)), & \bar{D}(t) &:= \text{diag}(D(t), D(b+a-t)), \\ \bar{B}(t) &:= \text{diag}(B(t), -B(b+a-t)), \\ \bar{K}_a &= \begin{pmatrix} K_a & K_b \\ 0 & 0 \end{pmatrix}, & \bar{K}_{\frac{a+b}{2}} &= \begin{pmatrix} 0 & 0 \\ I_m & -I_m \end{pmatrix}, & \bar{d} &= \begin{pmatrix} d \\ 0 \end{pmatrix},\end{aligned}$$

where I_m is the $m \times m$ identity matrix. Obviously, the original boundary value problem is equivalent to the BVP of doubled dimension on the halved interval $[a, (a+b)/2]$ with the above data. This latter problem for $\bar{x}$ is a BVP with separated boundary conditions, and all considerations of the previous subsection apply.

6. FINAL REMARKS

Remark 6. The homogeneous IVPs for (1.2) with our initial data $y(a) = y_{aa}$ and $y(b) = y_{bb}$ in Section 5 are always solvable. Thus, integrating system (1.2) from each of the interval ends up to an arbitrary common point t_0 , one obtains $y_a(t_0)$ and $y_b(t_0)$. In parallel, the IVPs for equation (4.18) are to be solved. One should compute (preserve) the values only at points $\hat{t}$ where the solution x is needed. At these points, the other two matrices, $G_{*2}(\hat{t})V(\hat{t})$ and $B^*(\hat{t})Q_{*0}(\hat{t})$, should also be calculated. If the linear system (5.6)–(5.9) is nonsingular at an arbitrary $\hat{t} = t_0$, then so is it for all $\hat{t}$, and one can establish the solvability and uniqueness of the solution and get the solution at all $\hat{t}$.

Remark 7. To go in line with this program, a reliable integrator for (1.2) is needed and all of the other coefficients occurring in (5.6)–(5.9) must be available. It is worth noting that in this system we need only $A^*(t)y(t) = A^*(t)\hat{P}_{*1}y(t)$, i. e., the solution of the inherent ODE of the adjoint equation. One may prefer solving this homogeneous inherent ODE instead of the homogeneous DAE (1.2). Practically, there is no difference in computational complexity. A reliable integrator for any DAE would use its inherent ODE to keep the numerical solution in the corresponding subspace at least implicitly [3].

Remark 8. Theoretically, a properly discretised version of the transfer method would yield an algorithm for the numerical solution of the BVPs for (1.1). However, the resulting procedure may be very sensitive to the accumulation of numerical errors. This phenomenon may appear even when the BVP is well-conditioned and the relevant subspaces vary slowly. Therefore, a modification of the transfer algorithm seems reasonable. That modification would rely on the orthonormalisation of basis vectors of the subspaces in question at the meshpoints or it would build a smoothly varying basis on the whole interval. These issues will be reported in another publication.

Remark 9. There is no gain in the complexity if one avoids using the adjoint equation and the method relies upon any kind of shooting. In that case the computational effort is spent on keeping either a fundamental matrix (not necessarily the maximal normalised fundamental matrix) of the homogeneous DAE (1.1) in the corresponding subspace or some solutions of the DAE (1.1) in the corresponding affine subspace. To achieve this aim, one must use projectors not simpler than those in our analysis.

REFERENCES

- [1] BALLA, K.: *Linear subspaces for linear DAEs of index 1*, Computers and Mathematics with Applications, **31** (1996), No. 4/5, 81–86.
- [2] BALLA, K. AND MÄRZ, R.: *A unified approach to linear differential algebraic equations and their adjoints*, Zeitschrift für Analysis und ihre Anwendungen, **21** (2002), No. 3, 783–802.
- [3] HIGUERAS, I. AND MÄRZ, R.: *Differential-algebraic equations with properly stated leading terms*, Computers and Mathematics with Applications, to appear.
- [4] MOSZYŃSKI, K.: *A method of solving the boundary value problem for a system of linear ordinary differential equations*, Algoritmy, **11** (1964), No. 3, 25–43.
- [5] ZIELKE, G.: *Motivation und Darstellung von verallgemeinerten Matrixinversen*, Beitr. Num. Math., **7** (1979), 177–218.

Authors' Addresses

Katalin Balla:

COMPUTER AND AUTOMATION RESEARCH INSTITUTE, HUNGARIAN ACADEMY OF SCIENCES, H-1518 BUDAPEST,
P. O. BOX 63, HUNGARY
E-mail address: balla@sztaki.hu

Roswitha März:

HUMBOLDT UNIVERSITY, INSTITUTE OF MATHEMATICS, D-10099 BERLIN, GERMANY
E-mail address: iam@mathematik.hu-berlin.de



CONGRUENCES IN TRANSITIVE RELATIONAL SYSTEMS

IVAN CHAJDA

[Received: May 5, 2003]

ABSTRACT. A transitive relational system means a pair (A, R) where $A \neq \emptyset$ and R is a transitive binary relation on A . We define a congruence θ on (A, R) and a factor relation R/θ on the factor set A/θ such that the factor system $(A/\theta, R/\theta)$ is also a transitive relational system. We show that these congruences are in a one-to-one correspondence with the so-called LU-morphisms whenever the relation R is a quasiorder on A .

Mathematics Subject Classification: 08A02, 08A30

Keywords: Relational system, quasiorder, transitive relation, congruence, LU-morphism

THE CONCEPT OF A RELATIONAL SYSTEM was introduced by A. I. Maltsev [5, 6]. We will restrict our consideration to relational systems with only one binary relation. Hence, by a *relational system* we will mean a pair $\mathcal{A} = (A, R)$, where $A \neq \emptyset$ and $R \subseteq A \times A$, i. e., R is a binary relation on A . Relational systems play an important role both in mathematics and in applications since every formal description of a real system can be done by means of relations. For these considerations we often ask about a certain factorisation of a relational system $\mathcal{A} = (A, R)$ because it enables us to introduce the method of abstraction on $\mathcal{A}$. Hence, if θ is an equivalence relation on A , we ask about a 'factor relation' R/θ on the factor set A/θ such that the factor system $(A/\theta, R/\theta)$ shares some of 'good' properties of $\mathcal{A}$.

In this paper, we are mostly interested in relational systems $\mathcal{A} = (A, R)$ where R is *transitive*, i.e. $\langle a, b \rangle \in R$ and $\langle b, c \rangle \in R$ imply $\langle a, c \rangle \in R$. Then $\mathcal{A}$ is called a *transitive system*. A transitive relation formalises the concept of an "ordering" so that, in a set A , one can thus ask what elements of A go "before" or "after" a given element of A . Our topic is to define a congruence θ on $\mathcal{A}$ and a factor relation R/θ such that

- (i) the system $(A/\theta, R/\theta)$ is also transitive, and if R is reflexive or symmetrical, then R/θ shares the same properties;
- (ii) a possible common bound is preserved by our construction.

Let us note that a similar task for ordered sets was already solved in [4], and we will try to modify that construction for transitive relational systems.

A *quasiordered system* will mean a relational system $\mathcal{A} = (A, R)$ where R is a *quasiorder* on A , i. e., R is a reflexive and transitive relation. Quasiorders on a given set A form an algebraic lattice, which was studied, e. g., in [3]. Here, we are interested in quasiordered systems where elements may have common “lower” and/or “upper” bounds. The systems where every two elements of A have also suprema and infima with respect to the quasiorder R are very important in applications; they were investigated by the author in [1, 2]. However, the lower and upper bounds can be defined also for general relational systems as follows.

Notation 1. Let $\mathcal{A} = (A, R)$ be a relational system and $a, b \in A$. Introduce the following notation:

$$L_A(a, b) = \{x \in A; \langle x, a \rangle \in R \text{ and } \langle x, b \rangle \in R\},$$

$$U_A(a, b) = \{x \in A; \langle a, x \rangle \in R \text{ and } \langle b, x \rangle \in R\}.$$

If $a = b$, we will write $L_A(a)$ or $U_A(a)$ instead of $L_A(a, a)$ or $U_A(a, a)$, respectively. Clearly, if R is reflexive, then $a \in L_A(a)$ and $a \in U_A(a)$ for each $a \in A$. It is easy to prove that if R is transitive, then $\langle a, b \rangle \in R$ iff $L_A(a, b) = L_A(a)$ iff $U_A(a, b) = U_A(a)$.

Naturally, if R is transitive and $a, b \in R$, then $L_A(a, b)$ is the set of all lower bounds of a, b and $U_A(a, b)$ is the set of all upper bounds of a, b with respect to R .

If $f : A \rightarrow B$ is a mapping and $P \subseteq A$, we put $f(P) = \{f(z) : z \in P\}$.

Definition. Let $\mathcal{A} = (A, R), \mathcal{B} = (B, Q)$ be two relational systems. A surjective mapping $f : A \rightarrow B$ is called an *LU-morphism* if

$$f(L_A(x, y)) = L_B(f(x), f(y))$$

and

$$f(U_A(x, y)) = U_B(f(x), f(y)) \quad \text{for all } x, y \in A.$$

A mapping f is called a *homomorphism* of $\mathcal{A}$ into $\mathcal{B}$ if

$$\langle a, b \rangle \in R \Rightarrow \langle f(a), f(b) \rangle \in Q.$$

A homomorphism f is called *strong* if, for arbitrary $a, b \in A$, there exist $c, d \in A$ such that $f(c) = f(a)$, $f(d) = f(b)$ and $\langle f(a), f(b) \rangle \in Q \Rightarrow \langle c, d \rangle \in R$.

Lemma 1. Let $\mathcal{A} = (A, R), \mathcal{B} = (B, Q)$ be transitive relational systems and f be an LU-morphism of $\mathcal{A}$ onto $\mathcal{B}$. Then f is a homomorphism of $\mathcal{A}$ onto $\mathcal{B}$. If R is, moreover, reflexive, then f is a strong homomorphism.

Proof. Suppose $\langle a, b \rangle \in R$. Since R is transitive, it implies $L_A(a, b) = L_A(a)$ and, therefore,

$$L_B(f(a), f(b)) = f(L_A(a, b)) = f(L_A(a)) = L_B(f(a)),$$

whence $\langle f(a), f(b) \rangle \in Q$; thus, f is a homomorphism. Suppose now that R is also reflexive. If $\langle f(a), f(b) \rangle \in Q$, then

$$f(L_A(a, b)) = L_B(f(a), f(b)) = L_B(f(a)) = f(L_A(a))$$

and, on account of reflexivity, we have $a \in L_A(a)$; thus, $f(a) \in f(L_A(a)) = f(L_A(a, b))$. Analogously, one can show that $f(b) \in f(U_A(a, b))$. Hence, there exist $c \in L_A(a, b)$ and $d \in U_A(a, b)$ such that $f(c) = f(a)$, $f(d) = f(b)$. The condition $c \in L_A(a, b)$ yields $\langle c, a \rangle \in R$ and $\langle c, b \rangle \in R$, and the condition $d \in U_A(a, b)$ implies that $\langle a, d \rangle \in R$ and $\langle b, d \rangle \in R$. Using the transitivity of R , we conclude that $\langle c, d \rangle \in R$. Hence, f is a strong homomorphism. $\square$

If $f : A \rightarrow B$ is a mapping, we denote by θ_f the so-called *induced equivalence* on A , i. e., $\langle x, y \rangle \in \theta_f$ iff $f(x) = f(y)$.

We say that relational systems $\mathcal{A}, \mathcal{B}$ are *isomorphic*, in symbols $\mathcal{A} \cong \mathcal{B}$, if there exists a bijection $f : A \rightarrow B$ such that both f and f^{-1} are homomorphisms.

Theorem 1. *Let $\mathcal{A} = (A, R)$, $\mathcal{B} = (B, Q)$ be quasiordered relational systems and $f : A \rightarrow B$ a surjective mapping. The following statements are equivalent:*

- (1) *f is an LU-morphism;*
- (2) *f is a homomorphism and, for arbitrary $x, y \in A$ with $\langle f(x), f(y) \rangle \in Q$, there exist $u, v \in A$ such that $\langle v, x \rangle \in R$, $\langle x, u \rangle \in R$ and $\langle v, y \rangle \in R$, $\langle y, u \rangle \in R$ and $f(u) = f(y)$, $f(v) = f(x)$.*

Proof. The implication (1) $\Rightarrow$ (2) follows directly by the same argument as in the proof of Lemma 1.

Let us prove the implication (2) $\Rightarrow$ (1). Let f be a homomorphism of $\mathcal{A}$ onto $\mathcal{B}$. Then $f(U_A(x, y)) \subseteq U_B(f(x), f(y))$ and $f(L_A(x, y)) \subseteq L_B(f(x), f(y))$. Let us prove the converse inclusions. Suppose that $z \in U_B(f(x), f(y))$. Then $z = f(w)$ for some $w \in A$ with $\langle f(x), f(w) \rangle \in Q$, $\langle f(y), f(w) \rangle \in Q$. By (2), there exist $c, d \in A$ such that $\langle x, c \rangle \in R$, $\langle w, c \rangle \in R$ and $\langle y, d \rangle \in R$, $\langle w, d \rangle \in R$ and $f(c) = f(w) = f(d)$. Applying the reflexivity of Q , we obtain $\langle f(c), f(d) \rangle \in Q$ and, by (2), there exists $u \in A$ such that $\langle c, u \rangle \in R$, $\langle d, u \rangle \in R$ and $f(u) = f(c) = f(w) = z$. Since R is transitive, it follows that $\langle x, u \rangle \in R$, $\langle y, u \rangle \in R$, thus $u \in U_A(x, y)$, i. e., $z = f(u) \in f(U_A(x, y))$. Analogously, it can be shown that the inclusion $f(L_A(x, y)) \supseteq L_B(f(x), f(y))$ is true. $\square$

Definition. Let $\mathcal{A} = (A, R)$ be a relational system and θ be an equivalence on A . Define a binary relation R/θ on the set A/θ as follows:

$$\langle [a]_\theta, [b]_\theta \rangle \in R/\theta \text{ iff there exist } x \in [a]_\theta \text{ and } y \in [b]_\theta \text{ with } \langle x, y \rangle \in R.$$

The system $\mathcal{A}/\theta = (A/\theta, R/\theta)$ will be called a *factor system* of $\mathcal{A}$ by θ .

The following statement is obvious.

Lemma 2. *Let $\mathcal{A} = (A, R)$ and θ be an equivalence on A . If R is reflexive or symmetrical, then R/θ also has this property.*

Definition. Let $\mathcal{A} = (A, R)$ be a relational system and θ be an equivalence on A . We say that θ is a *congruence* on $\mathcal{A}$ if $\theta = R \times R$ or

- (a) for arbitrary $x, y \in [a]_\theta$, there exists a $c \in [a]_\theta$ such that $\langle x, c \rangle \in R$ and $\langle y, c \rangle \in R$;
- (b) if $\langle v, a \rangle \in R$, $\langle v, b \rangle \in R$, and $\langle v, a \rangle \in \theta$, then there exists a $t \in A$ such that $\langle a, t \rangle \in R$, $\langle b, t \rangle \in R$, and $\langle b, t \rangle \in \theta$

and the conditions (a) and (b) hold for R^{-1} .

Theorem 2. Let $\mathcal{A} = (A, R)$ be a transitive relational system and θ be a congruence on $\mathcal{A}$. Then $\mathcal{A}/\theta = (A/\theta, R/\theta)$ is also a transitive relational system.

Proof. Suppose $\langle [a]_\theta, [b]_\theta \rangle \in R/\theta$ and $\langle [b]_\theta, [c]_\theta \rangle \in R/\theta$. Then there exist $x \in [a]_\theta$, $y, y' \in [b]_\theta$, and $z \in [c]_\theta$ such that $\langle x, y \rangle \in R$ and $\langle y', z \rangle \in R$. By (a), there exists an $u \in [b]_\theta$ such that $\langle y, u \rangle \in R$ and $\langle y', u \rangle \in R$. Since R is transitive and $\langle x, y \rangle \in R$, we also have $\langle x, u \rangle \in R$. By (b), there exists a $v \in A$ such that $\langle u, v \rangle \in R$, $\langle z, v \rangle \in R$ and $\langle z, v \rangle \in \theta$, i. e., $v \in [c]_\theta$. However, $\langle x, u \rangle \in R$ and $\langle u, v \rangle \in R$ yield $\langle x, v \rangle \in R$; thus, $\langle [a]_\theta, [c]_\theta \rangle \in R/\theta$. $\square$

Theorem 3. Let $\mathcal{A} = (A, R)$, $\mathcal{B} = (B, Q)$ be quasiordered relational systems. Then:

- (1) if $f : \mathcal{A} \rightarrow \mathcal{B}$ is an LU-morphism, then θ_f is a congruence on $\mathcal{A}$ and $\mathcal{A}/\theta_f \cong \mathcal{B}$;
- (2) if θ is a congruence on $\mathcal{A}$, then the canonical mapping $h : \mathcal{A} \rightarrow \mathcal{A}/\theta$ (given by the relation $h(a) = [a]_\theta$) is an LU-morphism.

Proof. (1) Suppose that $x, y \in [a]_{\theta_f}$. Then $f(x) = f(y)$ and, in view of the reflexivity of Q , we have $\langle f(x), f(y) \rangle \in Q$. By Theorem 1, there exists an $u \in A$ with $\langle x, u \rangle \in R$, $\langle y, u \rangle \in R$ and $f(x) = f(u) = f(y)$. Hence, $u \in [a]_{\theta_f}$. Analogously, one can show the existence of $v \in [a]_{\theta_f}$ with $\langle v, x \rangle \in R$, $\langle v, y \rangle \in R$, i. e., $[a]_{\theta_f}$ satisfies (a) and its dual (i. e., it is “directed”).

Let us prove (b). Let $\langle v, a \rangle \in R$, $\langle v, b \rangle \in R$ and $\langle v, a \rangle \in \theta_f$. Then $f(v) = f(a)$ and, therefore, $f(U_A(a, b)) = U_B(f(a), f(b)) = U_B(f(v), f(b)) = U_B(f(b)) = f(U_A(b))$. Hence, there exists a $t \in A$ such that $t \in U_A(a, b)$ and $f(t) = f(b)$, whence $\langle b, t \rangle \in \theta_f$ and $\langle a, t \rangle \in R$, $\langle b, t \rangle \in R$. We have thus shown that (b) holds. Analogously, the dual of (b) can be obtained.

(2) Suppose that $a, b \in A$ and $\langle a, b \rangle \in R$. Since $a \in [a]_\theta$, $b \in [b]_\theta$, we have $\langle h(a), h(b) \rangle = \langle [a]_\theta, [b]_\theta \rangle \in R/\theta$, i. e., h (the canonical mapping) is a surjective homomorphism. Let $x, y \in A$ and $\langle h(x), h(y) \rangle \in Q$. Then $\langle [x]_\theta, [y]_\theta \rangle \in R/\theta$; thus, there exist $c \in [x]_\theta$, $d \in [y]_\theta$ with $\langle c, d \rangle \in R$. By (a), there exists a $v \in A$ with $\langle v, x \rangle \in R$, $\langle v, c \rangle \in R$ and $v \in [x]_\theta$, and there exists $t \in A$ with $\langle d, t \rangle \in R$, $\langle y, t \rangle \in R$ and $t \in [y]_\theta$. By (b), there is an $u \in A$ such that $\langle t, u \rangle \in R$, $\langle x, u \rangle \in R$ and $\langle u, t \rangle \in \theta$. On account of the transitivity of R , we also have $\langle x, u \rangle \in R$, $\langle y, u \rangle \in R$, and $u \in [y]_\theta$, i. e., $h(u) = h(y)$. Analogously, there is an $s \in A$ such that $\langle s, x \rangle \in R$, $\langle s, y \rangle \in R$, and $h(s) = h(x)$. By Theorem 1, h is an LU-morphism. $\square$

Theorem 4. Let $\mathcal{A} = (A, R)$ be a quasiordered system and θ be an equivalence on A . Then θ is a congruence on $\mathcal{A}$ if and only if the following assertion is true: for every $a \in A$, $[a]_\theta$ is directed and

- (i) $\langle a, b \rangle \in R, \langle a, a_1 \rangle \in \theta \Rightarrow \exists b_1 \in A$ with $\langle a_1, b_1 \rangle \in R$ and $\langle b_1, b \rangle \in \theta$;
- (ii) $\langle a, b \rangle \in R, \langle b, b_1 \rangle \in \theta \Rightarrow \exists a_1 \in A$ with $\langle a_1, b_1 \rangle \in R$ and $\langle a_1, a \rangle \in \theta$.

Proof. (1) Suppose that $\langle a, b \rangle \in R$ and $\langle a, a_1 \rangle \in \theta$ for some $a, a_1, b \in A$. By (a), there exists $d \in [a]_\theta$ with $\langle d, a_1 \rangle \in R, \langle d, a \rangle \in R$ and, due to the transitivity, $\langle d, b \rangle \in R$. By (b), there exists $b_1 \in [b]_\theta$ such that $\langle a_1, b_1 \rangle \in R$. We have obtained (i). Analogously, it can be shown that (ii) is true.

(2) Let θ be an equivalence on A satisfying (i) and (ii). Clearly, (i) + (ii) yields property (b). $\square$

Corollary. Let $\mathcal{A} = (A, R)$ be a quasiordered system and θ be an equivalence on A . Then θ is a congruence on $\mathcal{A}$ if and only if:

- (i) R/θ is a quasiorder on A/θ ;
- (ii) $[L_A(x, y)]_\theta = L_{A/\theta}([x]_\theta, [y]_\theta)$ and $[U_A(x, y)]_\theta = U_{A/\theta}([x]_\theta, [y]_\theta)$ for arbitrary $x, y \in A$.

Proof. If θ is a congruence on $\mathcal{A}$, then by Theorem 2 and Lemma 2, we obtain (i). Applying Theorem 3, we have (ii). Conversely, let θ be an equivalence on A satisfying (i) and (ii). Then the canonical mapping $h : A \rightarrow A/\theta$ is an LU-morphism and, due to Theorem 3, we have $\theta = \theta_h$ is a congruence on $\mathcal{A}$. $\square$

REFERENCES

- [1] CHAJDA, I.: *Lattices in quasiordered sets*, Acta. Univ. Palack. Olom., Fac. Rer. Nat., Mathematica, **31** (1992), 6–12.
- [2] CHAJDA, I.: *An algebra of quasiordered logic*, Mathem. Bohemica, **119** (1994), 129–135.
- [3] CHAJDA, I. AND CZÉDLI, G.: *How to generate the involution lattice of quasiorders?*, Stud. Sci. Math. Hungar., **32** (1996), 415–427.
- [4] CHAJDA, I. AND SNÁŠEL, V.: *Congruences in ordered sets*, Czech. Math. Journal, **123** (1998), 95–100.
- [5] MALTSEV, A. I.: *Toward a general theory of algebraic systems*, Matem. Sb., **33** (1954), 3–20.
- [6] MALTSEV, A. I.: *Algebraic systems*, Springer-Verlag, Berlin, 1973.

Author's Address

Ivan Chajda:

DEPARTMENT OF ALGEBRA AND GEOMETRY, PALACKÝ UNIVERSITY OLOMOUČ, TOMKOVA 40, 779 00 OLOMOUČ, CZECH REPUBLIC

E-mail address: Chajda@risc.upol.cz



AUTOMATIC ERROR ANALYSIS WITH MILLER'S METHOD

ATTILA GÁTI

[Received: October 09, 2003]

ABSTRACT. The first considerable roundoff analyser software, which can provide global error bounds for a given numerical algorithm, was designed by Webb Miller and David Spooner in the second half of the 70s. Despite its abilities, the recently software is not widely used, and since its publication it has not been upgraded. The software was written in Fortran IV for the IBM 360/370 series. Without any modifications, it is not usable in a PC environment with the most widely used compilers based on the Fortran 77 standard (GNU compiler, Watcom, etc.), because there are several non-standard solutions in the source. In addition, the user is bound by extraordinary strict limitations of the size of the algorithm to analyse. We have upgraded the package in order to make it suit the Fortran 77 standard, and to expand the size limits. With the upgraded version we have analysed two variants of the ABS method: the implicit LU and Huang methods.

Mathematics Subject Classification: Primary 12G34; Secondary 09G87

Keywords: Roundoff analysis, automatic roundoff analysis, ABS method, implicit LU method, Huang method

1. INTRODUCTION

ON DIGITAL COMPUTERS, the arithmetic operations are performed in finite precision. The analyses of propagation of the rounding errors in numerical algorithms, i. e., the numerical stability, is one of the major tasks of numerical analysis. Basically, we can distinguish three different approaches to decide whether an algorithm is stable or unstable. The most obvious technique is testing the numerical algorithm upon sets of data for which the exact, or a relatively accurate, solution is known. Another approach is formal roundoff analysis. The best known result of that type is Wilkinson's theorem. According to it, the $\hat{x}$ computed solution of the $n \times n$ system of linear equations $Ax = b$ found by the Gaussian elimination satisfies the relation

$$(A + \Delta A) \hat{x} = b,$$

where $\|\Delta A\|_\infty \leq 8n^3 \varrho_n \|A\|_\infty u + O(u^2)$ (ϱ_n is the growth factor of the pivot entries, and u is the machine rounding unit).

Both of these two approaches have defects. Testing is sometimes insufficient, and the formal analysis needs a great deal of effort, which cannot always be afforded. Because of this, even at the dawn of scientific computing already techniques appeared in which computer is used to help us analyse the stability of a numerical computation. So the third approach of analysing the effects of finite precision arithmetic is the automatic roundoff analysis. Miller makes a distinction between local and global techniques of automatic roundoff analysis [1]. Local methods are used to bound, or estimate the error incurred in a single computation. Most local techniques use special systems of computer arithmetic (e.g. interval arithmetic, or unnormalised arithmetic) to monitor the error in each computed value. Using global techniques we can determine the propagation of rounding errors in a given numerical method for all permissible sets of data. In the beginning only local techniques were developed, the first considerable roundoff analyser software, which can provide global error bounds for a given numerical algorithm, was designed by Webb Miller and David Spooner in the second half of the 1970s.

2. THE MILLER-SPOONER ROUNDOFF ERROR ANALYSER SOFTWARE

The software can be found in the ACM TOMS library with serial number 532, and the Fortran language source code is available freely at the following location:

<http://www.netlib.org/toms/532>

Miller and his co-workers published the use and the theoretical background of the program in two articles [1, 2] and also in a book [3]. The software package is intended to help algorithm designers seeking numerically stable algorithms, or a user doing initial testing of a numerical algorithm proposed for use.

The software cannot deal with every kind of numerical algorithm. Instead, direct methods solving algebraic problems are most susceptible for analysis. The numerical method to be analysed must be expressed in a special Fortran-like language. The language allows for-loops and if-tests that are not based on the values of real variables. If an algorithm contains branching based on real variables, then the possible paths through comparisons must be treated separately. The parameters of the algorithm defining the size of the problem (sizes of vectors and matrices) must be fixed for analysis, and these parameters cannot be arbitrarily large due to memory limitations.

The method is based on the standard model of the floating point arithmetic, so we assume that the relative error of each arithmetic operation is bounded by the machine rounding unit, and we ignore the possibilities of overflow and underflow.

The basic idea of Miller's method is very simple. Given a numerical method to analyse, a number $\omega(d)$ is associated with each set d ($d \in \mathbb{R}^n$) of data. The function ω measures rounding error, i. e., $\omega(d)$ is large exactly when the method applied to d produces results which are excessively sensitive to rounding errors. A numerical maximiser is applied to search for large values of ω to provide information about the numerical properties of the method. Finding large values of ω can be interpreted to

mean that the given numerical method suffers from a specific kind of instability. With the software package we can also compare the numerical stability of two algorithms Q and R , which neglecting rounding errors, compute identical values. In this case ω is large exactly when Q produces a much larger error than R .

The computation of ω is based on the partial derivatives at the nodes of the program's computational graph. As we use derivatives, the second order effects of rounding errors are neglected. In general ω is not a differentiable function of data d , so a direct search method is used for maximising ω . As these methods are heuristic, failure of the maximiser to find large values of ω does not guarantee that none exists, so unstable methods may appear to be stable.

The software package consists of three programs, a minicompiler and two error analyser programs. The minicompiler takes as data the numerical method to analyse, written in a special programming language, and produces as output a translation of that program into a series of assignment statements. The output is presented in a readable form for the user, and in a coded form for use as input to the error analyser programs. One of the error analyser programs (i) is for deciding numerical stability of a single algorithm, and the other (ii) is for comparing two competing algorithms. Both (i) and (ii) form automatically the partial derivatives, and based on them, can compute various error-measuring numbers (various ways to assign ω).

First let us consider the error measuring numbers of program (i). The software provides four stability-measuring numbers, $JW_E(d)$, $JW_L(d)$, $WK_E(d)$, $WK_L(d)$, which measure the minimum problem perturbation equivalent to rounding error in a given program at data d . Other numbers $ER_E(d)$ and $ER_L(d)$ use a weaker comparison of computed and exact solutions. Let $Z : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a vector-valued function, and R an algorithm with k arithmetic operations for evaluating Z . If the operations are performed in floating point arithmetic, with the assumption of the standard model, then the computed value is a function $R(d, \delta)$, where δ is the vector of individual relative rounding errors on the k arithmetic operations ($\delta \in \mathbb{R}^k$, $\|\delta\|_\infty \leq u$).

$JW_E(d)$ is the smallest number for which the equality

$$R(d, \delta) = Z(d + \pi)$$

holds, for some $\pi \in \mathbb{R}^n$, $|\pi_i| \leq JW_E(d) \cdot |d_i| \cdot u$, $1 \leq i \leq n$, whenever $\|\delta\|_\infty \leq u$. So, $JW_E(d)$ measures the componentwise backward error. The normwise backward error is measured by $JW_L(d)$, in this case $\|\pi\|_\infty \leq JW_L(d) \cdot \|d\|_\infty \cdot u$. WK_X measures ($X = E$, or L) the mixed forward-backward error, so we also allow perturbations in the output space. $WK_E(d)$ is the smallest number for which the equality

$$R(d, \delta) = Z(d + \pi) + \phi$$

holds for some $|\pi_i| \leq WK_E(d) \cdot |d_i| \cdot u$, and $\phi \in \mathbb{R}^m$, $|\phi_i| \leq WK_E(d) \cdot |Z_i(d)| \cdot u$, whenever $\|\delta\|_\infty \leq u$. In the case of the normwise measuring number $WK_L(d)$, we have

$$\|\pi\|_\infty \leq WK_L(d) \cdot \|d\|_\infty \cdot u$$

and

$$\|\phi\|_\infty \leq WK_L(d) \cdot \|Z(d)\|_\infty \cdot u.$$

Note that

$$WK_L(d) \leq WK_E(d) \leq JW_E(d)$$

and

$$WK_L(d) \leq JW_L(d) \leq JW_E(d).$$

The numbers JW_X and WK_X require that the computed solution be obtained at (or close to) the exact solution. ER_X measures how much the data must be perturbed to create an error as large as that in the computed solution. So $ER_E(d)$ is the smallest number for which

$$\|R(d, \delta) - Z(d)\|_\infty = \|Z(d + \pi) - Z(d)\|_\infty$$

for some π , where $|\pi_i| \leq ER_E(d) \cdot |d_i| \cdot u$, $1 \leq i \leq n$. ER_L measures normwise, in this case the same equation holds for π , $\|\pi\|_\infty \leq ER_L(d) \cdot \|d\|_\infty \cdot u$. The inequalities

$$ER_L(d) \leq ER_E(d)$$

and

$$ER_X(d) \leq JW_X(d),$$

where X is either E or L , are obvious.

Now let us take a glance at the error-measuring numbers used by program (ii). Let Q and R be two algorithms for evaluating the function Z , and the computed values at data d , with rounding errors $\delta \in \mathbb{R}^k$ and $\gamma \in \mathbb{R}^l$ be $Q(d, \delta)$ and $R(d, \gamma)$. $JW_{Q/R}(d)$ is the smallest number for which the condition

$$Q(d, \delta) = R(d, \gamma)$$

holds whenever $\|\delta\|_\infty \leq u$, for some $\|\gamma\|_\infty \leq JW_{Q/R}(d) \cdot u$. $ER_{Q/R}(d)$ is the smallest number for which

$$\|Q(d, \delta) - Z(d)\|_\infty = \|R(d, \gamma) - Z(d)\|_\infty$$

for all $\|\delta\|_\infty \leq u$, for some $\|\gamma\|_\infty \leq ER_{Q/R}(d) \cdot u$. The inequalities

$$ER_{Q/R}(d) \leq JW_{Q/R}(d)$$

and

$$\frac{JW_X^Q}{JW_X^R} \leq JW_{Q/R}(d),$$

where $X = E$ or $X = L$, will also hold.

The input for the error-testing programs is arranged as follows: (1) the output of the minicompiler (for program (i), which is the compiled version of the algorithm to analyse; in the case of program (ii) we must provide the compiled version of Q followed by R), (2) a list of initial data for the numerical maximiser, (3) the code of the chosen error-measuring value, (4) target value for the maximiser. The programs return with the final set of data found by the maximiser routine and with the value

of the chosen error-measuring number at this set of data. If program (i) diagnoses instability (the value of the error-measuring number at the final set of data exceeds the target value), then the condition number is also computed at the final set of data.

3. UPGRADING THE SOFTWARE PACKAGE

The Miller–Spooner roundoff error analyser software was written in Fortran IV for the IBM 360/370 series, and its final version was issued in 1979. Without any modifications, it is not usable in a PC environment with the most widely used compilers based on the Fortran 77 standard (GNU compiler, Watcom, etc.), because there are several non-standard solutions in the source. In addition, the user is bound to extraordinarily strict limitations of the size of the algorithm to analyse. The number of inputs cannot exceed 30, the number of outputs must be smaller than 20, and the total number of inputs and real arithmetic operations may not exceed 300. We have upgraded the package in order to make it suit the Fortran 77 standard, and to expand the size limits. The upgraded version works correctly with the compilers mentioned earlier, and we can analyse algorithms with inputs up to 3000, outputs up to 1000, and the total number of inputs and operations can be maximum 50000.

In [3], the abilities of the software are demonstrated by 14 important case studies. In these cases the answers of the software are consistent with the known formal analytical results. The program shows correctly the stability properties of the Gaussian elimination without pivoting and with partial pivoting, the Gauss–Jordan elimination, the Cholesky factorisation, several QR factorisation methods (Householder transformation, Givens transformation, method by solving the normal equations), the classical and modified Gram–Schmidt methods and so on. We have tested the upgraded version on these case studies, and with small dimensions we have obtained results similar to those of Miller's, with larger dimensions the results were consistent.

4. ANALYSIS OF THE IMPLICIT LU AND HUANG METHODS

Using the software package we have analysed the numerical stability of solving the $Ax = b$ linear system ($A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$) with the implicit LU and Huang methods. These methods are special variants of the ABS method. The ABS method:

Algorithm 1. $x_1 \in \mathbb{R}^n$, $H_1 = I$, $V = [v_1, \dots, v_n] \in \mathbb{R}^{n \times n}$,
 $W = [w_1, \dots, w_n] \in \mathbb{R}^{n \times n}$, $Z = [z_1, \dots, z_n] \in \mathbb{R}^{n \times n}$
 for $k = 1, \dots, n$
 $p_k = H_k^T z_k \quad (z_k \in \mathbb{R}^n, p_k^T A^T v_k \neq 0)$
 $x_{k+1} = x_k - \frac{p_k v_k^T (Ax_k - b)}{v_k^T A p_k}$
 $H_{k+1} = H_k - \frac{H_k A^T v_k w_k^T H_k}{w_k^T H_k A^T v_k} \quad (w_k \in \mathbb{R}^n, w_k^T H_k A^T v_k \neq 0)$
 end
 $x_{n+1} = A^{-1}b$

4.1. The implicit LU method. The implicit LU algorithm is given by the choice $W = Z = I$. We consider the special case, when $V = I$ and $x_1 = 0$. So the algorithm to analyse is the following one:

Algorithm 2. $x_1 = 0, H_1 = I$

for $k = 1, \dots, n$

$$x_{k+1} = x_k - \frac{H_k^T e_k e_k^T (Ax_k - b)}{e_k^T A H_k^T e_k}$$

$$H_{k+1} = H_k - \frac{H_k A^T e_k e_k^T H_k}{e_k^T H_k A^T e_k}$$

end

Let the initial data for maximising be:

$$A_0 = \begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 4 & 1 & 1 \\ 1 & 1 & 5 & 1 \\ 1 & 1 & 1 & 6 \end{bmatrix}, \quad b_0 = \begin{bmatrix} 6 \\ 7 \\ 8 \\ 9 \end{bmatrix}.$$

We have chosen 10,000 as target value. The maximiser finds values greater than the target value for both WK_L and ER_L error-measuring numbers. Because of the inequalities held between the error-measuring numbers, a set of data exists for all error-measuring numbers, where their values exceed 10,000. In addition, at the final set of data the condition number is extremely low. In the case of WK_L it is 12.3, and in the case of ER_L it is 8.1. So the implicit LU algorithm is unstable even at very well-conditioned data.

Comparing the algorithm with the Gaussian elimination without pivoting, we obtain

$$JW_{LU/Gauss} > 10,000, \quad JW_{Gauss/LU} > 10,000, \quad ER_{LU/Gauss} > 10,000,$$

and

$$ER_{Gauss/LU} < 29.$$

So, the implicit LU has worse stability properties than the Gaussian elimination without pivoting.

4.2. The Huang method. The algorithm is obtained by the choice

$$w_i = z_i = A^T v_i.$$

We have analysed four variants of the method. With the given substitutions applied to the general ABS algorithm, we get the first variant:

Algorithm 3. $x_1 \in \mathbb{R}^n, H_1 = I$

for $k = 1, \dots, n$

$$p_k = H_k^T A^T e_k$$

$$x_{k+1} = x_k - \frac{p_k e_k^T (Ax_k - b)}{e_k^T A p_k}$$

$$H_{k+1} = H_k - \frac{H_k A^T e_k e_k^T A H_k}{e_k^T A H_k A^T e_k}$$

end

We have started the maximiser from the same initial data A_0, b_0 as in the case of the implicit LU method. Applying the maximiser, ER_L and WK_L have exceeded 10,000, so the method is unstable. Unlike the case of implicit LU, large values were found at ill-conditioned data (61,710 for WK_L and 32,365 for ER_L). So, the algorithm is more sensitive to ill-conditioning than it is reasonable. Comparing the algorithm with the modified Gram-Schmidt method, we get the following results:

$$JW_{\text{Huang}/\text{MGS}} > 10,000, \quad JW_{\text{MGS}/\text{Huang}} < 3,400, \quad ER_{\text{Huang}/\text{MGS}} > 10,000,$$

and $ER_{\text{MGS}/\text{Huang}} < 2050$. So, they are both sets of data where the Huang and sets of data where the MGS method give much poorer results.

If we consider that $H_i = H_i^T$, then the algorithm can be modified as follows:

Algorithm 4. $x_1 \in \mathbb{R}^n, H_1 = I$

for $k = 1, \dots, n$

$$p_k = H_k^T A^T e_k$$

$$x_{k+1} = x_k - \frac{p_k e_k^T (Ax_k - b)}{e_k^T A p_k}$$

$$H_{k+1} = H_k - \frac{p_k p_k^T}{e_k^T A p_k}$$

end

In this case, we have results very similar to the previous algorithm. Comparing the two methods (Algorithm 3 and Algorithm 4), we see that all the error-measuring numbers are bounded by 6.1. Thus, the two variants have the same stability properties.

The following version is based on the fact that $H_i^2 = H_i$:

Algorithm 5. $x_1 \in \mathbb{R}^n, H_1 = I$

for $k = 1, \dots, n$

$$p_k = H_k^T A^T e_k$$

$$x_{k+1} = x_k - \frac{p_k e_k^T (Ax_k - b)}{e_k^T A p_k}$$

$$H_{k+1} = H_k - \frac{p_k p_k^T}{p_k^T p_k}$$

end

We have values for ER_L and WK_L exceeding 10,000 also at ill-conditioned sets of data. Comparing the algorithm with the MGS method we get results similar to the previous two cases. If we compare the method with the second variant (Algorithm 4), we get that all error-measuring numbers are not bounded by 10,000 except that

$$ER_{\text{Huang } 3/\text{Huang } 2} < 1.5.$$

So, the third variant is more stable than the previous ones.

The fourth analysed algorithm was the so-called modified Huang method:

Algorithm 6. $x_1 \in \mathbb{R}^n$, $H_1 = I$

for $k = 1, \dots, n$

$$p_k = H_k (H_k^T A^T e_k)$$

$$x_{k+1} = x_k - \frac{p_k e_k^T (Ax_k - b)}{e_k^T A p_k}$$

$$H_{k+1} = H_k - \frac{p_k p_k^T}{p_k^T p_k}$$

end

In this case the maximiser cannot find a value for JW_L greater than 4.3, so the algorithm is backward stable. A comparison with the MGS method produces the following results:

$$JW_{\text{Huang/MGS}} < 561, \quad JW_{\text{MGS/Huang}} > 10,000, \quad ER_{\text{Huang/MGS}} < 20,$$

and

$$ER_{\text{MGS/Huang}} < 4,038.$$

So, the modified Huang method has better stability properties than the MGS method. If we compare the method with the previous (third) variant, we have that

$$ER_{\text{Huang 4/Huang 3}} < 1.23, \quad JW_{\text{Huang 4/Huang 3}} < 1.6,$$

and the other two numbers are not bounded by 10,000. So, the method is much more stable than the previous variants of the Huang method.

REFERENCES

- [1] MILLER, W.: *Software for roundoff analysis*, ACM Trans. Math. Software, **1** (1975), No. 2, 108–128.
- [2] MILLER, W. AND SPOONER, D.: *Software for roundoff analysis II*, ACM Trans. Math. Software, **4** (1978), No. 4, 388–390.
- [3] MILLER, W. AND WRATHALL, C.: *Software for Roundoff Analysis of Matrix Algorithms*, Academic Press, New York, 1980.
- [4] ABAFFY, J. AND SPEDICATO, E.: *ABS Projection Algorithms: Mathematical Techniques for Linear and Nonlinear Equations*, Ellis Horwood Limited, Chichester, 1989.

Author's Address

Attila Gáti:

INSTITUTE OF MATHEMATICS, UNIVERSITY OF MISKOLC, 3515 MISKOLC-EGYETEMVÁROS, HUNGARY

E-mail address: matgati@gold.uni-miskolc.hu



ON PERIODIC-TYPE BOUNDARY VALUE PROBLEMS FOR FUNCTIONAL DIFFERENTIAL EQUATIONS WITH A POSITIVELY HOMOGENEOUS OPERATOR

ROBERT HAKL

[Received: September 10, 2003]

ABSTRACT. Consider the problem

$$u'(t) = H(u)(t) + Q(u)(t), \quad u(a) - \lambda u(b) = h(u),$$

where $H, Q : C([a, b]; \mathbb{R}) \rightarrow L([a, b]; \mathbb{R})$ are continuous operators satisfying the Carathéodory condition, the operator H is positively homogeneous, $\lambda \in \mathbb{R}_+$, and $h : C([a, b]; \mathbb{R}) \rightarrow \mathbb{R}$ is a continuous functional. In this paper, efficient sufficient conditions guaranteeing the solvability and unique solvability of the problem considered are established.

Mathematics Subject Classification: 34K06, 34K10

Keywords: Functional differential equation, boundary value problem, periodic solution

1. INTRODUCTION

IN THIS PAPER, we establish efficient sufficient conditions guaranteeing the existence and uniqueness of a solution of a periodic-type boundary value problem for a scalar functional differential equation involving a positively homogeneous operator.

The following notation is used throughout the paper.

$\mathbb{R}$ is the set of all real numbers, $\mathbb{R}_+ = [0, +\infty[$, $[x]_+ = \frac{1}{2}(|x| + x)$, $[x]_- = \frac{1}{2}(|x| - x)$.

$C([a, b]; \mathbb{R})$ is the Banach space of continuous functions $u : [a, b] \rightarrow \mathbb{R}$ with the norm $\|u\|_C = \max\{|u(t)| : t \in [a, b]\}$.

$\tilde{C}([a, b]; \mathbb{R})$ is the set of absolutely continuous functions $u : [a, b] \rightarrow \mathbb{R}$.

$L([a, b]; \mathbb{R})$ is the Banach space constituted by the Lebesgue integrable functions $p : [a, b] \rightarrow \mathbb{R}$ with the norm $\|p\|_L = \int_a^b |p(s)| ds$.

$L([a, b]; \mathbb{R}_+) = \{p \in L([a, b]; \mathbb{R}) : p(t) \geq 0 \text{ for } t \in [a, b]\}$.

$\mathcal{M}_{ab}$ is the set of measurable functions $\tau : [a, b] \rightarrow [a, b]$.

$\mathcal{K}_{ab}$ is the set of continuous operators $F : C([a, b]; \mathbb{R}) \rightarrow L([a, b]; \mathbb{R})$ satisfying the Carathéodory condition, i. e., for every $r > 0$ there exists $q_r \in L([a, b]; \mathbb{R}_+)$ such

that

$$|F(v)(t)| \leq q_r(t) \quad \text{for } t \in [a, b], \quad v \in C([a, b]; \mathbb{R}), \quad \|v\|_C \leq r.$$

$\mathcal{H}_{ab}$ is the set of operators $H \in \mathcal{K}_{ab}$ which are positively homogeneous, i. e., such that, for an arbitrary number $\alpha > 0$, we have

$$H(\alpha v)(t) = \alpha H(v)(t) \quad \text{for } t \in [a, b], \quad v \in C([a, b]; \mathbb{R}).$$

$\mathcal{P}_{ab}$ is the set of operators $H \in \mathcal{H}_{ab}$ which are nondecreasing, i. e., for all $H \in \mathcal{P}_{ab}$ and $u, v \in C([a, b]; \mathbb{R})$ such that $u(t) \leq v(t)$ for $t \in [a, b]$, we have

$$H(u)(t) \leq H(v)(t) \quad \text{for } t \in [a, b].$$

$K([a, b] \times \mathbb{R}_+; \mathbb{R}_+)$ is the set of functions $\eta : [a, b] \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfying the Carathéodory conditions, i. e., $\eta(\cdot, x) : [a, b] \rightarrow \mathbb{R}_+$ is a measurable function for all $x \in \mathbb{R}_+$, $\eta(t, \cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a continuous function for almost all $t \in [a, b]$, and for every $r > 0$ there exists $q_r \in L([a, b]; \mathbb{R}_+)$ such that

$$|\eta(t, x)| \leq q_r(t) \quad \text{for } t \in [a, b], \quad 0 \leq x \leq r.$$

By a solution of the equation

$$u'(t) = H(u)(t) + Q(u)(t), \quad (1.1)$$

where $H \in \mathcal{H}_{ab}$ and $Q \in \mathcal{K}_{ab}$, we understand a function $u \in \tilde{C}([a, b]; \mathbb{R})$ satisfying equality (1.1) almost everywhere in $[a, b]$.

Consider the problem on the existence and uniqueness of a solution of (1.1) satisfying the condition

$$u(a) - \lambda u(b) = h(u), \quad (1.2)$$

where $h : C([a, b]; \mathbb{R}) \rightarrow \mathbb{R}$ is a continuous operator such that for every $r > 0$ there exists $M_r \in \mathbb{R}_+$ such that

$$|h(v)| \leq M_r \quad \text{for } v \in C([a, b]; \mathbb{R}), \quad \|v\|_C \leq r.$$

The differential equations with “maxima” are one of the special cases of equation (1.1). Various types of boundary value problems for this type of equations were studied, e. g., in [3, 4, 33, 35, 36, 38, 41].

In the case where H in (1.1) is a linear operator, the results presented here coincide with those obtained earlier in [6, 17] (see also [22]). Another type of conditions guaranteeing the solvability of (1.1), (1.2) with a linear operator H can be found, e. g., in [5, 8, 10–15, 18–21, 28, 34]. Conditions for the solvability and unique solvability for other types of boundary value problems for equation (1.1) with a linear operator H are established, e. g., in [9, 16, 22, 24, 27, 39, 42].

There are many interesting results concerning the solvability of general boundary value problems for functional differential equations (see, e. g., [1, 2, 7, 9, 22–26, 29–32, 37, 40] and references therein). In spite of this, the general theory of boundary value problems for functional differential equations is not still complete. Here, we try to fill this gap in a certain way. More precisely, in Section 2, we establish unimprovable,

in a certain sense, efficient conditions sufficient for the solvability of problem (1.1), (1.2). In Section 3, we prove some auxiliary propositions. Section 4 is devoted to the proofs of the main results.

The results are concretized for the boundary value problem of the form

$$u'(t) = p(t) \max \{u(s) : \tau_1(t) \leq s \leq \tau_2(t)\} + q(t), \quad (1.3)$$

$$u(a) - \lambda u(b) = c, \quad (1.4)$$

where $p, q \in L([a, b]; \mathbb{R})$, $\tau_1, \tau_2 \in \mathcal{M}_{ab}$, $\tau_1(t) \leq \tau_2(t)$ for $t \in [a, b]$, and $c \in \mathbb{R}$.

Definition 1.1. We will say that an operator $H \in \mathcal{K}_{ab}$ is subadditive if for any $u, v \in C([a, b]; \mathbb{R})$ we have

$$H(u + v)(t) \leq H(u)(t) + H(v)(t) \quad \text{for } t \in [a, b].$$

Definition 1.2. We will say that an operator $H \in \mathcal{K}_{ab}$ is superadditive if for any $u, v \in C([a, b]; \mathbb{R})$ we have

$$H(u + v)(t) \geq H(u)(t) + H(v)(t) \quad \text{for } t \in [a, b].$$

2. MAIN RESULTS

Throughout the paper, $\eta \in K([a, b] \times \mathbb{R}_+; \mathbb{R}_+)$ and $\eta_0 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ are such that

$$\lim_{x \rightarrow +\infty} \left(\frac{\eta_0(x)}{x} + \frac{1}{x} \int_a^b \eta(s, x) ds \right) = 0.$$

Theorem 2.1. Let $\lambda \in]0, 1]$ and let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore,

$$|Q(v)(t)| \leq \eta(t, \|v\|_C) \quad \text{for } t \in [a, b], \quad v \in C([a, b]; \mathbb{R}), \quad (2.1)$$

$$|h(v)| \leq \eta_0(\|v\|_C) \quad \text{for } v \in C([a, b]; \mathbb{R}). \quad (2.2)$$

If, moreover,

$$\int_a^b H_0(1)(s) ds < 1, \quad \int_a^b |H_0(-1)(s)| ds < 1, \quad (2.3)$$

$$\frac{\int_a^b H_0(1)(s) ds}{1 - \int_a^b H_0(1)(s) ds} - \frac{1 - \lambda}{\lambda} < \int_a^b H_1(1)(s) ds, \quad (2.4)$$

$$\frac{\int_a^b |H_0(-1)(s)| ds}{1 - \int_a^b |H_0(-1)(s)| ds} - \frac{1 - \lambda}{\lambda} < \int_a^b |H_1(-1)(s)| ds, \quad (2.5)$$

$$\begin{aligned}
\left[\int_s^t H_1(1)(\xi) d\xi - 1 \right]_+ \left[\int_I |H_1(-1)(\xi)| d\xi - \lambda \right]_+ &< \\
&< \left(1 - \int_s^t |H_0(-1)(\xi)| d\xi \right) \left(1 - \int_I H_0(1)(\xi) d\xi \right) \\
&\text{for } s, t \in [a, b], \quad s \leq t, \quad (2.6)
\end{aligned}$$

and

$$\begin{aligned}
\left[\int_s^t |H_1(-1)(\xi)| d\xi - 1 \right]_+ \left[\int_I H_1(1)(\xi) d\xi - \lambda \right]_+ &< \\
&< \left(1 - \int_s^t H_0(1)(\xi) d\xi \right) \left(1 - \int_I |H_0(-1)(\xi)| d\xi \right) \\
&\text{for } s, t \in [a, b], \quad s \leq t, \quad (2.7)
\end{aligned}$$

where $I = [a, b] \setminus]s, t[$, then problem (1.1), (1.2) has at least one solution.

The following assertion can be regarded as a supplement to the previous theorem in the case when $\lambda = 0$.

Theorem 2.1'. *Let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, (2.1) and (2.2) be fulfilled. If, moreover, the inequalities (2.3),*

$$\begin{aligned}
\int_I |H_1(-1)(\xi)| d\xi \left[\int_s^t H_1(1)(\xi) d\xi - 1 \right]_+ &< \\
&< \left(1 - \int_s^t |H_0(-1)(\xi)| d\xi \right) \left(1 - \int_I H_0(1)(\xi) d\xi \right) \\
&\text{for } s, t \in [a, b], \quad s \leq t, \quad (2.8)
\end{aligned}$$

and

$$\begin{aligned}
\int_I H_1(1)(\xi) d\xi \left[\int_s^t |H_1(-1)(\xi)| d\xi - 1 \right]_+ &< \\
&< \left(1 - \int_s^t H_0(1)(\xi) d\xi \right) \left(1 - \int_I |H_0(-1)(\xi)| d\xi \right) \\
&\text{for } s, t \in [a, b], \quad s \leq t \quad (2.9)
\end{aligned}$$

are fulfilled, where $I = [a, b] \setminus]s, t[$, then problem (1.1), (1.2) with $\lambda = 0$ has at least one solution.

Theorem 2.2. Let $\lambda \in]0, 1]$ and let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, (2.1) and (2.2) be fulfilled. If, moreover,

$$\int_a^b H_1(1)(s)ds < \lambda, \quad \int_a^b |H_1(-1)(s)|ds < \lambda, \quad (2.10)$$

$$\frac{1}{\lambda - \int_a^b H_1(1)(s)ds} - 1 < \int_a^b H_0(1)(s)ds, \quad (2.11)$$

$$\frac{1}{\lambda - \int_a^b |H_1(-1)(s)|ds} - 1 < \int_a^b |H_0(-1)(s)|ds, \quad (2.12)$$

$$\begin{aligned} \left[\int_I H_0(1)(\xi)d\xi - 1 \right]_+ \left[\int_s^t |H_0(-1)(\xi)|d\xi - 1 \right]_+ < \\ < \left(\lambda - \int_I |H_1(-1)(\xi)|d\xi \right) \left(1 - \int_s^t H_1(1)(\xi)d\xi \right) \\ \text{for } s, t \in [a, b], \quad s \leq t, \end{aligned} \quad (2.13)$$

and

$$\begin{aligned} \left[\int_I |H_0(-1)(\xi)|d\xi - 1 \right]_+ \left[\int_s^t H_0(1)(\xi)d\xi - 1 \right]_+ < \\ < \left(\lambda - \int_I H_1(1)(\xi)d\xi \right) \left(1 - \int_s^t |H_1(-1)(\xi)|d\xi \right) \\ \text{for } s, t \in [a, b], \quad s \leq t, \end{aligned} \quad (2.14)$$

where $I = [a, b] \setminus]s, t[$, then problem (1.1), (1.2) has at least one solution.

Theorem 2.3. Let $\lambda \in]0, 1]$ and let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, each of the operators H_0 and H_1 be either a subadditive or a superadditive operator, and let there exist $q \in L([a, b]; \mathbb{R})$ and $c \in \mathbb{R}$ such that

$$Q(v)(t) = q(t) \quad \text{for } t \in [a, b], \quad v \in C([a, b]; \mathbb{R}), \quad (2.15)$$

$$h(v) = c \quad \text{for } v \in C([a, b]; \mathbb{R}). \quad (2.16)$$

If, moreover,

$$\int_a^b \bar{H}_0(1)(s)ds < 1, \quad (2.17)$$

$$\frac{\int_a^b \bar{H}_0(1)(s)ds}{1 - \int_a^b \bar{H}_0(1)(s)ds} - \frac{1 - \lambda}{\lambda} < \int_a^b |\bar{H}_1(-1)(s)|ds, \quad (2.18)$$

$$\begin{aligned}
\left[\int_s^t \bar{H}_1(1)(\xi) d\xi - 1 \right]_+ \left[\int_I \bar{H}_1(1)(\xi) d\xi - \lambda \right]_+ &< \\
&< \left(1 - \int_s^t \bar{H}_0(1)(\xi) d\xi \right) \left(1 - \int_I \bar{H}_0(1)(\xi) d\xi \right) \\
&\text{for } s, t \in [a, b], \quad s \leq t, \quad (2.19)
\end{aligned}$$

where $I = [a, b] \setminus]s, t[$ and

$$\bar{H}_i(v)(t) \stackrel{\text{def}}{=} \begin{cases} H_i(v)(t) & \text{if } H_i \text{ is subadditive} \\ -H_i(-v)(t) & \text{if } H_i \text{ is superadditive} \end{cases} \quad (2.20)$$

for $t \in [a, b]$ and $i = 0, 1$, then problem (1.1), (1.2) has a unique solution.

The following assertion can be regarded as a supplement of the previous theorem in the case when $\lambda = 0$.

Theorem 2.3'. *Let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, each of the operators H_0 and H_1 be either a subadditive or a superadditive operator, and let there exist $q \in L([a, b]; \mathbb{R})$ and $c \in \mathbb{R}$ such that (2.15) and (2.16) are fulfilled. If, moreover, inequalities (2.17) and*

$$\begin{aligned}
\int_I \bar{H}_1(1)(\xi) d\xi \left[\int_s^t \bar{H}_1(1)(\xi) d\xi - 1 \right]_+ &< \\
&< \left(1 - \int_s^t \bar{H}_0(1)(\xi) d\xi \right) \left(1 - \int_I \bar{H}_0(1)(\xi) d\xi \right) \\
&\text{for } s, t \in [a, b], \quad s \leq t \quad (2.21)
\end{aligned}$$

hold, where $I = [a, b] \setminus]s, t[$, and $\bar{H}_i$ ($i = 0, 1$) are given by (2.20), then problem (1.1), (1.2) with $\lambda = 0$ has a unique solution.

Theorem 2.4. *Let $\lambda \in]0, 1]$ and let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, each of the operators H_0 and H_1 be either a subadditive or a superadditive operator, and let there exist $q \in L([a, b]; \mathbb{R})$ and $c \in \mathbb{R}$ such that (2.15) and (2.16) are fulfilled. If, moreover,*

$$\int_a^b \bar{H}_1(1)(s) ds < \lambda, \quad (2.22)$$

$$\frac{1}{\lambda - \int_a^b \bar{H}_1(1)(s) ds} - 1 < \int_a^b |\bar{H}_0(-1)(s)| ds, \quad (2.23)$$

$$\left[\int_I \bar{H}_0(1)(\xi) d\xi - 1 \right]_+ \left[\int_s^t \bar{H}_0(1)(\xi) d\xi - 1 \right]_+ < \\ < \left(\lambda - \int_I \bar{H}_1(1)(\xi) d\xi \right) \left(1 - \int_s^t \bar{H}_1(1)(\xi) d\xi \right) \\ \text{for } s, t \in [a, b], \quad s \leq t, \quad (2.24)$$

where $I = [a, b] \setminus]s, t[$, and $\bar{H}_i$ ($i = 0, 1$) are given by (2.20), then problem (1.1), (1.2) has a unique solution.

Corollary 2.1. Let $\lambda \in]0, 1]$ and let the inequalities

$$\int_a^b [p(s)]_+ ds < 1, \quad (2.25)$$

$$\frac{\int_a^b [p(s)]_+ ds}{1 - \int_a^b [p(s)]_+ ds} - \frac{1 - \lambda}{\lambda} < \int_a^b [p(s)]_- ds < 1 + \lambda + 2 \sqrt{1 - \int_a^b [p(s)]_+ ds} \quad (2.26)$$

be fulfilled. Then problem (1.3), (1.4) has a unique solution.

Corollary 2.1'. Let inequalities (2.25) and

$$\int_a^b [p(s)]_- ds < 1 + 2 \sqrt{1 - \int_a^b [p(s)]_+ ds}$$

be fulfilled. Then problem (1.3), (1.4) with $\lambda = 0$ has a unique solution.

Corollary 2.2. Let $\lambda \in]0, 1]$ and let the inequalities

$$\int_a^b [p(s)]_- ds < \lambda, \quad (2.27)$$

$$\frac{1}{\lambda - \int_a^b [p(s)]_- ds} - 1 < \int_a^b [p(s)]_+ ds < 2 + 2 \sqrt{\lambda - \int_a^b [p(s)]_- ds} \quad (2.28)$$

be fulfilled. Then problem (1.3), (1.4) has a unique solution.

Remark 2.1. Let $\lambda \in [1, +\infty[$. Define operator $\psi : L([a, b]; \mathbb{R}) \rightarrow L([a, b]; \mathbb{R})$ by the equality

$$\psi(w)(t) \stackrel{\text{def}}{=} w(a + b - t) \quad \text{for } t \in [a, b].$$

Let φ be the restriction of ψ to the space $C([a, b]; \mathbb{R})$. Put

$$\hat{H}_0(w)(t) \stackrel{\text{def}}{=} \psi(H_0(\varphi(w)))(t), \quad \hat{H}_1(w)(t) \stackrel{\text{def}}{=} \psi(H_1(\varphi(w)))(t) \quad \text{for } t \in [a, b],$$

$$\hat{Q}(w)(t) \stackrel{\text{def}}{=} -\psi(Q(\varphi(w)))(t) \quad \text{for } t \in [a, b],$$

and

$$\hat{h}(w) \stackrel{\text{def}}{=} -\frac{1}{\lambda} h(\varphi(w)).$$

It is clear that if u is a solution of problem (1.1), (1.2), then the function $v \stackrel{\text{def}}{=} \varphi(u)$ is a solution of the problem

$$v'(t) = \hat{H}_1(v)(t) - \hat{H}_0(v)(t) + \hat{Q}(v)(t), \quad v(a) - \frac{1}{\lambda}v(b) = \hat{h}(v), \quad (2.29)$$

and *vice versa*, if v is a solution of problem (2.29), then the function $u \stackrel{\text{def}}{=} \varphi(v)$ is a solution of problem (1.1), (1.2).

Furthermore, if H_i ($i \in \{0, 1\}$) is a subadditive (resp., superadditive) operator, then $\hat{H}_i$ ($i \in \{0, 1\}$) is a subadditive (resp., superadditive) operator, as well.

Theorems 2.5–2.8 and Corollaries 2.3 and 2.4 formulated below can be derived easily from Theorems 2.1–2.4 and Corollaries 2.1 and 2.2, respectively, by using the change of variables described in Remark 2.1.

Theorem 2.5. *Let $\lambda \in [1, +\infty[$ and let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, (2.1) and (2.2) be fulfilled. If, moreover,*

$$\begin{aligned} \int_a^b H_1(1)(s)ds &< 1, & \int_a^b |H_1(-1)(s)|ds &< 1, \\ \frac{\int_a^b H_1(1)(s)ds}{1 - \int_a^b H_1(1)(s)ds} - (\lambda - 1) &< \int_a^b H_0(1)(s)ds, \\ \frac{\int_a^b |H_1(-1)(s)|ds}{1 - \int_a^b |H_1(-1)(s)|ds} - (\lambda - 1) &< \int_a^b |H_0(-1)(s)|ds, \end{aligned}$$

and the inequalities

$$\begin{aligned} \left[\int_s^t H_0(1)(\xi)d\xi - 1 \right]_+ \left[\int_I |H_0(-1)(\xi)|d\xi - \frac{1}{\lambda} \right]_+ &< \\ &< \left(1 - \int_s^t |H_1(-1)(\xi)|d\xi \right) \left(1 - \int_I H_1(1)(\xi)d\xi \right) \\ &\text{for } s, t \in [a, b], \quad s \leq t, \end{aligned}$$

and

$$\begin{aligned} \left[\int_s^t |H_0(-1)(\xi)|d\xi - 1 \right]_+ \left[\int_I H_0(1)(\xi)d\xi - \frac{1}{\lambda} \right]_+ &< \\ &< \left(1 - \int_s^t H_1(1)(\xi)d\xi \right) \left(1 - \int_I |H_1(-1)(\xi)|d\xi \right) \\ &\text{for } s, t \in [a, b], \quad s \leq t, \end{aligned}$$

are satisfied, where $I = [a, b] \setminus]s, t[$, then problem (1.1), (1.2) has at least one solution.

Theorem 2.6. Let $\lambda \in [1, +\infty[$ and let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, (2.1) and (2.2) be fulfilled. If, moreover,

$$\begin{aligned} \int_a^b H_0(1)(s)ds &< \frac{1}{\lambda}, & \int_a^b |H_0(-1)(s)|ds &< \frac{1}{\lambda}, \\ \frac{\lambda}{1 - \lambda \int_a^b H_0(1)(s)ds} - 1 &< \int_a^b H_1(1)(s)ds, \\ \frac{\lambda}{1 - \lambda \int_a^b |H_0(-1)(s)|ds} - 1 &< \int_a^b |H_1(-1)(s)|ds, \end{aligned}$$

and the inequalities

$$\begin{aligned} \left[\int_I H_1(1)(\xi)d\xi - 1 \right]_+ \left[\int_s^t |H_1(-1)(\xi)|d\xi - 1 \right]_+ &< \\ &< \left(\frac{1}{\lambda} - \int_I |H_0(-1)(\xi)|d\xi \right) \left(1 - \int_s^t H_0(1)(\xi)d\xi \right) \\ &\text{for } s, t \in [a, b], \quad s \leq t, \end{aligned}$$

and

$$\begin{aligned} \left[\int_I |H_1(-1)(\xi)|d\xi - 1 \right]_+ \left[\int_s^t H_1(1)(\xi)d\xi - 1 \right]_+ &< \\ &< \left(\frac{1}{\lambda} - \int_I H_0(1)(\xi)d\xi \right) \left(1 - \int_s^t |H_0(-1)(\xi)|d\xi \right) \\ &\text{for } s, t \in [a, b], \quad s \leq t, \end{aligned}$$

are satisfied, where $I = [a, b] \setminus]s, t[$, then problem (1.1), (1.2) has at least one solution.

Theorem 2.7. Let $\lambda \in [1, +\infty[$ and let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, each of the operators H_0 and H_1 be either a subadditive or a superadditive operator, and let there exist $q \in L([a, b]; \mathbb{R})$ and $c \in \mathbb{R}$ such that (2.15) and (2.16) are fulfilled. If, moreover,

$$\begin{aligned} \int_a^b \bar{H}_1(1)(s)ds &< 1, \\ \frac{\int_a^b \bar{H}_1(1)(s)ds}{1 - \int_a^b \bar{H}_1(1)(s)ds} - (\lambda - 1) &< \int_a^b |\bar{H}_0(-1)(s)|ds, \end{aligned}$$

$$\begin{aligned}
\left[\int_s^t \bar{H}_0(1)(\xi) d\xi - 1 \right]_+ \left[\int_I \bar{H}_0(1)(\xi) d\xi - \frac{1}{\lambda} \right]_+ &< \\
&< \left(1 - \int_s^t \bar{H}_1(1)(\xi) d\xi \right) \left(1 - \int_I \bar{H}_1(1)(\xi) d\xi \right) \\
&\text{for } s, t \in [a, b], \quad s \leq t,
\end{aligned}$$

where $I = [a, b] \setminus]s, t[$, and $\bar{H}_i$ ($i = 0, 1$) are given by (2.20), then problem (1.1), (1.2) has a unique solution.

Theorem 2.8. Let $\lambda \in [1, +\infty[$ and let H admit the representation $H = H_0 - H_1$, where $H_0, H_1 \in \mathcal{P}_{ab}$. Let, furthermore, each of the operators H_0 and H_1 be either a subadditive or a superadditive operator, and let there exist $q \in L([a, b]; \mathbb{R})$ and $c \in \mathbb{R}$ such that (2.15) and (2.16) are fulfilled. If, moreover,

$$\begin{aligned}
&\int_a^b \bar{H}_0(1)(s) ds < \frac{1}{\lambda}, \\
&\frac{\lambda}{1 - \lambda \int_a^b \bar{H}_0(1)(s) ds} - 1 < \int_a^b |\bar{H}_1(-1)(s)| ds
\end{aligned}$$

and

$$\begin{aligned}
\left[\int_I \bar{H}_1(1)(\xi) d\xi - 1 \right]_+ \left[\int_s^t \bar{H}_1(1)(\xi) d\xi - 1 \right]_+ &< \\
&< \left(\frac{1}{\lambda} - \int_I \bar{H}_0(1)(\xi) d\xi \right) \left(1 - \int_s^t \bar{H}_0(1)(\xi) d\xi \right) \\
&\text{for } s, t \in [a, b], \quad s \leq t,
\end{aligned}$$

where $I = [a, b] \setminus]s, t[$, and $\bar{H}_i$ ($i = 0, 1$) are given by (2.20), then problem (1.1), (1.2) has a unique solution.

Corollary 2.3. Let $\lambda \in [1, +\infty[$ and let the inequalities

$$\begin{aligned}
&\int_a^b [p(s)]_- ds < 1, \\
&\frac{\int_a^b [p(s)]_- ds}{1 - \int_a^b [p(s)]_- ds} - (\lambda - 1) < \int_a^b [p(s)]_+ ds < 1 + \frac{1}{\lambda} + 2 \sqrt{1 - \int_a^b [p(s)]_- ds}
\end{aligned}$$

be fulfilled. Then problem (1.3), (1.4) has a unique solution.

Corollary 2.4. *Let $\lambda \in [1, +\infty[$ and let the inequalities*

$$\int_a^b [p(s)]_+ ds < \frac{1}{\lambda},$$

$$\frac{\lambda}{1 - \lambda \int_a^b [p(s)]_+ ds} - 1 < \int_a^b [p(s)]_- ds < 2 + 2 \sqrt{\frac{1}{\lambda} - \int_a^b [p(s)]_+ ds}$$

be fulfilled. Then problem (1.3), (1.4) has a unique solution.

3. AUXILIARY PROPOSITIONS

The following two lemmas are consequences of the result obtained in [31, Corollary 1.4].

Lemma 3.1. *Let $\lambda \in [0, 1]$ and let the problem*

$$u'(t) = \delta H(u)(t), \quad u(a) = \delta \lambda u(b)$$

have only the trivial solution for every $\delta \in [0, 1]$. Let, moreover, (2.1) and (2.2) be fulfilled. Then problem (1.1), (1.2) has at least one solution.

Lemma 3.2. *Let $\lambda \in]0, 1]$ and let the problem*

$$u'(t) = \delta H(u)(t), \quad \delta u(a) = \lambda u(b)$$

have only the trivial solution for every $\delta \in [0, 1]$. Let, moreover, (2.1) and (2.2) be fulfilled. Then problem (1.1), (1.2) has at least one solution.

Lemma 3.3. *Let $\lambda \in]0, 1]$, $F_i, G_i \in \mathcal{P}_{ab}$ ($i = 0, 1$), and let $w \in \tilde{C}([a, b]; \mathbb{R})$ be a non-trivial function satisfying*

$$w'(t) \leq \delta F_0(w)(t) - \delta F_1(w)(t) \quad \text{for } t \in [a, b], \quad (3.1)$$

$$w'(t) \geq \delta G_0(w)(t) - \delta G_1(w)(t) \quad \text{for } t \in [a, b], \quad (3.2)$$

$$w(a) = \delta \lambda w(b) \quad (3.3)$$

for some $\delta \in]0, 1]$. If, moreover,

$$\int_a^b F_0(1)(s) ds < 1, \quad \int_a^b |G_0(-1)(s)| ds < 1, \quad (3.4)$$

$$\frac{\int_a^b F_0(1)(s) ds}{1 - \int_a^b F_0(1)(s) ds} - \frac{1 - \lambda}{\lambda} < \int_a^b F_1(1)(s) ds, \quad (3.5)$$

$$\frac{\int_a^b |G_0(-1)(s)| ds}{1 - \int_a^b |G_0(-1)(s)| ds} - \frac{1 - \lambda}{\lambda} < \int_a^b |G_1(-1)(s)| ds, \quad (3.6)$$

then there exists $t_0 \in [a, b]$ such that

$$w(t_0) = 0. \quad (3.7)$$

Proof. Assume that, on the contrary, the function w has no zero. First suppose that

$$w(t) > 0 \quad \text{for } t \in [a, b]. \quad (3.8)$$

Put

$$M = \max\{w(t) : t \in [a, b]\}, \quad m = \min\{w(t) : t \in [a, b]\} \quad (3.9)$$

and choose $t_M, t_m \in [a, b]$ such that $t_M \neq t_m$ and

$$w(t_M) = M, \quad w(t_m) = m. \quad (3.10)$$

According to (3.8) and (3.9), we have

$$M > 0, \quad m > 0. \quad (3.11)$$

It is obvious that either

$$t_M < t_m \quad (3.12)$$

or

$$t_m < t_M. \quad (3.13)$$

If (3.12) is fulfilled, then the integration of (3.1) from a to t_M and from t_m to b , respectively, in view of (3.10), yields

$$\begin{aligned} M - w(a) &\leq \delta \int_a^{t_M} F_0(w)(s)ds - \delta \int_a^{t_M} F_1(w)(s)ds, \\ w(b) - m &\leq \delta \int_{t_m}^b F_0(w)(s)ds - \delta \int_{t_m}^b F_1(w)(s)ds. \end{aligned}$$

Hence, in view of (3.8), (3.9), and the assumptions $\delta \in]0, 1]$, $\lambda \in]0, 1]$, $F_0, F_1 \in \mathcal{P}_{ab}$, we get

$$\begin{aligned} M - w(a) &\leq M \int_a^{t_M} F_0(1)(s)ds, \\ \delta \lambda w(b) - m &\leq M \int_{t_m}^b F_0(1)(s)ds. \end{aligned}$$

Summing the last two inequalities, according to (3.3), (3.4), and (3.11), we obtain

$$0 < M \left(1 - \int_a^b F_0(1)(s)ds \right) \leq m. \quad (3.14)$$

If (3.13) is fulfilled, then the integration of (3.1) from t_m to t_M , in view of (3.10), yields

$$M - m \leq \delta \int_{t_m}^{t_M} F_0(w)(s)ds - \delta \int_{t_m}^{t_M} F_1(w)(s)ds.$$

Hence, on account of (3.8), (3.9), (3.11), and the assumptions $\delta \in]0, 1]$ and $F_0, F_1 \in \mathcal{P}_{ab}$, we get (3.14).

Thus, in both cases (3.12) and (3.13), inequality (3.14) holds.

On the other hand, the integration of (3.1) from a to b yields

$$w(b) - w(a) \leq \delta \int_a^b F_0(w)(s)ds - \delta \int_a^b F_1(w)(s)ds,$$

whence, in view of (3.3), (3.8), (3.9), (3.11), and the assumptions $\delta \in]0, 1]$, $\lambda \in]0, 1]$, and $F_0, F_1 \in \mathcal{P}_{ab}$, we obtain

$$0 \leq m\delta \frac{1-\lambda}{\lambda} \leq w(b) - w(a) \leq M\delta \int_a^b F_0(1)(s)ds - m\delta \int_a^b F_1(1)(s)ds.$$

Consequently,

$$0 \leq m \left(\int_a^b F_1(1)(s)ds + \frac{1-\lambda}{\lambda} \right) \leq M \int_a^b F_0(1)(s)ds. \quad (3.15)$$

Multiplying the terms from (3.14) by the respective terms from (3.15), on account of (3.11), we obtain

$$\int_a^b F_1(1)(s)ds \leq \frac{\int_a^b F_0(1)(s)ds}{1 - \int_a^b F_0(1)(s)ds} - \frac{1-\lambda}{\lambda},$$

which contradicts (3.5).

Now suppose that $w(t) < 0$ for $t \in [a, b]$. Then, according to (3.2) and (3.3), $v \stackrel{\text{def}}{=} -w$ is a non-trivial function satisfying

$$v'(t) \leq \delta \tilde{F}_0(v)(t) - \delta \tilde{F}_1(v)(t) \quad \text{for } t \in [a, b], \quad v(a) = \delta \lambda v(b),$$

where

$$\tilde{F}_i(v)(t) = -G_i(-v)(t) \quad \text{for } t \in [a, b] \quad (i = 0, 1).$$

Obviously, $\tilde{F}_0, \tilde{F}_1 \in \mathcal{P}_{ab}$, and $v(t) > 0$ for $t \in [a, b]$. Therefore, following the steps taken above we get

$$\int_a^b |G_1(-1)(s)|ds \leq \frac{\int_a^b |G_0(-1)(s)|ds}{1 - \int_a^b |G_0(-1)(s)|ds} - \frac{1-\lambda}{\lambda},$$

which contradicts (3.6). $\square$

Lemma 3.4. *Let $\lambda \in [0, 1]$, $F_i, G_i \in \mathcal{P}_{ab}$ ($i = 0, 1$), and let $w \in \tilde{C}([a, b]; \mathbb{R})$ be a non-trivial function satisfying (3.1)–(3.3) for some $\delta \in]0, 1]$. If, moreover, inequalities (3.4),*

$$\begin{aligned} & \left[\int_s^t G_1(1)(\xi)d\xi - 1 \right]_+ \left[\int_I |F_1(-1)(\xi)|d\xi - \lambda \right]_+ < \\ & < \left(1 - \int_s^t |G_0(-1)(\xi)|d\xi \right) \left(1 - \int_I F_0(1)(\xi)d\xi \right) \\ & \quad \text{for } s, t \in [a, b], \quad s \leq t, \end{aligned} \quad (3.16)$$

and

$$\begin{aligned} \left[\int_s^t |F_1(-1)(\xi)| d\xi - 1 \right]_+ \left[\int_I G_1(1)(\xi) d\xi - \lambda \right]_+ &< \\ &< \left(1 - \int_s^t F_0(1)(\xi) d\xi \right) \left(1 - \int_I |G_0(-1)(\xi)| d\xi \right) \\ &\text{for } s, t \in [a, b], \quad s \leq t, \end{aligned} \quad (3.17)$$

are fulfilled, where $I = [a, b] \setminus]s, t[$, then

$$w(t) \neq 0 \quad \text{for } t \in [a, b]. \quad (3.18)$$

Proof. Assume that, on the contrary, there exists a point $t_0 \in [a, b]$ such that (3.7) is fulfilled. Put

$$M = \max\{w(t) : t \in [a, b]\}, \quad m = -\min\{w(t) : t \in [a, b]\} \quad (3.19)$$

and choose $t_M, t_m \in [a, b]$ such that $t_M \neq t_m$ and

$$w(t_M) = M, \quad w(t_m) = -m. \quad (3.20)$$

According to (3.7) and (3.19) we have

$$M \geq 0, \quad m \geq 0, \quad M + m > 0. \quad (3.21)$$

Obviously, either (3.12) or (3.13) is fulfilled.

Suppose that (3.12) is satisfied. Then the integration of (3.1) from a to t_M and from t_m to b , and the integration of (3.2) from t_M to t_m , respectively, by virtue of (3.20), yields

$$\begin{aligned} M - w(a) &\leq \delta \int_a^{t_M} F_0(w)(s) ds - \delta \int_a^{t_M} F_1(w)(s) ds, \\ -m - M &\geq \delta \int_{t_M}^{t_m} G_0(w)(s) ds - \delta \int_{t_M}^{t_m} G_1(w)(s) ds, \\ w(b) + m &\leq \delta \int_{t_m}^b F_0(w)(s) ds - \delta \int_{t_m}^b F_1(w)(s) ds. \end{aligned}$$

Hence, in view of (3.19), (3.21), and the assumptions $\delta \in]0, 1]$, $\lambda \in [0, 1]$, $F_i, G_i \in \mathcal{P}_{ab}$ ($i = 0, 1$), we get

$$\delta M - w(a) \leq \delta M \int_a^{t_M} F_0(1)(s) ds + \delta m \int_a^{t_M} |F_1(-1)(s)| ds, \quad (3.22)$$

$$M + m \leq m \int_{t_M}^{t_m} |G_0(-1)(s)| ds + M \int_{t_M}^{t_m} G_1(1)(s) ds, \quad (3.23)$$

$$\delta \lambda (w(b) + m) \leq \delta M \int_{t_m}^b F_0(1)(s) ds + \delta m \int_{t_m}^b |F_1(-1)(s)| ds. \quad (3.24)$$

Summing (3.22) and (3.24), in view of (3.3), (3.4), (3.21), and the assumption $\delta > 0$, we obtain

$$0 \leq M \left(1 - \int_I F_0(1)(s) ds \right) \leq m \left[\int_I |F_1(-1)(s)| ds - \lambda \right]_+, \quad (3.25)$$

where $I = [a, b] \setminus]t_M, t_m[$. Further on, (3.23) according to (3.4) and (3.21), results in

$$0 \leq m \left(1 - \int_{t_M}^{t_m} |G_0(-1)(s)| ds \right) \leq M \left[\int_{t_M}^{t_m} G_1(1)(s) ds - 1 \right]_+. \quad (3.26)$$

Now from (3.25) and (3.26) it follows that $M > 0$, $m > 0$. Therefore, multiplying (3.25) and (3.26) we obtain

$$\begin{aligned} \left(1 - \int_I F_0(1)(s) ds \right) \left(1 - \int_{t_M}^{t_m} |G_0(-1)(s)| ds \right) &\leq \\ &\leq \left[\int_I |F_1(-1)(s)| ds - \lambda \right]_+ \left[\int_{t_M}^{t_m} G_1(1)(s) ds - 1 \right]_+, \end{aligned}$$

which contradicts (3.16).

Now suppose that (3.13) holds. Then, according to (3.1)–(3.3), $v \stackrel{\text{def}}{=} -w$ is a non-trivial function satisfying the relations

$$\begin{aligned} v'(t) &\leq \delta \tilde{F}_0(v)(t) - \delta \tilde{F}_1(v)(t) && \text{for } t \in [a, b], \\ v'(t) &\geq \delta \tilde{G}_0(v)(t) - \delta \tilde{G}_1(v)(t) && \text{for } t \in [a, b], \\ v(a) &= \delta \lambda v(b), \end{aligned}$$

where

$$\tilde{F}_i(v)(t) = -G_i(-v)(t), \quad \tilde{G}_i(v)(t) = -F_i(-v)(t) \quad \text{for } t \in [a, b] \quad (i = 0, 1).$$

Obviously, $\tilde{F}_i, \tilde{G}_i \in \mathcal{P}_{ab}$ ($i = 0, 1$), and

$$v(t_m) = \max\{v(t) : t \in [a, b]\}, \quad v(t_M) = \min\{v(t) : t \in [a, b]\}.$$

Therefore, the substitution $v \stackrel{\text{def}}{=} -w$ transforms the case considered into the previous one and, following the steps above, we get

$$\begin{aligned} \left(1 - \int_I |G_0(-1)(s)| ds \right) \left(1 - \int_{t_m}^{t_M} F_0(1)(s) ds \right) &\leq \\ &\leq \left[\int_I G_1(1)(s) ds - \lambda \right]_+ \left[\int_{t_m}^{t_M} |F_1(-1)(s)| ds - 1 \right]_+, \end{aligned}$$

where $I = [a, b] \setminus]t_m, t_M[$, which contradicts (3.17). $\square$

Lemma 3.5. *Let $\lambda \in]0, 1]$, $F_i, G_i \in \mathcal{P}_{ab}$ ($i = 0, 1$), and let $w \in \tilde{C}([a, b]; \mathbb{R})$ be a non-trivial function satisfying inequalities (3.1), (3.2), and*

$$\delta w(a) = \lambda w(b) \quad (3.27)$$

for some $\delta \in]0, 1]$. If, moreover,

$$\int_a^b G_1(1)(s)ds < \lambda, \quad \int_a^b |F_1(-1)(s)|ds < \lambda, \quad (3.28)$$

$$\frac{1}{\lambda - \int_a^b G_1(1)(s)ds} - 1 < \int_a^b G_0(1)(s)ds, \quad (3.29)$$

$$\frac{1}{\lambda - \int_a^b |F_1(-1)(s)|ds} - 1 < \int_a^b |F_0(-1)(s)|ds, \quad (3.30)$$

then there exists a point $t_0 \in [a, b]$ such that (3.7) holds.

Proof. Assume that, on the contrary, the function w has no zero. First suppose that (3.8) is fulfilled. Define numbers M and m by (3.9) and choose $t_M, t_m \in [a, b]$ such that $t_M \neq t_m$ and (3.10) is satisfied. According to (3.8) and (3.9) we have (3.11). Obviously, either (3.12) or (3.13) holds.

If (3.13) is fulfilled, then the integration of (3.2) from a to t_m and from t_M to b , respectively, in view of (3.10), yields

$$\begin{aligned} w(a) - m &\leq -\delta \int_a^{t_m} G_0(w)(s)ds + \delta \int_a^{t_m} G_1(w)(s)ds, \\ M - w(b) &\leq -\delta \int_{t_M}^b G_0(w)(s)ds + \delta \int_{t_M}^b G_1(w)(s)ds. \end{aligned}$$

Hence, in view of (3.8), (3.9), and the assumptions $\delta \in]0, 1]$, $\lambda \in]0, 1]$, $G_0, G_1 \in \mathcal{P}_{ab}$, we get

$$\begin{aligned} \delta(w(a) - m) &\leq M \int_a^{t_m} G_1(1)(s)ds, \\ \lambda(M - w(b)) &\leq M \int_{t_M}^b G_1(1)(s)ds. \end{aligned}$$

Summing the last two inequalities, and taking (3.11), (3.27), and (3.28) into account, we obtain

$$0 < M \left(\lambda - \int_a^b G_1(1)(s)ds \right) \leq \delta m. \quad (3.31)$$

If (3.12) is fulfilled, then the integration of (3.2) from t_M to t_m , in view of (3.10), yields

$$M - m \leq -\delta \int_{t_M}^{t_m} G_0(w)(s)ds + \delta \int_{t_M}^{t_m} G_1(w)(s)ds. \quad (3.32)$$

Suppose that $\lambda \leq \delta$. Then from (3.32), in view of (3.8), (3.9), and the assumptions $\delta \in]0, 1]$, $\lambda \in]0, 1]$, and $G_0, G_1 \in \mathcal{P}_{ab}$, it follows that

$$\lambda M - \delta m \leq \lambda(M - m) \leq M \int_a^b G_1(1)(s)ds,$$

whence we get (3.31). If $\delta < \lambda$, then from (3.32), in view of (3.8), (3.9), and the assumptions $\delta \in]0, 1]$, $\lambda \in]0, 1]$, and $G_0, G_1 \in \mathcal{P}_{ab}$, we obtain

$$0 < M \left(\lambda - \int_a^b G_1(1)(s)ds \right) \leq m.$$

Thus in both cases (3.12) and (3.13), we have

$$0 < M \left(\lambda - \int_a^b G_1(1)(s)ds \right) \leq \delta m \quad \text{if } \lambda \leq \delta, \quad (3.33)$$

$$0 < M \left(\lambda - \int_a^b G_1(1)(s)ds \right) \leq m \quad \text{if } \lambda > \delta. \quad (3.34)$$

On the other hand, the integration of (3.2) from a to b yields

$$w(a) - w(b) \leq -\delta \int_a^b G_0(w)(s)ds + \delta \int_a^b G_1(w)(s)ds. \quad (3.35)$$

If now $\lambda \leq \delta$, then from (3.35), on account of (3.8), (3.9), (3.11), (3.27), and the assumptions $\delta \in]0, 1]$, $\lambda \in]0, 1]$, and $G_0, G_1 \in \mathcal{P}_{ab}$, we obtain

$$-M(1 - \lambda) \leq w(b)(\lambda - 1) \leq w(a) - w(b) \leq -m\delta \int_a^b G_0(1)(s)ds + M \int_a^b G_1(1)(s)ds.$$

Therefore,

$$0 \leq m\delta \int_a^b G_0(1)(s)ds \leq M \left(\int_a^b G_1(1)(s)ds + 1 - \lambda \right) \quad \text{if } \lambda \leq \delta. \quad (3.36)$$

If $\lambda > \delta$, from (3.35), on account of (3.8), (3.11), (3.27), and the assumption $\lambda \in]0, 1]$, we have

$$-M\delta(1 - \lambda) \leq 0 \leq w(a) - w(b) \leq -m\delta \int_a^b G_0(1)(s)ds + M\delta \int_a^b G_1(1)(s)ds.$$

Hence, in view of the assumption $\delta > 0$, we get

$$0 \leq m \int_a^b G_0(1)(s)ds \leq M \left(\int_a^b G_1(1)(s)ds + 1 - \lambda \right) \quad \text{if } \lambda > \delta. \quad (3.37)$$

Multiplying the terms in (3.33) (resp., (3.34)) by the corresponding terms in (3.36) (resp., (3.37)), on account of (3.11) and the assumption $\delta > 0$, we obtain

$$\int_a^b G_0(1)(s)ds \leq \frac{1}{\lambda - \int_a^b G_1(1)(s)ds} - 1,$$

which contradicts (3.29).

Now suppose that $w(t) < 0$ for $t \in [a, b]$. Then, according to (3.1) and (3.27), $v \stackrel{\text{def}}{=} -w$ is a non-trivial function satisfying

$$v'(t) \geq \delta \tilde{G}_0(v)(t) - \delta \tilde{G}_1(v)(t) \quad \text{for } t \in [a, b], \quad \delta v(a) = \lambda v(b),$$

where

$$\tilde{G}_i(v)(t) = -F_i(-v)(t) \quad \text{for } t \in [a, b] \quad (i = 0, 1).$$

Obviously, $\tilde{G}_0, \tilde{G}_1 \in \mathcal{P}_{ab}$, and $v(t) > 0$ for $t \in [a, b]$. Therefore, following the steps taken above we get

$$\int_a^b |F_0(-1)(s)| ds \leq \frac{1}{\lambda - \int_a^b |F_1(-1)(s)| ds} - 1,$$

which contradicts (3.30). $\square$

Lemma 3.6. *Let $\lambda \in]0, 1]$, $F_i, G_i \in \mathcal{P}_{ab}$ ($i = 0, 1$), and let $w \in \tilde{C}([a, b]; \mathbb{R})$ be a non-trivial function satisfying (3.1), (3.2), and (3.27) for some $\delta \in]0, 1]$. If, moreover, inequalities (3.28),*

$$\begin{aligned} \left[\int_I F_0(1)(\xi) d\xi - 1 \right]_+ \left[\int_s^t |G_0(-1)(\xi)| d\xi - 1 \right]_+ &< \\ &< \left(\lambda - \int_I |F_1(-1)(\xi)| d\xi \right) \left(1 - \int_s^t G_1(1)(\xi) d\xi \right) \\ &\text{for } s, t \in [a, b], \quad s \leq t, \end{aligned} \quad (3.38)$$

and

$$\begin{aligned} \left[\int_I |G_0(-1)(\xi)| d\xi - 1 \right]_+ \left[\int_s^t F_0(1)(\xi) d\xi - 1 \right]_+ &< \\ &< \left(\lambda - \int_I G_1(1)(\xi) d\xi \right) \left(1 - \int_s^t |F_1(-1)(\xi)| d\xi \right) \\ &\text{for } s, t \in [a, b], \quad s \leq t, \end{aligned} \quad (3.39)$$

are fulfilled, where $I = [a, b] \setminus]s, t[$, then (3.18) holds.

Proof. Assume that, on the contrary, there exists $t_0 \in [a, b]$ such that (3.7) holds. Define numbers M and m by (3.19) and choose $t_M, t_m \in [a, b]$ such that $t_M \neq t_m$ and (3.20) is fulfilled. According to (3.7) and (3.19) we have (3.21). Obviously, either (3.12) or (3.13) is fulfilled.

Suppose that (3.12) is satisfied. Then the integration of (3.1) from a to t_M and from t_m to b , and the integration of (3.2) from t_M to t_m , respectively, by virtue of (3.20), yields

$$\begin{aligned} M - w(a) &\leq \delta \int_a^{t_M} F_0(w)(s) ds - \delta \int_a^{t_M} F_1(w)(s) ds, \\ -m - M &\geq \delta \int_{t_M}^{t_m} G_0(w)(s) ds - \delta \int_{t_M}^{t_m} G_1(w)(s) ds, \\ w(b) + m &\leq \delta \int_{t_m}^b F_0(w)(s) ds - \delta \int_{t_m}^b F_1(w)(s) ds. \end{aligned}$$

Hence, in view of (3.19), (3.21), and the assumptions $\delta \in]0, 1]$, $\lambda \in]0, 1]$, $F_i, G_i \in \mathcal{P}_{ab}$ ($i = 0, 1$), we get

$$\delta(M - w(a)) \leq \delta M \int_a^{t_M} F_0(1)(s)ds + \delta m \int_a^{t_M} |F_1(-1)(s)|ds, \quad (3.40)$$

$$M + m \leq m \int_{t_M}^{t_m} |G_0(-1)(s)|ds + M \int_{t_M}^{t_m} G_1(1)(s)ds, \quad (3.41)$$

$$\lambda(w(b) + \delta m) \leq \delta M \int_{t_m}^b F_0(1)(s)ds + \delta m \int_{t_m}^b |F_1(-1)(s)|ds. \quad (3.42)$$

Summing (3.40) and (3.42), according to (3.21), (3.27), (3.28), and the assumption $\delta > 0$, we obtain

$$0 \leq m \left(\lambda - \int_I |F_1(-1)(s)|ds \right) \leq M \left[\int_I F_0(1)(s)ds - 1 \right]_+, \quad (3.43)$$

where $I = [a, b] \setminus]t_M, t_m[$. Furthermore, (3.41) in view of (3.28) and (3.21), results in

$$0 \leq M \left(1 - \int_{t_M}^{t_m} G_1(1)(s)ds \right) \leq m \left[\int_{t_M}^{t_m} |G_0(-1)(s)|ds - 1 \right]_+. \quad (3.44)$$

Now from (3.43) and (3.44) it follows that $M > 0$, $m > 0$. Therefore, multiplying the terms in (3.43) by the corresponding terms in (3.44), we obtain

$$\begin{aligned} \left(\lambda - \int_I |F_1(-1)(s)|ds \right) \left(1 - \int_{t_M}^{t_m} G_1(1)(s)ds \right) &\leq \\ &\leq \left[\int_I F_0(1)(s)ds - 1 \right]_+ \left[\int_{t_M}^{t_m} |G_0(-1)(s)|ds - 1 \right]_+, \end{aligned}$$

which contradicts (3.38).

Now suppose that (3.13) holds. Then, according to (3.1), (3.2), and (3.27), $v \stackrel{\text{def}}{=} -w$ is a non-trivial function satisfying

$$\begin{aligned} v'(t) &\leq \delta \tilde{F}_0(v)(t) - \delta \tilde{F}_1(v)(t) & \text{for } t \in [a, b], \\ v'(t) &\geq \delta \tilde{G}_0(v)(t) - \delta \tilde{G}_1(v)(t) & \text{for } t \in [a, b], \\ \delta v(a) &= \lambda v(b), \end{aligned}$$

where

$$\tilde{F}_i(v)(t) = -G_i(-v)(t), \quad \tilde{G}_i(v)(t) = -F_i(-v)(t) \quad \text{for } t \in [a, b] \quad (i = 0, 1).$$

Obviously, $\tilde{F}_i, \tilde{G}_i \in \mathcal{P}_{ab}$ ($i = 0, 1$), and

$$v(t_m) = \max\{v(t) : t \in [a, b]\}, \quad v(t_M) = \min\{v(t) : t \in [a, b]\}.$$

Therefore, the substitution $v \stackrel{\text{def}}{=} -w$ transforms the case considered into the previous one and, arguing similarly to the considerations above, we get

$$\begin{aligned} \left(\lambda - \int_I G_1(1)(s)ds \right) \left(1 - \int_{t_m}^{t_M} |F_1(-1)(s)|ds \right) &\leq \\ &\leq \left[\int_I |G_0(-1)(s)|ds - 1 \right]_+ \left[\int_{t_m}^{t_M} F_0(1)(s)ds - 1 \right]_+, \end{aligned}$$

where $I = [a, b] \setminus]t_m, t_M[$, which contradicts (3.39). $\square$

4. PROOFS

Theorem 2.1 follows from Lemmas 3.1, 3.3, and 3.4, Theorem 2.1' follows from Lemmas 3.1 and 3.4, and Theorem 2.2 follows from Lemmas 3.2, 3.5, and 3.6.

Proof of Theorem 2.3. Let us first note that, for every operator $T \in \mathcal{H}_{ab}$ and arbitrary $t \in [a, b]$ and $v \in C([a, b]; \mathbb{R})$, we have

$$-T(-v)(t) \leq T(v)(t)$$

if T is subadditive, and

$$T(v)(t) \leq -T(-v)(t)$$

if T is superadditive. Therefore,

$$-\bar{H}_i(-1)(t) \leq \bar{H}_i(1)(t) \quad \text{for } t \in [a, b] \quad (i = 0, 1),$$

and, consequently, inequalities (2.17)–(2.19) yield inequalities (2.3)–(2.7). Thus, the assumptions of Theorem 2.1 are fulfilled, and problem (1.1), (1.2) has at least one solution.

We will show that (1.1), (1.2) has no more than one solution. Let u and v be solutions of (1.1), (1.2). Then, according to (2.15) and (2.16), the function $w(t) = u(t) - v(t)$ for $t \in [a, b]$ satisfies (3.1)–(3.3) with $\delta = 1$ and

$$\begin{aligned} F_0(w)(t) &= \bar{H}_0(w)(t), & F_1(w)(t) &= -\bar{H}_1(-w)(t) & \text{for } t \in [a, b], \\ G_0(w)(t) &= -\bar{H}_0(-w)(t), & G_1(w)(t) &= \bar{H}_1(w)(t) & \text{for } t \in [a, b]. \end{aligned}$$

According to Lemmas 3.3 and 3.4 we have $w \equiv 0$, i. e., $u \equiv v$. $\square$

Theorems 2.3' and 2.4 can be proved analogously.

Proof of Corollary 2.1. Let us put

$$\begin{aligned} H_0(v)(t) &\stackrel{\text{def}}{=} [p(t)]_+ \max\{v(s) : \tau_1(t) \leq s \leq \tau_2(t)\} & \text{for } t \in [a, b], \\ H_1(v)(t) &\stackrel{\text{def}}{=} [p(t)]_- \max\{v(s) : \tau_1(t) \leq s \leq \tau_2(t)\} & \text{for } t \in [a, b], \\ Q(v)(t) &\stackrel{\text{def}}{=} q(t) & \text{for } t \in [a, b], \end{aligned}$$

and $h(v) \stackrel{\text{def}}{=} c$. Then the assumptions of Theorem 2.3 are fulfilled because

$$\begin{aligned} \left[\int_s^t \bar{H}_1(1)(\xi) d\xi - 1 \right]_+ \left[\int_I \bar{H}_1(1)(\xi) d\xi - \lambda \right]_+ &\leq \\ &\leq \frac{1}{4} \left[\int_a^b [p(s)]_- ds - 1 - \lambda \right]_+^2 \quad \text{for } s, t \in [a, b], \quad s \leq t \end{aligned}$$

and

$$\begin{aligned} \left(1 - \int_s^t \bar{H}_0(1)(\xi) d\xi \right) \left(1 - \int_I \bar{H}_0(1)(\xi) d\xi \right) &\geq \\ &\geq 1 - \int_a^b [p(s)]_+ ds \quad \text{for } s, t \in [a, b], \quad s \leq t, \end{aligned}$$

where $I = [a, b] \setminus]s, t[$. □

In a similar way, it can be proved that Corollary 2.1' follows from Theorem 2.3' and Corollary 2.2 follows from Theorem 2.4.

ACKNOWLEDGEMENT

This work was supported by the Grant No. 201/00/D058 of the Grant Agency of the Czech Republic.

REFERENCES

- [1] AZBELEV, N. V., MAKSIMOV, V. P., AND RAKHMATULLINA, L. F.: *Introduction to the Theory of Functional Differential Equations* [in Russian], Nauka, Moscow, 1991.
- [2] AZBELEV, N. V. AND RAKHMATULLINA, L. F.: *Theory of linear abstract functional differential equations and applications*, Mem. Differential Equations Math. Phys., **8** (1996), 1–102.
- [3] BANTSUR, N. R. AND TROFIMCHUK, E. P.: *On existence and stability of periodic and almost-periodic solutions of quasilinear equations with maxima*, Nonlinear Oscillations, **1** (1998), No. 1, 1–5.
- [4] BANTSUR, N. R. AND TROFIMCHUK, E. P.: *On the existence of T-periodic solutions to essentially nonlinear scalar differential equations with maxima*, Ukrain. Math. J., **50** (1998), No. 6, 747–754.
- [5] BRAVYI, E., HAKL, R., AND LOMTATIDZE, A.: *On Cauchy problem for the first order nonlinear functional differential equations of non-Volterra's type*, Czechoslovak Math. J., **52** (2002), No. 4, 673–690.
- [6] BRAVYI, E., HAKL, R., AND LOMTATIDZE, A.: *Optimal conditions for unique solvability of the Cauchy problem for first order linear functional differential equations*, Czechoslovak Math. J., **52** (2002), No. 3, 513–530.
- [7] GELASHVILI, S. M.: *On a boundary value problem for systems of functional differential equations*, Arch. Math. (Brno), **20** (1984), 157–168.
- [8] HAKL, R.: *On some boundary value problems for systems of linear functional differential equations*, El. J. Qualitative Theory of Diff. Equations (1999), No. 10, 1–16.
- [9] HAKL, R., KIGURADZE I., AND PŮŽA, B.: *Upper and lower solutions of boundary value problems for functional differential equations and theorems on functional differential inequalities*, Georgian Math. J., **7** (2000), No. 3, 489–512.

- [10] HAKL, R. AND LOMTATIDZE, A.: *A note on the Cauchy problem for first order linear differential equations with a deviating argument*, Arch. Math. (Brno), **38** (2002), No. 1, 61–71.
- [11] HAKL, R., LOMTATIDZE, A., AND PŮŽA, B.: *New optimal conditions for unique solvability of the Cauchy problem for first order linear functional differential equations*, Math. Bohemica, **127** (2002), No. 4, 509–524.
- [12] HAKL, R., LOMTATIDZE, A., AND PŮŽA, B.: *On a boundary value problem for first order scalar functional differential equations*, Nonlin. Anal.: Theory, Meth. & Appl., **53** (2003), Nos. 3–4, 391–405.
- [13] HAKL, R., LOMTATIDZE, A., AND PŮŽA, B.: *On nonnegative solutions of first order scalar functional differential equations*, Mem. Differential Equations Math. Phys., **23** (2001), 51–84.
- [14] HAKL, R., LOMTATIDZE, A., AND PŮŽA, B.: *On periodic solutions of first order linear functional differential equations*, Nonlinear Anal.: Theory, Meth. & Appl., **49** (2002), No. 7, 929–945.
- [15] HAKL, R., LOMTATIDZE, A., AND PŮŽA, B.: *On periodic solutions of first order nonlinear functional differential equations of non–Volterra’s type*, Mem. Differential Equations Math. Phys., **24** (2001), 83–105.
- [16] HAKL, R., LOMTATIDZE, A., AND ŠREMR, J.: *On an antiperiodic type boundary value problem for first order linear functional differential equations*, Arch. Math. (Brno), **38** (2002), No. 2, 149–160.
- [17] HAKL, R., LOMTATIDZE, A., AND ŠREMR, J.: *On a periodic type boundary value problem for first order linear functional differential equations*, Nonlin. Oscillations, **5** (2002), No. 3, 416–432.
- [18] HAKL, R., LOMTATIDZE, A., AND ŠREMR, J.: *On a periodic–type boundary value problem for first–order nonlinear functional differential equations*, Nonlin. Anal.: Theory, Meth. & Appl., **51** (2002), No. 3, 425–447.
- [19] HAKL, R., LOMTATIDZE, A., AND ŠREMR, J.: *On constant sign solutions of a periodic type boundary value problems for first order scalar functional differential equations*, Mem. Differential Equations Math. and Phys., **26** (2002), 65–90.
- [20] HAKL, R., LOMTATIDZE, A., AND ŠREMR, J.: *Solvability and unique solvability of a periodic type boundary value problem for first order functional differential equations*, Georgian Math. J., **9** (2002), No. 3, 525–547.
- [21] HAKL, R., LOMTATIDZE, A., AND ŠREMR, J.: *Solvability of a periodic type boundary value problem for first order functional differential equations*, Arch. Math. (Brno), to appear.
- [22] HAKL, R., LOMTATIDZE, A., AND ŠREMR, J.: *Some Boundary Value Problems for First Order Scalar Functional Differential Equations*, Folia Facult. Scien. Natur. Univ. Masar. Brunensis, Brno, 2002.
- [23] HALE, J.: *Theory of Functional Differential Equations*, Springer–Verlag, New York–Heidelberg–Berlin, 1977.
- [24] KIGURADZE I. AND PŮŽA, B.: *Boundary Value Problems for Systems of Linear Functional Differential Equations*, Folia Facult. Scien. Natur. Univ. Masar. Brunensis, Brno, 2003.
- [25] KIGURADZE I. AND PŮŽA, B.: *Conti–Opial type theorems for systems of functional differential equations*, Differents. Uravn., **33** (1997), No. 2, 185–194.
- [26] KIGURADZE I. AND PŮŽA, B.: *On boundary value problems for functional differential equations*, Mem. Differential Equations Math. Phys., **12** (1997), 106–113.
- [27] KIGURADZE I. AND PŮŽA, B.: *On boundary value problems for systems of linear functional differential equations*, Czechoslovak Math. J., **47** (1997), No. 2, 341–373.
- [28] KIGURADZE I. AND PŮŽA, B.: *On periodic solutions of systems of linear functional differential equations*, Arch. Math. (Brno), **33** (1997), No. 3, 197–212.
- [29] KIGURADZE I. AND PŮŽA, B.: *On the solvability of boundary value problems for systems of nonlinear differential equations with deviating arguments*, Mem. Differential Equations Math. Phys., **10** (1997), 157–161.
- [30] KIGURADZE I. AND PŮŽA, B.: *On the solvability of nonlinear boundary value problems for functional differential equations*, Georgian Math. J., **5** (1998), No. 3, 251–262.

- [31] KIGURADZE I., PŮŽA, B., AND STAVROULAKIS, I. P.: *On singular boundary value problems for functional differential equations of higher order*, Georgian Math. J., **8** (2001), No. 4, 791–814.
- [32] KOLMANOVSKII, V. AND MYSHKIS, A.: *Introduction to the Theory and Applications of Functional Differential Equations*, Kluwer Academic Publishers, 1999.
- [33] LIZ, E. AND TROFIMCHUK, S.: *Existence and stability of almost periodic solutions for quasilinear delay systems and Halanay inequality*, J. Math. Anal. Appl., **248** (2000), 625–644.
- [34] LOMTATIDZE, A., PŮŽA, B., AND HAKL, R.: *On a periodic boundary value problem for the first order scalar functional differential equation*, Differents. Uravn., **39** (2003), No. 3, 320–327.
- [35] MAGOMEDOV, A. R.: *Theorems on existence and uniqueness of solutions of differential equations with maximal functional parameters*, Arch. Math. (Brno), **28** (1992), 139–154.
- [36] PINTO, M. AND TROFIMCHUK, S.: *Stability and existence of multiple periodic solutions for a quasilinear differential equation with maxima*, Proc. Roy. Soc. Edinburgh Sect. A, **130** (2000), 1103–1118.
- [37] PŮŽA, B.: *Some boundary value problems for nonlinear functional-differential equations*, Differents. Uravn., **37** (2001), No. 6, 761–770.
- [38] RONTO, A. N.: *On the periodic solutions of systems with “maxima”*: Dopov. Nats. Akad. Nauk Ukr., 1999, No. 12, 27–30.
- [39] RONTO, A. AND SAMOILENKO, A.: *On the unique solvability of some linear boundary value problems*. In the book “Nonlinear Analysis and Applications” (R. P. Agarwal and D. O’Regan, eds.), Vol. II, Kluwer Academic Publishers, Dordrecht, 2003.
- [40] SCHWABIK, Š, TVRDÝ, M., AND VEJVODA, O.: *Differential and Integral Equations: Boundary Value Problems and Adjoints*, Academia, Praha, 1979.
- [41] STEPANOV E.: *On solvability of some boundary value problems for differential equations with “maxima”*, Topol. Methods Nonlinear Anal., **8** (1996), 315–326.
- [42] TSITSKISHVILI R. A.: *Unique solvability and correctness of a linear boundary value problem for functional–differential equations*, Rep. Enlarged Sessions Sem. I. N. Vekua Inst. Appl. Math., **5** (1990), No. 3, 195–198.

Author’s Address

Robert Hakl:

MATHEMATICAL INSTITUTE, ACADEMY OF SCIENCES OF THE CZECH REPUBLIC, ŽIŽKOVA 22, 616 62 BRNO, CZECH REPUBLIC

E-mail address: hakl@ipm.cz



ON THE HIGH ORDER ASYMPTOTIC SOLUTION OF CERTAIN WAVE EQUATIONS

VU HOANG LINH

[Received: August 5, 2003]

ABSTRACT. The paper is devoted to the asymptotic approximation for solutions of a class of second order linear ordinary differential equations. A special form of the Liouville–Green type approximation is developed to obtain high order asymptotics. Asymptotic solutions can be constructed easily by using a recursive formula. An error estimation based on Olver’s explicit error bound is proposed. The result and its applications are discussed and illustrated by some examples.

Mathematics Subject Classification: 34E05, 34E20, 65D20

Keywords: Liouville–Green (or JWKB) approximation, wave equations, high order asymptotic solutions

1. INTRODUCTION

LET US CONSIDER the differential equation

$$y''(x) + U(x)y(x) = 0, \quad x_0 \leq x < \infty. \quad (1.1)$$

Assume that the real function $U(x)$ is continuous and has the asymptotics

$$U(x) \sim x^\sigma \sum_{j=0}^{\infty} U_j x^{-j/2}, \quad x \rightarrow \infty, \quad (1.2)$$

where σ and U_j ($j = 0, 1, 2, \dots$) are given; $U_0 \neq 0$; σ is a nonnegative integer. The solutions of problem (1.1), (1.2) are called wave functions. In this paper, we focus on the case when U_0 is positive. The corresponding solutions are called radial wave functions. From now on, we suppose that U_0 is positive. The case with negative U_0 will be discussed in Remark 3.1. Note that no extra condition on the smoothness of $U(x)$ is needed.

We face differential equations of type (1.1), (1.2) frequently when solving problems of mathematical physics, especially those arising in quantum mechanics, see [8–11]. Asymptotic analysis of linear differential equations with irregular singularities has been discussed widely in the literature, e. g., [4, Chapter 5] and [11, Chapter

7]. However, the construction as well as the proof of the existence of asymptotic solutions are very complicated, especially when σ is positive. Most of the results on high order asymptotic approximations were obtained by the singular perturbation method. In the present paper, another approach is used. The asymptotic formula obtained is of the form of the well-known Liouville–Green approximation [5, 11].

First, let us introduce the Liouville transformation, a basic tool of the Liouville–Green approximation theory. Suppose that f is a strictly positive, twice continuously differentiable function on the interval $[a, \infty)$, where $a \geq x_0$. Apply the transformation

$$\xi = \int_a^x f(t) dt, \quad z(\xi) = \sqrt{f(x(\xi))} y(x(\xi)) \quad (1.3)$$

to equation (1.1) on the interval $[a, \infty)$. The differential equation for the new function $z(\xi)$ is

$$z''(\xi) + (1 + \Phi(\xi)) z(\xi) = 0, \quad 0 \leq \xi < \infty, \quad (1.4)$$

where

$$\Phi = (U - f^2) \frac{1}{f^2} - \left(\frac{f'}{2f^{3/2}} \right)' \frac{1}{f^{3/2}}$$

and f' denotes the derivative of f with respect to variable x . Equation (1.4) is called the Liouville normal form of (1.1). If one could find an auxiliary function f such that $\Phi \equiv 0$ on the whole interval, then one would express the general solution of the original equation by the formula

$$y(x) = \frac{1}{\sqrt{f(x)}} \left[A \exp \left(i \int_a^x f(t) dt \right) + B \exp \left(-i \int_a^x f(t) dt \right) \right],$$

where A, B are arbitrary constants. Thus, a natural question arising here is whether the nonlinear second order differential equation

$$\frac{U - f^2}{f^2} - \left(\frac{f'}{2f^{3/2}} \right)' \frac{1}{f^{3/2}} = 0 \quad (1.5)$$

has solution(s) on the interval $[a, \infty)$. A closed formula for a solution f , even if it exists, seems impossible in general. In this paper, we first show the existence of a solution of (1.5). Then, based on the condition (1.2), we construct its asymptotic formula as a power series of the variable x .

The paper is organized in the following way. In the next section, we summarize some basic results in the Liouville–Green approximation theory. As a consequence, a simple but low order asymptotic approximation is constructed. In Section 3 we show the existence of high order asymptotic solutions to (1.1) based on a preliminary result on an auxiliary Riccati differential equation. A method of construction as well as an error estimation are given. We conclude the paper with the discussion of applications and some illustrative examples.

2. THE LIOUVILLE–GREEN APPROXIMATION

The following theorem deals with the so-called Liouville–Green approximation. The statement is a reformulation of Theorem 2.2 in [11, Chapter 6] with respect to large x . We refer to Problem 30 in [4, Chapter 3], where the asymptotic behaviour of the solutions of (1.4) is considered.

Let $f(x)$ be a positive, twice continuously differentiable function on a given interval $[a, \infty)$ and let the error-control function be defined by equation

$$S(x) = \frac{U(x) - f^2(x)}{f(x)} - \left(\frac{f'(x)}{2f^{3/2}(x)} \right)' \frac{1}{f^{1/2}(x)}.$$

Proposition 2.1. *Suppose that the inequality*

$$\int_a^\infty |S(t)| dt < \infty \quad (2.1)$$

holds. Then equation (1.1) has a pair of linearly independent solutions y_+ , y_- such that

$$y_\pm(x) = \frac{1}{\sqrt{f(x)}} \exp\left(\pm i \int_a^x f(t) dt\right) (1 + \varepsilon_\pm(x)), \quad x \rightarrow \infty. \quad (2.2)$$

These solutions form a complex conjugate pair. Moreover, the estimate

$$|\varepsilon_\pm(x)| \leq \exp\left\{\int_x^\infty |S(t)| dt\right\} - 1$$

holds for the error terms $\varepsilon_\pm(x)$.

Approximation (2.2) was obtained first by Liouville and Green, while the explicit error bound is due to Olver [5, 11].

A simple approximation of the solutions of (1.1), (1.2) is obtained as a direct consequence of Proposition 2.1. Let us introduce the function $V(x) = \sqrt{U(x)}$. It is well defined for sufficiently large arguments and has asymptotic representation

$$V(x) \sim x^{\sigma/2} \sum_{j=0}^{\infty} V_j x^{-j/2}, \quad x \rightarrow \infty.$$

From (1.2), the coefficients V_j can be calculated by the recurrence formula

$$V_0 = \sqrt{U_0}, \quad V_1 = U_1/2V_0, \quad V_j = \left(U_j - \sum_{k=1}^{j-1} V_k V_{j-k} \right) / 2V_0, \quad (j = 2, 3, \dots).$$

Let $g(x)$ be defined as the partial sum of the first $(\sigma+2)$ terms of the above asymptotic series

$$g(x) = x^{\sigma/2} \sum_{j=0}^{\sigma+2} V_j x^{-j/2}. \quad (2.3)$$

Since V_0 is positive, so is $g(x)$ on an interval $[a, \infty)$, where a is sufficiently large. Moreover, by elementary calculations, one can check that the relation

$$\left| \frac{U(x) - g^2(x)}{g(x)} - \left(\frac{g'(x)}{2g^{3/2}(x)} \right)' \frac{1}{g^{1/2}(x)} \right| = O\left(\frac{1}{x^{3/2}}\right), \quad x \rightarrow \infty,$$

holds. This means that function g satisfies condition (2.1). We obtain

Proposition 2.2. *Equation (1.1), (1.2) has a pair of independent solutions y_+ , y_- such that*

$$y_{\pm}(x) = \frac{1}{\sqrt{g(x)}} \exp(\pm i G(x)) \left(1 + O\left(\frac{1}{\sqrt{x}}\right) \right), \quad x \rightarrow \infty, \quad (2.4)$$

where $G(x)$ is an antiderivative of $g(x)$:

$$G(x) = x^{\sigma/2+1} \sum_{j=0}^{\sigma+1} \frac{2V_j}{\sigma+2-j} x^{-j/2} + V_{\sigma+2} \ln x.$$

Another proof of this proposition can be found in [7]. A basis of the real solution space of equation (1.1), (1.2) can be created by $\operatorname{Re} y_+$ and $\operatorname{Im} y_+$. Thus, it is easy to obtain an approximation of real solutions of (1.1), (1.2).

Corollary. *Let $y(x)$ be an arbitrary real solution of (1.1), (1.2). Then, there exists a pair of constants $(r_{\infty}, \delta_{\infty})$ such that*

$$y(x) = \frac{r_{\infty}}{\sqrt{g(x)}} \cos\left(G(x) + \delta_{\infty} + O\left(\frac{1}{\sqrt{x}}\right)\right), \quad x \rightarrow \infty, \quad (2.5)$$

where $r_{\infty} > 0$ is uniquely determined and δ_{∞} is unique up to a multiple of 2π . Moreover, any pair $(r_{\infty}, \delta_{\infty})$, $r_{\infty} > 0$, defines the solution uniquely.

Approximations (2.4), (2.5) are simple, but their accuracy is low. In order to improve the accuracy of the Liouville–Green approximation, one may use the so-called higher order phase integral solution (see [5, 11, 12] and references therein). It can be derived as follows. Suppose that $U(x)$ is positive and continuously differentiable as many times as needed. Instead of (1.1), we consider the singularly perturbed equation

$$y''(x) + \frac{U(x)}{\epsilon^2} y(x) = 0,$$

where ϵ is a positive small parameter. Then, a solution f of equation (1.5) is constructed as a power series of ϵ

$$f = \epsilon^{-1}(f_0 + \epsilon^2 f_1 + \epsilon^4 f_2 + \dots),$$

where

$$f_0 = \sqrt{U}, \quad f_1 = -\left(\frac{f_0'}{2f_0^{3/2}}\right)' \frac{1}{2f_0^{1/2}}, \quad \dots$$

If one takes the zero-order approximation, i. e., f_0 only and puts $\epsilon = 1$, then the standard JWKB approximation

$$y_{\pm}(x) \approx U^{-1/4}(x) \exp\left(\pm i \int_{x_0}^x \sqrt{U(t)} dt\right). \quad (2.6)$$

arises. The more terms are taken into consideration, the higher accuracy may be expected. However, the application of higher order approximations seems too complicated since higher derivatives of $U(x)$ are required. In addition, the integral occurring in the approximations in general, or even for the zero-order one cannot be calculated formally. This fact makes the computation of the high order approximations inconvenient and time-consuming.

In the next section, we will show that the expression on the right-hand side of (2.3) can be continued up to the asymptotic series of a solution of (1.5).

3. HIGH ORDER ASYMPTOTIC APPROXIMATIONS

From now on we consider one of the solutions defined in (2.4), say y_+ . One can check that the so-called radiation boundary condition

$$\lim_{x \rightarrow \infty} \frac{x^{-\sigma/2} y'_+(x) - i \sqrt{U_0} y_+(x)}{|x^{-\sigma/2} y'_+(x)| + |y_+(x)|} = 0 \quad (3.1)$$

holds. The following crucial result can be found in [2, 6].

Proposition 3.1. *For sufficiently large x , condition (3.1) can be replaced by an equivalent condition*

$$y'_+(x) = R(x) y_+(x).$$

Here, $R(x)$ is a solution of Cauchy problem

$$\begin{aligned} R'(x) + R^2(x) + U(x) &= 0, \\ \lim_{x \rightarrow \infty} x^{-\sigma/2} R(x) &= i \sqrt{U_0}. \end{aligned} \quad (3.2)$$

For large x , the solution of (3.2) is unique. In addition, it can be represented by an asymptotic series, namely,

$$R(x) \sim x^{\sigma/2} \sum_{j=0}^{\infty} \frac{R_j}{x^{j/2}} \quad x \rightarrow \infty,$$

where the coefficients R_j can be evaluated by the recurrence formula

$$\begin{aligned} R_0 &= i \sqrt{U_0}, \quad R_1 = \frac{-U_1}{2R_0}, \\ R_j &= \begin{cases} \frac{-U_j - \sum_{s=1}^{j-1} R_s R_{j-s}}{2R_0} & \text{if } j < \sigma + 2 \\ \frac{-U_j - \sum_{s=1}^{j-1} R_s R_{j-s} + (j/2 - \sigma - 1) R_{j-\sigma-2}}{2R_0} & \text{if } j \geq \sigma + 2, \end{cases} \end{aligned} \quad (3.3)$$

for $j = 2, 3, \dots$

Now, we arrive at the main results of the paper. Let $a(x), b(x)$ be the real and imaginary part of $R(x)$, respectively,

$$a(x) = \operatorname{Re}(R(x)), \quad b(x) = \operatorname{Im}(R(x)). \quad (3.4)$$

Obviously, these functions are continuously differentiable and they have asymptotic series also when x tends to infinity. Namely, the expansions

$$a(x) \sim x^{\sigma/2} \sum_{j=0}^{\infty} \frac{a_j}{x^{j/2}}, \quad b(x) \sim x^{\sigma/2} \sum_{j=0}^{\infty} \frac{b_j}{x^{j/2}},$$

are valid for large x , where $a_j = \operatorname{Re} R_j$, $b_j = \operatorname{Im} R_j$, $j = 0, 1, \dots$. The Riccati equation in (3.2) is equivalent to the system

$$\begin{aligned} b'(x) + 2a(x)b(x) &= 0, \\ a'(x) + a^2(x) - b^2(x) + U(x) &= 0. \end{aligned} \quad (3.5)$$

Theorem 3.1. *For sufficiently large x , function $b(x)$ defined by (3.4) is positive and it is an exact solution to equation (1.5). Moreover, this solution has asymptotic expansions*

$$b(x) \sim x^{\sigma/2} \sum_{j=0}^{\infty} \frac{b_j}{x^{j/2}}, \quad x \rightarrow \infty,$$

where the coefficients b_j can be evaluated by the recursive formulae

$$\begin{aligned} b_0 &= \sqrt{U_0}, \quad a_0 = 0, \\ b_j &= \frac{U_j + \sum_{s=1}^{j-1} a_s a_{j-s} - \sum_{s=1}^{j-1} b_s b_{j-s} - (j/2 - \sigma - 1)a_{j-\sigma-2}}{2b_0}, \quad j = 1, 2, \dots \end{aligned} \quad (3.6)$$

$$a_j = \frac{-2 \sum_{s=1}^{j-1} a_s b_{j-s} + (j/2 - \sigma - 1)b_{j-\sigma-2}}{2b_0}, \quad j = 1, 2, \dots \quad (3.7)$$

Here, we set $a_s = b_s = 0$ for $s < 0$; $\{a_j\}_{j=0}^{\infty}$ is considered an auxiliary sequence.

Before turning to the proof, we call attention to the fact that the first $(\sigma+3)$ terms of the asymptotic series of $b(x)$ and those of $\sqrt{U(x)}$ coincide. In general, the $(\sigma+4)$ -th coefficients, however, are different.

Proof. Due to the definition, function $b(x)$ behaves like $\sqrt{U_0}x^{\sigma/2}$ as x tends to infinity. Therefore, there exists x_1 such that $b(x) > 0$ for $x \geq x_1$. From the first equation in (3.5), it is clear that $b(x)$ is twice continuously differentiable. Using (3.5), one can check that

$$\frac{U(x) - b^2(x)}{b^2(x)} - \left(\frac{b'(x)}{2b^{3/2}(x)} \right)' \frac{1}{b^{3/2}(x)} = \frac{a'(x) + a^2(x) - b^2(x) + U(x)}{b^2(x)} = 0$$

for $x \geq x_1$. It means exactly that $b(x)$ is a solution to equation (1.5). Moreover, it is unique if the condition $\lim_{x \rightarrow \infty} x^{-\sigma/2} b(x) = \sqrt{U_0}$ is satisfied. Finally, we get the asymptotic series for $b(x)$ directly from (3.3) by taking the definition of $a(x)$ and $b(x)$ into account. $\square$

Due to the considerations in Section 1, the next results are obtained immediately.

Theorem 3.2. *On the interval $[x_1, \infty)$, equation (1.1), (1.2) has a pair of solutions*

$$y_{\pm}(x) = b^{-1/2}(x) \exp\left(\pm i \int_{x_1}^x b(t) dt\right).$$

Furthermore, this solution pair is the same (up to a constant factor) as that appearing in Proposition 2.2.

Let $B(x)$ be the antiderivative of $b(x)$ defined by the asymptotics

$$B(x) \sim x^{\sigma/2} \sum_{\substack{j=0 \\ j \neq \sigma+2}}^{\infty} \frac{2b_j}{(\sigma-j+2)x^{j/2-1}} + b_{\sigma+2} \ln x, \quad x \rightarrow \infty.$$

Corollary. *For any real solution $y(x)$ of (1.1), (1.2) defined on $[x_1, \infty)$, there exists a pair of constants $(r_{\infty}, \delta_{\infty})$, $r_{\infty} > 0$, such that*

$$y(x) = \frac{r_{\infty}}{\sqrt{b(x)}} \cos(B(x) + \delta_{\infty}).$$

If the cosine function is replaced by the sine, we obtain another linearly independent solution. One cannot make use of the above results yet because the closed formulae of function $b(x)$ and $B(x)$ are not available. However, their asymptotic series can be calculated as we saw above. Then, by fixing an arbitrary index $N \geq \sigma + 2$, there exist a pair of linearly independent solutions of problem (1.1), (1.2)

$$y_{\pm}(x) = \frac{1}{\sqrt{b_N(x)}} \exp(\pm i B_N(x)) (1 + \varepsilon_{\pm}(x)), \quad (3.8)$$

where

$$b_N(x) = x^{\sigma/2} \sum_{j=0}^N \frac{b_j}{x^{j/2}}, \quad B_N(x) = x^{\sigma/2} \sum_{\substack{j=0 \\ j \neq \sigma+2}}^N \frac{2b_j}{(\sigma-j+2)x^{j/2-1}} + b_{\sigma+2} \ln x,$$

and

$$\varepsilon_{\pm}(x) = O\left(x^{(\sigma-N+1)/2}\right), \quad x \rightarrow \infty.$$

It is clear that the order of accuracy of approximation (3.8) can be easily increased by choosing a larger and larger index N . In general, by setting $N \geq \sigma + 3$, we immediately get approximations of accuracy higher order than those of approximation (2.6).

For completeness, we give an estimation method for the relative error of the complex approximations based on Olver's explicit error bound. Bounds for the absolute

and relative error of real approximations can be derived without any difficulty. For the sake of brevity, we put

$$P(x) = \frac{3}{4}[b'_N(x)]^2 - \frac{1}{2}b''_N(x)b_N(x) - b_N^2(x)[b_N^2(x) - U(x)], \quad Q(x) = b_N^3(x).$$

It is easy to check that the error-control function satisfies the formula

$$S(x) = \frac{P(x)}{Q(x)} = O(x^{(\sigma-N-1)/2}), \quad x \rightarrow \infty$$

and its leading coefficient is $2b_{N+1}$. It yields immediately the asymptotic estimate for Olver's upper bound of the error function

$$\exp\left(\int_x^\infty |S(t)| dt\right) - 1 \sim \frac{2|b_{N+1}|}{N-1-\sigma} x^{(\sigma+1-N)/2}, \quad x \rightarrow \infty.$$

In order to get an explicit error bound, one can proceed as follows. First, we consider the case when $U(x)$ can be expressed as a finite sum, namely

$$U(x) = x^\sigma \sum_{j=0}^M U_j x^{-j/2}, \quad (3.9)$$

where M is a non-negative integer. We note that this case arises frequently in applications, see [3, 8–12]. Then, both the numerator $P(x)$ and the denominator $Q(x)$ can be expressed as finite sums consisting of the powers of x . The following method [1] provides simple estimates for $P(x)$ and $Q(x)$. Suppose that we have to estimate the sum

$$C(x) = \sum_{i=0}^K \frac{c_i}{x^{i/2}}.$$

Without loss of generality we assume the leading coefficient $c_0 > 0$. Let m and n be the numbers of positive and negative coefficients c_i , $i = 1, 2, \dots, K$, respectively. Take arbitrary parameters $\alpha > 0$ and β , $0 < \beta < 1$, and put

$$x_1 = \max_{c_i > 0} \left\{ \left(\frac{mc_i}{\alpha c_0} \right)^{2/i} \right\}; \quad x_2 = \max_{c_j < 0} \left\{ \left(\frac{nc_j}{-\beta c_0} \right)^{2/j} \right\}. \quad (3.10)$$

Then, we obtain the estimates

$$C(x) \leq (1 + \alpha)c_0 \quad \text{for } x \geq x_1$$

and

$$C(x) \geq (1 - \beta)c_0 \quad \text{for } x \geq x_2.$$

Thus, based on the coefficients of $P(x)$ and $Q(x)$, we can calculate $\hat{x}_1$, $\hat{x}_2$ such that the estimates

$$|P(x)| \leq 2(1 + \alpha)b_0^3 |b_{N+1}| x^{2\sigma-(N+1)/2}$$

and

$$|Q(x)| \geq (1 - \beta)^3 b_0^3 x^{3\sigma/2}$$

hold with an arbitrary pair α, β as above. Then, for $x \geq \max\{\hat{x}_1, \hat{x}_2\}$, we obtain the explicit estimate

$$|S(x)| \leq \frac{2(1+\alpha)}{(1-\beta)^3} |b_{N+1}| x^{(\sigma-N-1)/2}.$$

Hence, for sufficiently large x , we obtain an explicit bound for the error terms

$$|\varepsilon_{\pm}(x)| \leq \exp\left(\int_x^{\infty} |S(t)| dt\right) - 1 \lesssim \frac{4|b_{N+1}|(1+\alpha)}{(1-\beta)^3(N-\sigma-1)} x^{(1+\sigma-N)/2}.$$

When $U(x)$ has the general asymptotic expansion (1.2), we suppose that

$$U(x) = U_M(x) + r(x),$$

where $U_M(x)$ is as the right-hand side of (3.9) with a sufficiently large M and there exists a constant $\hat{r}$ such that the estimate

$$|r(x)| \leq \hat{r} x^{(2\sigma-M-1)/2}$$

holds for the remainder term. Then, we have

$$S(x) = \left[\frac{U_M(x) - b_N^2(x)}{b_N(x)} - \left(\frac{b'_N(x)}{2b_N^{3/2}(x)} \right)' \frac{1}{b_N^{1/2}(x)} \right] + \frac{r(x)}{b_N(x)}.$$

A procedure similar to the above gives an estimate for the first term, while the simple estimate

$$\left| \frac{r(x)}{b_N(x)} \right| \leq \frac{\hat{r}}{(1-\beta)b_0} x^{(\sigma-M-1)/2}$$

holds for the second term assuming x is large enough. Therefore, an explicit error bound can also be obtained in the general case.

Remark 3.1. A similar asymptotic analysis can be carried out for the exponential-type solutions of the equation

$$y''(x) - U(x)y(x) = 0, \quad x_0 \leq x < \infty,$$

where the same assumption on the function $U(x)$ holds as in (1.2). The new sequences $\{\bar{a}_i\}_{i=0}^{\infty}, \{\bar{b}_i\}_{i=0}^{\infty}$ can be generated with the change for

$$\begin{aligned} \bar{b}_0 &= -\sqrt{U_0}, \quad \bar{a}_0 = 0, \\ \bar{b}_j &= \frac{-U_j + \sum_{s=1}^{j-1} \bar{a}_s \bar{a}_{j-s} + \sum_{s=1}^{j-1} \bar{b}_s \bar{b}_{j-s} - (j/2 - \sigma - 1) \bar{a}_{j-\sigma-2}}{-2\bar{b}_0}, \quad j = 1, 2, \dots \\ \bar{a}_j &= \frac{2 \sum_{s=1}^{j-1} \bar{a}_s \bar{b}_{j-s} - (j/2 - \sigma - 1) \bar{b}_{j-\sigma-2}}{-2\bar{b}_0}, \quad j = 1, 2, \dots \end{aligned}$$

Here, we set $\bar{a}_s = \bar{b}_s = 0$ for $s < 0$. Taking the finite sums $\bar{b}_N(x), \bar{a}_N(x)$ defined analogously to those in the oscillatory case, one can show that equation (3.10) has a

pair of solutions approximated by

$$y_{\pm}(x) = \frac{1}{\sqrt{-b_N(x)}} \exp(\pm B_N(x)) \left(1 + O(x^{(\sigma-N+1)/2})\right), \quad x \rightarrow \infty.$$

Error estimates for these approximations can be obtained similarly to the oscillatory case.

4. APPLICATIONS

Among radial wave functions occurring in practical problems, a number of well-known special functions can be defined as solutions of differential equations of type (1.1), (1.2). Therefore, the results of Section 3 seem to be useful. For a particular solution, the value of r_{∞} is determined by some normalization condition in general. If such a condition is not prescribed or is not essential, one can fix $r_{\infty} = 1$. The phase shift δ_{∞} , which plays a very important role in applications, separates definitely a particular solution from others. However, the exact formula for δ_{∞} can only be obtained in several special cases. In most cases, δ_{∞} should be approximated somehow, for example, see [11, Chapter 12] or [9]. For the exact values of the pair $\{r_{\infty}, \delta_{\infty}\}$ in case of some well-known radial wave functions, such as the examples below, see [3].

Example 4.1 (Airy's differential equation). First, we consider the oscillatory case

$$y''(x) + xy(x) = 0, \quad x \geq x_0. \quad (4.1)$$

Using the algorithm for the computation of $\{b_j\}_{j=0}^{\infty}$ described in Section 3, we calculate several coefficients of the asymptotic series of $b(x)$ and get

$$b(x) \sim x^{1/2} + \frac{5}{32}x^{-5/2} - \frac{1105}{2048}x^{-11/2} + \dots, \quad x \rightarrow \infty.$$

Consequently, we have

$$B(x) \sim \frac{2}{3}x^{3/2} - \frac{10}{96}x^{-5/2} + \frac{2210}{18432}x^{-9/2} + \dots, \quad x \rightarrow \infty.$$

If one takes the first two terms of $b(x)$ and $B(x)$, one gets an approximate formula of the real solutions

$$y(x) \approx \frac{r_{\infty}}{\sqrt{x^{1/2} + \frac{5}{32}x^{-5/2}}} \cos\left(\frac{2}{3}x^{3/2} - \frac{10}{96}x^{-5/2} + \delta_{\infty}\right)$$

for large x . Each constant pair $\{r_{\infty}, \delta_{\infty}\}$ defines a solution uniquely. For estimating the error, we calculate

$$P(x) = -\frac{1105}{1024}x^{-4} - \frac{25}{128^2}x^{-7} - \frac{5^4}{32^4}x^{-10}, \quad Q(x) = \left(x^{1/2} + \frac{5}{32}x^{-5/2}\right)^3.$$

We obtain a simple estimate for the error-control function

$$S(x) \approx -\frac{1105}{1024}x^{-11/2},$$

for sufficiently large x .

Now, we consider the exponential case

$$y''(x) - xy(x) = 0, \quad x \geq x_0.$$

We calculate the sequences $\{\bar{a}_i\}_{i=0}^\infty, \{\bar{b}_i\}_{i=0}^\infty$ and obtain, for example, an approximation for exponentially decaying solutions as

$$y(x) \approx \frac{r_\infty}{\sqrt{x^{1/2} - \frac{5}{32}x^{-5/2}}} \exp\left(-\frac{2}{3}x^{3/2} - \frac{10}{96}x^{-5/2}\right).$$

Each nonzero constant r_∞ defines a solution uniquely. The error-control function has similar asymptotics as in the oscillatory case.

It is clear that in both the oscillatory and exponential cases, one can easily take some additional terms and obtain higher order approximate solutions.

Example 4.2 (Bessel's differential equation in normal form). Consider the equation

$$y''(x) + \left(1 - \frac{p^2 - 1/4}{x^2}\right)y(x) = 0, \quad 0 < x < \infty, \quad (4.2)$$

where $p \geq 0$ is a parameter. Using the algorithm for the computation of $\{b_j\}_{j=0}^\infty$ described in Section 3, we calculate several coefficients of the asymptotic series of $b(x)$ and get

$$b(x) \sim 1 - \frac{n-1}{8x^2} - \frac{(n-1)(n-25)}{128x^4} + \dots, \quad x \rightarrow \infty,$$

where $n = 4p^2$. Consequently,

$$B(x) \sim x + \frac{n-1}{8x} + \frac{(n-1)(n-25)}{384x^3} + \dots, \quad x \rightarrow \infty.$$

Then we obtain, for example, an approximate formula for the solutions of (4.2)

$$y(x) \approx \frac{r_\infty}{\sqrt{1 - \frac{n-1}{8x^2} - \frac{(n-1)(n-25)}{128x^4}}} \cos\left(x + \frac{n-1}{8x} + \frac{(n-1)(n-25)}{384x^3} + \delta_\infty\right),$$

which holds for large x .

We underline that if the expressions we obtained here are expanded using the asymptotic representations of $b(x)$ and $B(x)$, then one arrives at the other frequently used asymptotic expansions of Bessel functions, which can be found in [3, 11].

Example 4.3 (Equation of parabolic cylinder functions). Consider the equation

$$y''(x) + \left(\frac{x^2}{4} - a\right)y(x) = 0, \quad 0 \leq x < \infty, \quad (4.3)$$

where a is a parameter. Analogously to the above, we get asymptotic series for $b(x)$ and $B(x)$ as follows

$$b(x) \sim \frac{x}{2} - \frac{a}{x} - \frac{a^2 - 3/4}{x^3} - \frac{2a^3 - 19a/2}{x^5} \dots, \quad x \rightarrow \infty,$$

$$B(x) \sim \frac{x^2}{4} - a \ln x + \frac{a^2 - 3/4}{2x^2} + \frac{2a^3 - 19a/2}{4x^4} \dots, \quad x \rightarrow \infty.$$

Similarly to the previous examples, high order approximate formulae for the parabolic cylinder functions can be given.

Now let the pair $\{r_\infty, \delta_\infty\}$ of a particular solution be given. Then, for sufficiently large arguments, the solution can be approximated with high accuracy using (3.8). Even in the case when equations are more complicated than (4.1), (4.3), the coefficients b_j are available easily by the recursive formulae (3.6), (3.7). Evaluation of the approximations can be quickly achieved and in a less complicated way since no numerical integration is necessary. This advantage is extremely important in situations when a large number of function evaluations are needed, such as in computation of integrals containing solutions.

For sufficiently large arguments, the errors of the approximation can be estimated by computing some additional terms of the asymptotic series of $b(x)$ and $B(x)$. However, one should be very careful with the use of (3.8) because of its asymptotic nature: if the argument is not large enough, increasing the index N may actually lead to less accurate numerical results. To ensure the accuracy of numerical function evaluation, one should carry out the error estimation process as described in Section 3. For “small” arguments, approximation of the solutions can be obtained in other ways such as polynomial approximation or numerical integration.

For the numerical integration of the solutions of (1.1), (1.2), a variant of the so-called amplitude-phase methods (or the Prüfer methods [12]) has been proposed, see [7–9]. The application of the method to quantum mechanic problems was discussed in [8]. We do not go into detail here, just mention that a pair of auxiliary functions, the preconditioning function and the phase-shifting one, plays a very important role in the effective realization of the method [9]. One can set these functions using $b_N(x)$ and $B_N(x)$, respectively, with an appropriately chosen index N . The method then becomes very efficient: beside advantages gained from large admissible stepsizes in automatic integration, the interval of numerical integration can be essentially shortened. Consequently, a high accuracy of function evaluations can be easily achieved. This fact has been confirmed by numerical experiments (see, for instance, [8, 9]). In addition to the error estimation of the approximation (3.8), another error estimation with respect to the approximation of the amplitude and phase-shift constants $(r_\infty, \delta_\infty)$ was analyzed in [9, 10].

REFERENCES

- [1] ABRAMOV, A. A. AND BALLA, K.: *Approximate solutions, based on comparison theorems, of scalar and matrix Riccati equations in an infinite interval*, Comput. Math. Math. Phys., **33** (1993), No. 1, 29–43; translated from Zh. Vychisl. Mat. Mat. Fiz., **33** (1993), No. 1, 35–51.
- [2] ABRAMOV, A. A. AND KONYUKHOVA, N. B.: *Transfer of admissible boundary conditions from a singular point for systems of linear ordinary differential equations*, Sov. J. Numer. Anal. Math. Modelling, **1** (1986), No. 4, 245–265.
- [3] ABRAMOWITZ, A. AND STEGUN, I. A.: *Handbook of mathematical functions*, Dover, N.Y., 1968.
- [4] CODDINGTON, E. A. AND LEVINSON, N.: *Theory of ordinary differential equations*, McGraw-Hill, New York, 1955.
- [5] FEDORYUK, M. V.: *Asymptotic methods in analysis*. In: *Analysis I, Integral representation and asymptotic methods* (R. V. Gamkrelidze, ed.), Springer-Verlag, 1989.
- [6] KONYUKHOVA, N. B.: *On admissible boundary conditions at irregular singular point for systems of linear ordinary differential equations*, U.S.S.R. Comput. Math. Math. Phys., **23** (1983), No. 4, 23–35; translated from Zh. Vychisl. Mat. Mat. Fiz., **23** (1983), No. 4, 806–824.
- [7] KONYUKHOVA, N. B., MASALOVICH, S. YE., AND STAROVEROVA, I. B.: *Computation of rapidly oscillating eigenfunctions of continuous spectrum and their improper integrals*, Russian J. Comput. Math. Math. Phys., **35** (1995), No. 3, 287–302; translated from Zh. Vychisl. Mat. Mat. Fiz., **35** (1995), No. 3, 360–379.
- [8] KONYUKHOVA, N. B., LINH, V. H., AND STAROVEROVA, I. B.: *Modifications of the method of phase functions as applied to singular problems in quantum physics*, Comput. Math. Math. Phys., **39** (1999), No. 3, 468–498; translated from Zh. Vychisl. Mat. Mat. Fiz. **39** (1999), No. 3, 492–522.
- [9] LINH, V. H.: *On some questions arising in numerical realization of amplitude-phase methods*, J. Numerical Algorithms, **17** (1998), No. 1-2, 171–191.
- [10] LINH, V. H.: *Error estimates for the amplitude-phase method in the evaluation of radial wave functions*, Acta Sci. Math. (Szeged), **63** (1997), No. 3-4, 657–670.
- [11] OLVER, F. W. J.: *Asymptotics and special functions*, Academic Press, New York, 1974.
- [12] PRYCE, J. D.: *Numerical solution of Sturm–Liouville problems*, Clarendon Press, Oxford, New York & Tokyo, 1993.

Author's Address

Vu Hoang Linh:

FACULTY OF MATHEMATICS, MECHANICS, AND INFORMATICS, UNIVERSITY OF NATURAL SCIENCES, VIETNAM
 NATIONAL UNIVERSITY, 334 NGUYEN TRAI STR., THANH XUAN, HANOI
E-mail address: vhlinh@hn.vnn.vn



ON THE UNIQUE SOLVABILITY OF LINEAR EQUATIONS DETERMINED BY MONOTONE DECOMPOSABLE OPERATORS

ANDREI RONTÓ

[Received: January 5, 2004]

ABSTRACT. Conditions sufficient for the unique solvability of the linear equation

$$u = A_1 u - A_2 u + z$$

for arbitrary z are established, where $A_i : X \rightarrow X$, $i = 1, 2$, are compact linear operators preserving a wedge in the Banach space X . The conditions obtained have the form of one-sided order relations satisfied by the values of A_1 and A_2 at certain “strongly positive,” in a sense, elements of the space.

Mathematics Subject Classification: 46B40, 34K30, 46N20

Keywords: Wedge, operator preserving a wedge, spectral radius, test element

1. INTRODUCTION

IT WILL NOT BE AN EXAGGERATION to say that the majority of linear equations arising in applications can be formulated as an abstract equation

$$u = Au + z, \quad (1.1)$$

where $A : X \rightarrow X$ is a bounded linear operator in a suitably chosen Banach space X . Such an equation can often be represented in the form

$$u = A_1 u - A_2 u + z, \quad (1.2)$$

where A_1 and A_2 are monotone in some or another sense. A natural definition of monotonicity is based upon the notion of a wedge (or, alternatively, a linear semi-group) in a Banach space, which dates back to the works of M. Krein [5]; this approach will be used below. Thus, we consider here equation (1.2) on the assumption that the operators A_1 and A_2 preserve a certain wedge in the given space X (see Section 2 for the basic definitions).

The standard condition

$$r(A_1 - A_2) < 1 \quad (1.3)$$

Supported in part by the NATO Science Fellowships Programme for the Czech Republic, Grant 7/2003.

for the spectral radius $r(A_1 - A_2)$ of the operator $A_1 - A_2$, as is well-known, guarantees the existence and boundedness of the inverse operator

$$(\mathbf{1}_X - A_1 + A_2)^{-1} : X \rightarrow X \quad (1.4)$$

and, hence, the existence of a unique solution of equation (1.2) for an arbitrary z from X and its continuous dependence on z . This proposition is, in a sense, unimprovable, but has the disadvantage that the fulfilment of inequality (1.3) can be verified directly only in exceptional cases. It is, therefore, desirable to have some statements guaranteeing estimate (1.3) on the base of information which can be comparatively easily extracted from the definition of the operators considered. For positive, in a sense, linear operators $A : X \rightarrow X$, namely, for those preserving a suitable cone in X , there is a series of efficient theorems (see, e. g., [8, 4, 3]) allowing one to derive bounds for the spectral radius of A from relations involving the values of A at a single non-zero element of X (such an element, of course, should be suitably chosen, in accordance with the conditions assumed). This group of theorems is sometimes referred to as *Krein's method of test elements*.

The theorems mentioned above, however, have a certain limitation which restricts the field of their applicability. Namely, it is assumed that the operator under consideration preserves a cone satisfying some additional conditions (in particular, possessing the property of normality). Such conditions, despite their considerable generality, are however not satisfied in many important cases where the “positivity” is lacking. To overcome this difficulty, an approach different from those used in the works cited was suggested in [7, 6]. Some of the results obtained therein are used in the present note (see Section 2.3).

Here, we are interested in conditions sufficient for the unique solvability of equation (1.2) with A_1 and A_2 preserving a wedge P in the space X , i. e., such that

$$A_1(P) \cup A_2(P) \subset P. \quad (1.5)$$

The difference between the approach of [6] and that of the present note is that here conditions are imposed separately on the “positive” and “negative” parts of the equation, for which purpose a certain decomposition trick is applied (Section 3). The proofs of the results of Section 4 use the “positive” version of Theorem 4.1 from [6].

2. PRELIMINARIES

Let X be a Banach space over the field $\mathbb{R}$ and P be a wedge in X . Recall that a closed set $P \subset X$ is called a *wedge* [3] (or a *linear semigroup* [5]) if

$$\alpha_1 P + \alpha_2 P \subset P$$

for all $\{\alpha_1, \alpha_2\} \subset (0, +\infty)$, where, by definition,

$$\alpha_1 \Omega + \alpha_2 \Omega := \{\alpha_1 u_1 + \alpha_2 u_2 \mid \{u_1, u_2\} \subset \Omega\}. \quad (2.1)$$

for an arbitrary set $\Omega \subset X$ and all real α_1 and α_2 .

The linear manifold

$$\text{bl } P := P \cap (-P) \quad (2.2)$$

is referred to as the *blade* [3] of the wedge P . A wedge having the trivial blade is called a *cone* [5].

2.1. Preorderings. Every wedge P allows one to introduce a natural preordering “ $\geq_P$ ” in the space X by putting

$$u_1 \geq_P u_2 \text{ if, and only if } u_1 - u_2 \in P.$$

We shall also write $u_1 \leq_P u_2$ if, and only if $u_2 \geq_P u_1$. Note that the relations $u_1 \geq_P u_2$ and $u_2 \geq_P u_1$ satisfied simultaneously, generally speaking, do not imply the equality $u_1 = u_2$, unless $\text{bl } P = \{0\}$.

A wedge P is said to be *solid* [5] if its interior is non-empty. In this case, following [5], we define the relation “ $\gg_P$ ” by setting

$$u_1 \gg_P u_2 \text{ if, and only if } u_1 - u_2 \text{ is an interior element of } P.$$

An element u from X is said to be *f-measurable* [2, 3] with respect to P if there exists a constant $\beta \in [0, +\infty)$ such that

$$-\beta f \leq_P u \leq_P \beta f.$$

Here, f is a certain fixed non-zero element from X .

Proposition 2.1. *The set*

$$X_P(f) := \{u \in X \mid u \text{ is } f\text{-measurable with respect to } P\} \quad (2.3)$$

is non-empty if, and only if $f \geq_P 0$. Furthermore, for an arbitrary element f from P , the set $X_P(f)$ is a linear manifold in X .

This statement is readily obtained directly from the definitions formulated above.

2.2. Strict inequalities. Let H be a certain linear manifold in X (not necessarily a closed one).

Definition 2.2 ([6]). For $\{f_1, f_2\} \subset X$, we write $f_1 <_{P;H} f_2$ if, and only if the inclusion

$$X_P(f_2 - f_1) \supset H$$

is true.

Similarly, we write $f_1 >_{P;H} f_2$ if, and only if $f_2 <_{P;H} f_1$. Note that, by virtue of Proposition 2.1, $X_P(f)$ is a linear manifold for all f from P .

Definition 2.3 ([6]). For $\{f_1, f_2\} \subset X$, we write $f_1 <_P f_2$ if, and only if the equality

$$X_P(f_2 - f_1) = X$$

is true.

Analogously, we write $f_1 >_P f_2$ if, and only if $f_2 <_P f_1$.

Example 2.4. If, for instance, $X = C([a, b], \mathbb{R})$,* the wedge P is defined by the formula

$$P = \{u \in C([a, b], \mathbb{R}) \mid u(t) \geq 0 \text{ for all } t \text{ from } [a, b]\},$$

and f is the superlinear function given by the equality

$$f(t) = |t - \tau|^\alpha, \quad t \in [a, b],$$

with some $\tau \in [a, b]$ and $\alpha \in [1, +\infty)$, then, as is easy to see, the corresponding set (2.3) is described by the formula

$$X_P(f) = \left\{ u \in C([a, b], \mathbb{R}) \mid \sup_{t \in [a, b]} |(t - \tau)^{-\alpha} u(t)| < +\infty \right\}.$$

This set, clearly, forms a proper linear submanifold in $C([a, b], \mathbb{R})$. Every function belonging to the set mentioned, in particular, vanishes at the point τ .

The relations “ $<_{P;H}$ ” and “ $<_P$ ” provide us a kind of “strict inequalities” whose properties resemble, to some extent, those of the componentwise sign “ $>$ ” in $\mathbb{R}^n$, $n \in \mathbb{N}$. For example, it is not difficult to verify that

$$f >_P 0 \text{ implies that } f \geq_P 0$$

and

$$f \geq_P g \text{ and } g >_P 0 \text{ imply } f >_P 0.$$

An efficient condition sufficient for the relation

$$f >_P 0 \tag{2.4}$$

to be fulfilled is provided by the following lemma.

Lemma 2.5. *If P is a solid wedge in X and an element f is such that $f \gg_P 0$, then f satisfies relation (2.4).*

Proof. A statement equivalent to the equality $X_P(f) = X$ for f lying in the interior of P is well-known, e. g., from [5]. $\square$

2.3. Spectrum of monotone decomposable operators. Let us first introduce the following standard

Definition 2.6. An operator $A : X \rightarrow X$ is said to *preserve* the wedge P if $A(P) \subset P$.

The following statement has recently been obtained in [6].

Theorem 2.7 ([6]). *Let P be a wedge in X and $A_1 : X \rightarrow X$, $A_2 : X \rightarrow X$ be completely continuous linear operators preserving P , satisfying the condition*

$$\text{bl } P \subset \ker(A_1 - A_2), \tag{2.5}$$

and, moreover, such that

$$\text{im}(A_1 - A_2) \subset H,$$

*The symbol $C([a, b], \mathbb{R})$ here stands for the Banach space of all continuous functions $u : [a, b] \rightarrow \mathbb{R}$ equipped with the usual norm $\|u\| := \max_{t \in [a, b]} |u(t)|$.

where H is a certain given linear manifold from X . If, in addition, the inequality

$$\alpha f \geq_P A_1 f + A_2 f \quad (2.6)$$

holds with some constant $\alpha \in [0, +\infty)$ and element $f \in P \setminus \{0\}$ satisfying the condition

$$f \succ_{P;H} 0, \quad (2.7)$$

then the spectral radius $r(A_1 - A_2)$ of the operator $A_1 - A_2$ admits the estimate

$$r(A_1 - A_2) \leq \alpha. \quad (2.8)$$

Theorem 2.7 implies, in particular, the following corollaries.

Corollary 2.8. Let P be a solid wedge in X and $A_1 : X \rightarrow X$, $A_2 : X \rightarrow X$ be completely continuous linear operators preserving P , satisfying condition (2.5), and such that inequality (2.6) is true with some constant $\alpha \in [0, +\infty)$ and element $f \in P \setminus \{0\}$ possessing the property

$$f \gg_P 0. \quad (2.9)$$

Then the spectral radius $r(A_1 - A_2)$ of the operator $A_1 - A_2$ admits estimate (2.8).

Proof. In view of Lemma 2.5, it suffices to apply Theorem 2.7 with $H = X$. $\square$

Corollary 2.9. Let P be a wedge in X and $A : X \rightarrow X$ be a completely continuous linear operator preserving P and satisfying the conditions

$$\text{bl } P \subset \ker A \quad (2.10)$$

and

$$\text{im } A \subset H, \quad (2.11)$$

where H is a certain linear manifold in X . If, moreover, there exists an $\alpha \in [0, +\infty)$ and a certain element f satisfying condition (2.7) and the relation

$$\alpha f \geq_P Af, \quad (2.12)$$

then the spectral radius $r(A)$ of the operator A admits the estimate

$$r(A) \leq \alpha. \quad (2.13)$$

Proof. Application of Theorem 2.7 with $A_1 = A$ and $A_2 = 0$. $\square$

Corollary 2.10 ([7]). Assume that P is a solid wedge in X and $A : X \rightarrow X$ is a completely continuous linear operator preserving P , satisfying condition (2.10) and, moreover, such that relation (2.12) is true with certain $\alpha \in [0, +\infty)$ and an $f \in P$ possessing property (2.9). Then the spectral radius of A admits estimate (2.13).

Proof. It is sufficient to apply Corollary 2.8 with $A_1 = A$ and $A_2 = 0$. $\square$

3. A DECOMPOSITION TECHNIQUE

Let X be a Banach space, P be a wedge in X , and $A_1, A_2 : X \rightarrow X$ be bounded linear operators such that inclusion (1.5) is true.

3.1. The decomposition lemma. The following simple lemma allows one to rewrite equation (1.2) in a different form which may sometimes prove more convenient due to the presence of assumption (1.5).

Lemma 3.1. *If a pair $(u_1, u_2) \in X \times X$ is a solution of the system*

$$u_1 = A_1 u_1 + A_2 u_2 + z_1, \quad (3.1)$$

$$u_2 = A_2 u_1 + A_1 u_2 + z_2 \quad (3.2)$$

with some $(z_1, z_2) \in X \times X$, then the element

$$u = u_1 - u_2 \quad (3.3)$$

is a solution of equation (1.2) with $z := z_1 - z_2$. Conversely, if $u \in X$ is a solution of equation (1.2) with a certain $z \in X$, then there exist pairs $(u_1, u_2) \in X^2$ and $(z_1, z_2) \in X^2$ such that equalities (3.3), (3.1), and (3.2) are satisfied.

Proof. If (u_1, u_2) satisfies (3.1), (3.2), then, obviously,

$$u_1 - u_2 = (A_1 - A_2) u_1 + (A_2 - A_1) u_2 + z_1 - z_2$$

and, therefore, equality (1.2) is satisfied with $u := u_1 - u_2$ and $z := z_1 - z_2$.

Conversely, assume that u is a solution of equation (1.2). Let u_1 and u_2 be some elements of X such that equality (3.3) is true. Then (3.1) and (3.2) are satisfied with $z_1 := z + u_2 - A_1 u_2 - A_2 u_1$ and $z_2 := u_2 - A_1 u_2 - A_2 u_1$. $\square$

Despite its extreme simplicity, Lemma 3.1 nevertheless allows one to obtain meaningful results on the unique solvability of equation (1.2).

3.2. Unique solvability of system (3.1), (3.2). The following propositions provide conditions sufficient for the unique solvability of system (3.1), (3.2) with arbitrary values of z_1 and z_2 .

Proposition 3.2. *Let A_1 and A_2 be bounded linear operators satisfying the conditions*

$$r(A_1) < 1 \quad (3.4)$$

and

$$r(A_1 + A_2(\mathbf{1}_X - A_1)^{-1}A_2) < 1. \quad (3.5)$$

Then system (3.1), (3.2) is uniquely solvable for arbitrary $\{z_1, z_2\} \subset X$, and its solution (u_1, u_2) can be represented by the formula

$$u_1 = \hat{A}_1 z_1 + \hat{A}_1 A_2 (\mathbf{1}_X - A_1 - A_2 \hat{A}_1 A_2)^{-1} [z_2 + A_2 \hat{A}_1 z_1], \quad (3.6)$$

$$u_2 = (\mathbf{1}_X - A_1 - A_2 \hat{A}_1 A_2)^{-1} [z_2 + A_2 \hat{A}_1 z_1], \quad (3.7)$$

where

$$\hat{A}_1 := (\mathbf{1}_X - A_1)^{-1}. \quad (3.8)$$

If, moreover, A_1 and A_2 preserve the wedge P , then the solution of system (3.1), (3.2) satisfies the condition $(u_1, u_2) \in P^2$ whenever $(z_1, z_2) \in P^2$.

Proof. It follows from (3.4) that operator (3.8) is well-defined and, hence, (3.1) is equivalent to the equation

$$u_1 = \hat{A}_1 [z_1 + A_2 u_2]. \quad (3.9)$$

Substituting (3.9) into (3.2), we obtain

$$u_2 = A_2 \hat{A}_1 [z_1 + A_2 u_2] + A_1 u_2 + z_2$$

or, which is the same,

$$u_2 = [A_1 + A_2 \hat{A}_1 A_2] u_2 + z_2 + A_2 \hat{A}_1 z_1. \quad (3.10)$$

Condition (3.5) guarantees the existence of the inverse operator

$$(\mathbf{1}_X - A_1 - A_2 \hat{A}_1 A_2)^{-1} \quad (3.11)$$

and, therefore, equation (3.10) has a unique solution u_2 given by formula (3.7). Inserting (3.7) into (3.1) and taking (3.9) into account, we arrive at (3.6).

By virtue of (3.4) and (3.5), the bounded linear operators (3.8) and (3.11) are representable by the corresponding Neumann series

$$\hat{A}_1 = \mathbf{1}_X + A_1 + A_1^2 + \dots \quad (3.12)$$

and

$$(\mathbf{1}_X - A_1 - A_2 \hat{A}_1 A_2)^{-1} = \mathbf{1}_X + A_1 + A_2 \hat{A}_1 A_2 + (A_1 + A_2 \hat{A}_1 A_2)^2 + \dots,$$

whence it follows that operators (3.8) and (3.11) preserve the wedge P provided that A_1 and A_1 do so. This fact, in view of representation (3.6), (3.7) for the unique solution (u_1, u_2) of system (3.1), (3.2), implies that, under assumption (1.5), the solution mentioned satisfies the relations $u_1 \geq_P 0$ and $u_2 \geq_P 0$ whenever $z_1 \geq_P 0$ and $z_2 \geq_P 0$. $\square$

Proposition 3.3. *Let $A_1 : X \rightarrow X$ and $A_2 : X \rightarrow X$ be bounded linear operators such that inequality (3.4) is true and, moreover,*

$$r((\mathbf{1}_X - A_1)^{-1} A_2) < 1. \quad (3.13)$$

Then system (3.1), (3.2) is uniquely solvable for arbitrary $(z_1, z_2) \in X^2$, and its solution (u_1, u_2) admits representation in the form

$$u_1 = \hat{A}_1 z_1 + \hat{A}_1 A_2 (\mathbf{1}_X - A_2^2 \hat{A}_1^2)^{-1} \hat{A}_1 [z_2 + A_2 \hat{A}_1 z_1], \quad (3.14)$$

$$u_2 = (\mathbf{1}_X - A_2^2 \hat{A}_1^2)^{-1} \hat{A}_1 [z_2 + A_2 \hat{A}_1 z_1]. \quad (3.15)$$

If, moreover, A_1 and A_2 satisfy condition (1.5), then $u_1 \geq_P 0$ and $u_2 \geq_P 0$ whenever $z_1 \geq_P 0$ and $z_2 \geq_P 0$.

Proof. By virtue of (3.4), operator (3.8) is well-defined and, hence (3.2) can be rewritten as

$$u_2 = \hat{A}_1 [z_2 + A_2 u_1], \quad (3.16)$$

whereas equation (3.1) can be brought to form (3.9). Substituting (3.9) into (3.16), we get

$$\begin{aligned} u_2 &= \hat{A}_1[z_2 + A_2\hat{A}_1z_1 + A_2\hat{A}_1A_2u_2] \\ &= (\hat{A}_1A_2)^2u_2 + \hat{A}_1(z_2 + A_2\hat{A}_1z_1) \\ &= A_2^2\hat{A}_1^2u_2 + \hat{A}_1(z_2 + A_2\hat{A}_1z_1). \end{aligned} \quad (3.17)$$

For any bounded linear operator $B : X \rightarrow X$, the inequality $r(B^2) < 1$ holds if, and only if $r(B) < 1$. Therefore, in view of (3.8), assumption (3.13) guarantees the existence of a bounded inverse operator

$$(\mathbf{1}_X - A_2^2\hat{A}_1^2)^{-1}, \quad (3.18)$$

whence it follows that the unique element u_2 satisfying relation (3.17) is given by equality (3.15). Substituting (3.15) into (3.9), we obtain (3.14).

Finally, by virtue of (3.4) and (3.13), it follows from formulae (3.14) and (3.15) that, under condition (1.5), the inclusion $(z_1, z_2) \in P^2$ implies the property $(u_1, u_2) \in P^2$ for the unique solution (u_1, u_2) of system (3.1), (3.2). $\square$

Propositions 3.2 and 3.3 lead one to the following statement.

Proposition 3.4. *Let $A_1 : X \rightarrow X$ and $A_2 : X \rightarrow X$ be bounded linear operators satisfying condition (3.4) and one of relations (3.5) and (3.13). Then 1 is a regular value for the operator $A_1 - A_2$.*

Proof. According to Propositions 3.2 and 3.3, each of the pairs of conditions (3.4), (3.5) and (3.4), (3.13) guarantees the unique solvability of system (3.1), (3.2) for all $(z_1, z_2) \in X^2$. Formulae (3.6), (3.7) and (3.14), (3.15), which represent the unique solution of this system in the respective cases, imply, in particular, that the solution mentioned depends continuously on (z_1, z_2) . Therefore, in view of Lemma 3.1, the conditions indicated guarantee that equation (1.2) has a unique solution for an arbitrary $z \in X$, and the dependence of this solution on z is continuous. Thus, there exists a bounded inverse operator (1.4), i. e., 1 is a regular value for $A_1 - A_2$. $\square$

4. UNIQUE SOLVABILITY OF EQUATION (1.2)

Proposition 4.1. *Let P be a wedge in the space X and A_1 and A_2 be completely continuous linear operators in X preserving P , satisfying the condition*

$$\text{bl } P \subset \ker A_1 \cap \ker A_2, \quad (4.1)$$

and such that[†]

$$\text{im } A_1 \subset H_1, \quad (4.2)$$

$$\text{im } A_1 + \text{im } A_2 \subset H_2, \quad (4.3)$$

[†]Recall that we use notation (2.1). Inclusion (4.3) thus means that $A_1u_1 + A_2u_2 \in H_2$ for arbitrary u_1 and u_2 from X .

where H_1 and H_2 are certain linear manifolds in X . Assume also that the relations

$$A_1 f_1 \leq_P \alpha_1 f_1 \quad (4.4)$$

and

$$A_1 f_2 + A_2 (\mathbf{1}_X - A_1)^{-1} A_2 f_2 \leq_P \alpha_2 f_2 \quad (4.5)$$

are true with some $\{\alpha_1, \alpha_2\} \subset [0, 1)$ and $\{f_1, f_2\} \subset P$ such that

$$f_1 \succ_{P; H_1} 0 \quad (4.6)$$

and

$$f_2 \succ_{P; H_2} 0. \quad (4.7)$$

Then inequality (1.3) is true.

Remark 4.2. As follows from the proof below, the existence of the inverse operator

$$(\mathbf{1}_X - A_1)^{-1} : X \rightarrow X \quad (4.8)$$

involved in inequality (4.5) is guaranteed by the rest of the conditions assumed and, thus, relation (4.5) makes sense.

Proof of Proposition 4.1. By virtue of Corollary 2.9, assumptions (4.1), (4.2), (4.4), and (4.6) guarantee that operator (4.8) is well-defined.

It is obvious that the inclusion

$$\text{im} [A_1 + A_2 (\mathbf{1}_X - A_1)^{-1} A_2] \subset \text{im} A_1 + \text{im} A_2$$

is true and, hence, in view of (4.3), the operator

$$A := A_1 + A_2 (\mathbf{1}_X - A_1)^{-1} A_2 \quad (4.9)$$

satisfies the condition $\text{im} A \subset H_2$. Since conditions (4.1), (4.5), and (4.7) are assumed to be satisfied, we see that Corollary 2.9 can be applied with $f := f_2$, $\alpha := \alpha_2$, and A given by formula (4.9). Application of Corollary 2.9 implies inequality (3.5) for the spectral radius of operator (4.9), whence, by Proposition 3.2, it follows that system (3.1), (3.2) is uniquely solvable for arbitrary z_1 and z_2 . Formulae (3.6) and (3.7) representing the unique solution of system (3.1), (3.2) show that its dependence on (z_1, z_2) is continuous and, hence, taking Lemma 2.5 into account, we conclude that the spectral radius of the operator $A_1 - A_2$ satisfies the required inequality (1.3). $\square$

To obtain more constructive conditions sufficient for the unique solvability of equation (1.2), namely, Theorems 4.4 and 4.6 below, we need the following simple supplement to Corollary 2.9.

Lemma 4.3. *Let P be wedge in X and $A : X \rightarrow X$ be a completely continuous linear operator preserving P , satisfying conditions (2.10) and (2.11) with a certain linear manifold $H \subset X$ and, moreover, such that there exists an element $f \in P$ for which relations (2.7) and (2.12) are true with a certain constant α , $0 \leq \alpha < 1$.*

Then the inverse operator

$$(\mathbf{1}_X - A)^{-1} : X \rightarrow X \quad (4.10)$$

exists, is bounded, preserves the wedge P , and, furthermore, the relation

$$(\mathbf{1}_X - A)^{-1} f \leq_P \frac{1}{1 - \alpha} f \quad (4.11)$$

is true.

Proof. Corollary 2.9 guarantees that operator (4.10) is well-defined and bounded. By assumption, A preserves P and, hence, inequality (2.12) implies that $A^i f \leq_P \alpha^i f$ for all $i = 0, 1, 2, \dots$. Using now the Neumann series representation of operator (4.10), we arrive at the relation

$$(\mathbf{1}_X - A)^{-1} f = \sum_{i=0}^{+\infty} A^i f \leq_P \sum_{i=0}^{+\infty} \alpha^i f,$$

which, clearly, yields (4.11). $\square$

Theorem 4.4. Let $A_1 : X \rightarrow X$, $A_2 : X \rightarrow X$ be completely continuous linear operators preserving a wedge P in X and satisfying condition (4.1). Assume the existence of an element $f \in P$ satisfying condition (2.4), possessing the property

$$A_1^i A_2 f \leq_P A_2 A_1^i f \quad \text{for all } i = 0, 1, 2, \dots, \quad (4.12)$$

and such that the relations

$$A_1 f \leq_P \alpha f, \quad (4.13)$$

$$A_2^2 f \leq_P \gamma(1 - \alpha) f \quad (4.14)$$

are true with certain constants $\alpha \in [0, 1)$ and $\gamma \in [0, 1 - \alpha)$.

Then 1 is a regular value of the operator $A_1 - A_2$.

Proof. Let us show that the assumptions of Proposition 4.1 are satisfied with $f_1 = f$, $f_2 = f$, $\alpha_1 = \alpha$, $\alpha_2 = \gamma + \alpha$, $H_1 = X$, and $H_2 = X$.

Indeed, (4.13) is nothing but (4.4) with α_1 replaced by α . Condition (4.5) with $f_2 = f$ has the form

$$A_1 f + A_2 (\mathbf{1}_X - A_1)^{-1} A_2 f \leq_P \alpha_2 f. \quad (4.15)$$

Considering the Neumann series (3.12) for operator (3.8), we get

$$(\mathbf{1}_X - A_1)^{-1} A_2 f = \sum_{i=0}^{+\infty} A_1^i A_2 f,$$

whence, by taking assumptions (4.12), (4.13), and (1.5) into account and applying Lemma 4.3 with $H = X$ and $A = A_1$, we obtain

$$(\mathbf{1}_X - A_1)^{-1} A_2 f \leq_P A_2 \sum_{i=0}^{+\infty} A_1^i f = A_2 (\mathbf{1}_X - A_1)^{-1} f \leq_P \frac{1}{1 - \alpha} A_2 f.$$

Therefore, for relation (4.15) to be true, it would suffice that

$$A_1 f + \frac{1}{1-\alpha} A_2^2 f \leq_P \alpha_2 f. \quad (4.16)$$

In view of (4.13), relation (4.16) is satisfied whenever

$$\alpha f + \frac{1}{1-\alpha} A_2^2 f \leq_P \alpha_2 f,$$

i. e., if

$$A_2^2 f \leq_P (1-\alpha)(\alpha_2-\alpha) f. \quad (4.17)$$

Condition (4.17) coincides with (4.14) if we put $\alpha_2 := \gamma + \alpha$. Note that this value of α_2 is such that $0 \leq \alpha_2 < 1$ because, by assumption, $0 \leq \gamma < 1 - \alpha$. Finally, according to Definition 2.3, condition (2.4) is a particular case of (4.6) with $f_1 = f$ and $H_1 = X$, and it remains to refer to Proposition 4.1 to obtain the conclusion desired. $\square$

Remark 4.5. Condition (4.12) is satisfied for every $f \geq_P 0$ if A_1 and A_2 are such that[‡]

$$A_1 A_2 u \leq_P A_2 A_1 u \quad \text{for all } u \in P. \quad (4.18)$$

Another kind of conditions ensuring the unique solvability of equation (1.2) can be obtained by using Proposition 3.3.

Theorem 4.6. *Let $A_1 : X \rightarrow X$, $A_2 : X \rightarrow X$ be completely continuous linear operators preserving a wedge $P \subset X$ and satisfying condition (4.1) and the relation*

$$\text{im } A_1 \subset H \quad (4.19)$$

with a certain linear manifold $H \subset X$. Let, moreover, the relations

$$A_1 A_2 g \leq_P \alpha A_2 g, \quad (4.20)$$

$$A_2^2 g \leq_P \beta (1-\alpha) g \quad (4.21)$$

hold, where $\{\alpha, \beta\} \subset [0, 1)$ and $g \in P$ is an element having the following properties:

$$g \succ_P 0 \quad (4.22)$$

and

$$A_2 g \succ_{P;H} 0. \quad (4.23)$$

Then 1 is a regular value of the operator $A_1 - A_2$.

[‡]Condition (4.18) implies, in particular, that the operator $[A_1, A_2] = A_1 A_2 - A_2 A_1$ preserves the blade $\text{bl } P$ of the wedge P . In the case where P is a cone (i. e., if $\text{bl } P = \{0\}$), a pair of operators (A_1, A_2) satisfying (4.18) is sometimes referred to as *semicommuting* [1].

Proof. Conditions (4.19), (4.20), and (4.23) guarantee that the inverse operator (4.8) is well-defined and bounded (it suffices to apply Corollary 2.9 with $A := A_1$ and $f := A_2g$). Let us show that the rest of assumptions ensure that relation (3.13) holds.

Indeed, according to Corollary 2.9, it will suffice to show that the relation

$$(\mathbf{1}_X - A_1)^{-1} A_2 f \leq_P \gamma f \quad (4.24)$$

is true with some $\gamma \in [0, 1)$ and f satisfying (2.4). However, by virtue of Lemma 4.3 applied with $f = A_2g$, we have

$$(\mathbf{1}_X - A_1)^{-1} A_2 A_2 g \leq_P \frac{1}{1 - \alpha} A_2 g$$

and, hence, (4.24) will be satisfied for $f = A_2g$ provided that

$$\frac{1}{1 - \alpha} A_2^2 g \leq_P \gamma A_2 g. \quad (4.25)$$

Note now that (4.21) is nothing but (4.25) with $\gamma := \beta$. $\square$

Remark 4.7. To every statement appearing in Sections 3.2 and 4, a dual one corresponds, with the rôles of A_1 and A_2 interchanged.

REFERENCES

- [1] ESAJAN, A. R. AND SROŽIDDINOV, H. M.: *The comparison of the spectral radii of semicommuting operators*, Sibirsk. Mat. Zh., **12** (1971), 332–345.
- [2] KRASNOSELSKII, M. A.: *Positive Solutions of Operator Equations*, Noordhoff, Groningen, 1964.
- [3] KRASNOSELSKII, M. A., LIFSHITS, E. A., AND SOBOLEV, A. V.: *Positive Linear Systems. Method of Positive Operators*, Heldermann, Berlin, 1989.
- [4] KRASNOSELSKII, M. A., VAINIKKO, G. M., ZABREIKO, P. P., RUTITSKII, YA. B., AND STETSENKO, V. YA.: *Approximate Solution of Operator Equations*, Noordhoff, Groningen, 1972.
- [5] KREIN, M. G. AND RUTMAN, M. A.: *Linear Operators Leaving Invariant a Cone in a Banach Space*, No. 26 in Amer. Math. Soc. Transl., Amer. Math. Soc., R. I., 1950.
- [6] RONTÓ, A.: *Bounds for the spectrum of compact linear operators in a preordered Banach space*, submitted.
- [7] RONTÓ, A.: *On substantial eigenvalues of linear operators leaving invariant a closed wedge*, In: *CDDE-2002 Proceedings*, vol. 13 of *Folia FSN Universitatis Masarykianae, Mathematica*, Masaryk University, 2003, pp. 235–246.
- [8] ZABREIKO, P. P., KRASNOSELSKII, M. A., AND STETSENKO, V. Y.: *Estimates of the spectral radius of positive linear operators*, Mat. Zametki, **1** (1967), 461–468.

Author's Address

Andrei Rontó:

INSTITUTE OF MATHEMATICS, NATIONAL ACADEMY OF SCIENCES OF UKRAINE, 3 TERESCHENKOVSKAYA ST., 01601 KIEV, UKRAINE

Current address: Mathematical Institute, Czech Academy of Sciences, Žitkova 22, CZ-61662 Brno, Czech Republic

E-mail address: ronto@ipm.cz



COOPERATION IN PATCHY ENVIRONMENT WITH CROSS-DIFFUSION

ALY A. SHABAN

[Received: September 16, 2003]

ABSTRACT. In this paper, we formulate a Lotka–Volterra cooperative system in two patches in which the *per capita* migration rate of each species is influenced not only by its own but also by the other one's density, i. e., there is cross-diffusion present. Numerical studies show that at a critical value of the bifurcation parameter the system undergoes a Turing bifurcation and the cross-migration response is an important factor that should not be ignored when a pattern emerges.

Mathematics Subject Classification: 92A15

Keywords: Cross-diffusion, diffusive instability, pattern formation

1. INTRODUCTION

THE TURING BIFURCATION (see [8]) is the basic bifurcation generating a spatial pattern, wherein an equilibrium of a nonlinear system is asymptotically stable in the absence of diffusion but unstable in the presence of diffusion. This lies at the heart of almost all mathematical models for patterning in ecology, embryology, and elsewhere in biology and chemistry (see [1, 2]). Since the relation between organisms and space seems to be essential to stability of an ecological system, the effect of diffusion on the possibility of species coexistence in an ecological community has been an important subject in population biology (see [5–8]). We consider a two-species cooperative Lotka–Volterra system living in a habitat of two identical patches linked by migration and we show that at a critical value of the bifurcation parameter the system undergoes a Turing bifurcation, i. e., the stable constant steady state loses its stability and spatially non-constant stationary solutions, a pattern emerges.

This paper is organized as follows. In Section 2, the model is built, in Section 3 its linearization is treated and the conditions for the Turing bifurcation are established (these are the main results of the paper), in Section 4 we consider an example to illustrate what can be expected, and in Section 5 we summarize the main conclusions of the study.

Research partially supported by the Hungarian National Foundation for Scientific Research.

2. THE MODEL

We consider a two-species cooperative Lotka–Volterra system living in a habitat of two identical patches linked by migration.

Let $u_i(t, j)$ be the density of species i in patch j at time t , $i = 1, 2$; $j = 1, 2$; $t \in \mathbb{R}$. The cooperation between two species is described by the system of differential equations

$$\begin{aligned}
 \dot{u}_1(t, 1) &= u_1(t, 1)(r_1 - a_{11}u_1(t, 1) - a_{12}u_2(t, 1)) \\
 &\quad + d_1(\varrho_1(u_2(t, 2))u_1(t, 2) - \varrho_1(u_2(t, 1))u_1(t, 1)), \\
 \dot{u}_2(t, 1) &= u_2(t, 1)(r_2 - a_{21}u_1(t, 1) - a_{22}u_2(t, 1)) \\
 &\quad + d_2(\varrho_2(u_1(t, 2))u_2(t, 2) - \varrho_2(u_1(t, 1))u_2(t, 1)), \\
 \dot{u}_1(t, 2) &= u_1(t, 2)(r_1 - a_{11}u_1(t, 2) - a_{12}u_2(t, 2)) \\
 &\quad + d_1(\varrho_1(u_2(t, 1))u_1(t, 1) - \varrho_1(u_2(t, 2))u_1(t, 2)), \\
 \dot{u}_2(t, 2) &= u_2(t, 2)(r_2 - a_{21}u_1(t, 2) - a_{22}u_2(t, 2)) \\
 &\quad + d_2(\varrho_2(u_1(t, 1))u_2(t, 1) - \varrho_2(u_1(t, 2))u_2(t, 2)),
 \end{aligned} \tag{2.1}$$

where

$$a_{11}, a_{22} > 0, \quad a_{12}, a_{21} < 0, \quad d_i \geq 0, \quad \varrho_i > 0 \quad \text{for } i, k = 1, 2.$$

Here, a_{11} and a_{22} represent the strength of the intraspecific competition (the competition within the species, $\frac{r_1}{a_{11}}, \frac{r_2}{a_{22}}$ are the carrying capacities for the respective species), $|a_{12}|$ and $|a_{21}|$ represent the strength of interspecific cooperation, $d_i > 0$ ($i = 1, 2$) are the diffusion coefficients, and $\varrho_1 \in C^1$ is a positive decreasing function of u_2 , with analogous conditions on ϱ_2 . The idea is that these migration functions describe the inclination of individuals of one species to stay at a certain patch due to the attraction by the other species in the patch (see [1, 3, 4]).

The case to be considered is where each species survives if left alone and follows the logistic dynamics, that is, the intrinsic growth rates of the respective species are positive, $r_1, r_2 > 0$, which is called facultative cooperation.

We consider the kinetic system without migration, i. e., the case where d_1 and d_2 are equal to zero:

$$\begin{aligned}
 \dot{u}_1(t, 1) &= u_1(t, 1)(r_1 - a_{11}u_1(t, 1) - a_{12}u_2(t, 1)), \\
 \dot{u}_2(t, 1) &= u_2(t, 1)(r_2 - a_{21}u_1(t, 1) - a_{22}u_2(t, 1)), \\
 \dot{u}_1(t, 2) &= u_1(t, 2)(r_1 - a_{11}u_1(t, 2) - a_{12}u_2(t, 2)), \\
 \dot{u}_2(t, 2) &= u_2(t, 2)(r_2 - a_{21}u_1(t, 2) - a_{22}u_2(t, 2)).
 \end{aligned} \tag{2.2}$$

We assume that

$$\det A = a_{11}a_{22} - a_{21}a_{12} > 0. \tag{2.3}$$

Then system (2) has a positive equilibrium

$$(u_1(t, 1), u_1(t, 2), u_2(t, 1), u_2(t, 2)) \equiv (\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2), \tag{2.4}$$

where

$$\bar{u}_1 = \frac{r_1 a_{22} - r_2 a_{12}}{\det A}, \quad \bar{u}_2 = \frac{r_2 a_{11} - r_1 a_{21}}{\det A}. \quad (2.5)$$

The Jacobian matrix of the system without diffusion linearized at $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ is

$$J_k = \begin{pmatrix} -a_{11}\bar{u}_1 & -a_{12}\bar{u}_1 & 0 & 0 \\ -a_{21}\bar{u}_2 & -a_{22}\bar{u}_2 & 0 & 0 \\ 0 & 0 & -a_{11}\bar{u}_1 & -a_{12}\bar{u}_1 \\ 0 & 0 & -a_{21}\bar{u}_2 & -a_{22}\bar{u}_2 \end{pmatrix}. \quad (2.6)$$

The corresponding characteristic polynomial has the form

$$D_4(\lambda) = (D_2(\lambda))^2, \quad D_2(\lambda) = \lambda^2 + \lambda(a_{11}\bar{u}_1 + a_{22}\bar{u}_2) + \bar{u}_1\bar{u}_2 \det A. \quad (2.7)$$

Since $a_{11}\bar{u}_1 + a_{22}\bar{u}_2 > 0$ and $\det A > 0$, we see that the coexistence equilibrium point $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ is linearly asymptotically stable.

3. THE LINEARIZED PROBLEM

Returning to system (2.1), we see that $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ is also a spatially homogeneous equilibrium of the system with diffusion. The Jacobian matrix of the system with diffusion at $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ is written as

$$J_D = \begin{pmatrix} -a_{11}\bar{u}_1 - d_1\varrho_1 & -a_{12}\bar{u}_1 - d_1\varrho'_1\bar{u}_1 & d_1\varrho_1 & d_1\varrho'_1\bar{u}_1 \\ -a_{21}\bar{u}_2 - d_2\varrho'_2\bar{u}_2 & -a_{22}\bar{u}_2 - d_2\varrho_2 & d_2\varrho'_2\bar{u}_2 & d_2\varrho_2 \\ d_1\varrho_1 & d_1\varrho'_1\bar{u}_1 & -a_{11}\bar{u}_1 - d_1\varrho_1 & -a_{12}\bar{u}_1 - d_1\varrho'_1\bar{u}_1 \\ d_2\varrho'_2\bar{u}_2 & d_2\varrho_2 & -a_{21}\bar{u}_2 - d_2\varrho'_2\bar{u}_2 & -a_{22}\bar{u}_2 - d_2\varrho_2 \end{pmatrix}, \quad (3.1)$$

where ϱ_1 and ϱ'_1 are to be taken at $\bar{u}_2$ and ϱ_2, ϱ'_2 at $\bar{u}_1$. We have

$$\det(J_D - \lambda I) = \begin{vmatrix} -a_{11}\bar{u}_1 - d_1\varrho_1 - \lambda & -a_{12}\bar{u}_1 - d_1\varrho'_1\bar{u}_1 & d_1\varrho_1 & d_1\varrho'_1\bar{u}_1 \\ -a_{21}\bar{u}_2 - d_2\varrho'_2\bar{u}_2 & -a_{22}\bar{u}_2 - d_2\varrho_2 - \lambda & d_2\varrho'_2\bar{u}_2 & d_2\varrho_2 \\ d_1\varrho_1 & d_1\varrho'_1\bar{u}_1 & -a_{11}\bar{u}_1 - d_1\varrho_1 - \lambda & -a_{12}\bar{u}_1 - d_1\varrho'_1\bar{u}_1 \\ d_2\varrho'_2\bar{u}_2 & d_2\varrho_2 & -a_{21}\bar{u}_2 - d_2\varrho'_2\bar{u}_2 & -a_{22}\bar{u}_2 - d_2\varrho_2 - \lambda \end{vmatrix}. \quad (3.2)$$

Using the properties of determinant, we get

$$\begin{aligned} \det(J_D - \lambda I) &= \begin{vmatrix} -a_{11}\bar{u}_1 - \lambda & -a_{12}\bar{u}_1 & d_1\varrho_1 & d_1\varrho'_1\bar{u}_1 \\ -a_{21}\bar{u}_2 & -a_{22}\bar{u}_2 - \lambda & d_2\varrho'_2\bar{u}_2 & d_2\varrho_2 \\ 0 & 0 & -a_{11}\bar{u}_1 - 2d_1\varrho_1 - \lambda & -a_{12}\bar{u}_1 - 2d_1\varrho'_1\bar{u}_1 \\ 0 & 0 & -a_{21}\bar{u}_2 - 2d_2\varrho'_2\bar{u}_2 & -a_{22}\bar{u}_2 - 2d_2\varrho_2 - \lambda \end{vmatrix} \\ &= D_2(\lambda)(\lambda^2 + \lambda(a_{11}\bar{u}_1 + a_{22}\bar{u}_2 + 2(d_1\varrho_1 + d_2\varrho_2)) + \bar{u}_1\bar{u}_2 \det A \\ &\quad + 2\bar{u}_1d_2(a_{11}\varrho_2 - a_{12}\varrho'_2\bar{u}_2) + 2\bar{u}_2d_1(a_{22}\varrho_1 - a_{21}\varrho'_1\bar{u}_1) \\ &\quad + 4d_1d_2(\varrho_1\varrho_2 - \bar{u}_1\bar{u}_2\varrho'_1\varrho'_2)). \end{aligned}$$

We know that $D_2(\lambda)$ has two roots with negative real parts. The other polynomial will have a negative and a positive root if the constant term is negative. Clearly,

$$(\varrho_1\varrho_2 - \bar{u}_1\bar{u}_2\varrho'_1\varrho'_2) = \varrho_1\varrho_2 \left(1 - \bar{u}_1\bar{u}_2 \frac{\varrho'_1\varrho'_2}{\varrho_1\varrho_2} \right) < 0$$

if $\frac{\varrho'_1\varrho'_2}{\varrho_1\varrho_2}$ is large enough. If we have achieved this, we may increase d_1 and/or d_2 and the constant term becomes negative (see [3]). These calculations lead us to the following theorem.

Theorem. *The equilibrium $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ of system (2.1) is asymptotically stable if $\frac{\varrho'_1\varrho'_2}{\varrho_1\varrho_2}$, d_1 and d_2 are sufficiently small. If $\frac{\varrho'_1\varrho'_2}{\varrho_1\varrho_2}$ and either d_1 or d_2 are sufficiently large, then $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ loses its stability by a Turing bifurcation.*

Remark. The situation is different if the cooperation is obligatory, $r_1, r_2 < 0$; the condition of having a point of intersection in the positive quadrant is

$$\det A = a_{11}a_{22} - a_{21}a_{12} < 0. \quad (3.3)$$

The characteristic polynomial of the linearized system (2.6) without diffusion at $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ has the form

$$D_4(\lambda) = (D_2(\lambda))^2, \quad D_2(\lambda) = \lambda^2 + \lambda(a_{11}\bar{u}_1 + a_{22}\bar{u}_2) + \bar{u}_1\bar{u}_2 \det A. \quad (3.4)$$

Since $a_{11}\bar{u}_1 + a_{22}\bar{u}_2 > 0$ and $\det A < 0$, we see that the coexistence equilibrium point $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ is a saddle point and diffusion never stabilizes an equilibrium which is unstable for the kinetic system.

4. NUMERICAL INVESTIGATION

In this section, we apply our analytical approach of Section 3 to a specific migration function and we are looking for conditions which imply the Turing instability (diffusion driven instability). Namely, we choose

$$\varrho_1(u_2) = m_1 \exp(-u_2/m_1), \quad \varrho_2(u_1) = m_2 \exp(-u_1/m_2), \quad m_1, m_2 > 0. \quad (4.1)$$

If $r_1 = 2$, $r_2 = 1$, $a_{11} = 5$, $a_{22} = 4$, $a_{12} = -4$, $a_{21} = -3$, $m_1 = 1$, $m_2 = 1$, and $d_1 = 1$, then

$$\det A = a_{11}a_{22} - a_{21}a_{12} = 8$$

and

$$\bar{u}_1 = \frac{r_1 a_{22} - r_2 a_{12}}{\det A} = \frac{3}{2}, \quad \bar{u}_2 = \frac{r_2 a_{11} - r_1 a_{21}}{\det A} = \frac{11}{8}.$$

At $d_2 = d_{2\text{crit}}$,

$$d_{2\text{crit}} = \frac{-\frac{33}{2} + \frac{11}{8 \exp(\frac{11}{8})}}{-\frac{17}{4 \exp(\frac{23}{8})} - \frac{3}{2 \exp(\frac{3}{2})}} \cong 28.11725408,$$

we have four eigenvalues λ_i ($i = 1, 2, 3, 4$) such that $\lambda_i < 0$ ($i = 1, 2, 3$) and $\lambda_4 = 0$.

d_2	$u_1(t, 1)$	$u_2(t, 1)$	$u_1(t, 2)$	$u_2(t, 2)$
28	1.500000000	1.375000000	1.500000000	1.375000000
28.2	1.524440807	1.403943678	1.474516967	1.344807960
	1.500000000	1.375000000	1.500000000	1.375000000
	1.474516967	1.344807960	1.524440807	1.403943678
29	1.574289053	1.462917090	1.415068046	1.274335600
	1.500000000	1.375000000	1.500000000	1.375000000
	1.415068046	1.274335600	1.574289053	1.462917090
30	1.602257674	1.495956494	1.376189993	1.228230906
	1.500000000	1.375000000	1.500000000	1.375000000
	1.376189993	1.228230906	1.602257674	1.495956494
40	1.683294096	1.591314489	1.225824281	1.049861530
	1.500000000	1.375000000	1.500000000	1.375000000
	1.225824281	1.049861530	1.683294096	1.591314489
50	1.704327936	1.615899482	1.169559919	.9830950704
	1.500000000	1.375000000	1.500000000	1.375000000
	1.169559919	.9830950704	1.704327936	1.615899482
60	1.714081033	1.627253651	1.138343964	.9460399293
	1.500000000	1.375000000	1.500000000	1.375000000
	1.138343964	.9460399293	1.714081033	1.627253651
80	1.723234442	1.637870036	1.104276166	.9055833674
	1.500000000	1.500000000	1.500000000	1.500000000
	1.104276166	.9055833674	1.723234442	1.637870036
100	1.727536907	1.642841684	1.085923814	.8837801086
	1.500000000	1.500000000	1.500000000	1.500000000
	1.085923814	.8837801086	1.727536907	1.642841684

TABLE 1. The equilibria before and after bifurcation (Section 4).

If $d_2 < d_{2\text{crit}}$, then $\lambda_i < 0$ ($i = 1, 2, 3, 4$), and $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ is asymptotically stable. If $d_2 > d_{2\text{crit}}$, then $\lambda_i < 0$ ($i = 1, 2, 3$) and $\lambda_4 > 0$; hence, $(\bar{u}_1, \bar{u}_2, \bar{u}_1, \bar{u}_2)$ is unstable in this case.

Thus, as d_2 is increased through $d_2 = d_{2\text{crit}}$, then the spatially homogeneous equilibrium loses its stability. Numerical calculations show that two new spatially non-constant equilibria emerge, and these equilibria are asymptotically stable, so that this is a pitchfork bifurcation. Table 1 contains some numerical data, whereas Figure 2 shows the graphs* of the coordinate $u_1(t, 1)$ of solutions corresponding to the respective initial conditions

$$\begin{aligned} & (1.80, 1.60, 1.50, 1.25), (1.20, 1.10, 1.59, 1.47), (1.58, 1.45, 1.36, 1.22), \\ & (1.00, 1.10, 1.585, 1.47), (1.65, 1.100, 1.320, 1.500) \end{aligned} \quad (4.2)$$

in the case where $d_2 = 30$. The three solutions on Figure 2 tend to 1.6022576 and two solutions tend to 1.376189. Figure 1 shows the corresponding solutions in a “pre-bifurcation case” (for $d_2 = 28$).

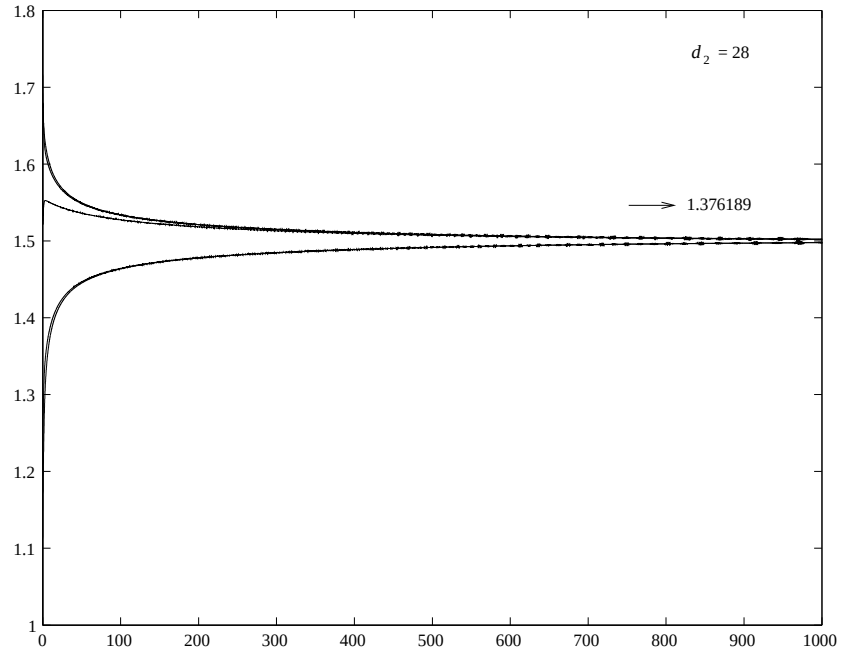


FIGURE 1. Graphs of the coordinate $u_1(t, 1)$ of the solutions in a “pre-bifurcation case” (Section 4, $d_2 = 28$)

*The graphs were produced by using PHASER.

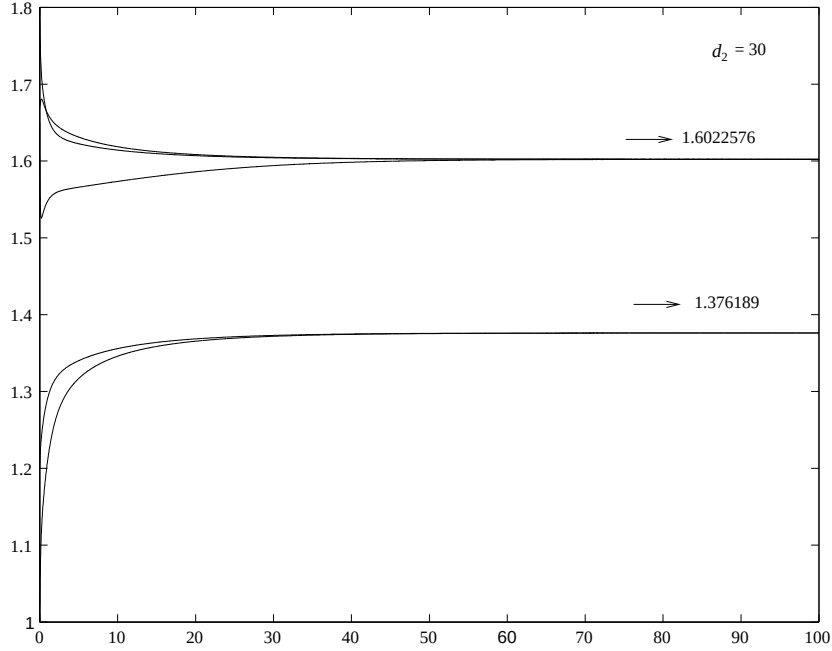


FIGURE 2. Graphs of the coordinate $u_1(t, 1)$ of the five solutions (Section 4, $d_2 = 30$; see Table 1) corresponding to the respective initial conditions (4.2).

It should be noted that, after the bifurcation, the sum of the stable equilibrium values of species 1 at the two patches (and, similarly, that of species 2) is equal to the double of its spatially homogeneous equilibrium value $\bar{u}_1$ (resp., $\bar{u}_2$).

5. CONCLUSIONS

In the present paper, our interest is to study a Lotka–Volterra cooperative system in two patches in which the *per capita* migration rate of each species is influenced not only by its own but also by the other one's density, i. e., there is cross-diffusion present. We show that, at a critical value of the bifurcation parameter, the system undergoes a Turing bifurcation, and the cross-migration response is an important factor that should not be ignored when a pattern emerges. As d_2 is increased through $d_2 = d_{2\text{crit}}$, the spatially homogeneous equilibrium loses its stability, and two new stable equilibria emerge. Diffusion never stabilizes an equilibrium which is unstable for the kinetic system.

REFERENCES

- [1] FARKAS, M.: *Two ways of modeling cross diffusion*, Nonlinear Analysis. Theory, Methods & Appl., **30** (1997), 1225–1233.
- [2] FARKAS, M.: *Dynamical Models in Biology*, Academic Press, San Diego, 2001.
- [3] HUANG, Y. AND DIEKMANN, O.: *Interspecific influence on mobility and Turing instability*, Bull. Math. Biol., **65** (2003), 143–156.
- [4] JANSEN, V. A. A. AND ALUN LLOYD, L.: *Local stability analysis of spatially homogeneous solutions of multi-patch systems*, J. Math. Biol., **41** (2000), 232–252.
- [5] MURRAY, J. D.: *Mathematical Biology*, Springer-Verlag, Berlin 1989.
- [6] SHABAN, A. AND FARKAS, M.: *Competition in patchy environment with cross-diffusion*, submitted.
- [7] TAKEUCHI, Y.: *Global Dynamical Properties of Lotka–Volterra Systems*, World Scientific, River Edge, NJ, 1996.
- [8] TURING, A. M.: *The chemical basis of morphogenesis*, Philos. Trans. Roy. Soc. London, B237 (1953) 37–72; reprinted in Bull. Math. Biol., **52** (1990), 153–197.

*Author's Address***Aly A. Shaban:**

INSTITUTE OF MATHEMATICS, BUDAPEST UNIVERSITY OF TECHNOLOGY, BUDAPEST, HUNGARY, H-1521
E-mail address: shaban@math.bme.hu



A NONLOCAL SINGULAR BOUNDARY VALUE PROBLEM FOR SECOND-ORDER DIFFERENTIAL EQUATIONS

SVATOSLAV STANĚK

ABSTRACT. We discuss the singular differential equation $(g(x'))' = f(t, x, x')$ together with the nonlocal boundary conditions $\max\{x(t) : t \in [0, T]\} = A$, $x(0) = x(T)$. Here, $g \in C^0(\mathbb{R})$ is an increasing and odd function, whereas the negative f satisfying the local Carathéodory conditions on $[0, T] \times \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ may be singular at the value 0 of its second phase variable and $A \in \mathbb{R}$. An existence result for the above boundary value problem is proved by the regularization and sequential techniques. The proofs use the Leray–Schauder degree principle and the Vitali convergence theorem.

Mathematics Subject Classification: 34B16, 34B15

Keywords: Singular boundary value problem, singular second-order differential equation, nonlocal boundary condition, Leray–Schauder degree, Vitali convergence theorem

1. INTRODUCTION

LET T BE A POSITIVE NUMBER, $J = [0, T]$, and let $A \in \mathbb{R}$. Consider the boundary value problem

$$(g(x'(t)))' = f(t, x(t), x'(t)), \quad (1.1)$$

$$x(0) = x(T), \quad \max\{x(t) : t \in J\} = A. \quad (1.2)$$

Here, $g \in C^0(\mathbb{R})$ is an increasing and odd function, and the function f satisfying local Carathéodory conditions on $J \times \mathbb{R} \times (\mathbb{R} \setminus \{0\})$ ($f \in \text{Car}(J \times \mathbb{R} \times (\mathbb{R} \setminus \{0\}))$) is negative, and it may be singular at the value 0 of its second phase variable in the following sense: $\lim_{y \rightarrow 0} f(t, x, y) = -\infty$ for a. e. $t \in J$ and every $x \in \mathbb{R}$.

We say that a function $x \in C^1(J)$ is a *solution of the boundary value problem* (1.1), (1.2) if $g(x') \in AC(J)$ (absolutely continuous functions on J), x satisfies the boundary conditions (1.2), and (1.1) holds a. e. on J .

The aim of this paper is to give conditions guaranteeing the solvability of the boundary value problem (1.1), (1.2). We note that the derivative of any solution of the boundary value problem (1.1), (1.2) ‘goes through’ the singularity of f somewhere

Supported by Grant No. 201/01/1451 of the Grant Agency of the Czech Republic and by the Council of the Czech Government J14/98:153100011.

inside of J . Moreover, we consider solutions x of the boundary value problem (1.1), (1.2) having the ‘maximal’ smoothness, that is, $x \in C^1(J)$ and $g(x') \in AC(J)$. Till the present, various types of nonlocal boundary conditions have been considered for differential equations, systems of differential equations, and systems of functional differential equations with continuous or Carathéodory nonlinearities (see, e. g., [4–6, 8–10, 13, 16–18] and references therein) or nonlinearities having singularities in the time variable t (see, e. g., [12, 14, 15] and references therein). There are only a few papers (see [1, 2, 19, 20]) dealing with boundary value problems for nonlocal boundary conditions and second-order differential equations having singularities in phase variables.

As usual, $\|x\| = \max\{|x(t)| : t \in J\}$ and $\|x\|_L = \int_0^T |x(t)| dt$ stand for the norm in $C^1(J)$ and $L_1(J)$, respectively. For every measurable set $M \subset \mathbb{R}$, its Lebesgue measure is denoted by the symbol $\mu(M)$.

Throughout the paper, the following assumptions are used.

(H₁) $g \in C^0(\mathbb{R})$ is an increasing and odd function, $\lim_{u \rightarrow \infty} g(u) = \infty$;

(H₂) $f \in \text{Car}(J \times \mathbb{R} \times (\mathbb{R} \setminus \{0\}))$ and there exists a positive constant a such that

$$f(t, x, y) \leq -a$$

for a. e. $t \in J$ and all $(x, y) \in \mathbb{R} \times (\mathbb{R} \setminus \{0\})$;

(H₃) For a. e. $t \in J$ and all $(x, y) \in \mathbb{R} \times (\mathbb{R} \setminus \{0\})$,

$$f(t, x, y) \geq -q(x)(\omega_1(|y|) + \omega_2(|y|))$$

with $q \in C^0(\mathbb{R})$ positive, $\omega_1 \in C^0([0, \infty))$ non-negative, $\omega_2 \in C^0((0, \infty))$ positive and non-increasing and

$$\int_0^1 \omega_2(g^{-1}(s)) ds < \infty.*$$

A further assumption (H₄) will be given in Section 2.

The plan of the paper is as follows: In Section 2, we first define a sequence of auxiliary regular boundary value problems with nonlinearities f_n such that $f_n(t, x, y) = f(t, x, y)$ for a. e. $t \in J$ and for $x \in \mathbb{R}$, $|y| \geq 1/n$, $n \in \mathbb{N}$. Then we consider properties of their solutions (Lemma 2.1) and give them *a priori* bounds (Lemma 2.2). By using the Leray–Schauder degree principle (see [7]), the existence of a solution x_n of auxiliary boundary value problem with nonlinearity f_n is proved (Lemma 2.3). Finally, we show that the sequence $\{f_n(t, x_n(t), x'_n(t))\}$ is uniformly absolutely continuous on J (Lemma 2.5). It is this result which is used in the next section when passing to the limit because it is impossible to find a Lebesgue majorant for $\{f_n(t, x_n(t), x'_n(t))\}$ which is necessary for the application of the Lebesgue theorem on dominated convergence.

*Since g^{-1} is increasing by (H₁) and ω_2 is positive and non-increasing by (H₃), it follows that $\int_0^c \omega_2(g^{-1}(s)) ds < \infty$ for every $c \in (0, \infty)$.

Section 3 presents the main existence result for singular boundary value problem (1.1), (1.2) (Theorem 3.1). It is proved by the Arzelà–Ascoli theorem and the Vitali convergence theorem (see [3, 11]).

The results are demonstrated on an example (Example 3.2).

2. AUXILIARY REGULAR BOUNDARY VALUE PROBLEMS

For every $n \in \mathbb{N}$, we define $f_n \in \text{Car}(J \times \mathbb{R}^2)$ by

$$f_n(t, x, y) = \begin{cases} f(t, x, y) & \text{for } (t, x) \in J \times \mathbb{R}, |y| \geq \frac{1}{n} \\ \frac{n}{2} \left[f\left(t, x, \frac{1}{n}\right)\left(y + \frac{1}{n}\right) - f\left(t, x, -\frac{1}{n}\right)\left(y - \frac{1}{n}\right) \right] & \text{for } (t, x) \in J \times \mathbb{R}, |y| < \frac{1}{n}. \end{cases}$$

By (H_2) and (H_3) ,

$$-a \geq f_n(t, x, y) \geq -q(x)(\omega_{*1}(|y|) + \omega_2(|y|)) \quad (2.1)$$

for a. e. $t \in J$ and all $(x, y) \in \mathbb{R}^2$, where

$$\omega_{*1}(x) = \begin{cases} \omega_1(1) & \text{for } 0 \leq x \leq 1 \\ \omega_1(x) & \text{for } x > 1. \end{cases}$$

Consider the family of regular differential equations

$$(g(x'(t)))' = \lambda f_n(t, x(t), x'(t)) \quad (E)_n^\lambda \quad (2.2)$$

depending on the parameters $n \in \mathbb{N}$ and $\lambda \in [0, 1]$. We note that $x = A$ is the unique solution of the boundary value problem $(E)_n^0, (1.2)$ for every $n \in \mathbb{N}$.

Lemma 2.1. *Let assumptions $(H_1) - (H_3)$ be satisfied and let $n \in \mathbb{N}$, $\lambda \in (0, 1]$. Assume that x is a solution of the boundary value problem $(E)_n^\lambda, (1.2)$. Then*

$$x(T) = x(0) = \min\{x(t) : t \in J\} < A, \quad (2.2)$$

$$A - x(0) \geq \int_0^{T/2} g^{-1}(\lambda a(s)) ds, \quad (2.3)$$

$x'(\xi) = 0$ for a unique $\xi \in (0, T)$, and

$$\begin{aligned} x'(t) &\geq g^{-1}(\lambda a(\xi - t)) \quad \text{for } t \in [0, \xi], \\ x'(t) &\leq -g^{-1}(\lambda a(t - \xi)) \quad \text{for } t \in [\xi, T]. \end{aligned} \quad (2.4)$$

Proof. Since

$$(g(x'(t)))' \leq -\lambda a \quad \text{for a. e. } t \in J, \quad (2.5)$$

by (2.1), we see that $g(x')$ is decreasing on J and, therefore, x' is decreasing as well. Now from (1.2) we deduce that $x'(\xi) = 0$ for a $\xi \in (0, T)$, and so ξ is unique, $\max\{x(t) : t \in J\} = x(\xi)$, $x' > 0$ on $[0, \xi]$, $x' < 0$ on $(\xi, T]$, and (2.2) is also true. Integrating (2.5) over $[t, \xi] \subset [0, \xi]$ and over $[\xi, t] \subset [\xi, T]$, we get

$$g(x'(t)) \geq \lambda a(\xi - t), \quad t \in [0, \xi]$$

and

$$g(x'(t)) \leq -\lambda a(t - \xi), \quad t \in [\xi, T],$$

respectively. Hence, (2.4) is true.

Finally, it follows from (2.4) and the equality $x(\xi) = A$ that

$$A - x(0) = x(\xi) - x(0) = \int_0^\xi x'(t) dt \geq \int_0^\xi g^{-1}(\lambda a(\xi - t)) dt = \int_0^\xi g^{-1}(\lambda a s) ds$$

and

$$\begin{aligned} x(T) - A &= x(T) - x(\xi) = \int_\xi^T x'(t) dt \leq - \int_\xi^T g^{-1}(\lambda a(t - \xi)) dt \\ &= - \int_0^{T-\xi} g^{-1}(\lambda a s) ds. \end{aligned}$$

Therefore,

$$\begin{aligned} (A - x(T)) - (A - x(0)) &\geq \max \left\{ \int_0^\xi g^{-1}(\lambda a s) ds, \int_0^{T-\xi} g^{-1}(\lambda a s) ds \right\} \\ &\geq \int_0^{T/2} g^{-1}(\lambda a s) ds, \end{aligned}$$

which proves (2.3). $\square$

We are going to give *a priori* bounds for the solution of the boundary value problem $(E)_n^1$, (1.2). Define the functions $H \in C^0([0, \infty))$ and $Q_A \in C^0((-\infty, A])$, $A \in \mathbb{R}$, by the relations

$$H(u) = \int_0^{g(u)} \frac{g^{-1}(s)}{\omega_{*1}(g^{-1}(s)) + \omega_2(g^{-1}(s))} ds$$

and

$$Q_A(u) = \int_u^A q(s) ds.$$

Then H is increasing on $[0, \infty)$, Q_A is decreasing on $(-\infty, A]$, and $H^{-1}(Q_A(A)) = 0$. If the function $H^{-1}(Q_A(u))$ satisfies the assumption[†]

$$(H_4) \quad \int_{A-1}^A \frac{1}{H^{-1}(Q_A(s))} ds < \infty \quad \text{for } A \in \mathbb{R},$$

then we define, for every $A \in \mathbb{R}$, the function $S_A \in C^0((-\infty, A])$ by putting

$$S_A(u) = \int_u^A \frac{1}{H^{-1}(Q_A(s))} ds. \quad (2.6)$$

[†]Since the function $H^{-1}(Q_A)$ is decreasing on $(-\infty, A]$, the assumption (H_4) guarantees that $\int_u^A \frac{ds}{H^{-1}(Q_A(s))} < \infty$ for every $u \in (-\infty, A)$.

Then $S_A(A) = 0$ and S_A is decreasing on $(-\infty, A]$. Set

$$\mathcal{A} = \left\{ A : A \in \mathbb{R}, \lim_{u \rightarrow -\infty} S_A(u) > \frac{T}{2} \right\}. \quad (2.7)$$

We are now in a position to give *a priori* bounds for the solution of the boundary value problem $(E)_n^\lambda$, (1.2).

Lemma 2.2. *Let assumptions $(H_1) - (H_4)$ be satisfied and let $A \in \mathcal{A}$. Let x be a solution of the boundary value problem $(E)_n^\lambda$, (1.2), $\lambda \in (0, 1]$, $n \in \mathbb{N}$. Then*

$$x(0) = x(T) \geq \Lambda_A \quad (2.8)$$

and

$$\|x\| \leq \max \{|\Lambda_A|, |A|\}, \quad \|x'\| \leq H^{-1}(Q_A(\Lambda_A)), \quad (2.9)$$

where

$$\Lambda_A = S_A^{-1}\left(\frac{T}{2}\right) \quad (2.10)$$

and S_A is defined by (2.6).

Proof. By Lemma 2.1, $x'(\xi) = 0$ for a unique $\xi \in (0, T)$, $x' > 0$ on $[0, \xi)$, and $x' < 0$ on $(\xi, T]$. Then (see (2.1))

$$(g(x'(t)))' x'(t) \geq -q(x(t))[\omega_{*1}(x'(t)) + \omega_2(x'(t))]x'(t) \quad \text{for a. e. } t \in [0, \xi]$$

and

$$(g(x'(t)))' x'(t) \leq -q(x(t))[\omega_{*1}(-x'(t)) + \omega_2(-x'(t))]x'(t) \quad \text{for a. e. } t \in [\xi, T].$$

Integrating the inequality

$$\frac{(g(x'(t)))' x'(t)}{\omega_{*1}(x'(t)) + \omega_2(x'(t))} \geq -q(x(t))x'(t) \quad \text{for a. e. } t \in [0, \xi]$$

from $t \in [0, \xi]$ to ξ and the inequality

$$\frac{(g(x'(t)))' x'(t)}{\omega_{*1}(-x'(t)) + \omega_2(-x'(t))} \leq -q(x(t))x'(t) \quad \text{for a. e. } t \in [\xi, T]$$

from ξ to $t \in [\xi, T]$ and using the relation $x(\xi) = A$, we get

$$\int_0^{g(x'(t))} \frac{g^{-1}(s)}{\omega_{*1}(g^{-1}(s)) + \omega_2(g^{-1}(s))} ds \leq \int_{x(t)}^A q(s) ds, \quad t \in [0, \xi]$$

and

$$\int_0^{g(-x'(t))} \frac{g^{-1}(s)}{\omega_{*1}(g^{-1}(s)) + \omega_2(g^{-1}(s))} ds \leq \int_{x(t)}^A q(s) ds, \quad t \in [\xi, T].$$

Therefore,

$$H(x'(t)) \leq Q_A(x(t)) \quad \text{for } t \in [0, \xi] \quad (2.11)$$

and

$$H(-x'(t)) \leq Q_A(x(t)) \quad \text{for } t \in [\xi, T]. \quad (2.12)$$

Now integrating the relation

$$\frac{x'(t)}{H^{-1}(Q_A(x(t)))} \leq 1, \quad t \in [0, \xi]$$

from 0 to ξ and

$$-\frac{x'(t)}{H^{-1}(Q_A(x(t)))} \leq 1, \quad t \in [\xi, T]$$

from ξ to T , we have (see (2.6)) $S_A(x(0)) \leq \xi$ and $S_A(x(T)) \leq T - \xi$. Hence $(S_A(x(T)) =) S_A(x(0)) \leq T/2$, and consequently $(x(T) =) x(0) \geq S_A^{-1}(T/2) = \Lambda_A$ because S_A is decreasing. We have thus proved that (2.8) is true and, therefore,

$$\|x\| \leq \max\{|\Lambda_A|, |A|\}.$$

Now (2.11) with $t = 0$ and (2.12) with $t = T$ yield

$$x'(0) \leq H^{-1}(Q_A(x(0))) \leq H^{-1}(Q_A(\Lambda_A)),$$

$$-x'(T) \leq H^{-1}(Q_A(x(T))) \leq H^{-1}(Q_A(\Lambda_A)),$$

and since $\|x'\| = \max\{x'(0), -x'(T)\}$, we have

$$\|x'\| \leq H^{-1}(Q_A(\Lambda_A)),$$

as required. $\square$

Lemma 2.3. *Let assumptions $(H_1) - (H_4)$ be satisfied. Let $A \in \mathcal{A}$ and $n \in \mathbb{N}$. Then there exists a solution x of the boundary value problem $(E)_n^1$, (1.2) satisfying (2.8) and (2.9), where Λ_A is defined by (2.10).*

Proof. Let us put

$$\Omega = \{(x, c) : (x, c) \in C^1(J) \times \mathbb{R}, \|x\| < \max\{|\Lambda_A|, |A|\} + 1, \\ \|x'\| \leq H^{-1}(Q_A(\Lambda_A)) + 1, |c| < \max\{|\Lambda_A|, |A|\} + 1\}.$$

Then Ω is a bounded, open, and symmetric with respect to $(0, 0)$ subset of the Banach space $C^1(J) \times \mathbb{R}$ with the norm $\|(x, c)\|_* = \|x\| + \|x'\| + |c|$. Define the mapping $\mathcal{F} : \overline{\Omega} \times [0, 1] \rightarrow C^1(J) \times \mathbb{R}$ by the equality

$$\mathcal{F}(x, c, \lambda) = (c, c + \max\{x(t) : t \in J\} - \lambda A).$$

Then $\mathcal{F}$ is a compact operator. Let I be the identity operator on $C^1(J) \times \mathbb{R}$ and let us define $\mathcal{G} : \overline{\Omega} \rightarrow C^1(J) \times \mathbb{R}$ by putting $\mathcal{G} = I - \mathcal{F}(\cdot, \cdot, 0)$. Then

$$\mathcal{G}(x, c) = (x - c, -\max\{x(t) : t \in J\}).$$

Assume now that

$$\mathcal{G}(-x_0, -c_0) = \chi_0 \mathcal{G}(x_0, c_0)$$

for some $(x_0, c_0) \in \partial\Omega$ and $\chi_0 \in [1, \infty)$. Then

$$(-x_0 + c_0, -\max\{-x_0(t) : t \in J\}) = \chi_0(x_0 - c_0, -\max\{x_0(t) : t \in J\}),$$

and, thus,

$$(x_0(t) - c_0)(1 + \chi_0) = 0 \quad \text{for } t \in J, \quad (2.13)$$

$$\max\{-x_0(t) : t \in J\} = \chi_0 \max\{x_0(t) : t \in J\}. \quad (2.14)$$

It follows from (2.13) that $x_0(t) = c_0$ for $t \in J$, and (2.14) then yields $(\chi_0 + 1)c_0 = 0$. Hence, $(x_0, c_0) = (0, 0)$, contrary to the inclusion $(x_0, c_0) \in \partial\Omega$. Now Theorem 8.3 from [7] implies that

$$D(I - \mathcal{F}(\cdot, \cdot, 0), \Omega, 0) \neq 0, \quad (2.15)$$

where “ D ” stands for the Leray–Schauder degree.

If $\mathcal{F}(x_*, c_*, \lambda_*) = (x_*, c_*)$ for some $(x_*, c_*) \in \partial\Omega$ and $\lambda_* \in [0, 1]$, then $x_* = c_*$ and $\max\{x_*(t) : t \in J\} = \lambda_* A$. Hence, $c_* = \lambda_* A$ and, consequently, $|c_*| \leq |A|$, $\|x_*\| \leq |A|$, and $\|x'\| = 0$, which contradicts the relation $(x_*, c_*) \in \partial\Omega$. We have thus proved that

$$\mathcal{F}(x, c, \lambda) \neq (x, c)$$

for $(x, c) \in \partial\Omega$ and $\lambda \in [0, 1]$ and, therefore, by virtue of the homotopy property and (2.15), we obtain

$$D(I - \mathcal{F}(\cdot, \cdot, 1), \Omega, 0) = D(I - \mathcal{F}(\cdot, \cdot, 0), \Omega, 0) \neq 0. \quad (2.16)$$

Finally, let us define the operator $\mathcal{K} : \overline{\Omega} \times [0, 1] \rightarrow C^1(J) \times \mathbb{R}$ by the relation

$$\mathcal{K}(x, c, \lambda) = \left(c + \int_0^t g^{-1} \left(K + \lambda \int_0^s f_n(v, x(v), x'(v)) dv \right) ds, \right. \\ \left. c + \max\{x(t) : t \in J\} - A \right)$$

where $K = K(x, \lambda) \in \mathbb{R}$ is a solution of the equation

$$p(X; x, \lambda) = 0 \quad (2.17)$$

with

$$p(X; x, \lambda) = \int_0^T g^{-1} \left(X + \lambda \int_0^t f_n(s, x(s), x'(s)) ds \right) dt. \quad (2.18)$$

Since $p(\cdot; x, \lambda)$ is continuous and increasing on $\mathbb{R}$ and $\lim_{X \rightarrow \pm\infty} p(X; x, \lambda) = \pm\infty$ by virtue of (H_1) , it follows that equation (2.17) has a unique solution for every $(x, \lambda) \in \overline{\Omega} \times [0, 1]$. Assume that (x, c) is a fixed point of the operator $\mathcal{K}(\cdot, \cdot, 1)$. Then

$$x(t) = c + \int_0^t g^{-1} \left(K_* + \int_0^s f_n(v, x(v), x'(v)) dv \right) ds, \quad t \in J,$$

and $\max\{x(t) : t \in J\} = A$, where K_* is the unique solution of equation $p(X; x, 1) = 0$. Hence, x is a solution of the boundary value problem $(E)_n^1$, (1.2) and $c = x(0)$. We see that, to prove our lemma, it is sufficient to verify that

$$D(I - \mathcal{K}(\cdot, \cdot, 1), \Omega, 0) \neq 0. \quad (2.19)$$

Since

$$\mathcal{K}(x, c, 0) = (c, c + \max\{x(t) : t \in J\} - A),$$

we have $\mathcal{K}(\cdot, \cdot, 0) = \mathcal{F}(\cdot, \cdot, 1)$, and then (2.16) yields the inequality

$$D(\mathcal{I} - \mathcal{K}(\cdot, \cdot, 0), \Omega, 0) \neq 0.$$

By the homotopy property, the fulfilment of (2.19) will be proved if we show that

- (i) $\mathcal{K}$ is a compact operator and
- (ii) $\mathcal{K}(x, c, \lambda) \neq (x, c)$ for $(x, c) \in \partial\Omega$ and $\lambda \in [0, 1]$.

We first show that $\mathcal{K}$ is a continuous operator. Let $\{(x_m, c_m, \lambda_m)\} \subset \overline{\Omega} \times [0, 1]$ be a convergent sequence and $\lim_{m \rightarrow \infty} (x_m, c_m, \lambda_m) = (x_0, c_0, \lambda_0)$. Let K_m be the unique solution of equation $p(X; x_m, \lambda_m) = 0$. Since $f_n \in \text{Car}(J \times \mathbb{R}^2)$, there exists $\gamma \in L_1(J)$ such that

$$|f_n(t, x, y)| \leq \gamma(t) \quad \text{for a. e. } t \in J \text{ and all } (x, y) \\ \text{such that } |x| \leq \max\{|\Lambda_A|, |A|\} + 1 \text{ and } |y| \leq H^{-1}(Q_A(\Lambda_A)) + 1. \quad (2.20)$$

Therefore,

$$p(\|\gamma\|_L; x_m, \lambda_m) = \int_0^T g^{-1} \left(\|\gamma\|_L + \lambda_m \int_0^t f_n(v, x_m(v), x'_m(v)) dv \right) dt \geq 0$$

for $m \in \mathbb{N}$.

Now from $p(0; x_m, \lambda_m) \leq 0$ for $m \in \mathbb{N}$ we deduce that $\{K_m\}$ is bounded and $0 \leq K_m \leq \|\gamma\|_L$. If $\{K_m\}$ is not convergent, there exist subsequences $\{k_m\}$ and $\{l_m\}$ of $\{m\}$ such that $\lim_{m \rightarrow \infty} K_{k_m} = V_1$, $\lim_{m \rightarrow \infty} K_{l_m} = V_2$ and $V_1 \neq V_2$. Then $p(V_j; x_0, \lambda_0) = 0$ ($j = 1, 2$) by the Lebesgue dominated convergence theorem, contrary to the fact that $p(\cdot; x_0, \lambda_0)$ increasing. Hence $\{K_m\}$ is convergent and the continuity of $\mathcal{K}$ follows again from the Lebesgue dominated convergence theorem. By (2.20) and the Bolzano–Weierstrass theorem we can verify that $\mathcal{K}(\overline{\Omega} \times [0, 1])$ is relatively compact in $C^1(J) \times \mathbb{R}$. Hence, property (i) is valid.

Assume that $\mathcal{K}(\bar{x}, \bar{c}, \bar{\lambda}) = (\bar{x}, \bar{c})$ for some $(\bar{x}, \bar{c}) \in \overline{\Omega}$ and $\bar{\lambda} \in [0, 1]$. If $\bar{\lambda} = 0$, then from the first part of our proof we have $\bar{x} = \bar{c} = A$, and so $(\bar{x}, \bar{c}) \notin \partial\Omega$. Let $\bar{\lambda} \in (0, 1]$. Then it can be readily verified that $\bar{x}$ is a solution of the boundary value problem $(E)_n^{\bar{\lambda}}$, (1.2) and $\bar{c} = \bar{x}(0)$. Hence, by Lemma 2.2, $\|\bar{x}\| \leq \max\{|\Lambda_A|, |A|\}$, $\|\bar{x}'\| \leq H^{-1}(Q_A(\Lambda_A))$ and, therefore,

$$|\bar{c}| \leq \max\{|\Lambda_A|, |A|\}.$$

We have proved that $(\bar{x}, \bar{c}) \notin \partial\Omega$, which shows that condition (ii) is satisfied. $\square$

Lemma 2.4. *Let assumptions (H_1) – (H_4) be satisfied and $A \in \mathcal{A}$. Let x_n be a solution of the boundary value problem $(E)_n^1$, (1.2) and $x'_n(\xi_n) = 0$ for a (unique) $\xi \in (0, T)$. Then there exist constants c_1, c_2 such that*

$$0 < c_1 \leq \xi_n \leq c_2 < T \quad \text{for } n \in \mathbb{N}. \quad (2.21)$$

Proof. Assume that there is a subsequence $\{\xi_{k_n}\}$ of $\{\xi_n\}$ such that $\lim_{n \rightarrow \infty} \xi_{k_n} = 0$. Then

$$A - x_{k_n}(0) = x_{k_n}(\xi_{k_n}) - x_{k_n}(0) = x'_{k_n}(\tau_n)\xi_{k_n},$$

where $\tau_n \in (0, \xi_{k_n})$. By Lemma 2.1, we have

$$A - x_{k_n}(0) \geq \int_0^{T/2} g^{-1}(as) ds,$$

and, therefore,

$$\lim_{n \rightarrow \infty} x'_{k_n}(\tau_n) = (A - x_{k_n}(0))/\xi_{k_n} = \infty,$$

contrary to the inequality $\|x'_n\| \leq H^{-1}(Q_A(\Lambda_A))$ for $n \in \mathbb{N}$ (see Lemma 2.2). Hence $\xi_n \geq c_1$ for $n \in \mathbb{N}$ with a positive constant c_1 . The inequalities $\xi_n \leq c_2 < T$, $n \in \mathbb{N}$, with a constant c_2 can be proved analogously. $\square$

Lemma 2.5. *Let assumptions (H_1) – (H_4) be satisfied and $A \in \mathcal{A}$. Let x_n be a solution of the boundary value problem $(E)_n^1$, (1.2). Then the sequence*

$$\{f_n(t, x_n(t), x'_n(t))\} \subset L_1(J)$$

is uniformly absolutely continuous (UAC) on J , that is, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\int_{\mathcal{M}} |f_n(t, x_n(t), x'_n(t))| dt < \varepsilon, \quad n \in \mathbb{N}$$

whenever $\mathcal{M} \subset J$ is measurable and $\mu(\mathcal{M}) < \delta$.

Proof. By Lemma 2.2,

$$\|x_n\| \leq \max\{|\Lambda_A|, |A|\}, \quad \|x'_n\| \leq H^{-1}(Q_A(\Lambda_A)) \quad \text{for } n \in \mathbb{N}, \quad (2.22)$$

where the constant Λ_A is defined by (2.10), and then (see (2.1))

$$\begin{aligned} |f_n(t, x_n(t), x'_n(t))| &\leq \max\{q(u) : |u| \leq \max\{|\Lambda_A|, |A|\}\} \times \\ &\times \left(\max\{\omega_{*1}(u) : 0 \leq u \leq H^{-1}(Q_A(\Lambda_A))\} + \omega_2(|x'_n(t)|) \right) \end{aligned} \quad (2.23)$$

for a. e. $t \in J$ and every $n \in \mathbb{N}$. In addition, $x'(\xi_n) = 0$ for a unique $\xi_n \in (0, T)$,

$$x'_n(t) \geq g^{-1}(a(\xi_n - t)) \quad \text{for } t \in [0, \xi_n], \quad (2.24)$$

$$|x'_n(t)| \geq g^{-1}(a(t - \xi_n)) \quad \text{for } t \in [\xi_n, T] \quad (2.25)$$

for $n \in \mathbb{N}$ by Lemma 2.1, and (2.21) is true with constants c_1, c_2 . Now (2.23) yields that $\{f_n(t, x_n(t), x'_n(t))\}$ is UAC on J if $\{\omega_2(|x'_n(t)|)\}$ is UAC on J .

Due to the structure of the measurable sets in $\mathbb{R}$, it follows that $\{\omega_2(|x'_n(t)|)\}$ is UAC on J if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any at most countable set $\{(a_j, b_j)\}_{j \in \mathbb{J}}$ of the mutually disjoint intervals $(a_j, b_j) \subset J$, $\sum_{j \in \mathbb{J}} (b_j - a_j) < \delta$, we have

$$\sum_{j \in \mathbb{J}} \int_{a_j}^{b_j} \omega_2(|x'_n(t)|) dt < \varepsilon, \quad n \in \mathbb{N}.$$

For this reason, let $\{(a_j, b_j)\}_{j \in \mathbb{J}}$ be an at most countable set of mutually disjoint intervals $(a_j, b_j) \subset J$. Let us put

$$\mathbb{J}_n^1 = \{j : j \in \mathbb{J}, (a_j, b_j) \subset (0, \xi_n)\}$$

and

$$\mathbb{J}_n^2 = \{j : j \in \mathbb{J}, (a_j, b_j) \subset (\xi_n, T)\}.$$

Then, for $j \in \mathbb{J}_n^1$ and $i \in \mathbb{J}_n^2$, in view of (2.24) and (2.25), we have

$$\int_{a_j}^{b_j} \omega_2(|x'_n(t)|) dt \leq \int_{a_j}^{b_j} \omega_2(g^{-1}(a(\xi_n - t))) dt = \frac{1}{a} \int_{a(\xi_n - b_j)}^{a(\xi_n - a_j)} \omega_2(g^{-1}(s)) ds$$

and

$$\int_{a_i}^{b_i} \omega_2(|x'_n(t)|) dt \leq \int_{a_i}^{b_i} \omega_2(g^{-1}(a(t - \xi_n))) dt = \frac{1}{a} \int_{a(a_i - \xi_n)}^{a(b_i - \xi_n)} \omega_2(g^{-1}(s)) ds.$$

If $\mathbb{J}_n^1 \cup \mathbb{J}_n^2 \neq \mathbb{J}$ and $\{j_0\} = \mathbb{J} \setminus (\mathbb{J}_n^1 \cup \mathbb{J}_n^2)$, then

$$\begin{aligned} \int_{a_{j_0}}^{b_{j_0}} \omega_2(|x'_n(t)|) dt &\leq \int_{a_{j_0}}^{\xi_n} \omega_2(g^{-1}(a(\xi_n - t))) dt + \int_{\xi_n}^{b_{j_0}} \omega_2(g^{-1}(a(t - \xi_n))) dt \\ &= \frac{1}{a} \left[\int_0^{a(\xi_n - a_{j_0})} \omega_2(g^{-1}(s)) ds + \int_0^{a(b_{j_0} - \xi_n)} \omega_2(g^{-1}(s)) ds \right]. \end{aligned}$$

Let us set

$$\Delta_n^1 = \bigcup_{j \in \mathbb{J}_n^1} (a(\xi_n - b_j), a(\xi_n - a_j)) + \mathcal{E}_n^1$$

and

$$\Delta_n^2 = \bigcup_{j \in \mathbb{J}_n^2} (a(a_j - \xi_n), a(b_j - \xi_n)) + \mathcal{E}_n^2$$

where

$$\mathcal{E}_n^1 = \begin{cases} \emptyset & \text{if } \mathbb{J} = \mathbb{J}_n^1 \cup \mathbb{J}_n^2 \\ (0, a(\xi_n - a_{j_0})) & \text{if } \{j_0\} = \mathbb{J} \setminus (\mathbb{J}_n^1 \cup \mathbb{J}_n^2) \end{cases}$$

and

$$\mathcal{E}_n^2 = \begin{cases} \emptyset & \text{if } \mathbb{J} = \mathbb{J}_n^1 \cup \mathbb{J}_n^2 \\ (0, a(b_{j_0} - \xi_n)) & \text{if } \{j_0\} = \mathbb{J} \setminus (\mathbb{J}_n^1 \cup \mathbb{J}_n^2). \end{cases}$$

Then

$$\mu(\Delta_n^i) \leq a \sum_{j \in \mathbb{J}} (b_j - a_j), \quad i = 1, 2 \quad (2.26)$$

and

$$\sum_{j \in \mathbb{J}} \int_{a_j}^{b_j} \omega_2(|x'_n(t)|) dt \leq \frac{1}{a} \left[\int_{\Delta_n^1} \omega_2(g^{-1}(t)) dt + \int_{\Delta_n^2} \omega_2(g^{-1}(t)) dt \right]. \quad (2.27)$$

Since $\omega_2(g^{-1}) \in L_{\text{loc}}([0, \infty))$, we conclude from (2.26) and (2.27) that the sequence $\{\omega_2(|x'_n(t)|)\}$ is UAC on J . $\square$

3. EXISTENCE RESULT AND AN EXAMPLE

Theorem 3.1. *Let assumptions $(H_1) - (H_4)$ be satisfied. Then boundary value problem (1.1), (1.2) has a solution for every $A \in \mathcal{A}$.*

Proof. Let us fix $A \in \mathcal{A}$. By Lemmas 2.1–2.4, there exists a solution x_n of the boundary value problem $(E)_n^1$, (1.2) and inequalities (2.21)–(2.25) are satisfied, where Λ_A is defined by (2.10), $\xi_n \in (0, T)$ is the unique zero of x'_n and c_1, c_2 are constants. In addition, by Lemma 2.5, the sequence $\{f_n(t, x_n(t), x'_n(t))\}$ is UAC on J , which implies that the sequence

$$\{g(x'_n(t))\} = \left\{ \int_{\xi_n}^t f_n(s, x_n(s), x'_n(s)) ds \right\}$$

is equicontinuous on J . Then, in view of the relation

$$|x'_n(t_2) - x'_n(t_1)| = |g^{-1}(g(x'_n(t_2))) - g^{-1}(g(x'_n(t_1)))|, \quad t_1, t_2 \in J, \quad n \in \mathbb{N};$$

and the fact that g^{-1} is continuous and increasing on $\mathbb{R}$, we conclude that $\{x'_n(t)\}$ is equicontinuous on J . Hence, $\{x_n\}$ is bounded in $C^1(J)$, $\{x'_n(t)\}$ is equicontinuous on J and since $\{\xi_n\}$ is bounded, using the Arzelà–Ascoli theorem and the Bolzano–Weierstrass theorem, we can assume without loss of generality that $\{x_n\}$ is convergent in $C^1(J)$ and $\{\xi_n\}$ is convergent in $\mathbb{R}$.

Let $\lim_{n \rightarrow \infty} x_n = x$ and $\lim_{n \rightarrow \infty} \xi_n = \xi$. Then $x \in C^1(J)$, $x(0) = x(T)$, $\max\{x(t) : t \in J\} = A$, $0 < c_1 \leq \xi \leq c_2 < T$, and

$$x'(t) \geq g^{-1}(a(\xi - t)) \quad \text{for } t \in [0, \xi], \quad |x'(t)| \geq g^{-1}(a(t - \xi)) \quad \text{for } t \in [\xi, T].$$

Therefore,

$$\lim_{n \rightarrow \infty} f_n(t, x_n(t), x'_n(t)) = f(t, x(t), x'(t)) \quad \text{for a. e. } t \in J,$$

Thus, $f(t, x(t), x'(t)) \in L_1(J)$ and

$$\lim_{n \rightarrow \infty} \int_0^t f_n(s, x_n(s), x'_n(s)) ds = \int_0^t f(s, x(s), x'(s)) ds, \quad t \in J,$$

by the Vitali convergence theorem. Passing to the limit as $n \rightarrow \infty$ in the equalities

$$g(x'_n(t)) = \int_{\xi_n}^t f_n(s, x_n(s), x'_n(s)) ds, \quad t \in J, \quad n \in \mathbb{N}$$

we get

$$g(x'(t)) = \int_{\xi}^t f(s, x(s), x'(s)) ds, \quad t \in J.$$

Hence, $g(x') \in AC(J)$ and x is a solution of boundary value problem (1.1), (1.2). $\square$

Example 3.2. Consider the differential equation

$$(|x'(t)|^p \operatorname{sgn} x'(t))' = -(2 + \sin(tx(t)) + |x(t)|) \left(|x'(t)|^\alpha + \frac{1}{|x'(t)|^\beta} \right) \quad (3.1)$$

where $p > 1$, $0 \leq \alpha < p - 1$ and $0 < \beta < p$.

Let us set

$$f(t, x, y) = -(2 + \sin(tx) + |x|) \left(|y|^\alpha + \frac{1}{|y|^\beta} \right)$$

for $(t, x, y) \in J \times \mathbb{R} \times (\mathbb{R} \setminus \{0\})$. Then (H_1) is satisfied with $g(u) = |u|^p \operatorname{sgn} u$, and f satisfies (H_2) and (H_3) with $a = 1$, $q(x) = 3 + |x|$, $\omega_1(u) = u^\alpha$ and $\omega_2(u) = 1/u^\beta$. Therefore,

$$\omega_{*1}(u) = \begin{cases} 1 & \text{for } 0 \leq u \leq 1 \\ u^\alpha & \text{for } u > 1 \end{cases}$$

and

$$Q_A(u) = \int_u^A q(s) ds = \int_u^A (3 + |s|) ds \geq 3(A - u), \quad (3.2)$$

$$Q_A(u) = \int_u^A (3 + |s|) ds \leq (3 + \max\{|A|, |u|\})(A - u) \quad (3.3)$$

for $u \leq A$. Since

$$\begin{aligned} H(u) &= \int_0^{g(u)} \frac{g^{-1}(s)}{\omega_{*1}(g^{-1}(s)) + \omega_2(g^{-1}(s))} ds = \int_0^{u^p} \frac{s^{1/p}}{1 + s^{-\beta/p}} ds \\ &\leq \int_0^{u^p} s^{(1+\beta)/p} ds = \frac{p}{1+\beta+p} u^{1+\beta+p} \end{aligned}$$

for $u \in [0, 1]$, we have that

$$H^{-1}(u) \geq \left(\frac{1+\beta+p}{p} u \right)^{1/(1+\beta+p)}$$

for $u \in [0, H(1)]$ and, hence (see (3.2)),

$$\int_{A-\varepsilon}^A \frac{1}{H^{-1}(Q_A(s))} ds \leq \int_{A-\varepsilon}^A \left(\frac{3(1+\beta+p)}{p} (A-s) \right)^{-1/(1+\beta+p)} ds < \infty$$

for $\varepsilon = \min\{1, A - Q_A^{-1}(H(1))\}$ because $1/(1+\beta+p) < 1$. This shows that (H_4) is satisfied.

We are going to verify that $\mathcal{A} = \mathbb{R}$, where $\mathcal{A}$ is given by (2.7). For $u \geq 1$, we have

$$\begin{aligned} H(u) &= H(1) + \int_1^{u^p} \frac{s^{1/p}}{s^{\alpha/p} + s^{-\beta/p}} ds = H(1) + \int_1^{u^p} \frac{s^{(1+\beta)/p}}{1 + s^{(\alpha+\beta)/p}} ds \\ &\geq H(1) + \frac{1}{2} \int_1^{u^p} s^{(1-\alpha)/p} ds = H(1) + \frac{p}{2(1+p-\alpha)} (u^{1+p-\alpha} - 1) \end{aligned}$$

and, consequently,

$$H^{-1}(u) \leq \left(\frac{2(1+p-\alpha)}{p} (u - H(1)) + 1 \right)^{1/(1+p-\alpha)}$$

for $u \in [H(1), \infty)$. Hence, we have (see (3.3))

$$\begin{aligned} \int_u^A \frac{1}{H^{-1}(Q_A(s))} ds &= \int_{Q_A^{-1}(H(1))}^A \frac{1}{H^{-1}(Q_A(s))} ds + \int_u^{Q_A^{-1}(H(1))} \frac{1}{H^{-1}(Q_A(s))} ds \\ &> \int_u^{Q_A^{-1}(H(1))} \left[\frac{2(1+p-\alpha)}{p} (3 + \max\{|A|, |s|\}) (A - s) - H(1) + 1 \right]^{-\frac{1}{1+p-\alpha}} ds \end{aligned}$$

for $u < Q_A^{-1}(H(1))$. Since $2/(1+p-\alpha) < 1$, we see that

$$\lim_{u \rightarrow -\infty} \int_u^A \frac{1}{H^{-1}(Q_A(s))} ds = \infty$$

for every $A \in \mathbb{R}$ and, hence, we have proved that $\mathcal{A} = \mathbb{R}$. Applying Theorem 3.1, we conclude that the boundary value problem (3.1), (1.2) has a solution for all $A \in \mathbb{R}$.

REFERENCES

- [1] AGARWAL, R. P., O'REGAN, D., AND STANĚK, S.: *Singular Lidstone boundary value problem with given maximal values for solutions*, to appear.
- [2] AGARWAL, R. P., O'REGAN, D., AND STANĚK, S.: *Solvability of singular Dirichlet boundary value problems with given maximal values for positive solutions*, to appear.
- [3] BARTLE, R. G.: *A Modern Theory of Integration*, Amer. Math. Soc., Providence, R. I., 2001.
- [4] BRYKALOV, S. A.: *Solutions with a prescribed minimum and maximum*, Differents. Uravn., **29** (1993), 638–642 (in Russian). English transl. in Differential Equations, **29** (1993), 802–805.
- [5] BRYKALOV, S. A.: *Solvability of problems with monotone boundary conditions*, Differents. Uravn., **29** (1993), 744–750 (in Russian). English transl. in Differential Equations, **29** (1993), 633–639.
- [6] BRYKALOV, S. A.: *Multiplicity results for problems with uniform norms in boundary conditions*, Nonlinear Anal., **30** (1997), 4781–4788.
- [7] DEIMLING, K.: *Nonlinear Functional Analysis*, Springer-Verlag, Berlin, Heidelberg, etc., 1985.
- [8] FEČKAN, M.: *The interaction of linear boundary value and nonlinear functional conditions*, Ann. Polon. Math., **58** (1993), 299–310.
- [9] HAKI, R., KIGURADZE, I., AND PŮŽA, B.: *Upper and lower solutions to boundary value problems for functional-differential equations and theorems on functional-differential inequalities*, Georgian Math. J., **7** (2000), 489–512.
- [10] HAŠÁK, A., PIRČ, V., AND STANĚK, S.: *On boundary value problems for third-order functional differential equations*, Nonlinear Anal., **42** (2000), 1077–1089.

- [11] HEWITT, E. AND STROMBERG, S.: *Real and Abstract Analysis*, Springer-Verlag, Berlin, Heidelberg, etc., 1965.
- [12] KIGURADZE, I. AND PŮŽA, B.: *On boundary value problems for functional differential equations*, Mem. Differential Equations Math. Phys., **12** (1997), 106–113.
- [13] KIGURADZE, I. AND PŮŽA, B.: *On the solvability of nonlinear boundary value problems for functional differential equations*, Georgian Math. J., **5** (1998), 251–262.
- [14] KIGURADZE, I. T. AND SHEKHTER, B. L.: *Singular boundary value problems for second order ordinary differential equations*, In: “Current Problems in Math.: Newest Results,” Moscow, **30**, 105–201 (in Russian). English transl. in J. Soviet Math., **43** (1988), 2340–2417.
- [15] LOMTATIDZE, A. AND MALAGUTI, L.: *On a nonlocal boundary value problem for second order nonlinear singular differential equations*, Georgian Math. J., **7** (2000), 133–154.
- [16] LEPIN, A. YA. AND MYSHKIS, A. D.: *General nonlocal nonlinear boundary value problem for differential equation of 3rd order*, Nonlinear Anal., **28** (1997), 1533–1543.
- [17] STANĚK, S.: *Multiplicity results for second order nonlinear problems with maximum and minimum*, Math. Nachr., **192** (1998), 225–237.
- [18] STANĚK, S.: *Existence and multiplicity of solutions for functional boundary value problems*, Nonlinear Anal., **47** (2001), 3925–3936.
- [19] STANĚK, S.: *Two functional boundary value problems with singularities in phase variables*, Nonlinear Oscillations, **5** (2002), 387–415.
- [20] STANĚK, S.: *On nonlocal boundary value problem with singularities in phase variables*, submitted.

Author's Address

Svatoslav Staněk:

DEPARTMENT OF MATHEMATICAL ANALYSIS, FACULTY OF SCIENCE, PALACKÝ UNIVERSITY, TOMKOVA 40, 779 00
OLOMOUČ, CZECH REPUBLIC

E-mail address: stanek@risc.upol.cz



PARABOLA–PARABOLA COMBINED METHOD

Z. SZABÓ

ABSTRACT. By using the always convergent method of tangential parabolas (but not using any interval-arithmetic tool), a combined root-finding iterative algorithm is given which provides a quadratically convergent descending sequence of compact real intervals J_n containing a simple zero of a twice differentiable real function defined on J_0 .

Mathematics Subject Classification: 65H05

Keywords: Nonlinear equation, iteration, always convergent method, tangential parabola, decreasing sequence of compact intervals

1. PRELIMINARIES, CONVERGENCE, ERROR ESTIMATION

IN [1] AND [2], we described two combined root-finding algorithms based on Newton's method and the method of tangential parabolas which is always convergent. The information used by the method in [1] comprises 4 Horner units: $f(a_n)$, $f'(a_n)$, $f(b_n)$, and $f'(b_n)$. This number is reduced to 2 in [2]. Keeping this number 2 of Horner's units used, we aim here at finding a faster combined method generating an interval sequence with the known features. The conditions (2), (3), and (6) in [2] will be replaced by the assumption

$$m_2 \leq |f''(x)| \text{ in } I = [a, b] \subset \mathbb{R}, \text{ for some } m_2 > 0, \quad (1.1)$$

provided f'' does not change its sign in I .

In this new algorithm, the tangential straight line in the “Newton–Parabola” combined method described in [2] will be replaced by another tangential parabola lying “outside” the graph of f (the area between the curve of f and the x -axis is considered to be “inside”). For this “outer parabola,” we first prove our

Lemma 1. Assume that the function $f : I = [a, b] \rightarrow \mathbb{R}$ is twice differentiable in the compact interval $I \subset \mathbb{R}$, assumption (1.1) holds, and $f(x_0) \neq 0$ for some x_0 in I . Then, for the tangential parabola

$$p(x) = f(x_0) + f'(x_0)(x - x_0) - \frac{1}{2}m_2 \operatorname{sign}(f(x_0))(x - x_0)^2,$$

Supported by the Hungarian Ministry of Culture and Education under Grant No. FKFP0429.

the difference $p(x) - f(x)$ is non-negative (resp., non-positive), provided $f(x_0)$ is positive (resp., negative), and f is concave from below (resp., above) in I .

In other words, we claim that $f(x_0)(p(x) - f(x)) \geq 0$ in I , provided the assumptions are fulfilled.

Proof of Lemma 1. Let $f(x_0) > 0$. Then we have

$$\begin{aligned} p(x) &= f(x_0) + f'(x_0)(x - x_0) - \frac{1}{2}m_2(x - x_0)^2, \\ f(x) &= f(x_0) + f'(x_0)(x - x_0) - \frac{1}{2}f''(c)(x - x_0)^2 \end{aligned}$$

for some $c \in (x_0, x)$, and $0 < m_2 \leq -f''(x)$, $x \in I$. Thus,

$$p(x) - f(x) = \frac{1}{2}(-f''(c) - m_2)(x - x_0)^2 \geq 0, \quad x \in I,$$

because $-f''(c) = |f''(c)| \geq m_2$. When $f(x_0) < 0$, we have $m_2 \leq f''(x)$, $x \in I$, and the proof is similar, namely, in $p(x) - f(x)$, the term $-f''(c)$ will be replaced by $-m_2$. $\square$

The convexity assumptions in Lemma 1 can be summarized as follows:

$$0 < m_2 \leq -\text{sign}(f(x_0)) f''(x), \quad x \in I.$$

We need now a more accurate notation for the iteration function of the “tangential parabola” method (briefly, TP-method) described in [1, pp. 581-582]. Let

$$F_{TP}(x, r, M) := \begin{cases} x + sf'(x)/M + r(2|f(x)|/M + (f'(x)/M)^2)^{1/2} & \text{if } s = f(x_0) \neq 0; \\ x & \text{if } s = 0. \end{cases}$$

In order to define our combined “parabola-parabola” method (PP-method for short), we assume that the following condition is satisfied:*

- (A) The nonlinear function $f : I = [a, b] \rightarrow \mathbb{R}$ is twice differentiable on $I^\dagger$, the inequalities $|f''(x)| \leq M_2 \neq 0$, $x \in I$, and $f(a)f(b) < 0$ are fulfilled, and f'' does not change its sign on I .[‡]

In addition, we assume that (1.1) holds as well. Let, e. g. $f(a) > 0$ and $f''(x) < 0$. Thus, $f(a)f''(x) < 0$. (It does not matter whether $f'(a)$ is negative or non-negative.) Starting with the points $a_0 = a$ and $b_0 = b$, we construct the sequences

$$a_{n+1} = F_{TP}(a_n, 1, M_2), \quad b_{n+1} = \min \{F_{TP}(a_n, 1, m_2), b\}, \quad (1.2)$$

*Actually, assumption (A) is the same as (1) in [2].

[†]At the endpoints of I , the one-sided derivatives are to be considered.

[‡]It follows from (A) that f' is monotonic on I , f has a unique zero (say, α) in I , and α is a simple zero: $f(\alpha) = 0$, $f'(\alpha) \neq 0$.

for $n = 0, 1, 2, \dots$. Since α is the only zero of f in (a, b) and, according to [1, pp. 581-582], F_{TP} is an always convergent iteration function: $F_{TP} \in A(f, I)$, and we have $\lim_{n \rightarrow \infty} a_n = \alpha$. So, after finitely many (say, N) steps,

$$b_N = \min \{F_{TP}(a_{N-1}, 1, m_2), b\} = F_{TP}(a_{N-1}, 1, m_2) \leq b$$

and, for $n \geq N$, the formulae (1.2) take the form

$$a_{n+1} = F_{TP}(a_n, 1, M_2), \quad b_{n+1} = F_{TP}(a_n, 1, m_2), \quad n = 0, 1, 2, \dots \quad (1.3)$$

In this manner, we obtain the sequence of compact real intervals

$$J_n = [a_n, b_n]; \quad n = 0, 1, 2, \dots; \quad (J_0 = I).$$

It follows that

$$J_{n+1} \subset J_n, \quad n = 0, 1, 2, \dots, \quad \text{and} \quad \bigcap_{n=0}^{\infty} J_n = \alpha \in I, \quad f(\alpha) = 0, \quad f'(\alpha) \neq 0.$$

In general, we start our iteration from the endpoint $E \in \{a, b\}$ for which the inequality

$$f(E)f''(x) < 0 \quad (1.4)$$

is fulfilled. If, by using formulae similar to (1.3), a_n (or b_n) does not belong to I , then we choose the nearest endpoint of I to be a_n (or b_n). This nearest endpoint will be $\hat{E} = \{a, b\} \setminus \{E\}$. In particular, if $f(a)f''(x) < 0$, then $E = a$, $\hat{E} = b$, $r = 1$, and (1.2) remains the same; after a finite number of steps, we reach a point of iteration a_n such that f' keeps its sign in $[a_n, b]$; if $f(b)f''(x) < 0$, then $E = b$, $\hat{E} = a$, $r = -1$, and (1.2) becomes

$$a_{n+1} = \max \{F_{TP}(b_n, -1, m_2), a\}, \quad b_{n+1} = F_{TP}(b_n, -1, M_2), \quad (1.5)$$

for $n = 0, 1, 2, \dots$. In this case, after finitely many steps, we reach a point b_n such that f' keeps its sign in $[a, b_n]$. On the other hand, we have $r = \text{sign}((a+b)/2 - E) = \text{sign}(a+b-2E)$. An error estimate for our PP-method is contained in the following

Lemma 2. Assume that conditions (1.1) and (A) are fulfilled, and f' keeps its sign in $I = J_0$. Then, for the diameters of the intervals J_n , we have the estimate

$$d(J_{n+1}) \leq C(d(J_n))^2, \quad n = 0, 1, 2, \dots,$$

where

$$C = \frac{1}{2}u^2 \left(M_2 v^{-3} - m_2 (u^2 + 2|f(E)|m_2)^{-3/2} \right),$$

$$u = |f'(\hat{E})| = \max \{|f'(a)|, |f'(b)|\},$$

$$v = |f'(E)| = \min \{|f'(a)|, |f'(b)|\}$$

and $E, \hat{E}$ are defined by the relations $\{E, \hat{E}\} = \{a, b\}$ and (1.4).

Proof. Let, e. g., $f(a) > 0$, $f''(x) < 0$. Then $E = a$, $\hat{E} = b$. By using the formulas

$$a_{n+1} = a_n + f'(a_n)/M_2 + (2f(a_n)/M_2 + (f'(a_n)/M_2)^2)^{1/2}$$

and

$$b_{n+1} = a_n + f'(a_n)/m_2 + (2f(a_n)/m_2 + (f'(a_n)/m_2)^2)^{1/2},$$

we get

$$\begin{aligned} d(J_{n+1}) &= b_{n+1} - a_{n+1} = f'(a_n)(1/m_2 - 1/M_2) \\ &\quad + \left(2f(a_n)/m_2 + (f'(a_n)/m_2)^2\right)^{1/2} \\ &\quad - \left(2f(a_n)/M_2 + (f'(a_n)/M_2)^2\right)^{1/2} \\ &= 1/m_2(y + F(m_2)) - 1/M_2(y + F(M_2)) \end{aligned} \quad (1.6)$$

where $y = f'(a_n) = -|y| < 0$, $q = 2f(a_n)$ and $F(m) = (y^2 + qm)^{1/2}$. We take $F'(m) = \frac{1}{2}q(y^2 + qm)^{-1/2}$ and $F''(m) = -\frac{1}{4}q^2(y^2 + qm)^{-3/2}$. By using Taylor's formula, we obtain

$$\begin{aligned} F(m) &= F(0) + F'(0)m + \frac{1}{2}F''(w)m^2 \\ &= |y| + qm/(2|y|) - \frac{1}{8}q^2m^2(y^2 + qw)^{-3/2}, \end{aligned}$$

where $w \in (0, m)$. Then (1.6) takes the form

$$\begin{aligned} d(J_{n+1}) &= \frac{1}{m_2} \left(y + |y| + q \frac{m_2}{2|y|} - \frac{1}{8}q^2m_2^2(y^2 + qw)^{-3/2} \right) \\ &\quad - \frac{1}{M_2} \left(y + |y| + q \frac{M_2}{2|y|} - \frac{1}{8}q^2M_2^2(y^2 + qW)^{-3/2} \right) \\ &= \frac{q^2}{8} \left(M_2(y^2 + qW)^{-3/2} - m_2(y^2 + qw)^{-3/2} \right) \\ &= \frac{1}{2}f^2(a_n) \left(M_2(f'^2(a_n) + 2f(a_n)W)^{-3/2} - m_2(f'^2(a_n) + 2f(a_n)w)^{-3/2} \right), \end{aligned}$$

where $w \in (0, m_2)$ and $W \in (0, M_2)$. The factor $f^2(a_n)$ on the right-hand side can be replaced by $f'^2(t_n)(\alpha - a_n)^2$ for some $t_n \in (a_n, \alpha)$ because, by the mean value theorem, $f(\alpha) - f(a_n) = f'(t_n)(\alpha - a_n)$. Thus, one can estimate $d(J_{n+1})$ as follows:

$$\begin{aligned} d(J_{n+1}) &\leq \frac{1}{2}f'^2(t_n)(\alpha - a_n)^2 \left(M_2/|f'(a_n)|^3 - m_2(f'^2(a_n) + 2f(a_n)m_2)^{-3/2} \right) \\ &\leq \frac{1}{2}f'^2(b)(\alpha - a_n)^2 \left(M_2/|f'(a)|^3 - m_2(f'^2(b) + 2f(a)m_2)^{-3/2} \right) \\ &= \frac{1}{2}u^2(\alpha - a_n)^2 \left(M_2/v^3 - m_2(u^2 + 2|f(E)|m_2)^{-3/2} \right) \\ &\leq C(b_n - a_n)^2 = C(d(J_n))^2. \end{aligned}$$

In the remaining three cases corresponding to the signs of $f(a)$ and $f''(x)$, the proofs are quite similar. $\square$

Remark 1. Assumptions (1.1) and (A) imply that

$$v < u < \left(u^2 + 2|f(E)|m_2\right)^{1/2},$$

so $C > 0$.

Remark 2. A geometrical representation shows that our PP-method is faster than the combined method described in [2]. (Namely, the “outer parabola” lies below (resp., above) the tangential line of Newton’s method if $f(a_n) > 0$ (resp., $f(a_n) < 0$). This fact is also reflected in the error estimates because

$$C = K - \frac{1}{2}u^2m_2\left(u^2 + 2|f(E)|m_2\right)^{-3/2}.$$

Remark 3. If $|f(a_n)| \ll 1$ (i. e., a_n is close to α), then

$$C \approx \frac{1}{2}u^2\left(M_2/v^3 - m_2/u^3\right) = K - \frac{1}{2}m_2/u,$$

where K is the error constant in [2].

Remark 4. For the endpoint $\hat{E}$, we have $f(\hat{E})f''(x) > 0$. According to the Fourier conditions, the Newton method can be used, starting at $\hat{E}$ (like in [1]).

The results above can be summarized in our main

Theorem 1. *If assumptions (1.1) and (A) hold, then the sequence of compact real intervals $J_n = [a_n, b_n]$, $n = 0, 1, 2, \dots$, generated by the combined PP-method described above has the following properties:*

- 1° $J_{n+1} \subset J_n$, $n = 0, 1, 2, \dots$;
- 2° $\bigcap_{n=0}^{\infty} J_n = \alpha \in I$, $f(\alpha) = 0$, $f'(\alpha) \neq 0$;
- 3° if f' keeps its sign on $I = J_0$, then, for the diameters of the resulting intervals, we have the estimate

$$d(J_{n+1}) \leq C(d(J_n))^2, \quad n = 0, 1, 2, \dots,$$

where

$$C = \frac{1}{2}u^2\left(M_2v^{-3} - m_2\left(u^2 + 2|f(E)|m_2\right)^{-3/2}\right),$$

$$u = |f'(\hat{E})| = \max\{|f'(a)|, |f'(b)|\}, \quad v = |f'(E)| = \min\{|f'(a)|, |f'(b)|\},$$

and $E, \hat{E}$ are defined by the relations $\{E, \hat{E}\} = \{a, b\}$ and (1.4).

2. ALGORITHM AND SOME NUMERICAL EXAMPLES

The algorithm for the PP-method can be given by the following Boolean function procedure *PPM*. The input parameters are $a, b, \text{eps}, m_2, M_2, f$ and f_1 (corresponding to f'). If they fulfil the requirements $a < b$, $\text{eps} > 0$, $0 < m_2 \leq M_2$ and $f(a)f(b) \leq 0$, then *PPM*, the identifier of our Boolean procedure, will have the logical value *true*, otherwise *PPM* will be *false*. The other output parameters are the endpoints of the resulting interval $[A, B]$ including α (the zero of f) and having a diameter less than eps , provided *PPM* = *true*. Otherwise, $[A, B]$ will be the initial interval $[a, b]$.

The subroutines/procedures for the functions f and f' have to be defined by the user.

SUB is a real function procedure with the formal input parameter L . The actual input parameters are m_2 and M_2 . This procedure actually calculates the value of $F_{TP}(z, r, L)$, the tangential parabola iterate of z for $L \in \{m_2, M_2\}$, and puts it into the output parameter *SUB*.

The integer variable k contains the information on convexity of the graph of f . If $k = 1$, then $(f(a) + f(b))/2 > f((a+b)/2)$ and f is convex; if $k = -1$, then f is concave. Thus, $k = \text{sign}(f(a) + f(b) - 2f((a+b)/2))$. (We have $k \neq 0$, i. e., f is non-linear because $m_2 > 0$.)

```

Boolean procedure PPM ( $a, b, \text{eps}, m_2, M_2, A, B$ );
begin integer  $k, r, s$ ; real  $a, b, \text{eps}, m_2, M_2, A, B, F, H, P, y, z$ ;
  real procedure  $f(x)$ ; real  $x$ ;      (user should define  $f$ )
  real procedure  $f_1(x)$ ; real  $x$ ;      (user should define  $f_1 = f'$ )
  real procedure SUB( $L$ ); real  $L$ ;
  begin real  $T, S$ ;  $T := 2 \star \text{abs}(F)/L$ ;  $S := H/L$ ;
  SUB :=  $z + s \star S + r \star \text{sqr}(T + S \star S)$  end SUB;
   $A := a$ ;  $B := b$ ;
  if  $a < b$  and  $\text{eps} > 0$  and  $0 < m_2$  and  $m_2 \leq M_2$  then goto gd;
bd: PPM := false; goto E;
gd: PPM := true;
   $F := f(a)$ ; if  $F = 0$  then begin  $B := a$ ; goto E end;
   $P := f(b)$ ; if  $P = 0$  then begin  $A := b$ ; goto E end;
  if  $F \star P > 0$  then goto bd;  $k := \text{sign}(F + P - 2 \star f((a+b)/2))$ ;
  if  $F \star k < 0$  then begin  $z := a$ ;  $r := 1$ ; end
  else begin  $z := b$ ;  $F := P$ ;  $r := -1$ ; end;
   $H := f_1(z)$ ;  $s := \text{sign}(F)$ ;  $y := \text{SUB}(m_2)$ ;
  if  $a > y$  then  $y := a$ ; if  $y > b$  then  $y := b$ ;
W:  $z := \text{SUB}(M_2)$ ; if  $\text{abs}(z - y) < \text{eps}$  then goto T;
   $F := f(z)$ ;  $H := f_1(z)$ ;  $y := \text{SUB}(m_2)$ ; goto W;
T: if  $r > 0$  then begin  $A := z$ ;  $B := y$ ; end
  else begin  $A := y$ ;  $B := z$ ; end;
E: end PPM;

```

Some numerical examples can be found in the following tables, where the correct digits are underlined.

Example 1. $f(x) = x - e^{-x}$, $a = 0$, $b = 1 (= E)$, $M_2 = 1$, $m_2 = 0.35$;

n	a_n	b_n	$b_n - a_n$
0	<u>0</u>	1	1
1	<u>0.56238349331149966899</u>	<u>0.59719164168881961091</u>	$3 \cdot 10^{-2}$
2	<u>0.56708373561334769584</u>	<u>0.56727015271662188072</u>	$2 \cdot 10^{-4}$
3	<u>0.56714328929501556755</u>	<u>0.56714329263260011212</u>	$3 \cdot 10^{-9}$
4	<u>0.56714329040978387265</u>	<u>0.56714329040978387368</u>	$1 \cdot 10^{-18}$
5	<u>0.56714329040978387300</u>	<u>0.56714329040978387300</u>	$1 \cdot 10^{-37}$

Example 2. $f(x) = \tan^{-1} x - 2.6 + \sqrt{x}$, $a = 1$, $b = 4 (= E)$, $M_2 = 0.75$, $m_2 = 0.035$;

n	a_n	b_n	$b_n - a_n$
0	1	4	3
1	<u>1.89970378394449937319</u>	<u>2.96088085705371547709</u>	1
2	<u>2.06567277560842922080</u>	<u>2.36679176536415266599</u>	$3 \cdot 10^{-1}$
3	<u>2.13894682376643847337</u>	<u>2.16810797543185165243</u>	$3 \cdot 10^{-2}$
4	<u>2.14658693492719685263</u>	<u>2.14689875120987042863</u>	$3 \cdot 10^{-4}$
5	<u>2.14666632870554397592</u>	<u>2.14666636586609630990</u>	$4 \cdot 10^{-8}$
6	<u>2.14666633811284909659</u>	<u>2.14666633811284962657</u>	$5 \cdot 10^{-16}$
7	<u>2.14666633811284923074</u>	<u>2.14666633811284923074</u>	$1 \cdot 10^{-31}$

Example 3. $f(x) = 1 - x - \sin x$, $a = 0.01$, $b = 1 (= E)$, $M_2 = 0.842$, $m_2 = 0.0099$;

n	a_n	b_n	$b_n - a_n$
0	0.01	1	1
1	<u>0.45465326096563166766</u>	<u>0.51736453936087952833</u>	$6 \cdot 10^{-2}$
2	<u>0.51096815380042764464</u>	<u>0.51097723467313242901</u>	$9 \cdot 10^{-6}$
3	<u>0.51097342938671630865</u>	<u>0.51097342938993405418</u>	$3 \cdot 10^{-12}$
4	<u>0.51097342938856910952</u>	<u>0.51097342938856910952</u>	$4 \cdot 10^{-25}$

REFERENCES

- [1] SZABÓ, Z.: *Combined iteration method for solving equations*, Colloquia Math. Soc. J. Bolyai, **50**, Numerical Methods, Miskolc, 1986, pp. 581–587.
- [2] SZABÓ, Z.: *Newton-parabola combined method for solving equations*, In: *Computational & Applied Mathematics, I. Algorithms and Theory*, North-Holland, Amsterdam, 1992, pp. 447–452.

Author's Address

Z. Szabó:

SULTAN QABOOS UNIVERSITY, DEPARTMENT OF MATHEMATICS AND STATISTICS, P. O. Box 36., AL-KHODH
P. C. 123, SULTANATE OF OMAN

E-mail address: szabo@squ.edu.om



THE SECOND INTERNATIONAL WORKSHOP “CONSTRUCTIVE METHODS FOR NON-LINEAR BOUNDARY VALUE PROBLEMS”

MIKLÓS RONTÓ

LAST SUMMER, more precisely, from 4th to 7th June, 2003, the Second International Workshop

“CONSTRUCTIVE METHODS FOR NON-LINEAR BOUNDARY VALUE PROBLEMS”

took place in Miskolc (Hungary). The workshop had been organised by the Institute of Mathematics, University of Miskolc, in cooperation with the Regional Committee of the Hungarian Academy of Sciences in Miskolc.



Participants of the Second International Workshop “CONSTRUCTIVE METHODS FOR NON-LINEAR BOUNDARY VALUE PROBLEMS” (June 4–7, 2003, Miskolc, Hungary)

The topics of the meeting covered the theory of boundary value problems in a broad sense (including, e. g., initial value problems, periodic-type problems for ordinary differential and functional differential equations, two-point problems for equations with impulses, non-local problems for higher-order equations, and applications). More than 30 researchers from Belgium, Czech Republic, Egypt, Hungary, Roumania, Slovakia, Ukraine, and Vietnam participated in the meeting with 20 and 30-minute lectures.

The Organising Committee consisting of G. Bognár, A. Galántai, and the author of this note (chairman), gratefully acknowledges the financial support provided by the following funds and institutions:

- Hungarian National Foundation for Scientific Research (OTKA);
- Peregrinatio Fund, University of Miskolc;
- Faculty of Mechanical Engineering, University of Miskolc;
- Institute of Mathematics, University of Miskolc.

The workshop of year 2003 follows the tradition established in 2000, when the first workshop with this title, participated by 25 mathematicians from 8 countries, took place in Miskolc. When organising this kind of short meetings, our aim is to bring together the leading mathematicians working in the theory of boundary value problems and to promote their active long-term cooperation. We are glad to note that they have already produced useful results and we hope that they will be fruitful in the future.

Author's Address

Miklós Rontó:

INSTITUTE OF MATHEMATICS, UNIVERSITY OF MISKOLC, 3515 MISKOLC-EGYETEMVÁROS, HUNGARY
E-mail address: matronto@gold.uni-miskolc.hu

EDITOR-IN-CHIEF

MIKLÓS RONTÓ, Institute of Mathematics,
University of Miskolc; H-3515 Miskolc–
Egyetemváros, Hungary;
matronto@gold.uni-miskolc.hu

EDITORIAL BOARD

GABRIELLA BOGNÁR, Institute of Mathematics,
University of Miskolc; H-3515 Miskolc–
Egyetemváros, Hungary;
matvbog@gold.uni-miskolc.hu

MÁTYÁS DOMOKOS, A. Rényi Institute of
Mathematics, Hungarian Academy of Sciences;
P. O. Box 127, H-1053 Budapest, Hungary;
domokos@renyi.hu

MIKLÓS FARKAS, School of Mathematics,
Budapest University of Technology and
Economics; 3 Műegyetem rkp. H-1111 Budapest,
Hungary; fm@math.bme.hu

MICHAL FEČKAN, Comenius University,
Department of Mathematical Analysis;
Mlynská dolina, 842 15 Bratislava, Slovakia;
Michal.Feckan@fmph.uniba.sk

AURÉL GALÁNTAI, Institute of Mathematics,
University of Miskolc; H-3515 Miskolc–
Egyetemváros, Hungary;
matgal@gold.uni-miskolc.hu

CHAITAN P. GUPTA, Department of Mathematics,
University of Nevada; Reno, NV 89557, USA;
chaitang@hotmail.com

ISTVÁN GYŐRI, Department of Mathematics
and Computing, University of Veszprém;
P. O. Box 158, H-8201 Veszprém, Hungary;
gyori@almos.vein.hu

LÁSZLÓ HATVANI, Bolyai Institute, University
of Szeged; 1 Aradi vértanúk tere, H-6720 Szeged,
Hungary;
hatvani@math.u-szeged.hu

JÓZSEF KOLUMBÁN, “Babeş-Bolyai”
University of Cluj-Napoca, Faculty of

Mathematics and Computer Science; 1B Mihail
Kogalniceanu St., 3400 Cluj-Napoca,
Romania; kolumban@math.ubbcluj.ro

NIKOLAI PERESTYUK, Kiev Shevchenko
University, Faculty of Mechanics and
Mathematics; 64 Vladimirska St., 01006
Kiev, Ukraine; pmo@mechmat.kiev.ua

ANDREI RONTÓ, Institute of Mathematics,
National Academy of Sciences of Ukraine; 3
Tereshenkovskaya St., 01601 Kiev, Ukraine;
ar@imath.kiev.ua

NIKOLAI ROZOV, Moscow State University,
Faculty of Mechanics and Mathematics;
The Main Building, Vorobyovy Gory,
119899 Moscow, Russia;
rozov@rozov.mccme.ru

ANATOLY SAMOILENKO, Institute of
Mathematics, National Academy of Sciences
of Ukraine; 3 Tereshenkovskaya St.,
01601 Kiev, Ukraine; sam@imath.kiev.ua

EMILIO SPEDICATO, University of Bergamo,
Department of Mathematics, Statistics,
Informatics and Applications; Sede di
Caniana, via dei Caniana 2, 24127 Bergamo
BG, Italy; emilio@unibg.it

GISBERT STOYAN, Eötvös Lóránd University
of Sciences, Department of Numerical
Analysis; 1D Pázmány P., H-1117 Budapest,
Hungary; stoyan@cs.elte.hu

GYÖRGY SZEIDL, University of Miskolc,
Department of Mechanics; H-3515 Miskolc–
Egyetemváros, Hungary;
mechszgy@gold.uni-miskolc.hu

JENŐ SZIGETI, Institute of Mathematics,
University of Miskolc; H-3515 Miskolc–
Egyetemváros, Hungary;
matszj@gold.uni-miskolc.hu

ZSOLT TUZA, Computer and Automation
Institute, Hungarian Academy of Sciences;
13-17 Kende St., H-1111 Budapest, Hungary;
tuza@sztaki.hu

Miskolc Mathematical Notes

Abstracted in *Zentralblatt für Mathematik*, *Mathematical Reviews*, and Периферативный журнал „Математика”

Responsible for publication: Rector of the University of Miskolc
Published by the Miskolc University Press under the leadership of Dr. József Péter
Responsible for duplication: Works manager Mária Kovács
Number of copies printed: 150
Put to the Press on February 26, 2004
Number of permission: TU 2004-180-ME

HU e-ISSN 1787-2413