HU e-ISSN 1787-2413

MATHEMATICAL NOTES

A Publication of the University of Miskolc

VOLUME 1, NUMBER 2 (2000)

MISKOLC UNIVERSITY PRESS

Mathematical Notes

Volume 1, Number 2 (2000)

Contents

Contributed Papers

Doi: 10.18514/MMN.2000.22

Mathematical Notes, Miskolc, Vol. 1., No. 2., (2000), pp. 87–98

HOW TO CHARACTERIZE SOME PROPERTIES OF MEASURABLE FUNCTIONS

N. Kwami Agbeko
Department of Applied Mathematics, University of Miskolc
3515 Miskolc – Egyetemváros, Hungary
matagbek@gold.uni-miskolc.hu

[Received May 23, 2000]

Dedicated to the memory of Prof. J. Mogyoródi

Abstract. Making use of the so-called optimal measures dealt with in [1-2], we characterize the boundedness of measurable functions, the uniform boundedness and some well-known asymptotic behaviours of sequences of measurable functions (such as discrete, equal as well as pointwise types of convergence). The so-called quasi-uniform convergence is also characterized in the fourth section.

Mathematical Subject Classification: 28A20, 28E10
Keywords: measurable functions and their properties

1. Introduction

Making use of the so-called 'optimal measures' (cf. [1-2]), our main goal here is to characterize some well-known notions in Analysis such as the boundedness of measurable functions, the uniform boundedness as well as some commonly used asymptotic behaviors of sequences of measurable functions. We should like to mention that our results in [1-2] and in this article might interest everyone who handles measurable functions. Before we tackle our paper, let us first recall the following results (as we need them later on).

All along $(\Omega, \mathcal{F})$ will stand for any measurable space (where the elements of $\mathcal{F}$ are referred to as measurable sets).

By an optimal measure we mean a set function $p : \mathcal{F} \to [0,1]$ fulfilling the following axioms:

P1. $p(\emptyset) = 0$ and $p(\Omega) = 1$.

P2. $p(B \bigcup E) = p(B) \vee p(E)$ for all measurable sets B and E.

P3. $p\left(\bigcap\limits_{n=1}^{\infty} E_n\right) = \lim\limits_{n\to\infty} p(E_n)$, for every decreasing sequence of measurable sets (E_n).

(The symbols $\bigvee$ and $\bigwedge$ will stand for the maximum and minimum respectively.)

First we shall summarize the background of the Theory of Optimal Measures .

Let $s = \sum_{i=1}^{n} b_i \chi(B_i)$ be an arbitrary nonnegative measurable simple function, where $\{B_i : i = 1, \ldots, n\} \subset \mathcal{F}$ is a partition of Ω. Then the so-called optimal average of s is defined by $\wr_\Omega s dp = \bigvee_{i=1}^{n} b_i p(B_i)$, where $\chi(B)$ is the indicator function of the measurable set B. We note that this quantity does not depend on the decompositions of s (cf. [1], *Theorem 1.0.*, page 135).

The optimal average of a measurable function f is defined by $\wr_\Omega |f| dp = \sup \wr_\Omega s dp$, where the supremum is taken over all measurable simple functions $s \geq 0$ for which $s \leq |f|$. (From now on *m.f.*'s will stand for measurable functions.)

Let f be any m.f. We shall say that f belongs to:

1. $\mathcal{A}_\infty(p)$ if $p(|f| \leq b) = 1$ for some constant $b \in (0, \infty)$.
2. $\mathcal{A}_\alpha(p)$ if $\wr_\Omega |f|^\alpha dp < \infty$, $\alpha \in [1, \infty)$.

For any $\alpha \in [1, \infty]$, the space $\mathcal{A}_\alpha(p)$ endowed with the norm $\|\cdot\|_\alpha$, defined by

$$\|f\|_\alpha = \begin{cases} \inf\{b \in (0, \infty) : p(|f| \leq b) = 1\}, & \text{if } f \in \mathcal{A}_\infty(p), \alpha = \infty \\ \sqrt[\alpha]{\wr_\Omega |f|^\alpha dp}, & \text{if } f \in \mathcal{A}_\alpha(p), \alpha \in [1, \infty) \end{cases}$$

is a Banach space. (For more about this refer to [1].)

In [2] we have obtained the following results for all optimal measures p.

By (p-)atom we mean a measurable set H, $p(H) > 0$ such that whenever $B \in \mathcal{F}$, $B \subset H$, then $p(B) = p(H)$ or $p(B) = 0$.

A p-atom H is decomposable if there exists a subatom $B \subset H$ such that $p(B) = p(H) = p(H \setminus B)$. If no such subatom exists, we shall say that H is indecomposable.

Fundamental Optimal Measure Theorem. *Let $(\Omega, \mathcal{F})$ be a measurable space and p an optimal measure on it. Then there exists a collection $\mathcal{H}(p) = \{H_n : n \in J\}$ of disjoint indecomposable p-atoms, where J is some countable (i.e. finite or countably infinite) index-set such that for any measurable set B, with $p(B) > 0$, we have that*

$$p(B) = \max\left\{p\left(B \bigcap H_n\right) : n \in J\right\}.$$

Moreover the only limit point of the set $\{p(H_n) : n \in J\}$ is 0 provided that J is a countably infinite set. ($\mathcal{H}(p)$ is referred to as p-generating countable system.)

In proving the *Fundamental Optimal Measure Theorem* we used *Zorn's lemma* which, as we know, is equivalent to the *Axiom of Choice.* It is worth noting that in [6] the above structure theorem has been proven without *Zorn's lemma.*

By a quasi-optimal measure we mean a set function $q : \mathcal{F} \to [0, \infty)$ satisfying the axioms *P1.-P3.* with the hypothesis $q(\Omega) = 1$ in *P1.* replaced by $0 < q(\Omega) < \infty$.

We say that a quasi-optimal measure q is absolutely continuous relative to an optimal measure p (abbreviated $q << p$) if $q(B) = 0$ whenever $p(B) = 0$, $B \in \mathcal{F}$.

Remark A. Every m.f. is constant almost surely on each indecomposable atom (cf. [2], page 84, *Remark 2.1.*).

Theorem B. (cf. [1] page 139, *Theorem 3.1.*)

1. *If* (f_n) *is an increasing sequence of nonnegative m.f.'s, then* $\lim_{n\to\infty} \wr_\Omega f_n dp = \wr_\Omega \left(\lim_{n\to\infty} f_n\right) dp$.

2. *If* (g_n) *is a decreasing sequence of nonnegative m.f.'s with* $g_1 \leq b$ *for some* $b \in (0, \infty)$, *then* $\lim_{n\to\infty} \wr_\Omega g_n dp = \wr_\Omega \left(\lim_{n\to\infty} g_n\right) dp$.

Lemma C. *Let q be a quasi-optimal measure, absolutely continuous relative to an optimal measure p. Then $\mathcal{H}_*(p) = \{H \in \mathcal{H}(p) : q(H) > 0\}$ is a q-generating countable system (where $\mathcal{H}(p)$ denotes a p-generating countable system).*

Lemma D. (cf. [1] page 141, *Lemma 3.2.***)** *If (f_n) and (h_n) are sequences of nonnegative m.f.'s, then for every optimal measure p, we have that*

1. $\wr_\Omega \left(\liminf_{n\to\infty} f_n\right) dp \leq \liminf_{n\to\infty} \wr_\Omega f_n dp$;

2. $\limsup_{n\to\infty} \wr_\Omega h_n dp \leq \wr_\Omega \left(\limsup_{n\to\infty} h_n\right) dp$, *provided that (h_n) is a uniformly bounded sequence.*

NOTATIONS.

1. $\mathcal{P}$ will denote the set of all optimal measures defined on $(\Omega, \mathcal{F})$.

2. $\mathcal{P}_{<\infty}$ is the collection of all optimal measures whose generating systems are finite.

3. $\mathcal{P}_\infty$ is the set of all optimal measures whose generating systems are countably infinite.

4. For every fixed $\omega \in \Omega$, the optimal measure p_ω (defined on $(\Omega, \mathcal{F})$ by $p_\omega(B) = 1$ if $\omega \in B$, and 0 if $\omega \notin B$) will be referred to as ω-concentrated optimal measure.

5. $|E|$ stands for the cardinality of the measurable set E.

6. $\mathbb{N}$ will stand for the set of positive integers.

2. Some preliminary results

We say that a nonempty measurable set E is closely related to some sequence $(\omega_n) \subset \Omega$ if

$$\left|E \bigcap \{\omega_n : n \in \mathbb{N}\}\right| = \begin{cases} \infty, & \text{if } |E| = \infty \\ |E|, & \text{if } |E| < \infty \end{cases}$$

(that is, if E is infinite, then infinitely many members of the sequence belong to E, otherwise all of its elements are members of the sequence).

Definition 2.1. Let E be closely related to a sequence $(\omega_n) \subset \Omega$, and let $(\alpha_n) \subset [0, 1]$ be any fixed sequence tending decreasingly to 0. The optimal measure $p_E : \mathcal{F} \to [0, 1]$, defined by $p_E(B) = \max\{\alpha_n : \omega_n \in B\}$, will be called 1st-type E-dependent optimal measure.

Proposition 2.1. *Let $p \in \mathcal{P}$ and f be any m.f. Then*

$$\wr_\Omega |f|\, dp = \sup\left\{\wr_{H_n} |f|\, dp : n \in J\right\},$$

where $\mathcal{H}(p) = \{H_n : n \in J\}$ is a p-generating countable system.

Moreover if $f \in \mathcal{A}_1(p)$, then $\wr_\Omega |f|\, dp = \sup\{c_n \cdot p(H_n) : n \in J\}$, where $c_n = f(\omega)$ for almost all $\omega \in H_n$, $n \in J$.

(The proof is straightforward.) The following remark is worth being noted.

Remark 2.1. Let p, $q \in \mathcal{P}$, $\mathcal{H}(p) = \{H_n : n \in J\}$ be a p-generating countable system and f any m.f. Suppose that $q << p$ and $q(H) \le p(H)$ for every $H \in \mathcal{H}(p)$. Then $\wr_\Omega |f|\, dq \le \wr_\Omega |f|\, dp$, provided that $\wr_\Omega |f|\, dp < \infty$.

(This is immediate from *Lemma C* and *Proposition 2.1.*)

Remark 2.2. If (x_n) is a sequence of real numbers such that $\limsup_{n \to \infty} |x_n| < \infty$, then for each of its subsequences (x_{n_k}) we have that $\limsup_{k \to \infty} |x_{n_k}| < \infty$.

NOTICE. For every fixed m.f. f, the mapping $z_f : \mathcal{P} \to [0, \infty]$, defined by $z_f(p) = \wr_\Omega |f|\, dp$, is a function.

Lemma 2.2. *Let $\omega \in \Omega$ be fixed. Then for every m.f. f, we have that $z_f(p_\omega) = |f(\omega)|$.*

Proof. Let $0 \le s = \sum_{i=1}^{k} b_i \chi(B_i)$ be a measurable simple function. Then it is obvious that $z_s(p_\omega) = s(\omega)$. Let (s_n) be a sequence of nonnegative measurable simple functions tending increasingly to $|f|$. Then by *Theorem D* it ensues that

$$z_f(p_\omega) = \lim_{n \to \infty} z_{s_n}(p_\omega) = \lim_{n \to \infty} s_n(\omega) = |f(\omega)|$$

which was to be proved. **q.e.d.**

Theorem 2.3. *Let f be any m.f. The following assertions are equivalent.*

1. *f is bounded.*
2. *$\lim_{x \to \infty} \wr_{(|f| \ge x)} |f|\, dp = 0$ for all $p \in \mathcal{P}_\infty$.*
3. *There exists a constant $b > 0$ such that $\wr_\Omega |f|\, dp \ne b$ for all $p \in \mathcal{P}_\infty$.*

(The proof will be carried out in two steps. In *Proposition 2.4* we shall show the equivalence 1. $\longleftrightarrow$ 2. and then the equivalence $\rceil$1. $\longleftrightarrow\rceil$3. in *Proposition 2.5.*)

Proposition 2.4. *A m.f. f is bounded if and only if* $\lim\limits_{x\to\infty} \wr_{(|f|\geq x)} |f|\, dp = 0$ *for all* $p \in \mathcal{P}_\infty$.

Proof. Suppose that f is bounded, and write $b > 0$ for its bound. Then for every $p \in \mathcal{P}_\infty$, we have that $\wr_{(|f|\geq x)} |f|\, dp \leq b \cdot p\,(|f| \geq x) \to 0$, as $x \to \infty$.

Conversely, assume that $\lim\limits_{k\to\infty} \wr_{(|f|\geq k)} |f|\, dp = 0$ *for all* $p \in \mathcal{P}_\infty$, but for every $n \in \mathbb{N}$ we have that $(|f| \geq n-1) \neq \emptyset$. It obviously ensues that

$$(|f| \geq n-1) \setminus (|f| \geq n) = H_n \neq \emptyset$$

for infinitely many $n \in \mathbb{N}$. (Suppose without loss of generality that $H_n \neq \emptyset$, $n \in \mathbb{N}$.) Further let $(\omega_n) \subset \Omega$ be such that $\omega_n \in H_n$ for all $n \in \mathbb{N}$. Define $p \in \mathcal{P}_\infty$ by $p\,(B) = \max\left\{\frac{1}{n} : \omega_n \in B\right\}$. Clearly (H_n) is a generating system for p. Then by assumption it follows that $\lim\limits_{k\to\infty} \wr_{(|f|\geq k)} |f|\, dp = 0$. Now note that $(|f| \geq k) = \bigcup\limits_{i=k+1}^{\infty} H_i$ for all $k \in \mathbb{N}$. Hence *Proposition 2.1.* entails that $\wr_{(|f|\geq k)} |f|\, dp = \sup\limits_{i\geq k+1} \wr_{H_i} |f|\, dp$. It is not difficult to check that $\wr_{H_i} |f|\, dp \geq 1 - \frac{1}{i}$, $i \geq k+1$. Consequently it results that $\wr_{(|f|\geq k)} |f|\, dp \geq 1 - \frac{1}{k+1}$ $(k \in \mathbb{N})$, leading to $0 = \lim\limits_{k\to\infty} \wr_{(|f|\geq k)} |f|\, dp \geq 1$, which is absurd. This contradiction concludes on the validity of the sufficiency, ending the proof. **q.e.d.**

Proposition 2.5. *Let f be a finite m.f. Then f is unbounded if and only if for every constant $c > 0$, there exists some $p_c \in \mathcal{P}_\infty$ such that*

$$z_f\,(p_c) = c. \tag{1.1}$$

Proof. *Necessity.* Assume that f is unbounded. For every $n \in \mathbb{N}$, write $E_n = (c\cdot(n-1) \leq |f| < c\cdot n)$ where $c > 0$ is an arbitrarily fixed constant. Clearly the members of the sequence (E_n) are pairwise disjoint and $\Omega = \bigcup\limits_{n=1}^{\infty} E_n$. Fix a sequence $(\omega_n) \subset \Omega$ in the following way: $\omega_n \in E_n$, $n \in \mathbb{N}$. Define $p_c \in \mathcal{P}_\infty$ by $p_c\,(B) = \max\left\{\frac{1}{n} : \omega_n \in B\right\}$. It is obvious that sequence (E_n) is a p_c-generating system such that $z_f\,(p_c) = \sup\limits_{n\geq 1} \wr_{E_n} |f|\, dp_c$, because of *Proposition 2.1.* But as $\left(1 - \frac{1}{n}\right) c \leq \wr_{E_n} |f|\, dp_c < c$ (for all $n \in \mathbb{N}$), it ensues that $c = \sup\limits_{n\geq 1} \wr_{E_n} |f|\, dp_c = z_f\,(p_c)\,.$

Sufficiency. Suppose that for every constant $c > 0$, identity (1.1) holds with a suitable $p \in \mathcal{P}_\infty$. Assume that f is bounded (and denote by b its bound). Now let $c > b$ be any fixed constant with a corresponding $p_c \in \mathcal{P}_\infty$ satisfy (1.1). Then we trivially obtain that $z_f\,(p_c) \leq b$. Hence we must have that $c \leq b$, which is in contradiction with the choice of c. This absurdity allows us to conclude on the validity of the proposition. **q.e.d.**

Lemma 2.6. *Let $p \in \mathcal{P}_\infty$ and (B_n) be a sequence of measurable sets tending increasingly to a measurable set $B \neq \emptyset$. Then there exists some $n_0 \in \mathbb{N}$ such that $p(B) = p(B_n)$ whenever $n \geq n_0$.*

(See the proof of *Lemma 0.1.*, [1] page 134.)

We shall next give a set of measurable functions including uniformly bounded ones.

Definition 2.2. We say that a sequence of measurable functions (f_n) is uniformly bounded starting from an index if there can be found a real number $b > 0$ and some positive integer n_0 such that $(f_n > b) = \emptyset$ for all integers $n > n_0$. (We shall simply say that (f_n) is *i-uniformly bounded.*)

The following two results are just the extensions of *Theorem B/2* and *Lemma D/2.* We shall omit their proofs as they can be similarly carried out.

Lemma 2.7. *Let (g_n) be a decreasing sequence of nonnegative m.f.'s and $\lim\limits_{n\to\infty} g_n = g$ such that $(g_m \leq b) = \Omega$ for some $m \geq 1$ and some constant $b > 0$. Then $\lim\limits_{n\to\infty} \wr_\Omega g_n dp = \wr_\Omega g dp$ for all $p \in \mathcal{P}$.*

Lemma 2.8. *Let (f_n) be an i-uniformly bounded sequence of nonnegative m.f.'s. Then $\limsup\limits_{n\to\infty} \wr_\Omega f_n dp \leq \wr_\Omega \left(\limsup\limits_{n\to\infty} f_n\right) dp$ for every $p \in \mathcal{P}$.*

Theorem 2.9. *Let (f_n) be an arbitrary sequence of m.f.'s. Then*

1. (f_n) is i-uniformly bounded,

if and only if the following two assertions hold simultaneously:

2. $z_f(p) \leq c$ for some constant $c > 0$ and all $p \in \mathcal{P}_\infty$;

3. $\limsup\limits_{n\to\infty} z_n(p) \leq z_f(p)$, for all $p \in \mathcal{P}_\infty$ (where $f = \limsup\limits_{n\to\infty} |f_n|$ and $z_n(p) = \wr_\Omega |f_n| dp$ with $n \in \mathbb{N}$, $p \in \mathcal{P}_\infty$).

Proof. *Necessity.* We just note that the implication 1. $\to$ 2. is obvious and on the other hand the implication 1. $\to$ 3. is no more than *Lemma 2.8.*

Sufficiency. Assume that 2. and 3. hold. Let us suppose further that 1. is false, i.e. for every real number $b > 0$ and any positive integer n_0 there is some integer $m > n_0$ such that $(|f_m| > b) \neq \emptyset$. Then we can choose by recurrence a sequence (n_k) of positive integers as follows. Write $n_1 = 1$ and $n_2 = \min\{m > n_1 : (|f_m| > n_1) \neq \emptyset\}$. If n_k has been defined, then write $n_{k+1} = \min\{m > n_k : (|f_m| > k \cdot n_k) \neq \emptyset\}$. Clearly the sequence (n_k) tends increasingly to infinity and for all positive integers $k \in \mathbb{N}$, $\left(\left|f_{n_{k+1}}\right| > k \cdot n_k\right) \neq \emptyset$. Now set $E = \bigcup\limits_{k=1}^{\infty} B_{n_k}$, where $B_{n_k} = \left(\left|f_{n_{k+1}}\right| > k \cdot n_k\right)$, $k \in \mathbb{N}$. Write $H_1 = B_{n_1}$, and $H_k = \left(\bigcup\limits_{j=1}^{k} B_{n_j}\right) \setminus \left(\bigcup\limits_{j=1}^{k-1} B_{n_j}\right)$, $k > 2$. It is obvious that (H_k)

is a sequence of pairwise disjoint measurable sets with $E = \bigcup_{k=1}^{\infty} H_k$. Let $p \in \mathcal{P}_\infty$ be a 1st-type E-dependent optimal measure defined by $p(B) = \max\left\{\frac{1}{k} : \omega_k \in B\right\}$, where $(\omega_k) \subset \Omega$ is a fixed sequence so that $\omega_k \in H_k$ $(k \in \mathbb{N})$. It is clear that $\mathcal{H}(p) = \{H_k : k \in \mathbb{N}\}$ is a p-generating system. Then via 2. and 3. we have that $c \geq \wr_\Omega \left(\limsup_{n\to\infty} |f_n|\right) dp \geq \limsup_{n\to\infty} \wr_\Omega |f_n|\, dp$ and hence $b > \limsup_{k\to\infty} \wr_\Omega \left|f_{n_{k+1}}\right| dp$ for some $b > 0$ (this is true because of *Remark 2.2*). Consequently, as $p(H_k) = \frac{1}{k}$ for every $k \in \mathbb{N}$, we must have

$$\begin{aligned} b > \limsup_{k\to\infty} \wr_\Omega \left|f_{n_{k+1}}\right| dp &= \limsup_{k\to\infty} \wr_E \left|f_{n_{k+1}}\right| dp \\ &\geq \limsup_{k\to\infty} \wr_{H_k} \left|f_{n_{k+1}}\right| dp \\ &\geq \limsup_{k\to\infty} k \cdot n_k \cdot p(H_k) = \infty, \end{aligned}$$

which is absurd. This contradiction justifies the validity of the theorem. **q.e.d.**

3. The case of some well-known types of convergence

Definition 3.1. Let X be an arbitrary nonempty set. We say that a sequence of real-valued functions (h_n) converges to a real-valued function h:

(i) discretely if for every $x \in X$ there exists a positive integer $n_0(x)$ such that $h_n(x) = h(x)$, whenever $n > n_0(x)$;

(ii) equally if there is a sequence (b_n) of positive numbers tending to 0 and for every $x \in X$ there can be found an $n_0(x)$ such that $|h_n(x) - h(x)| < b_n$ whenever $n > n_0(x)$.

(For more about these notions, cf. [3 - 5].)

Theorem 3.0. *Let f and f_n $(n \in \mathbb{N})$ be any m.f.'s. Then (f_n) tends to f uniformly if and only if (z_n) tends to 0 uniformly on $\mathcal{P}_\infty$, where $z_n(p) = \wr_\Omega |f_n - f|\, dp$ with $n \in \mathbb{N}$, $p \in \mathcal{P}_\infty$.*

Proof. *Sufficiency.* Suppose that (z_n) tends to 0 uniformly. To prove the sufficiency it is enough to show that for every number $b > 0$, there can be found some $n_0(b) \in \mathbb{N}$ such that $(|f - f_n| \geq b) = \emptyset$ whenever $n \geq n_0(b) + 1$. In fact, let us assume that the contrary holds. Then for some $b_0 > 0$ and all $n_0 \in \mathbb{N}$, there is an integer $m > n_0$ such that $(|f - f_m| \geq b_0) \neq \emptyset$. Define

$$n_1 = \min\{m > n_1 : (|f - f_m| \geq b_0) \neq \emptyset\}$$

when $n_0 = 1$. If n_k has been selected, define

$$n_{k+1} = \min\{m > n_k : (|f - f_m| \geq b_0) \neq \emptyset\}$$

when $n_0 = n_k$. It is clear that sequence (n_k) tends increasingly to infinity alongside with k, so that $(|f - f_{n_k}| \geq b_0) \neq \emptyset$, $k \in \mathbb{N}$. Then by assumption some $n_m \in \{n_k : k \in \mathbb{N}\}$ exists such that $z_{n_k}(p) < b_0$, for all $k \geq m$ and $p \in \mathcal{P}_\infty$. Now let $E_m = \bigcup\limits_{k=m}^{\infty} B_{n_k}$, (where $B_{n_k} = (|f - f_{n_k}| \geq b_0)$, $k \in \mathbb{N}$). Write $H_{n_m} = B_{n_m}$ and for $k \geq m+1$, set $H_{n_k} = \left(\bigcup\limits_{j=m}^{k} B_{n_j}\right) \setminus \left(\bigcup\limits_{j=m}^{k-1} B_{n_j}\right)$. Clearly $\mathcal{H} = \{H_{n_k} : k \geq m\}$ is a sequence of pairwise disjoint measurable sets with $E_m = \bigcup\limits_{k=m}^{\infty} H_{n_k}$. Fix a sequence $(\omega_k) \in \Omega$ so that $\omega_k \in H_{n_k}$ whenever $k \geq m$. Next, let $p_0 \in \mathcal{P}_\infty$ be a 1st-type E_m-dependent optimal measure defined by $p_0(B) = n_m \cdot \max\left\{\frac{1}{n_k} : \omega_k \in B\right\}$. It is obvious that $\mathcal{H}$ is a p_0-generating system. Hence we have on the one hand that $z_{n_m}(p_0) < b_0$. Nevertheless on the other hand we also obtain that $z_{n_m}(p_0) \geq \int_{H_{n_m}} |f_{n_m} - f|\, dp_0 \geq b_0$, since $p_0(H_{n_m}) = 1$. As these last two inequalities contradict each other, the sufficiency is thus proved.

Necessity. Assume that $f_n \to f$ uniformly, as $n \to \infty$. Then for every $b \in (0, \infty)$, there is some $n_0(b) \in \mathbb{N}$ such that $\left(|f_n - f| < \frac{b}{2}\right) = \Omega$ whenever $n > n_0(b)$. Consequently, for every $p \in \mathcal{P}_\infty$, it ensues that $z_n(p) \leq \frac{b}{2} < b$, $n > n_0(b)$. This completes the proof of the theorem. **q.e.d.**

Lemma 3.1. *Let f and f_n ($n \in \mathbb{N}$) be any m.f.'s. If (f_n) tends to f pointwise (equally or discretely), then $\limsup\limits_{n\to\infty} B_n = \emptyset$, where $B_n = (|f_n - f| = \infty)$, $n \in \mathbb{N}$.*

Proof. It is enough to prove the lemma for the pointwise convergence, since proving the remaining cases is similarly done. Assume that $\limsup\limits_{n\to\infty} B_n \neq \emptyset$. Let us pick an arbitrary $\omega \in \limsup\limits_{n\to\infty} B_n$. Then it is clair that $\limsup\limits_{n\to\infty} |f_n(\omega) - f(\omega)| = \infty$ and hence $\bigwedge\limits_{n=k}^{\infty} \bigvee\limits_{j=n}^{\infty} |f_j(\omega) - f(\omega)| = \infty$ for every $k \in \mathbb{N}$. But since (f_n) tends to f pointwise we must have that for every constant $b > 0$ there is a positive integer $m_0 = m_0(b, \omega)$ such that $|f_n(\omega) - f(\omega)| < b$ whenever $n > m_0$. Hence $b \geq \bigwedge\limits_{n=m_0}^{\infty} \bigvee\limits_{j=n}^{\infty} |f_j(\omega) - f(\omega)| = \infty$, which is absurd, completing the proof. **q.e.d.**

Theorem 3.2. *Let (f_n) be any sequence of m.f.'s. Then (f_n) tends to a m.f. f pointwise if and only if (z_n) tends to 0 pointwise on $\mathcal{P}_{<\infty}$, where for every $n \in \mathbb{N}$, z_n is defined on $\mathcal{P}_{<\infty}$ by $z_n(p) = \int_\Omega |f_n - f|\, dp$.*

Proof. *Sufficiency.* Assume that for all $b > 0$ and $p \in \mathcal{P}_{<\infty}$ there is a positive integer $n_0 = n_0(b, p)$ such that $z_n(p) < b$ whenever $n > n_0$. Then since for every fixed $\omega \in \Omega$ the ω-concentrated measure p_ω depends solely upon $\omega \in \Omega$, index $n_0(b, p_\omega)$ also depends on ω. Hence via *Lemma 2.2* we have for all $n \geq n_0(b, \omega) = n_0(b, p_\omega)$ that $|f_n(\omega) - f(\omega)| = z_n(p_\omega) < b$.

Necessity. Suppose that for all $a > 0$ and $\omega \in \Omega$, there can be found some positive integer $m_0 = m_0(a, \omega)$ such that $|f_n(\omega) - f(\omega)| < a$, whenever $n \geq m_0$. Assume further that there is some $b > 0$ and some $p \in \mathcal{P}_{<\infty}$ such that for every $n \in \mathbb{N}$, there exists some $m \geq n$ with the property that $z_n(p) \geq b$. Let $H_1, \ldots, H_k$ be a p-generating system. Via *Lemma 3.1*, there is some $n_0 \in \mathbb{N}$, big enough so that $f_n - f$ is finite on Ω whenever $n \geq n_0$. Then for every $n \geq n_0$, a measurable set $A_n^{(i)}$ exists with $A_n^{(i)} \subset H_i$ and $p\left(A_n^{(i)}\right) = 0$ such that $f_n - f$ is constant on $H_i \backslash A_n^{(i)}$, $i = 1, \ldots, k$ (because of *Remark A*). Clearly $p\left(\bigcup\limits_{j=n_0}^{\infty} A_j^{(i)}\right) = 0$, so that the identity $p\left(H_i \backslash \bigcup\limits_{j=n_0}^{\infty} A_j^{(i)}\right) = p(H_i)$ holds. Hence $f_n - f$ is constant on $H_i \backslash \bigcup\limits_{j=n_0}^{\infty} A_j^{(i)}$ whenever $i \in \{1, \ldots, k\}$ and $n \geq n_0$. Fix $\omega_i \in H_i \backslash \bigcup\limits_{j=n_0}^{\infty} A_j^{(i)}$, $i \in \{1, \ldots, k\}$. Then by assumption there must be some positive integer $k_0^{(i)} = k_0(b, \omega_i)$ such that $|f_n(\omega_i) - f(\omega_i)| < b$, $n > k_0^{(i)}$. Thus for all $n \geq k_0$ (where $k_0 = \bigvee\limits_{i=1}^{k} k_0^{(i)}$), we have that $\bigvee\limits_{i=1}^{k} |f_n(\omega_i) - f(\omega_i)| < b$. Now write $k^* = \max(k_0, n_0)$. Then some integer $m > k^*$ exists such that $z_m(p) \geq b$. Therefore (via *Proposition 2.1* and *Remark A*) we obtain that

$$b \leq z_m(p) = \bigvee_{i=1}^{k} c_i \cdot p(H_i) \leq \bigvee_{i=1}^{k} c_i = \bigvee_{i=1}^{k} |f_m(\omega_i) - f(\omega_i)| < b$$

where for $i \in \{1, \ldots, k\}$, $c_i = |f_m(\omega) - f(\omega)|$ if $\omega \in H_i \backslash \bigcup\limits_{j=n_0}^{\infty} A_j^{(i)}$. However, this is absurd, a contradiction which ends the proof of the theorem. **q.e.d.**

Theorem 3.3. *A sequence of m.f.'s* (f_n) *converges to some m.f.* f *equally if and only if* (z_n) *converges to* 0 *equally on* $\mathcal{P}_{<\infty}$*, where for every* $n \in \mathbb{N}$, z_n *is defined on* $\mathcal{P}_{<\infty}$ *by* $z_n(p) = \wr_\Omega |f_n - f|\, dp$.

Proof. *Necessity.* Suppose that there exists a sequence $(b_n) \subset (0, \infty)$ tending to 0 and for every $\omega \in \Omega$ there can be found a positive integer $n_0(\omega)$ such that $|f_n(\omega) - f(\omega)| < b_n$ for all $n \geq n_0(\omega)$. It is enough to show that the equal convergence of (z_n) holds true for this sequence (b_n). In fact, assume that for this sequence (b_n), there is some $p \in \mathcal{P}_{<\infty}$ such that for all $j \in \mathbb{N}$ an integer $m = m(p) > j$ can be found with the property that $z_m(p) \geq b_m$. Let $H_1, \ldots, H_k$ be a p-generating system. Via *Lemma 3.1*, there is some $n_0 \in \mathbb{N}$, big enough so that $f_n - f$ is finite on Ω whenever $n \geq n_0$. Then for every $n \geq n_0$, a measurable set $A_n^{(i)}$ exists with $A_n^{(i)} \subset H_i$ and $p\left(A_n^{(i)}\right) = 0$ such that $f_n - f$ is constant on $H_i \backslash A_n^{(i)}$, $i = 1, \ldots, k$. But as $p\left(\bigcup\limits_{j=n_0}^{\infty} A_j^{(i)}\right) = 0$, we can easily observe that $p\left(H_i \backslash \bigcup\limits_{j=n_0}^{\infty} A_j^{(i)}\right) = p(H_i)$, $i \in \{1, \ldots, k\}$. Hence $f_n - f$ is constant on $H_i \backslash \bigcup\limits_{j=n_0}^{\infty} A_j^{(i)}$ for all $i \in \{1, \ldots, k\}$ and

$n \geq n_0$. Fix $\omega_i \in H_i \setminus \bigcup_{j=n_0}^{\infty} A_j^{(i)}$, $i \in \{1, \ldots, k\}$. Then by assumption there must be some positive integer $k_0^{(i)} = k_0(\omega_i)$ such that $|f_n(\omega_i) - f(\omega_i)| < b_n$, $n > k_0^{(i)}$. Thus for all $n \geq k_0$ (where $k_0 = \bigvee_{i=1}^{k} k_0^{(i)}$), we have that $\bigvee_{i=1}^{k} |f_n(\omega_i) - f(\omega_i)| < b_n$. Consequently we have on the one hand that $z_m(p) \geq b_m$. But on the other hand, *Proposition 2.1* yields that

$$z_m(p) = \bigvee_{i=1}^{k} c_i \cdot p(H_i) \leq \bigvee_{i=1}^{k} c_i = \bigvee_{i=1}^{k} |f_m(\omega_i) - f(\omega_i)| < b_m$$

(where for $i \in \{1, \ldots, k\}$, $c_i = |f_m(\omega) - f(\omega)|$ if $\omega \in H_i \setminus \bigcup_{j=n_0}^{\infty} A_j^{(i)}$), meaning that $b_m < b_m$, which is, however, absurd. This contradiction concludes the proof of the necessity.

Sufficiency. Assume that there is a sequence (b_n) of positive numbers tending to 0 and for every $p \in \mathcal{P}_{<\infty}$ there exists a positive integer $n_0(p)$ such that $z_n(p) < b_n$ whenever $n > n_0(p)$. Then for each fixed $\omega \in \Omega$, *Lemma 2.2.* entails that $|f_n(\omega) - f(\omega)| = z_n(p_\omega) < b_n$ whenever $n > n_0(p_\omega) = n_0(\omega)$. The sufficiency is thus proved, which completes the proof of the theorem. **q.e.d.**

Theorem 3.4. *A sequence of m.f.'s* (f_n) *converges to some m.f.* f *discretely if and only if* (z_n) *converges to* 0 *discretely on* $\mathcal{P}_{<\infty}$, *where for every* $n \in \mathbb{N}$, z_n *is defined on* $\mathcal{P}_{<\infty}$ *by* $z_n(p) = \wr_\Omega |f_n - f|\, dp$.

(The proof is omitted as it can be carried out "mutatis mutandis" as in *Theorems 3.3* and *3.4*)

4. Quasi-uniform convergence

Definition 4.1. A sequence of real-valued functions (g_n), defined on a nonempty set X, is said to converge quasi-uniformly to a real-valued function g if for every given number $\epsilon \in (0,\ 1)$ there exists some nonempty set $B_\epsilon \subset X$ and some positive integer $n_0 = n_0(\epsilon)$ such that $|g_n(x) - g(x)| < \epsilon$ whenever $n > n_0$ and $x \in B_\epsilon$.

Example 1. Every uniformly convergent sequence of m.f.'s also converges quasi-uniformly.

Example 2. Let us endow the real line $\mathbb{R}$ with the Borel σ-algebra $\mathcal{B}$ and let (f_n) be a sequence of Borel measurable functions defined by

$$f_n(x) = \frac{x}{n},\ n \in \mathbb{N},\ x \in \mathbb{R}.$$

It is not difficult to see that (f_n) converges to zero pointwise but not uniformly. We show that (f_n) converges to zero quasi-uniformly. In fact, pick an arbitrary number

$\varepsilon \in (0,\, 1)$ and fix any number $x \in \mathbb{R}$. Clearly with the choice $n_0 = n_0\,(\varepsilon,\, x) = \left[\frac{|x|}{\varepsilon}\right]+1$, we have that $\frac{|x|}{n} < \varepsilon$ for all $n \geq n_0$. Now define the set $B_\varepsilon = \{t \in \mathbb{R} : \left[\frac{|t|}{\varepsilon}\right] < \left[\frac{|x|}{\varepsilon}\right]\}$. Obviously we have that $\frac{|t|}{n} < \varepsilon$ for all $t \in B_\varepsilon$. Therefore (f_n) converges to zero quasi-uniformly.

Lemma 4.1. *Let f be any m.f., $p \in \mathcal{P}_\infty$, H some indecomposable p-atom with $p\,(H) = 1$ and $\epsilon \in (0,\, 1)$ any number. Then $p\,(H \bigcap (|f| \geq \epsilon)) = 0$ if and only if $\wr_H\, |f|\, dp < \epsilon$.*

Proof. As the necessity is obvious we shall just show the sufficiency. Suppose that $\wr_H\, |f|\, dp < \epsilon$ but $p\,(H \bigcap (|f| \geq \epsilon)) > 0$. Then

$$\begin{aligned}
\epsilon &> \wr_H\, |f|\, dp \\
&= \left(\wr_{H\cap(|f|<\epsilon)}\, |f|\, dp\right) \bigvee \left(\wr_{H\cap(|f|\geq\epsilon)}\, |f|\, dp\right) \\
&\geq \wr_{H\cap(|f|\geq\epsilon)}\, |f|\, dp \\
&\geq \epsilon \cdot p\,(H \textstyle\bigcap (|f| \geq \epsilon))\,.
\end{aligned}$$

But since H is an indecomposable p-atom and $p\,(H \bigcap (|f| \geq \epsilon)) > 0$, it ensues that $p\,(H \bigcap (|f| \geq \epsilon)) = p\,(H) = 1$. Consequently we must have that $\epsilon > \epsilon$, which is absurd, indeed. This contradiction concludes the proof. **q.e.d.**

Theorem 4.2. *Let f and f_n $(n \in \mathbb{N})$ be any m.f.'s. Then (f_n) tends to f quasi-uniformly if and only if (z_n) tends to 0 quasi-uniformly on $\mathcal{P}_\infty$, where $z_n\,(p) = \wr_\Omega\, |f_n - f|\, dp$ with $n \in \mathbb{N}$, $p \in \mathcal{P}_\infty$.*

Proof. *Sufficiency.* Assume the quasi-uniform convergence of (f_n), i.e. for every $\epsilon \in (0,\, 1)$ we can find some nonempty measurable set $B_{\frac{\epsilon}{2}}$ and some positive integer $n_0 = n_0\,(\epsilon)$ such that

$$B_{\frac{\epsilon}{2}} \subset \left(|f_n - f| < \frac{\epsilon}{2}\right), n > n_0.$$

Write $\mathcal{P}_\infty\,(\epsilon) = \left\{p \in \mathcal{P}_\infty : p\left(\Omega \backslash B_{\frac{\epsilon}{2}}\right) = 0\right\}$. We note that $\mathcal{P}_\infty\,(\epsilon) \neq \emptyset$, since each 1st-type $B_{\frac{\epsilon}{2}}$-dependent optimal measure belongs to $\mathcal{P}_\infty\,(\epsilon)$. Clearly for all $n > n_0$ and $p \in \mathcal{P}_\infty\,(\epsilon)$

$$\wr_\Omega\, |f_n - f|\, dp = \wr_{B_{\frac{\epsilon}{2}}}\, |f_n - f|\, dp \leq \frac{\epsilon}{2} < \epsilon.$$

Necessity. Assume the quasi-uniform convergence of (z_n), but (f_n) fails to converge quasi-uniformly to f. Then the latter assumption means that for some $\epsilon_* \in (0,\, 1)$, all nonempty measurable sets B and every positive integer m_0 there exists an $M \geq m_0$ such that $(|f_M - f| \geq \epsilon_*) \bigcap B \neq \emptyset$. Nevertheless, because of the former assumption there can be found some $\mathcal{P}_\infty\,(\epsilon_*) \subset \mathcal{P}_\infty$ and some integer $m_* = m_*\,(\epsilon_*) \geq 1$ such that $z_m(p) < \epsilon_*$ for all $m \geq m_*$ and $p \in \mathcal{P}_\infty(\epsilon_*)$. Let us fix some $p \in \mathcal{P}_\infty(\epsilon_*)$ with (H_k) its generating system so that $z_m(p) < \epsilon_*$ whenever $m \geq m_*$. Then *Proposition 2.1* entails that $\wr_{H_k} |f_m - f| \, dp < \epsilon_*$, for all $k \geq 1$ and $m \geq m_*$. As the *Fundamental Theorem* guarantees that $\lim\limits_{k\to\infty} p(H_k) = 0$, there must exist some integer j such that $p(H_j) = 1$. Next, noting that the conditions of *Lemma 4.1* are met, it results that $p((|f_m - f| \geq \epsilon_*) \bigcap H_j) = 0$, $m \geq m_*$. Write

$$S = H_j \backslash \left(\bigcup_{m=m_*}^{\infty} \left((|f_m - f| \geq \epsilon_*) \bigcap H_j \right) \right) = H_j \backslash \left(\bigcup_{m=m_*}^{\infty} (|f_m - f| \geq \epsilon_*) \right).$$

It is not difficult to see that $(|f_m - f| < \epsilon_*) \bigcap S = S$, $m \geq m_*$. Consequently, since we have rather assumed the negation of the conclusion, some integer $i > m_*$ must exist so that $(|f_i - f| \geq \epsilon_*) \bigcap S \neq \emptyset$. This, however, is in contradiction with $(|f_i - f| < \epsilon_*) \bigcap S = S$, which ends the proof of the theorem. **q.e.d.**

5. Concluding remarks

I would like to simply note that when preparing those two works (see [1-2]) I was not aware of the existence of the so-called 'maxitive measures' proposed by N. Shilkret. in [7]. Hereafter one can find a briefing of his work.

By *a-maxitive measures*, i.e. set functions $m : \mathcal{F} \to [0, \infty)$, satisfying the conditions $m(\emptyset) = 0$ and $m\left(\bigcup\limits_{i\in I} E_i\right) = \sup\limits_{i\in I} m(E_i)$ for every collection $\{E_i\}_{i\in I} \subset \mathcal{F}$, where $\mathcal{F}$ is a ring of subsets of an arbitrary nonempty set Ω; it is called a maxitive measure if I is finite or countably infinite. Shilkret realized that maxitive measures are not in general continuous from above and he proved that: *A maxitive measure m is continuous from above if and only if the following assertion is false "There exist some $k \in \mathbb{N}$ and some sequence of measurable sets $\{E_i\} \subset \mathcal{F}$ such that $\frac{1}{k} < m(E_i) < k$, $i \in \mathbb{N}$."*

REFERENCES

[1] Agbeko, N. K.: *On optimal averages*, Acta Math. Hung., **63,** (1994), 1–15.

[2] Agbeko, N. K.: *On the structure of optimal measures and some of its applications*, Publ. Math. Debrecen, **46,** (1995), 79–87.

[3] Császár, Á. and Laczkovich, M.: *Discrete and equal convergence*, Studia Sci. Math. Hungar., **10,** (1975), 463–472.

[4] Császár, Á. and Laczkovich, M.: *Some remarks on discrete Baire classes*, Acta Math. Acad. Sci. Hungar., **33,** (1979), 51–70.

[5] Császár, Á. and Laczkovich, M.: *Discrete and equal Baire classes*, Acta Math. Hung., **55,** (1990), 165–178.

[6] Fazekas, I.: *A note on 'optimal measures'*, Publ. Math. Debrecen, **51**(3-4), (1997), 273–277.

[7] Shilkret, N.: *Maxitive measure and integration*, Indagationes Math., **33,** (1971), 109–116.

Doi: 10.18514/MMN.2000.23

Mathematical Notes, Miskolc, Vol. 1., No. 2., (2000), pp. 99–107

PERIODIC AND CHAOTIC SOLUTIONS IN INFINITE DIMENSIONAL SYSTEMS

Michal Fečkan
Department of Mathematical Analysis, Comenius University
Mlynská dolina, 842 48 Bratislava, Slovakia
Michal.Feckan@fmph.uniba.sk

[Received June 20, 2000]

Abstract. Periodic, subharmonic and chaotic weak solutions are investigated for certain damped and undamped forced nonlinear hyperbolic partial differential equations of a beam.

Mathematical Subject Classification: 35B10, 35B20
Keywords: Periodic and chaotic solutions, weak solutions, infinite dimensional systems

1. Introduction

This note is a survey of our recent results [3, 4, 7]. We here focus on concrete problems and we refer the reader for proofs of results and for an abstract setting of problems to the above mentioned papers.

P. Holmes and J. Marsden in [10] studied a forced partial differential equation of a beam of the form

$$\begin{gathered} w_{tt} + w_{xxxx} + \Gamma w_{xx} - \kappa \left(\int_0^1 w_x^2(s,t)\,ds \right) w_{xx} = \epsilon\,(q\cos\omega t - \delta w_t) \\ w(0,\cdot) = w(1,\cdot) = w_{xx}(0,\cdot) = w_{xx}(1,\cdot) = 0\,, \end{gathered} \tag{1.1}$$

where $\Gamma =$ external load, $\kappa =$ stiffness due to "membrane" effect, $\delta =$ damping, and ϵ is small. They shown the existence of a Smale horseshoe. Consequently, (1.1) possesses an infinite number of subharmonic weak solutions. The new way in this direction is established in a series of papers by D.W. McLaughlin, J. Shatah, E.A. Overman II, S. Wiggins, C. Xiong and Y. Li [14–16]. They studied perturbed nonlinear Schrödinger equations and perturbed sine–Gordon equations.

We study in [3] a forced sine–beam partial differential equation given by

$$\begin{gathered} u_{tt} + \alpha u_{xxxx} + \sin u + \tau(x) = \epsilon \sin t \\ u(x+1,t) = u(x,t)\,, \end{gathered} \tag{1.2}$$

where $\alpha > 0$ is a constant, τ is continuous, 1–periodic satisfying $\int_0^1 \tau(x)\,dx = 0$ and $\epsilon \in \mathbf{R}$ is a small parameter.

A modified version of (1.1) is investigated in [4] of the form

$$
\begin{aligned}
u_{tt} + u_{xxxx} + \Gamma u_{xx} + p\left(\int_0^1 u^2(s,t)\,ds, \int_0^1 u_x^2(s,t)\,ds\right) D_{xx}^{\xi} u = \epsilon q(x) \cos \tfrac{2\pi t}{T} \\
u(0,\cdot) = u(1,\cdot) = u_{xx}(0,\cdot) = u_{xx}(1,\cdot) = 0\,,
\end{aligned}
\tag{1.3}
$$

where $D_{xx}u = -u_{xx}$, D_{xx}^{ξ} is the ξ–power of D_{xx} in $L^2(0,1)$, $0 \le \xi \le 1$, $\Gamma \in \mathbf{R}$, $p \in C^2(\mathbf{R}\times\mathbf{R},\mathbf{R})$, $p(0,0)=0$, $q \in H^2(0,1)\cap H_0^1(0,1)$, $T>0$ and $\epsilon \in \mathbf{R}$ is a small parameter.

We assume that α is sufficiently large and also consider a limit ordinary differential equation of (1.2) as $\alpha \to \infty$ of the form

$$
\ddot{u} + \sin u = \epsilon \sin t\,. \tag{1.4}
$$

It is well–known [8, 18] that for $\epsilon \neq 0$ sufficiently small, (1.4) exhibits a chaotic behaviour. In particular, it has an infinite number of subharmonic librational solutions with periods tending to infinity.

Similarly, if we put

$$
u(x,t) = z(t)\sqrt{2}\sin \pi x
$$

in (1.3) with $\epsilon = 0$ [10, 11] we arrive at an ordinary differential equation

$$
\ddot{z} + \pi^2(\pi^2 - \Gamma)z + p(z^2, \pi^2 z^2)\pi^{2\xi} z = 0\,. \tag{1.5}
$$

By assuming

$$
\Gamma > \pi^2, \quad p(z, \pi^2 z) = \kappa z \quad \forall z \ge 0 \tag{1.6}
$$

for a constant $\kappa > 0$, (1.5) becomes the Duffing equation [8]. It is well–known that the Duffing equation possesses a one–parametric bounded family of periodic solutions with periods tending to infinity.

We note that (1.2) and (1.3) are undamped and (1.1) is damped for $\epsilon \neq 0$. For this reason, we show in [3, 4] only the existence of finitely, but arbitrarily many subharmonic weak solutions of (1.2) or (1.3) persisting from (1.4) or (1.5) when $\alpha \to \infty$ or ϵ is sufficiently small and $\int\limits_0^1 q(x)\sin \pi x\,dx \neq 0$, respectively.

A damped case of (1.1) is studied in [7] of the following form: The boundary value problem for planar deflections of an elastic beam with a compressive axial load Γ and pinned ends is

$$
\begin{aligned}
\ddot{u} = -u'''' - \Gamma u'' + \left[\int_0^{\pi} \left(u'(s)\right)^2 \mathrm{d}s\right] u'' - 2\mu_2 \dot{u} + \mu_1 \cos \omega_0 t, \\
u(0,t) = u(\pi,t) = u''(0,t) = u''(\pi,t) = 0\,,
\end{aligned}
\tag{1.7}
$$

where $u(x,t)$ is the transverse deflection at a distance x from one end at time t. In (1.7), a superior dot denotes differentiation with respect to t and prime differentiation with respect to x. We consider the μ_i terms as perturbations.

Our first step is to consider the linearized, unperturbed problem and compute the eigenvalues at the origin which are $\lambda_n = n^2(n^2 - \Gamma)$ and corresponding eigenfunctions $\varphi_n(x) = \sin nx$ for $n = 1, 2, \cdots$. For small Γ the origin is a center with the first bifurcation occurring at $\Gamma = 1$, the first Euler buckling load. The corresponding eigenfunction, $\varphi_1(x) = \sin x$, is referred to as the first buckled mode. When Γ is sufficiently large, (1.7) can exhibit chaotic behavior. The first work on this was done by P. Holmes [9] and extended by P. Holmes and J. Marsden [10, 11]. Some more recent work on the full equation is by H.M. Rodrigues and M. Silveira [19], S.W. Shaw [21] and by M. Berti and C. Carminati [2].

In (1.7) substitute $u(x,t) = \sum_{k=1}^{\infty} u_k(t) \sin kx$, multiply by $\sin nx$ and integrate from 0 to π. This yields the infinite set of ordinary differential equations

$$\ddot{u}_n = n^2(\Gamma - n^2)u_n - \tfrac{\pi}{2} n^2 \left[\sum_{k=1}^{\infty} k^2 u_k^2 \right] u_n - 2\mu_2 \dot{u}_n + 2\mu_1 \cos \omega_0 t$$

$$n = 1, 2, \cdots .$$

We see that the linear parts of these equations are uncoupled and the equations divide into two types. The system of equations where $1 \leq n^2 < \Gamma$ has a hyperbolic equilibrium at the origin whereas, for $n^2 \geq \Gamma$ this equilibrium is a center. Infinite sets of ordinary differential equations are also investigated in [20].

For simplicity let us assume $1 < \Gamma < 4$. Then only the equation with $n = 1$ is hyperbolic while the remaining equations have a center. To emphasize this let us define $p = u_1$ and $q_n = u_{n+1}$, $n = 1, 2, \cdots$. The preceding equations now take the form

$$\begin{aligned} \ddot{p} &= a^2 p - \frac{\pi}{2} \left[p^2 + \sum_{k=1}^{\infty} (k+1)^2 q_k^2 \right] p - 2\mu_2 \dot{p} + \tfrac{4}{\pi} \mu_1 \cos \omega_0 t \,, \\ \ddot{q}_n &= -\omega_n^2 q_n - \tfrac{\pi}{2} (n+1)^2 \left[p^2 + \sum_{k=1}^{\infty} (k+1)^2 q_k^2 \right] q_n \\ &\quad -2\mu_2 \dot{q}_n + 2\mu_1 \left[\frac{1 - (-1)^{n+1}}{\pi(n+1)} \right] \cos \omega_0 t, \quad n = 1, 2, \cdots , \end{aligned} \tag{1.8}$$

where we have defined $a^2 = \Gamma - 1$ and $\omega_n^2 = (n+1)^2(n+1)^2 - \Gamma$ In (1.8) we project onto the hyperbolic subspace by setting $q = 0$ in the first equation of (1.8) to obtain what we shall call the reduced equation. In our example this is

$$\ddot{p} = a^2 p - \frac{\pi}{2} p^3 - 2\mu_2 \dot{p} + \frac{4}{\pi} \mu_1 \cos \omega_0 t \,. \tag{1.9}$$

We see that this is the forced, damped Duffing equation with negative stiffness for which standard theory yields chaotic dynamics. The purpose of [7] is to show that the chaotic dynamics of (1.9) are, in some sense, shadowed in the dynamics of the full equation (1.8).

2. The problem (1.2)

By using the results of [18, p. 253], we know that $(-\pi, 0)$ and $(\pi, 0)$ are hyperbolic fixed points of the ordinary differential equation

$$\dot{x} = y, \quad \dot{y} = -\sin x \tag{2.1}$$

joined by the upper separatrix

$$(\gamma, \dot{\gamma}), \quad \gamma = \pi - 4\tan^{-1}(\mathrm{e}^{-t}) .$$

We consider (1.4) as an ordinary differential equation on the circle $S^{2\pi} = \mathbf{R}/[0, 2\pi]$. Then (2.1) is defined on the cylinder $S^{2\pi} \times \mathbf{R} \ni (x, y)$ and $(-\pi, 0)$, $(\pi, 0)$ are glued to a hyperbolic fixed point of (2.1) joined by the homoclinic orbit $(\gamma, \dot{\gamma})$.

We note that $\lambda_k = 16\pi^4 k^4$, $k \in \mathbf{Z}_+ = \{0\} \cup \mathbf{N}$ are eigenvalues of u_{xxxx} with the periodic boundary value condition of (1.2), and

$$\sum_{\lambda_k > 0} \left(\frac{3}{\sqrt{\lambda_k}} + \frac{1}{\lambda_k} \right) = \sum_{m \geq 1} \left(\frac{3}{4\pi^2 m^2} + \frac{1}{16\pi^4 m^4} \right) = \frac{181}{1440} .$$

We need the following lemma.

Lemma 2.1. ([3]) *Let $D \in \mathbf{N}$. If $S(d)$ is the set of all $\omega \geq 1$ satisfying*

$$\left| \omega - \frac{n}{4\pi^2 k^2} \right| \geq \frac{d}{16\pi^4 k^4} \quad \forall\, k > 0, \forall\, n \in \mathbf{Z}_+$$

with $0 < d < 16\pi^4$, then

$$m\left(S(d) \cap [D, D+1]\right) \geq 1 - \frac{181}{720} d ,$$

where $m(\cdot)$ is the Lebesque measure.

Now we can state the main result of [3] concerning (1.2).

Theorem 2.2. ([3]) *Let $8 > \rho > 5$ and $0 < d < 720/181$. There are constants $K_1 > 0$, $K_2 > 0$ such that for any $1 > |\epsilon| > 0$, $m \in \mathbf{N}$, $0 < \eta \leq 1$ satisfying*

$$\frac{1}{|\epsilon|^{1/2}} < m < \frac{K_1}{2\pi |\epsilon|^{(\rho-4)/2}} , \quad \frac{m}{\eta |\epsilon|^{\rho/2}} \in S(d) ,$$

(1.2) has a $2\pi m$–periodic weak solution u_m with $\alpha = \frac{1}{\eta^2 |\epsilon|^\rho}$ satisfying

$$\max_{-\pi m \leq t \leq \pi m} \left| \int_0^1 u_m(x, t)\, dx - \gamma(t - \delta_m) \right| \leq K_2 |\epsilon|$$

for some $\delta_m \in [-K_2, K_2]$.

The above results can be modified also to the equations

$$\begin{gathered} u_{tt} + \alpha u_{xxxx} + \sin u + \tau(x) = \epsilon \cos t, \quad u(x,t) = u(1-x,t) \\ u_{xx}(0,t) = u_{xx}(1,t) = u_{xxx}(0,t) = u_{xxx}(1,t) = 0\,, \end{gathered} \tag{2.2}$$

and

$$\begin{gathered} u_{tt} + \alpha u_{xxxx} + 2u^3 - u + \tau(x) = \epsilon \cos t \\ u(x+1,t) = u(x,t)\,, \end{gathered} \tag{2.3}$$

where $\alpha > 0$ is a constant, τ is continuous satisfying $\int_0^1 \tau(x)\,dx = 0$ and $\epsilon \in \mathbf{R}$ is a small parameter. Moreover, τ is either symmetric with respect to $x = 1/2$ for (2.2) or 1–periodic for (2.3). We note that the limit equation of (2.3) as $\alpha \to \infty$ is the forced Duffing equation given by

$$\ddot{u} + 2u^3 - u = \epsilon \cos t\,.$$

We end this section with an existence result following directly from [3, Theorem 2.9].

Theorem 2.3. *Let* $0 < d < 720/181$. *The equation*

$$\begin{gathered} u_{tt} + \alpha u_{xxxx} + u + \theta \sin u + \tau(x) = \epsilon \sin t \\ u(x+1,t) = u(x,t)\,, \end{gathered}$$

where τ *is the same as in* (1.2) *and* $0 < \theta < 1$, *has a unique* 2π*–periodic weak solution for any* $\epsilon \in \mathbf{R}$ *and* α *satisfying*

$$\alpha > (1+)^2/d^2 \qquad \textit{and} \quad \sqrt{\alpha} \in S(d)\,.$$

3. The problem (1.3)

In this section we recall some results of [4] concerning (1.3). We note that

$$\lambda_j = \pi^2 j^2(\pi^2 j^2 - \Gamma), \quad j \in \mathbf{N}$$

are the eigenvalues of the linear part of (1.3) given by $u_{xxxx} + \Gamma u_{xx}$ with the corresponding boundary value conditions.

Lemma 3.1. ([4]) *Let* $D \geq 0$ *and* $\rho > 1/4$. *If* $S(c)$ *is the set of all* $T > 0$ *satisfying*

$$\lambda_j^\rho \left| \sin \frac{\sqrt{\lambda_j} T}{2} \right| \geq c \quad \forall j \in \mathbf{N} \,:\, j^2 > \Gamma/\pi^2\,,$$

where $c > 0$ is a constant, then the Lebesque measure of the complement

$$(\mathbf{R} \setminus S(c)) \cap [D, D+1]$$

satisfies

$$m\left((\mathbf{R} \setminus S(c)) \cap [D, D+1]\right) \leq \sum_{k^2 > \Gamma/\pi^2} \left(\frac{2c\pi}{\sqrt{\lambda_k}\lambda_k^\rho} + \frac{c}{\lambda_k^\rho} + \frac{2c^2\pi}{\sqrt{\lambda_k}\lambda_k^{2\rho}} \right).$$

Now we can state the main result of [4] concerning (1.3).

Theorem 3.2. ([4]) *Consider* (1.3) *with* $0 \leq \xi \leq 1$, $\pi^2 < \Gamma < 3\pi^2$. *Let* $1 \geq c > 0$ *and let* (1.6) *hold. If* $\int_0^1 q(x) \sin \pi x \, dx \neq 0$ *then for any* $p \in \mathbf{N}$, T *satisfying*

$$\frac{\sqrt{2}}{\sqrt{\Gamma - \pi^2}} \leq pT < \frac{1}{4^\xi \sqrt{8 + 2(4\pi^4)^{1-\xi}}} \frac{c}{\Gamma - \pi^2}$$

and $pT \in S(c)$, *the equation* (1.3) *has a* pT*–periodic weak solution for any* ϵ *sufficiently small. Moreover, if* $0 \leq \xi < 1/2$, *then* $m\left(S(c) \cap [D, D+1]\right) \to 1$ *as* $c \to 0_+$ *uniformly with respect to* $D \geq 0$.

Roughly speaking, Theorem 3.2 asserts that (1.3) with $0 \leq \xi < 1/2$ and $\pi^2 < \Gamma < 3\pi^2$, has many subharmonics for ϵ sufficiently small when c and $\pi^2 - \Gamma$ are sufficiently small. On the other hand, since we do not know the structure of the set $S(c)$, we do not know a concrete case for pT. This result can be related to the KAM theory [13]. Furthermore, Lemma 3.1 gives that the Lebesque measure of $S(c)$ tends to 1 as $c \to 0_+$. We note [4] that now $\rho = (1 - \xi)/2$ in Lemma 3.1. Consequently, we put $c = (\Gamma - \pi^2)^{1/4}$ in Theorem 3.2, and the assumptions of Theorem 3.2 take the following form

$$\frac{\sqrt{2}}{\sqrt{\Gamma - \pi^2}} \leq pT < \frac{1}{4^\xi \sqrt{8 + 2(4\pi^4)^{1-\xi}}} \frac{1}{(\Gamma - \pi^2)^{3/4}}, \quad pT \in S\left((\Gamma - \pi^2)\right).$$

Summarizing we obtain the following result.

Theorem 3.3. ([4]) *Let* $1/2 > \xi \geq 0$ *and* $\pi^2 < \Gamma < 3\pi^2$. *If* (1.6) *holds and* $\int\limits_0^1 q(x) \sin \pi x \, dx \neq 0$, *then the smaller* $\Gamma - \pi^2 > 0$, *the larger number of subharmonic weak solutions of* (1.5) *persists for* (1.3) *when* ϵ *is sufficiently small.*

The above results deal with the bifurcation of subharmonic weak solutions. Now we formulate some results about the existence of periodic weak solutions for (1.3).

Theorem 3.4. ([4]) *Let* $1/2 > \xi \geq 0$ *and* $\Gamma \in \mathbf{R} \setminus \{\pi^2 j^2 \mid j \in \mathbf{N}\}$ *be arbitrari. For almost all* $T > 0$ *(in the sense of the Lebesque measure), there is a small* T*–periodic weak solution of* (1.3) *for any* ϵ *sufficiently small.*

We do not know the form of T in the above result. Particular results with concrete T and Γ are given bellow. Related results are proved in [1, 12, 17, 22, 23].

Theorem 3.5. ([4]) *Consider* (1.3) *with* $1 \geq \xi \geq 0$*. If* $T\pi \in \mathbf{Q}$ *and* $\Gamma/\pi^2 \notin \mathbf{Q}$*, then there is a small* T*–periodic weak solution of* (1.3) *for any* ϵ *sufficiently small.*

Theorem 3.6. ([4]) *Consider* (1.3) *with* $\xi = 0$*. If* $T = 2\Omega/\pi$, $\Gamma = 2m\pi^2/\Omega$*, where* $m \in \{-1, 0\}$ *and* $\Omega > 0$ *has the infinite continuous fraction decomposition* [17] $\Omega = [a_0, a_1, a_2, \cdots]$ *such that* $a_i \leq M \forall i \in \mathbf{Z}_+$ *for a constant* $M \in \mathbf{N}$ *satisfying* $m^2(M+2) < 2\Omega$*, then there is a small* T*–periodic weak solution of* (1.3) *for any* ϵ *sufficiently small.*

We remark that Theorem 3.4 provides a more general result than Theorems 3.5 and 3.6, while Theorems 3.5 and 3.6 give concrete forms of T and Γ for the existence of periodic solutions of (1.3). For instance, Theorem 3.6 is valid for $\Omega = \left(M + \sqrt{M^2 + 4M}\right)/2$, $M \in \mathbf{N}$ and $m \in \{-1, 0\}$.

Results like above are obtained in [4] also for the equation

$$\begin{aligned} u_{tt} + u_{xxxx} + \Gamma u_{xx} + p\left(\int_0^2 u^2(s,t)\,ds, \int_0^2 u_x^2(s,t)\,ds\right) D_{xx}^{\xi} u \\ = \epsilon\left(q_1(x)\cos\tfrac{2\pi t}{T} + q_2(x)\sin\tfrac{2\pi t}{T}\right), \\ u(x,\cdot) = u(x+2,\cdot) \quad \forall x \in \mathbf{R}, \end{aligned}$$

where ξ, Γ, p are like in (1.3), D_{xx}^{ξ} is the ξ–power of D_{xx} in $L^2(\mathbf{R}/2\mathbf{Z})$, $q_{1,2} \in H^2(0,2)$ and $q_{1,2}$ are 2–periodic satisfying

$$\begin{aligned} \int_0^2 q_1(x)\cos\pi x\,dx = 0, \quad \int_0^2 q_2(x)\sin\pi x\,dx = 0, \\ \int_0^2 q_1(x)\sin\pi x\,dx \neq 0, \quad \int_0^2 q_2(x)\cos\pi x\,dx \neq 0. \end{aligned}$$

4. The problem (1.7)

The reduced equations of (1.9) are

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= a^2 x_1 - \tfrac{\pi}{2} x_1^3 - 2\mu_2 x_2 + \tfrac{4}{\pi}\mu_1 \cos\omega_0 t \end{aligned} \tag{4.1}$$

obtained by setting $q = 0$ in the hyperbolic part of (1.8). When $\mu = 0$, (4.1) has a homoclinic solution given by $\gamma = (r, \dot{r})$ where $r(t) = (2a/\sqrt{\pi})\operatorname{sech} at$. The Melnikov homoclinic function M for (4.1) [7, 18] becomes

$$M(\mu,\alpha) = \left[\frac{8\omega_0}{\sqrt{\pi}}\sin\omega_0\alpha \operatorname{sech}\frac{\pi\omega_0}{2a}\right]\mu_1 - \frac{16a^3}{3\pi}\mu_2 .$$

Thus, the conditions $M(\mu_0, \alpha_0) = 0$, $(\partial M/\partial\alpha)(\mu_0, \alpha_0) \neq 0$ are satisfied for all μ_0 such that

$$\left|\frac{\mu_{0,2}}{\mu_{0,1}}\right| < \frac{3\sqrt{\pi}\omega_0}{2a^3}\operatorname{sech}\frac{\pi\omega_0}{2a} .$$

We now obtain the following result from [7].

Theorem 4.1. *There exists a constant* $C \geq 0$ *so that whenever* μ_0 *satisfies* $C \leq \left|\frac{\mu_{0,2}}{\mu_{0,1}}\right| < \frac{3\sqrt{\pi}\omega_0}{2a^3}\operatorname{sech}\frac{\pi\omega_0}{2a}$ *there exists a corresponding* $\bar{\xi}_0 > 0$ *such that if* $0 < \xi \leq \bar{\xi}_0$, *if the parameters in* (1.7) *are given by* $\mu = \xi\mu_0$ *and* $\mu_{0,2} \neq 0$ *then* (1.7) *is chaotic. In the nonresonant case, i.e. if* $\omega_0 \neq \omega_n$ *for all* n, *then* $C = 0$.

It is interesting to look at some history of this problem. The first work was by P. Holmes [9] in which he started with the partial differential equation and carried out the Galerkin expansion but restricted his analysis to the reduced equation. The significance of that work is that it introduced the idea of Melnikov analysis. In subsequent work [10, 11] P. Holmes and J. Marsden extended the results to infinite dimension but abandoned the Galerkin approach in favor of nonlinear semigroup techniques directly in infinite dimensions. In our work [7] we go back to the original, simpler analysis of the reduced equation and then show that the results apply to the original partial differential equation. Some advantages to this are that the Galerkin projection is a technique familiar to many engineers and physicists and, also, we are able to utilize general Melnikov results in Section 3 of [7] and Palmer's important results of [18]. This is illustrated further in [7] for the following generalizations:

- Nonplanar motion of a symmetric beam with one buckled mode.
- Nonplanar, nonsymmetric beam with one buckled mode in each plane.
- Planar motion of a symmetric beam with multiple buckled modes.

Acknowledgement: Partially supported by Grant GA-MS 1/6179/00. This paper is in final form and no version of it will be submitted for publication elsewhere.

REFERENCES

[1] Ben-Naoum, A. K. and Mawhin, J.: *The periodic-Dirichlet problem for some semilinear wave equations*, J. Differential Equations **96**, (1992), 340-354.

[2] Berti, M. and Carminati, C.: *Chaotic dynamics for perturbations of infinite dimensional Hamiltonian systems*, ISAS - International School for Advanced Studies, Technical Report (Ref. SISSA 65/99/M), 1999.

[3] Fečkan, M.: *Periodic oscillations of abstract wave equations*, J. Dyn. Diff. Equations, **10**, (1998), 605-617.

[4] Fečkan, M.: *Bifurcation of periodics and subharmonics in abstract nonlinear undamped wave equations*, J. Differential Equations, **153**, (1999), 41-60.

[5] Fečkan, M.: *Periodic solutions of certain abstract wave equations*, Proc. Amer. Math. Soc., **123**, (1995), 465-470.

[6] Fečkan, M.: *On the existence of periodic solutions for a certain type of nonlinear equations*, J. Differential Equations, **121**, (1995), 28-41.

[7] GRUENDLER, J. and Fečkan, M.: *The existence of chaos in infinite dimensional systems*, preprint (2000).

[8] Guckenheimer, J. and Holmes, P.: *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer–Verlag, New York, 1983.

[9] Holmes, P.: *A nonlinear oscillator with a strange attractor*, Phil. Trans. Roy. Soc. A, **292**, (1979), 419-448.

[10] Holmes, P. and Marsden, J.: *A partial differential equation with infinitely many periodic orbits: chaotic oscillations of a forced beam*, Arch. Rational Mech. Anal., **76**, (1981), 135-165.

[11] Holmes, P. and Marsden, J.: *Bifurcation to divergence and flutter in flow–induced oscillations: an infinite dimensional analysis*, Automatica, **14**, (1978), 367-384.

[12] Kolesov, A. Yu., Mischchenko, E. F. and Rozov, N. Kh.: *Asymptotic Methods of Investigation of Periodic Solutions of Nonlinear Hyperbolic Equations*, Proc. V.A. Steklov Inst. Math. **222**, Nauka Publ.-MAIK Nauka, Moscow, 1998 (in Russian).

[13] Kuksin, S. B.: *Nearly Integrable Infinite–Dimensional Hamiltonian Systems*, Lec. Not. Math. Vol. 1556, Springer–Verlag, Berlin, 1993.

[14] McLaughlin, D.W. and Overman II, E.A.: *Whiskered tori for integrable pde's: chaotic behavior in near integrable pde's*, Surveys Appl. Math., **1**, (1995), 83-203.

[15] McLaughlin, D.W., Overman II, E.A., Wiggins, S. and XIONG, C.: *Homoclinic orbits in a four–dimensional model of a perturbed NLS equation: a geometric singular perturbed study*, Dynamics Reported, (1995), 190-287.

[16] McLaughlin, D.W. and Shatan, J.: *Homoclinic orbits for pde's*, preprint (1997).

[17] McKenna, P. J.: *On solutions of a nonlinear wave equation when the ratio of the period to the length of the interval is irrational*, Proc. Amer. Math. Soc., **93**, (1985), 59-64.

[18] Palmer, K. J.: *Exponential dichotomies and transversal homoclinic points*, J. Differential Equations, **55**, (1984), 225-256.

[19] Rodrigues, H. M. and Silveira, M.: *Properties of bounded solutions of linear and nonlinear evolution equations: homoclinics of a beam equation*, J. Differential Equations, **70**, (1987), 403-440.

[20] Samoilenko, A.M. and Teplinskii, Yu. V.: *Countable Systems of Differential Equations*, Kiev, 1993 (in Russian).

[21] Shaw, S.W.: *Chaotic dynamics of a slender beam rotating about its longitudinal axis*, J. Sound Vibr., **124**, (1988), 329-343.

[22] Vejvoda, O., et al.: *Partial Differential Equations: Time-Periodic Solutions*, Noordhoff, Groningen, 1981.

[23] Yamaguchi, M.: *Existence of periodic solutions of second order nonlinear evolution equations and applications*, Funk. Ekvacioj, **38**, (1995), 519-538.

Doi: 10.18514/MMN.2000.24

Mathematical Notes, Miskolc, Vol. 1., No. 2., (2000), pp. 109–118

COMPONENTWISE PERTURBATION BOUNDS FOR THE LU, LDU AND LDL^T DECOMPOSITIONS

Aurél Galántai
Institute of Mathematics, University of Miskolc
3515 Miskolc – Egyetemváros, Hungary
matgal@gold.uni-miskolc.hu

[Received June 1, 2000]

Abstract. We improve a componentwise perturbation bound of Sun for the LU factorization and derive a new perturbation bound for the LDU factorization. The latter bound also improves a result of Sun given for the LDL^T factorization.

Mathematical Subject Classification: 15A23, 15A45, 65F99.
Keywords: LU, LDU, LDL^T factorization, perturbation.

1. Introduction

Perturbation bounds for the LU, LDL^T factorizations are given by many authors (e.g., see [1], [9], [7], [8], [2]). Here we improve the componentwise LU perturbation bound of Sun [9] and derive a new perturbation bound for the LDU decomposition. These bounds are used to investigate the stability of full rank factorizations produced by Egerváry's rank reduction procedure [4], [3]. The LDU perturbation bounds are then applied to positive definite symmetric matrices. The result is shown to be better than the LDL^T perturbation result of Sun [9].

We need the following notations. Let $A = [a_{ij}]_{i,j=1}^n$. Then $|A| = [|a_{ij}|]_{i,j=1}^n$,

$$diag\,(A) = diag\,(a_{11}, a_{22}, \dots, a_{nn}),$$

$tril\,(A, l) = [\alpha_{ij}]_{i,j=1}^n$ and $triu\,(A, l) = \left[\beta_{ij}\right]_{i,j=1}^n$, where $0 \le |l| < n$ and

$$\alpha_{ij} = \begin{cases} a_{ij}, & i \ge j - l \\ 0, & i < j - l \end{cases}, \qquad \beta_{ij} = \begin{cases} a_{ij}, & i \le j - l \\ 0, & i > j - l \end{cases}.$$

We also use the special notations $tril\,(A) = tril\,(A, 0)$, $tril^*\,(A) = tril\,(A, -1)$, $triu\,(A) = triu\,(A, 0)$ and $triu^*\,(A) = triu\,(A, 1)$. The spectral radius of A will be denoted by $\rho\,(A)$. For two matrices $A, B \in R^{n\times n}$ the relation $A \le B$ holds if and only if $a_{ij} \le b_{ij}$ for all $i, j = 1, \dots, n$. Let $\widetilde{I}_k = \sum_{i=1}^k e_i e_i^T$ ($e_i \in R^n$ is the ith unit vector) for $1 \le k \le n$, $\widetilde{I}_k = 0$ for $k \le 0$ and $\min\,(A, B) = [\min\,(a_{ij}, b_{ij})]_{i,j=1}^n$.

In Sections 2 and 3 we derive the perturbation bound for the LU and LDU factorizations. A numerical example is shown in Section 4.

2. The LU factorization

We first prove the following

Lemma 1 *Assume that $A, B, C \in R^{n\times n}$ are such that $A, B, C \geq 0$ and $\rho(B) < 1$. The maximal solution of the inequality $A \leq C + Btriu(A, l)$ $(l \geq 0)$ is A^* $(A^* \geq C)$, where $A^* e_k = \left(I - B\widetilde{I}_{k-l}\right)^{-1} Ce_k$ $(k = 1, \ldots, n)$. A^* is the unique solution of the fixed point problem $A = f(A) = C + Btriu(A, l)$. If $A_0 = (I - B)^{-1} C$, then $A_i = f(A_{i-1})$ converges to A^* monotonically decreasing as $i \to +\infty$ and $0 \leq A_i - A^* \leq (I - B)^{-1} B^i (A_0 - A_1)$ $(i \geq 1)$.*

Proof. It follows from $A \leq C + Btriu(A, l) \leq C + BA$ that $(I - B) A \leq C$. As $I - B$ is a nonsingular M-matrix by assumption we obtain the upper bound $A \leq A_0 = (I - B)^{-1} C$. As

$$\left|f(A) - f\left(\widetilde{A}\right)\right| = \left|B\left(triu(A, l) - triu\left(\widetilde{A}, l\right)\right)\right| \leq B\left|A - \widetilde{A}\right|$$

for any two $n \times n$ matrices A and $\widetilde{A}$, the map $f(A)$ is a B-contraction [6] on $R^{n\times n}$ and there is a unique fixed point $A^* = f(A^*)$. Let $X_0 \in R^{n\times n}$ be arbitrary and $X_k = f(X_{k-1})$ $(k \geq 1)$. Then $|A^* - X_k| \leq (I - B)^{-1} B^k |X_1 - X_0|$ $(k \geq 1)$. As for any $0 \leq A \leq \widetilde{A}$, $f(A) \leq f\left(\widetilde{A}\right)$ holds and

$$A_1 = C + Btriu\left((I - B)^{-1} C, l\right) \leq C + B(I - B)^{-1} C = (I - B)^{-1} C = A_0,$$

the sequence $A_i = f(A_{i-1})$ tends to A^* and is monotonically decreasing. We prove that A^* is the maximal solution of the inequality. Assume that a solution $\widetilde{A}$ exists such that $\widetilde{A} \geq A^*$. Then $\widetilde{A} = A^* + L + U$, where $triu(U, l) = U$ and $tril(L, l - 1) = L$. Then

$$\widetilde{A} = A^* + L + U \leq C + Btriu(A^* + L + U, l) \leq C + Btriu(A, l) + BU$$

must hold implying that $L + U \leq BU$ and $0 \leq U \leq -(I - B)^{-1} L \leq 0$. Hence $U = L = 0$. The kth column of A^* can be written as $A^* e_k = Ce_k + Btriu(A^*, l) e_k$, where $triu(A^*, l) e_k = \widetilde{I}_{k-l} A^* e_k$. Hence we obtain $A^* e_k = \left(I - B\widetilde{I}_{k-l}\right)^{-1} Ce_k$. ■

Remark 2 *The sequence $\{A_i\}_{i\geq 0}$ gives an improving sequence of upper estimates for the maximal solution A^* of the inequality.*

We will use the following notations: $A^* = \phi(B, C, l)$, $A_i = \phi_i(B, C, l)$, $\phi_0(B, C, l) = (I - B)^{-1} C$ and $\phi_i(B, C, l) = C + Btriu\left(\phi_{i-1}(B, C, l), l\right)$ $(i \geq 1)$. Notice that for any diagonal matrix $\widetilde{D}$, $\phi\left(B, C\widetilde{D}, l\right) = \phi(B, C, l)\widetilde{D}$ and $\phi_i\left(B, C\widetilde{D}, l\right) = \phi_i(B, C, l)\widetilde{D}$.

Remark 3 *Consider the inequality* $A \leq C + tril\,(A,-l)\,B$ $(l \geq 0)$ *with* $0 \leq A,B,C \in R^{n\times n}$ *and* $\rho\,(B) < 1$. *By transposition we obtain* $A^T \leq C^T + B^T tril\,(A,-l)^T = C^T + B^T triu\left(A^T,l\right)$ *the maximal solution of which is given by* $\phi\left(B^T,C^T,l\right)$. *The sequence* $\phi_i\left(B^T,C^T,l\right)$ *tends to* $\phi\left(B^T,C^T,l\right)$ *and is monotonically decreasing. Hence for the original inequality we have the maximal solution* $\phi\left(B^T,C^T,l\right)^T$ *and the monotone decreasing sequence* $\phi_i\left(B^T,C^T,l\right)^T$ *converging to* $\phi\left(B^T,C^T,l\right)^T$.

The next theorem improves the componentwise estimate of Sun [9].

Theorem 4 *Assume that the* $n\times n$ *matrix* A *has the* LU *decomposition* $A = L_1U$, *where* L_1 *is unit lower triangular and* U *is upper triangular. Also assume that the perturbed matrix* $A+\delta_A$ *has the* LU *decomposition* $A+\delta_A = (L_1+\delta_{L_1})\,(U+\delta_U)$, *where* $L_1+\delta_{L_1}$ *is unit lower triangular and* $U+\delta_U$ *is upper triangular. Finally assume that* $\rho\left(\left|L_1\delta_A U^{-1}\right|\right) < 1$. *Then we have*

$$|\delta_{L_1}| \leq |L_1|\,tril^*\left(\phi\left(\left|L_1^{-1}\delta_A U^{-1}\right|,\left|L_1^{-1}\delta_A U^{-1}\right|,0\right)\right), \tag{2.1}$$

$$|\delta_U| \leq triu\left(\phi\left(\left|L_1^{-1}\delta_A U^{-1}\right|^T,\left|L_1^{-1}\delta_A U^{-1}\right|^T,1\right)^T\right)|U|. \tag{2.2}$$

Proof. Using the relation

$$\delta_U\,(U+\delta_U)^{-1} + L_1^{-1}\delta_{L_1} = L_1^{-1}\delta_A\,(U+\delta_U)^{-1},$$

where $L_1^{-1}\delta_{L_1}$ is a strict lower triangular matrix, while $\delta_U\,(U+\delta_U)^{-1}$ is upper triangular, we can establish the relations

$$tril^*\left(L_1^{-1}\delta_A\,(U+\delta_U)^{-1}\right) = L_1^{-1}\delta_{L_1}, \tag{2.3}$$

$$triu\left(L_1^{-1}\delta_A\,(U+\delta_U)^{-1}\right) = \delta_U\,(U+\delta_U)^{-1}. \tag{2.4}$$

From relation

$$L_1^{-1}\delta_A\,(U+\delta_U)^{-1} = L_1^{-1}\delta_A U^{-1} - L_1^{-1}\delta_A U^{-1}\delta_U\,(U+\delta_U)^{-1} \tag{2.5}$$

we obtain the inequality

$$\left|L_1^{-1}\delta_A\,(U+\delta_U)^{-1}\right| \leq \left|L_1^{-1}\delta_A U^{-1}\right| + \left|L_1^{-1}\delta_A U^{-1}\right| triu\left(\left|L_1^{-1}\delta_A\,(U+\delta_U)^{-1}\right|\right).$$

Applying Lemma 1 we obtain the bound

$$\left|L_1^{-1}\delta_A\,(U+\delta_U)^{-1}\right| \leq A^* = \phi\left(\left|L_1^{-1}\delta_A U^{-1}\right|,\left|L_1^{-1}\delta_A U^{-1}\right|,0\right).$$

Hence $\left|L_1^{-1}\delta_{L_1}\right| \leq tril^*\left(A^*\right)$ and $\left|\delta_{L_1}\right| \leq \left|L_1\right| tril^*\left(A^*\right)$.

Using the relation

$$\delta_U U^{-1} + \left(L_1 + \delta_{L_1}\right)^{-1}\delta_{L_1} = \left(L_1 + \delta_{L_1}\right)^{-1}\delta_A U^{-1},$$

where $\left(L_1 + \delta_{L_1}\right)^{-1}\delta_{L_1}$ is a strict lower triangular matrix, while $\delta_U U^{-1}$ is upper triangular, we can also establish the relations

$$tril^*\left(\left(L_1 + \delta_{L_1}\right)^{-1}\delta_A U^{-1}\right) = \left(L_1 + \delta_{L_1}\right)^{-1}\delta_{L_1} \tag{2.6}$$

and

$$triu\left(\left(L_1 + \delta_{L_1}\right)^{-1}\delta_A U^{-1}\right) = \delta_U U^{-1}. \tag{2.7}$$

>From relation

$$\left(L_1 + \delta_{L_1}\right)^{-1}\delta_A U^{-1} = L_1^{-1}\delta_A U^{-1} - \left(L_1 + \delta_{L_1}\right)^{-1}\delta_{L_1} L_1^{-1}\delta_A U^{-1} \tag{2.8}$$

we obtain the inequality

$$\left|\left(L_1 + \delta_{L_1}\right)^{-1}\delta_A U^{-1}\right| \leq \left|L_1^{-1}\delta_A U^{-1}\right| + tril^*\left(\left|\left(L_1 + \delta_{L_1}\right)^{-1}\delta_A U^{-1}\right|\right)\left|L_1^{-1}\delta_A U^{-1}\right|$$

the maximal solution of which is

$$\left|\left(L_1 + \delta_{L_1}\right)^{-1}\delta_A U^{-1}\right| \leq \widetilde{A}^* = \phi\left(\left|L_1^{-1}\delta_A U^{-1}\right|^T, \left|L_1^{-1}\delta_A U^{-1}\right|^T, 1\right)^T.$$

Hence $\left|\delta_U U^{-1}\right| \leq triu\left(\widetilde{A}^*\right)$ and $\left|\delta_U\right| \leq triu\left(\widetilde{A}^*\right)\left|U\right|$. This completes the proof. ■

Remark 5 *If function ϕ is replaced by ϕ_0 in (2.1)-(2.2) we obtain the theorem of Sun [9], Thm. 5.1). Hence, our result is sharper.*

Remark 6 *Assume that the LU factorizations $A = LU_1$ and*

$$A + \delta_A = \left(L + \delta_L\right)\left(U_1 + \delta_{U_1}\right)$$

are such that U_1 and $U_1 + \delta_{U_1}$ are upper unit triangular. If $A^T = U_1^T L^T$ and $A^T + \delta_A^T = \left(U_1^T + \delta_{U_1}^T\right)\left(L^T + \delta_L\right)$ satisfy the conditions of the previous theorem we may write

$$\left|\delta_L\right| \leq \left|L\right| tril\left(\phi\left(\left|L^{-1}\delta_A U_1^{-1}\right|, \left|L^{-1}\delta_A U_1^{-1}\right|, 1\right)\right), \tag{2.9}$$

and

$$\left|\delta_{U_1}\right| \leq triu^*\left(\phi\left(\left|L^{-1}\delta_A U_1^{-1}\right|^T, \left|L^{-1}\delta_A U_1^{-1}\right|^T, 0\right)^T\right)\left|U_1\right|. \tag{2.10}$$

Hence, Theorem 4 is also true for the case $A = LU_1$ with unit upper triangular U_1. Notice, however, that we have here tril and triu instead of tril* and triu, respectively. This is due to the change of the unit triangular part in the LU factorization.*

3. The LDU factorization

Consider the LDU factorization $A = L_1 D U_1$ with unit lower triangular L_1, diagonal D and unit upper triangular U_1. Assume that $A + \delta_A$ can be factorized so that

$$A + \delta_A = (L_1 + \delta_{L_1})(D + \delta_D)(U_1 + \delta_{U_1})$$

where $L_1 + \delta_{L_1}$ is unit lower triangular and $U + \delta_{U_1}$ is unit upper triangular. For δ_{L_1} and δ_{U_1} we have the bounds (2.1) and (2.10), respectively. We now look for an estimate of δ_D. We use the relation

$$L_1^{-1}\delta_A (U_1 + \delta_{U_1})^{-1} = D\delta_{U_1}(U + \delta_{U_1})^{-1} + \delta_D + L_1^{-1}\delta_{L_1}(D + \delta_D),$$

where the matrix $D\delta_{U_1}(U + \delta_{U_1})^{-1}$ is strict upper triangular, δ_D is diagonal, and $L_1^{-1}\delta_{L_1}(D + \delta_D)$ is strict lower triangular. Hence

$$tril^*\left(L_1^{-1}\delta_A (U_1 + \delta_{U_1})^{-1}\right) = L_1^{-1}\delta_{L_1}(D + \delta_D), \tag{3.1}$$

$$diag\left(L_1^{-1}\delta_A (U_1 + \delta_{U_1})^{-1}\right) = \delta_D, \tag{3.2}$$

$$triu^*\left(L_1^{-1}\delta_A (U_1 + \delta_{U_1})^{-1}\right) = D\delta_{U_1}(U_1 + \delta_{U_1})^{-1}. \tag{3.3}$$

>From relation

$$L_1^{-1}\delta_A (U_1 + \delta_{U_1})^{-1} = L_1^{-1}\delta_A U_1^{-1} - L_1^{-1}\delta_A U_1^{-1}\delta_{U_1}(U_1 + \delta_{U_1})^{-1} \tag{3.4}$$

we obtain the inequality

$$\begin{aligned}\left|L_1^{-1}\delta_A (U_1 + \delta_{U_1})^{-1}\right| \quad \leq \quad & \left|L_1^{-1}\delta_A U_1^{-1} D^{-1}\right| |D| + \\ & + \left|L_1^{-1}\delta_A U_1^{-1} D^{-1}\right| triu^*\left(\left|L_1^{-1}\delta_A (U_1 + \delta_{U_1})^{-1}\right|\right)\end{aligned}$$

the maximal solution of which is given by the bound

$$\left|L_1^{-1}\delta_A (U_1 + \delta_{U_1})^{-1}\right| \leq \phi\left(\left|L_1^{-1}\delta_A U_1^{-1} D^{-1}\right|, \left|L_1^{-1}\delta_A U_1^{-1} D^{-1}\right| |D|, 1\right).$$

Hence $|\delta_D| \leq |D|\, diag\left(\phi\left(\left|L_1^{-1}\delta_A U_1^{-1} D^{-1}\right|, \left|L_1^{-1}\delta_A U_1^{-1} D^{-1}\right|, 1\right)\right)$.

We may get another estimate by using the expression

$$(L_1 + \delta_{L_1})^{-1}\delta_A U_1^{-1} = (D + \delta_D)\delta_{U_1} U_1^{-1} + \delta_D + (L_1 + \delta_{L_1})^{-1}\delta_{L_1} D,$$

where the matrix $(D+\delta_D)\,\delta_{U_1}U_1^{-1}$ is strict upper triangular, δ_D is diagonal, and $(L_1+\delta_{L_1})^{-1}\,\delta_{L_1}D$ is strict lower triangular. Hence

$$tril^*\left((L_1+\delta_{L_1})^{-1}\,\delta_A U_1^{-1}\right) = (L_1+\delta_{L_1})^{-1}\,\delta_{L_1}D, \tag{3.5}$$

$$diag\left((L_1+\delta_{L_1})^{-1}\,\delta_A U_1^{-1}\right) = \delta_D, \tag{3.6}$$

$$triu^*\left((L_1+\delta_{L_1})^{-1}\,\delta_A U_1^{-1}\right) = (D+\delta_D)\,\delta_{U_1}U_1^{-1}. \tag{3.7}$$

>From relation

$$(L_1+\delta_{L_1})^{-1}\,\delta_A U_1^{-1} = L_1^{-1}\delta_A U_1^{-1} - (L_1+\delta_{L_1})^{-1}\,\delta_{L_1}L_1^{-1}\delta_A U_1^{-1} \tag{3.8}$$

we obtain the inequality

$$\begin{aligned}\left|(L_1+\delta_{L_1})^{-1}\,\delta_A U_1^{-1}\right| \;&\leq\; |D|\left|D^{-1}L_1^{-1}\delta_A U_1^{-1}\right| + \\ &+tril^*\left(\left|(L_1+\delta_{L_1})^{-1}\,\delta_A U_1^{-1}\right|\right)\left|D^{-1}L_1^{-1}\delta_A U_1^{-1}\right|.\end{aligned}$$

It has the maximal solution

$$\left|(L_1+\delta_{L_1})^{-1}\,\delta_A U_1^{-1}\right| \leq \phi\left(\left|D^{-1}L_1^{-1}\delta_A U_1^{-1}\right|^T, \left|D^{-1}L_1^{-1}\delta_A U_1^{-1}\right|^T |D|, 1\right)^T.$$

Hence $|\delta_D| \leq |D|\,diag\left(\phi\left(\left|D^{-1}L_1^{-1}\delta_A U_1^{-1}\right|^T, \left|D^{-1}L_1^{-1}\delta_A U_1^{-1}\right|^T, 1\right)^T\right)$. We now have two estimates for $|\delta_D|$. As in general $|AD| \neq |DA|$ these two estimates are different. We can establish

Theorem 7 *Assume that the* $n\times n$ *matrix* A *has the LDU decomposition* $A = L_1DU_1$, *where* L_1 *is unit lower triangular,* D *is diagonal and* U_1 *is unit upper triangular. Also assume that the perturbed matrix* $A+\delta_A$ *has the LDU decomposition* $A+\delta_A = (L_1+\delta_{L_1})(D+\delta_D)(U_1+\delta_{U_1})$, *where* $L_1+\delta_{L_1}$ *is unit lower triangular and* $U_1+\delta_{U_1}$ *is unit upper triangular. Finally assume that* $\max(\rho(\Gamma_{L_1}), \rho(\Gamma_{U_1})) < 1$ *holds with* $\Gamma_{L_1} = \left|L_1^{-1}\delta_A U_1^{-1}D^{-1}\right|$ *and* $\Gamma_{U_1} = \left|D^{-1}L_1^{-1}\delta_A U_1^{-1}\right|$. *Then the following inequalities are satisfied:*

$$|\delta_{L_1}| \leq |L_1|\,tril^*\left(\phi\left(\Gamma_{L_1}, \Gamma_{L_1}, 0\right)\right), \tag{3.9}$$

$$|\delta_D| \leq |D|\min\left\{diag\left(\phi\left(\Gamma_{L_1}, \Gamma_{L_1}, 1\right)\right), diag\left(\phi\left(\Gamma_{U_1}^T, \Gamma_{U_1}^T, 1\right)^T\right)\right\}, \tag{3.10}$$

$$|\delta_{U_1}| \leq triu^*\left(\phi\left(\Gamma_{U_1}^T, \Gamma_{U_1}^T, 0\right)^T\right)|U_1|. \tag{3.11}$$

Remark 8 *If ϕ is replaced by ϕ_0, we obtain the following weaker estimates:*

$$|\delta_{L_1}| \leq |L_1| \, tril^* \left((I - \Gamma_{L_1})^{-1} \Gamma_{L_1} \right), \tag{3.12}$$

$$|\delta_{U_1}| \leq triu^* \left(\Gamma_{U_1} (I - \Gamma_{U_1})^{-1} \right) |U_1|, \tag{3.13}$$

$$|\delta_D| \leq |D| \min \left(diag \left((I - \Gamma_{L_1})^{-1} \Gamma_{L_1} \right), diag \left(\Gamma_{U_1} (I - \Gamma_{U_1})^{-1} \right) \right). \tag{3.14}$$

Next we specialize the above result for symmetric and positive definite matrices. In such a case $\Gamma_{L_1} = \Gamma_{U_1}^T$ ($\Gamma_{L_1} = \left| L_1^{-1} \delta_A L_1^{-T} D^{-1} \right|$, $\Gamma_{U_1} = \left| D^{-1} L_1^{-1} \delta_A L_1^{-T} \right|$) and we have the following

Corollary 9 *Assume that A is symmetric and positive definite and its perturbation δ_A is such that $A + \delta_A$ remains symmetric and positive definite. If A and $A + \delta_A$ are written in the forms $A = L_1 D L_1^T$ ($D \geq 0$) and*

$$A + \delta_A = (L_1 + \delta_{L_1}) (D + \delta_D) \left(L_1^T + \delta_{L_1}^T \right),$$

respectively, then

$$|\delta_{L_1}| \leq |L_1| \, tril^* \left(\phi \left(\Gamma_{L_1}, \Gamma_{L_1}, 0 \right) \right) \tag{3.15}$$

and

$$|\delta_D| \leq D diag \left(\phi \left(\Gamma_{L_1}, \Gamma_{L_1}, 1 \right) \right). \tag{3.16}$$

Replacing ϕ by the weaker estimate ϕ_0, we obtain the following bounds:

$$|\delta_{L_1}| \leq |L_1| \, tril^* \left((I - \Gamma_{L_1})^{-1} \Gamma_{L_1} \right) \tag{3.17}$$

and

$$|\delta_D| \leq D diag \left((I - \Gamma_{L_1})^{-1} \Gamma_{L_1} \right). \tag{3.18}$$

We recall that Sun ([9], Thm. 3.1) for symmetric positive definite matrices proved that

$$|\delta_{L_1}| \leq |L_1| \, tril^* \left(E_{ld} \left(I - diag \left(D^{-1} E_{ld} \right) \right)^{-1} D^{-1} \right), \tag{3.19}$$

$$|\delta_D| \leq diag \left(E_{ld} \right) \tag{3.20}$$

with

$$E_{ld} = \left(I - \left| L_1^{-1} \delta_A L_1^{-T} \right| D^{-1} \right)^{-1} \left| L_1^{-1} \delta_A L_1^{-T} \right|. \tag{3.21}$$

We compare now estimates (3.17)-(3.18) and (3.19)-(3.20), respectively. We exploit the fact that for any diagonal matrix D, $|AD| = |A|\,|D|$ and $\mathrm{diag}(AD) = diag\,(A)\,D$ hold. We can write

$$(I-\Gamma_{L_1})^{-1}\,\Gamma_{L_1} = \left(I - \left|L_1^{-1}\delta_A L_1^{-T}\right| D^{-1}\right)^{-1}\left|L_1^{-1}\delta_A L_1^{-T}\right| D^{-1} = E_{ld}D^{-1}$$

and then estimate (3.18) yield

$$|\delta_D| \le diag\left(\left(I - \left|L_1^{-1}\delta_A L_1^{-T}\right| D^{-1}\right)^{-1}\left|L_1^{-1}\delta_A L_1^{-T}\right|\right) = diag\,(E_{ld})\,.$$

As $\left(I - diag\left(D^{-1}E_{ld}\right)\right)^{-1} \ge I$ and $E_{ld}\left(I - diag\left(D^{-1}E_{ld}\right)\right)^{-1} D^{-1} \ge E_{ld}D^{-1}$, the bound (3.19) satisfies

$$|L_1|\,tril^*\left(E_{ld}\left(I - diag\left(D^{-1}E_{ld}\right)\right)^{-1} D^{-1}\right) \ge |L_1|\,tril^*\left((I-\Gamma_{L_1})^{-1}\,\Gamma_{L_1}\right).$$

Thus it follows that Theorem 7 improves the special LDL^T perturbation result of Sun ([9], Thm. 3.1).

4. Final remarks

Computer experiments on symmetric positive definite MATLAB test matrices indicate that estimate ϕ_1 is often so good as ϕ itself. We could observe significant difference between the estimates if Γ_{L_1} was relatively large. A typical result is shown in Figure 4.1.

Here we display the maximum difference between the components of the bound and the true error matrix for Example 6.1 of [9] to which we added 20 random symmetric matrices with elements of the magnitude 5×10^{-3}. Hence, the line marked with $+$ denotes estimate (3.19) of Sun, the line with triangles denotes the estimate (3.17), the solid line denotes estimate ϕ_1, while the line with circles denotes the best estimate.

The estimates of Theorems 4 and 7 are optimal, if one accepts inequalities of the form $A \le C + Btriu\,(A,l)\quad (A,B,C \ge 0)$ in the estimation process. We can solve, however, the equation $A = C + Btriu\,(A,l)$ without any nonnegativity condition. Hence we can give exact expressions for the perturbation errors. For example, in case of Theorem 4 we can prove the following result.

Theorem 10 *For $k = 1,\ldots,n$ we have*

$$\delta_{L_1}e_k = \begin{bmatrix} 0 \\ -\left(L_2^{(k)}\left(L_1^{(k)}\right)^{-1}\delta_1^{(k)} - \delta_2^{(k)}\right)\left(L_1^{(k)}U_1^{(k)} + \delta_1^{(k)}\right)^{-1}\widetilde{e}_k \end{bmatrix}$$

and

$$e_k^T\delta_{U_1} = \left[0, -\widetilde{e}_k^T\left(L_1^{(k)}U_1^{(k)} + \delta_1^{(k)}\right)^{-1}\left(\delta_1^{(k)}\left(U_1^{(k)}\right)^{-1}U_2^{(k)} - \delta_4^{(k)}\right)\right],$$

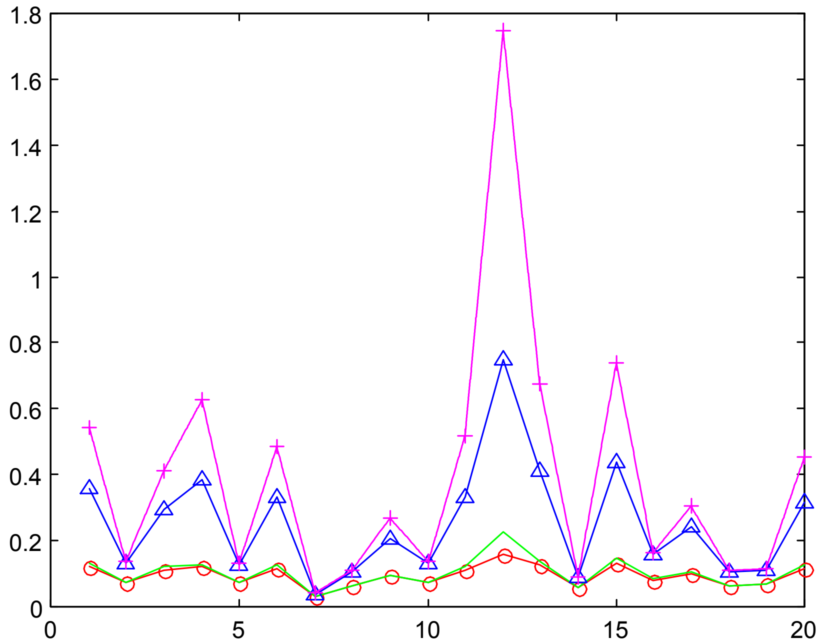

Figure 1. Perturbation bounds for the LDL^T factorization

where $\widetilde{e}_k \in R^k$ *is the kth unit vector,*

$$L_1 = \begin{bmatrix} L_1^{(k)} & 0 \\ L_2^{(k)} & L_3^{(k)} \end{bmatrix}, \qquad U = \begin{bmatrix} U_1^{(k)} & U_2^{(k)} \\ 0 & U_3^{(k)} \end{bmatrix}, \qquad \delta_A = \begin{bmatrix} \delta_1^{(k)} & \delta_4^{(k)} \\ \delta_2^{(k)} & \delta_3^{(k)} \end{bmatrix}$$

and $L_1^{(k)}, U_1^{(k)}, \delta_1^{(k)} \in R^{k\times k}$.

It does not seem easy to find componentwise estimates better than those of Theorem 4. We can obtain, however, better result than those of Chang and Paige [2].

Finally we remark that either from Theorem 4 or Theorem 7 we can easily obtain normwise perturbation estimates slightly weaker than those of Barrlund [1] by simply using the relation $\||A|\| = \|A\|_F$ and ϕ_0 instead of ϕ.

Acknowledgement: The author is grateful to Prof. E. Spedicato of the University of Bergamo for his helpful remarks and a supportive working environment. The author is also indebted to Prof. C. Broyden for pointing out reference [2].

REFERENCES

[1] Barrlund, A.: *Perturbation bounds for the LDL^H and LU decompositions*, BIT, **31**, (1991), 358-363.

[2] Chang, X.-W. and Paige, C.C.: *On the sensitivity of the LU factorization*, BIT, **38**, (1998), 486-501.

[3] Galántai A.: *Perturbation theory for full rank factorizations*, Quaderni DMSIA, University of Bergamo, 1999.

[4] Galántai A.: *Rank reduction and conjugation*, Mathematical Notes, Miskolc, **1**(1), (2000), 11-33.

[5] Higham, N.J.: *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, 1996

[6] Ortega, J.M. and Rheinboldt, W.C.: *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970

[7] Stewart, G.W.: *On the perturbation of LU, Cholesky and QR factorizations*, SIAM J. Matrix. Anal. Appl., **14**, (1993), 1141-1145.

[8] Stewart, G.W.: *On the perturbation of LU and Cholesky factors*, IMA J. Numer. Anal., **17**, (1997), 1-6.

[9] Sun, J.-G.: *Componentwise perturbation bounds for some matrix decompositions*, BIT, **32**, (1992), 702-714.

Doi: 10.18514/MMN.2000.25

Mathematical Notes, Miskolc, Vol. 1., No. 2., (2000), pp. 119–126

PARTITIONING THE N-SPACE INTO COLLINEAR SETS OF ORTHANTS

Mihály Hujter
Department of Applied Mathematics, University of Miskolc
3515 Miskolc – Egyetemváros, Hungary
mathm@gold.uni-miskolc.hu

[Received June 29, 2000]

Dedicated to my father on his seventieth birthday.

Abstract. How many n-orthants can be intersected in the n-dimensional Euclidean space simultaneously by a single straight line? In this sense of collinearity, how can we partition the entire space into collinear sets of n-orthants? How could it happen that a nice and innocent-looking conjecture is false in high dimensions? How are these questions related to the dominating sets in graph theory? What is the relationship to the illumination theory in intuitive geometry? We study such questions in this interdisciplinary paper.

Mathematical Subject Classification: 05C69, 52B11, 52C17, 52C35
Keywords: dominating sets in graphs, n-dimensional polytopes, packing and covering in n-dimensions, arrangements of flats

1. Introduction

We consider the n-dimensional Euclidean space, denoted R^n, for $n = 2, 3, \dots$. The elements of such a space are the n-dimensional columns vectors x with real number components $x_1, \dots, x_n \in R$. Another name of such a vector is *point*. By a *sign vector* we mean such a vector $s \in R^n$ for which each $|s_i| = 1$. In other words, the set of the sign vectors is $\{-1, +1\}^n$. Obviously, the number of different sign vectors is 2^n. If s is a sign vector, we associate with s the following set, called *open n-orthant*: $\{x \in R^n : x_1 s_1 > 0, \dots, x_n s_n > 0\}$. Note that the 2-orthants are known as *quadrants*, and the 2-quadrant associated to $\binom{+1}{+1}$ was named the *first quadrant* by Monge when he invented *descriptive geometry* and pioneered the development of *analytical geometry* two centuries ago. Obviously, the open n-orthants are pairwise disjoint open convex sets in n dimensions, and their number is 2^n. Each such open n-orthant contains exactly one sign vector, the one that the open n-orthant is associated with.

By a *straight line in general position* we mean a set of vectors $x \in R^n$ in the equation form $x = a + tv$ where t runs in R and $a, v \in R^n$ are given fixed vectors such that each $a_i \neq 0$ and each $v_i \neq 0$, furthermore for $i = 1, 2, \dots, n$ the unique solutions $t_i \in R$ of the equations $0 = a_i + t_i v_i$ are all distinct. Geometrically speaking, a straight line in general position is such a line which is neither parallel to the *coordinate hyperplanes* (these are the n hyperplanes with equations $x_i = 0$, where i is one of $1, 2, ..., n$), nor goes through two coordinate hyperplanes at the same point. For example, in 2 dimensions the straight lines in general position are those whose equations are $\alpha_1 x_1 + \alpha_2 x_2 = \beta$ for $\alpha_1, \alpha_2, \beta \in R$, $\alpha_1\alpha_2\beta \neq 0$.

Obviously, in n dimensions almost all straight lines are in general position. If a straight line is given in the equation form $x = a + tv$ where t runs in R and $a, v \in R^n$ are fixed, then obviously, at least one component of v is nonzero. We can change all components a_i and v_i a little bit for $i = 1, ..., n$ such a way that the new values all become nonzeros and the t_i numbers determined by the equations $0 = a_i + t_i v_i$ all become distinct. In this sense any straight line is a *"point of accumulation"* to the set of the straight lines in general position. It is also obvious, that if an arbitrary straight line *meets* (i.e., *intersects*) an arbitrary open n-orthant, then by changing the a_i and v_i components just a little bit we gain a straight line in general position that meets the same open n-orthant. In other words, for any fixed n-orthant any straight line (not necessarily in general position) which meets the n-orthant in question is a "point of accumulation" to the set of straight lines which are in general position and meet the n-orthant in question. As a corollary, if each open n-orthant in a subset of the set of all open n-orthants is met simultaneously by an arbitrary straight line, then a straight line in general position can also be found to meet each element of the subset.

A straightforward question is this: Given a straight line in general position, how many open n-orthants can it intersect? Equivalently, how many n-orthants can be shot down with one shot from a point a into direction v? "How many birds can be killed with one stone?" In 2 dimensions the answer is easy because any straight line in general position meets all but one open quadrants. If, for example, a is in the first open quadrant, and v is, for example, in the second open quadrant (the open quadrant associated with $\binom{+1}{-1}$), then the straight line in the form $x = a+tv$ is going to intersect the first, the second and the fourth open quadrants (these are the 2-orthants associated with $\binom{+1}{+1}$, $\binom{+1}{-1}$, and $\binom{-1}{+1}$).

In the next section we will answer the above question in n dimensions. Our Proposition 1 will state that any straight line in general position meets the same number of open n-orthants, and this number is exactly $n+1$. From the above argument and from the proof of this proposition it easily follows that in case of a straight line *not* in general position this number must be smaller, moreover the set of those open n-orthants which are met by a straight line not in general position is always a proper subset of a straight line in general position.

In the above discussed sense we can talk about *collinear* sets of open n-orthants. Such a set of open n-orthants is collinear whose elements can be met simultaneously in a single straight line in general position. It is clear that one or two open n-orthants are always collinear. (The empty set is also collinear.) If a subset of the 2^n open n-orthants is not collinear, we say that the subset is *noncollinear*.

The collinear sets of n-orthants have a simple structure if $n = 2$. Namely, any set of at most 3 open quadrants is collinear, but the set of all 4 open quadrants is not. In 3 dimensions the situation is a little bit more complicated. By inspection one can see that the set of all 8 open 3-orthants can be partitioned into 2 collinear subsets, and there are 12 such partitions. This way we gain 12 pairs of collinear 4-element subsets, and all collinear 4-element subsets are among them. However, the total number of 4-element subsets is $\binom{8}{4} = 70$, and so the number of the noncollinear 4-element subsets is 46. Even a 3-element subset is not necessarily collinear; the number of the 3-element subsets is $\binom{8}{3} = 56$, and there are 8 noncollinear 3-element subsets. An example of such a noncollinear 3-element subset is the one whose elements are associated to those sign vectors in 3 dimensions which have exactly one -1 component. Those subsets which have more than 4 elements are all noncollinear. It is worth mentioning that in 2 dimensions the collinear subsets of the n-orthants form a *matroid* but in higher dimensions they do not.

2. The maximum number of collinear n-orthants

In this section we study the structure of the inclusion-maximal collinear sets of open n-orthants.

Proposition 1. *A straight line in general position in R^n intersects exactly $n+1$ open n-orthants.*

Proof. We prove by induction on n. The case $n = 2$ is obvious by inspection. Assume that $n \geq 3$, and that for smaller dimensions the statement has already been proved. Consider, in n dimensions, a straight line in general position given in the form of the equation $x = a + tv$. Consider the largest t_i among the t_i numbers which were used in the definition of the straight lines in general position. Without loss of generality we may assume that the largest one among the distinct t_i values is t_n. Let $\varepsilon_n > 0$ be chosen such that all $t_i < t_n - \varepsilon_n$, $i = 1, \ldots, n-1$. Now observe that for $x' = (x_1, ..., x_{n-1})^T$, $a' = (a_1, ..., a_{n-1})^T$, $v' = (v_1, ..., v_{n-1})^T$ the equation $x' = a' + tv'$ defines a straight line in general position in $n-1$ dimensions. By the induction assumption, this line intersects exactly n open $(n-1)$-orthants in $n-1$ dimensions. Since t_n was the largest among the distinct t_i values, in n dimensions the straight line with equation $x = a + tv$ also intersects n open n-orthants for $t < t_n$. Note that the last components all have the same sign in these open n-orthants. For $t = t_n$ the point $x = a + tv$ is on the nth coordinate hyperplane. However, for $t > t_n$, there is only one open n-orthant in which all $a+tv$ points are, and this open n-orthant is associated with the sign vector $(s_1, ..., s_{n-1}, -s_n)^T$ where the sign vector associated with the point $a+(t_n-\varepsilon_n)v$ is $(s_1, ..., s_{n-1}, s_n)^T$. So the entire straight line $x = a+tv$ intersects exactly $n+1$ open n-orthants. □

3. Hypercubes and the associated graphs

By the *n-dimensional hypercube* in standard position we mean the convex hull of the set $\{-1, +1\}^n \subseteq R^n$. The vertices of the cube are the sign vectors, i.e. the elements of $\{-1, +1\}^n$. Given two such vertices, their *Hamming distance* is the number of components where the two vertices, as sign vectors, are different. So the famous *Manhattan distance* is exactly 2^n times larger that the Hamming distance. The *graph* of the hypercube has the same vertex set as the hypercube, and the graph edges are those pairs of vertices whose Hamming distance is exactly one. So the geometrical vertices and the graph theoretical vertices coincide. Similarly, the geometrical hypercube edges and the graph edges also coincide. So the number of vertices in the graph is 2^n, and the number of edges is $n2^{n-1}$. We *color* the edges of the graph with colors $1, 2, \ldots, n$ as follows: an edge gets color i if the vertices of the edge are different in the ith component and only in that one. So for each color there are exactly 2^{n-1} edges. Two edges, as line segments with endpoints at the vertices of the edge, are parallel (in the usual geometrical sense) if and only if they have the same color; otherwise the two edges are geometrically perpendicular to each other.

In Figure 1 we show the hypercube graph for $n = 5$. The black dots represent the 32 vertices. The 16 horizontal lines represent the edges of color 1. The 16 vertical lines represent the edges of color 2. The edges represented by the lines on the diagonals of the big square are all of color 3. There are 8 such edges. However, there are 8 more edges of color 3. Those are all among the sides of the central 20-gon. The tangents of these 8 lines are $+2$, -2, $+1/2$, or $-1/2$. The remaining 32 edges form 8 deltoids. There are 2 such deltoids in each corner of the figure. The edges of these deltoids are alternatively of color 4 and color 5 in such a way that the almost vertical edges are all of color 4. (The expression "almost vertical" is understood to mean that the tangent of such a line is at least 2 in absolute value. Similarly, the "almost horizontal" lines are those, whose tangents are between 0 and 1/2.

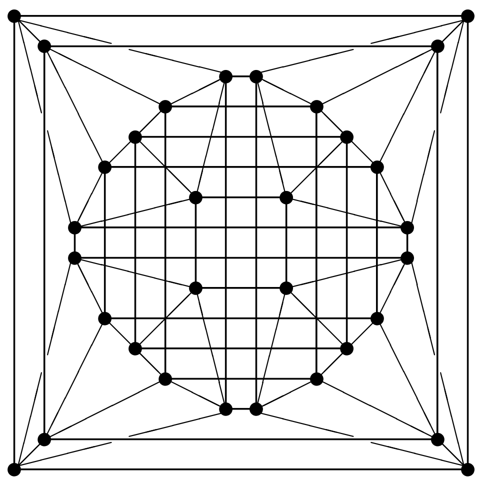

Figure 1. The graph of the 5-dimensional hypercube

We will use some usual notions of graph theory such as a *star* or a *tree*. By a *super* *tree* we mean such a tree, as an induced subgraph in the n-dimensional hypercube graph, that has exactly n edges with n different colors. Such a super tree has $n+1$ vertices. By a *super star* we mean such a super tree which is a star itself. In other words, a super star contains a vertex, called *central vertex*, and all those vertices, that are neighbors of the central vertex in the graph.

For example, for $n = 2$ there are four super trees in the 4 vertex hypercube graph. The vertex sets of these trees can be listed as follows: $\{\binom{-1}{+1}, \binom{+1}{+1}, \binom{+1}{-1}\}$, $\{\binom{+1}{+1}, \binom{+1}{-1}, \binom{-1}{-1}\}$, $\{\binom{+1}{-1}, \binom{-1}{-1}, \binom{-1}{+1}\}$, $\{\binom{-1}{-1}, \binom{-1}{+1}, \binom{+1}{+1}\}$. All these super trees are super stars.

For $n \geq 3$, not all super trees are super stars. For example, for $n = 3$ we have 8 super stars and 12 super trees that are *path graphs* in the graph theoretical sense. (A super tree is a *path* if exactly $n-1$ vertices have 2 neighbors.) The larger the value of n the smaller the percentage the super stars represent among the super trees.

4. The main result

Our main theorem can be formulated as follows:

Theorem 2. *Consider an arbitrary path, with edges of different colors, in the n-dimensional hypercube graph. There exists such a straight line in general position that intersects all n-orthants that are associated to the vertices of the path.*

Proof. We prove by induction on n. Without loss of generality we may assume that the given path has n edges (or equivalently, it has $n+1$ vertices) because in any other case we can add extra vertices and edges. Actually we will prove a stronger theorem. Given a straight line in general position in R^n in the form $x = a + tv$, consider the distinct real numbers t_i, $i = 1, 2, ..., n$, as above. Let $t_{\min}$ and $t_{\max}$ denote the smallest and largest t_i, respectively. We claim that the straight line in general position may have such parameters for which $a + (t_{\min} - 1)v$ and $a + (t_{\max} + 1)v$ are in the open n-orthants associated with the endvertices of the path. Moreover, we can prescribe that which endvertex should corresponds $t_{\max}$.

The case $n = 2$ is easy by inspection. We may assume that $n \geq 3$ and that the statement has already been proved in smaller dimensions. Consider one endvertex of the given path. Let it be denoted as P_n. Let P_{n-1} denote its neighbor in the path. Without loss of generality we may assume that the edge $\{P_n, P_{n-1}\}$ is of color n. Also, without loss of generality we may assume that $P_n = (+1, +1, ..., +1)^T$ as a sign vector. In other words, P_n is the vertex which in the first open n-orthant. Now the other vertices of the graph all have -1 in their last coordinate because color n can never occur among the colors of the path edges. Similarly to the proof of Proposition 1 we can decrease the dimension number by one by cutting off the nth coordinates. By the induction hypothesis we gain that there exists at least one straight line in general position in R^{n-1} which line intersects all open $(n-1)$-orthants that are associated with the first n vertices of our path. Consider such a straight line in the equation form $x' = a' + tv'$ where $x' = (x_1, ..., x_{n-1})^T$, $a' = (a_1, ..., a_{n-1})^T$, $v' = (v_1, ..., v_{n-1})^T$. By the induction hypothesis we may assume that P_{n-1} is in the same $(n-1)$-orthant as $a' + (t_{\max} + 1)v'$. Now let us fix $\tau_n \in R$ arbitrarily such that $\tau_n > t_{\max} + 2$ and let $v_n = 1$, $a_n = -\tau_n$, $a = (a_1, ..., a_{n-1}, a_n)^T$, $v = (v_1, ..., v_{n-1}, v_n)^T$. We can choose such a value of t_n for which the n-dimensional straight line with equation $x = a + tv$ will be in general position, and for this straight line all requirements are fulfilled. Here t_n will be τ_n, and thus for $t > t_n$ the points $x + tv$ all will be in the open n-orthant associated with P_n. At the same time for $t < t_n$, $t \neq t_i$, $i = 1, 2, ..., n-1$, the points $x + tv$ will be in the open n-orthants associated with the vertex $P \neq P_n$ of the n-dimensional path in question. □

5. Killing all birds with as few stones as possible

Consider the open n-orthants in n dimensions to be birds in the air. We have to kill all birds. A straight line can be considered to be the orbit of a stone. We have shown that one stone can kill $n + 1$ birds. How many stones are needed to kill all birds?

Problem 3. *How many straight lines in general position can together meet all open n-orthants.*

Let $\kappa(n)$ denote the minimum number of straight lines in general position that can together meet all open n-orthants. In a more general form and by using different terminology Gy. Kiss conjectured the following in [2].

Conjecture 4. (Gy. Kiss [2]) *The vertex set of the n-dimensional hypercube graph can be covered by using the vertex sets of $\kappa(n)$ super stars.*

Here we tell some reasons to convince the reader that this conjecture is a reasonable one. From the results of Kiss [2] it easily follows that there is bijection from the vertex set of the n-dimensional hypercube to the set of all open n-orthant with the following property: The image of the vertex set of a super star is an $(n + 1)$-element collinear set of open n-orthants. On the other hand, from the proofs of our Proposition 1 and Theorem 2 we can see that there is a natural correspondence between the superstars and the paths formed by the sign vectors of the $(n + 1)$-element collinear sets of open n-orthants. Namely, let P an endvertex of the path, and consider such the uniquely determined superstar whose 2 vertices are P and its neighbor in the path in question. So the central vertex of the super star is the only neighbor of P in the path. Furthermore, for any fixed pair of such a path and such a super star, the bijection of Kiss can be chosen such that the image of this super star is exactly the set of those open n-orthants who are associated to the vertices of the path in question. Interestingly enough, those pairs of edges which corresponds to each other, always have the same color in the hypercube graph.

Another fact to convince the reader is this: We can easily check that the conjecture of Kiss is valid for $n = 2, 3$. In both cases the entire set of all open n-orthants can easily be partitioned into 2 collinear subsets. On the other hand, the two super stars with central vertices s and $-s$, where each component of s is $+1$, obviously cover the entire vertex set of the hypercube graph.

The conjecture is also valid for $n = 4$. One can easily see that the minumum number of super stars is the required covering which is nothing else but the famous *dominating number* of the n-dimensional hypercube graph. (Cf. [1] for the importance of the dominating number). It is an easy exercise to check that the dominating number of the 4-dimensional hypercube graph is 4. The total number of the vertices is 16, and each super star has 5 vertices. So we need at least 4 super stars. Here are the 4 central vertices of a feasible super star covering: $(+1,+1,+1,-1)^T$, $(-1,-1,-1,-1)^T$, $(+1,+1,+1,+1)^T$, $(-1,-1,-1,+1)^T$. On the other hand, the following tables show the column vectors of three sets of collinear open 4-orthants (more precisely, the associated sign vectors):

+1	+1	+1	+1	−1
+1	+1	+1	−1	−1
+1	+1	−1	−1	−1
+1	−1	−1	−1	−1

−1	−1	−1	−1	+1
+1	+1	+1	−1	−1
+1	+1	−1	−1	−1
−1	+1	+1	+1	+1

−1	−1	−1	−1
−1	−1	−1	−1
−1	−1	−1	−1
−1	−1	−1	−1

Only 2 sign vectors are missing, but we know that a 2-element set of open n-orthants is always collinear. This completes the proof of Conjecture 4 for $n = 4$.

Interestingly enough, Conjecture 4 is not valid for $n = 5$.

Theorem 5. *We have $\kappa(5) = 6$, and the vertices of the 5-dimensional hypercube cannot be covered by the vertex sets of 6 superstars.*

Proof. Here are the vertices of 6 paths in the 5-dimensional hypercube (only the signs are showed):

−1	+1	+1	+1	+1	+1
+1	+1	+1	−1	−1	−1
+1	+1	−1	−1	−1	−1
+1	+1	+1	+1	+1	−1
−1	−1	−1	−1	+1	+1

+1	−1	−1	−1	−1	−1
−1	−1	−1	−1	−1	+1
+1	+1	+1	−1	−1	−1
+1	+1	−1	−1	−1	−1
−1	−1	−1	−1	+1	+1

−1	−1	−1	+1	+1	+1
+1	+1	+1	+1	−1	−1
−1	−1	+1	+1	+1	+1
+1	+1	+1	+1	+1	−1
−1	+1	+1	+1	+1	+1

−1	−1	−1	−1	−1	+1
−1	−1	−1	−1	+1	+1
−1	−1	+1	+1	+1	+1
+1	+1	+1	−1	−1	−1
−1	+1	+1	+1	+1	+1

+1	+1	+1	+1	+1
−1	+1	+1	+1	+1
+1	+1	−1	−1	−1
−1	−1	−1	−1	+1
−1	−1	−1	+1	+1

+1	−1	−1	−1
−1	−1	+1	+1
−1	−1	−1	+1
−1	−1	−1	−1
−1	−1	−1	−1

On the other hand, it is well-known in graph theory that the so-called *dominating number* of the 5-dimension hypercube graph is larger than 6. This fact can be formulated simply as follows: One cannot take 6 vertices from the graph of Figure 1 such that any other vertex is the neighbor of at least one chosen vertex. By using the symmetry of Figure 1 we can easily check that this statement is true. □

6. Concluding remarks

We have proved Conjecture 4 for $n = 2, 3, 4$, and we have disproved it for $n = 5$. We have not studied the conjecture in the cases $n = 6, 7, \ldots$. Kiss [2] proved such a result which implies that Conjecture 4 is valid for infinite different values of n.

Acknowledgement: The support provided by the Hungarian National Research Foundation (grant T026575) and by the the "FKFP" Research Fund in Hungary, grant 0486/1999, is gratefully acknowledged.

REFERENCES

1. FOULDS, L. R.: *Graph Theory Application*, Springer, 1992.
2. KISS, GY.: *Illumination problems and codes*, Periodica Mathematica Hungarica, **39**(1–3), (1999), 65–71..

Doi: 10.18514/MMN.2000.26

Mathematical Notes, Miskolc, Vol. 1., No. 2., (2000), pp. 127–133

COMPUTING VOLUME OF SOLIDS BOUNDED BY BÉZIER SURFACES

Imre Juhász
Department of Descriptive Geometry, University of Miskolc
3515 Miskolc – Egyetemváros, Hungary
agtji@gold.uni-miskolc.hu

[Received June 22, 2000]

Abstract. We provide an optimized closed formula for the volume of the solid, bounded by a Bézier surface and the cones determined by its boundary curves and the origin. The obtained formula is easy to implement in programming languages, since it contains only basic arithmetic operations.

Mathematical Subject Classification: 51M25, 68U07
Keywords: Bézier surface, volume

1. Introduction

Bézier curves and surfaces play a significant role in Computer Aided Geometric Design (CAGD), therefore their specification, modification and different properties are of interest. Specification and modification of Bézier objects are well explored fields of CAGD but there are some special problems to be solved, such as to find the optimal way of computing the volume of solids bounded by Bézier surfaces. Following an elementary approach, we derive an exact closed formula which can substitute approximating numerical procedures. This task can be considered as a spatial generalization of the problem of computing the signed area of a plane figure bounded by Bézier curves, cf. [3].

In this paper we consider Bézier surfaces of the following definition:

Definition 1 *The tensor product surface defined by the equation*

$$\mathbf{b}\left(u,v\right)=\sum_{i=0}^{n}\sum_{j=0}^{m}\mathbf{b}_{ij}B_{i}^{n}\left(u\right)B_{j}^{m}\left(v\right),u\in\left[0,1\right],v\in\left[0,1\right] \tag{1.1}$$

is called Bézier surface, where $\mathbf{b}_{ij}\in\mathbb{R}^{3}$ *are the control points of the surface and* $B_{i}^{n}\left(u\right)=\left(\begin{array}{c}n\\ i\end{array}\right)u^{i}\left(1-u\right)^{n-i}$, $(i=0,\dots,n)$ *is the* i^{th} *Bernstein polynomial of degree* n.

2. Volume calculation

Theorem 2 *The signed volume of the solid bounded by the Bézier surface (1.1) and the cones determined by its boundary curves and the origin is*

$$V = \frac{1}{3\,(3n-1)\,(3m-1)} \sum_{r=0}^{z-3} \sum_{s=r+1}^{z-2} \sum_{t=s+1}^{z-1} \beta_{ijklpq}^{mn} \left(\mathbf{b}_{ij}, \mathbf{b}_{kl}, \mathbf{b}_{pq}\right)$$

where $(\mathbf{b}_{ij}, \mathbf{b}_{kl}, \mathbf{b}_{pq})$ *denotes the mixed product of the specified vectors,* $z = (n+1)\,(m+1)$,

$$\begin{aligned} j &= r \bmod (m+1)\,,\ i = (r-j)\,/\,(m+1)\,, \\ l &= s \bmod (m+1)\,,\ k = (s-l)\,/\,(m+1)\,, \\ q &= t \bmod (m+1)\,,\ p = (t-q)\,/\,(m+1)\,, \end{aligned}$$

and

$$\beta_{ijklpq}^{mn} = \frac{l\,(i-p) + q\,(k-i) + j\,(p-k)}{\binom{3n-2}{i+k+p-1}\binom{3m-2}{j+l+q-1}} \binom{n}{i}\binom{n}{k}\binom{n}{p}\binom{m}{j}\binom{m}{l}\binom{m}{q}.$$

Proof. First we determine the signed volume V of the solid bounded by the surface $\mathbf{s}(u,v)\,, u \in [u_0, u_1]\,,\ v \in [v_0, v_1]$ and the cones determined by its boundary curves and the origin. For this purpose we divide the interval $[u_0, u_1]$ into c, and $[v_0, v_1]$ into d equal parts. The corresponding parameter lines divide the surface into rectangular patches. We approximate the volume of the solid determined by such a patch and the origin with tetrahedra of Figure 1. Their volumes are

$$\begin{aligned} V_{ij} &= \frac{1}{6}\left(\mathbf{s}(u_{i+1}, v_j) - \mathbf{s}(u_i, v_j)\,, \mathbf{s}(u_i, v_{j+1}) - \mathbf{s}(u_i, v_j)\,, \mathbf{s}(u_i, v_j)\right) \text{ and} \\ V^{ij} &= \frac{1}{6}\left(\mathbf{s}(u_{i+1}, v_{j+1}) - \mathbf{s}(u_{i+1}, v_j)\,, \mathbf{s}(u_{i+1}, v_{j+1}) - \mathbf{s}(u_i, v_{j+1})\,, \mathbf{s}(u_{i+1}, v_{j+1})\right). \end{aligned}$$

Denoting the sum of the appropriate tetrahedra by V_{cd} and V^{cd} we obtain

$$V_{cd} = \sum_{i=0}^{c-1}\sum_{j=0}^{d-1} V_{ij},\ V^{cd} = \sum_{i=0}^{c-1}\sum_{j=0}^{d-1} V^{ij} \text{ and } \lim_{\substack{c\to\infty \\ d\to\infty}} \left(V_{cd} + V^{cd}\right) = V,$$

moreover, the equalities

$$\lim_{\substack{c\to\infty \\ d\to\infty}} V_{cd} = \lim_{\substack{c\to\infty \\ d\to\infty}} V^{cd} = \frac{1}{6}\int_{u_0}^{u_1}\int_{v_0}^{v_1} \left(\frac{\partial}{\partial u}\mathbf{s}(u,v)\,, \frac{\partial}{\partial v}\mathbf{s}(u,v)\,, \mathbf{s}(u,v)\right) dudv$$

hold, thus

$$V = \frac{1}{3}\int_{u_0}^{u_1}\int_{v_0}^{v_1}\left(\frac{\partial}{\partial u}\mathbf{s}\left(u,v\right),\frac{\partial}{\partial v}\mathbf{s}\left(u,v\right),\mathbf{s}\left(u,v\right)\right)dudv.$$

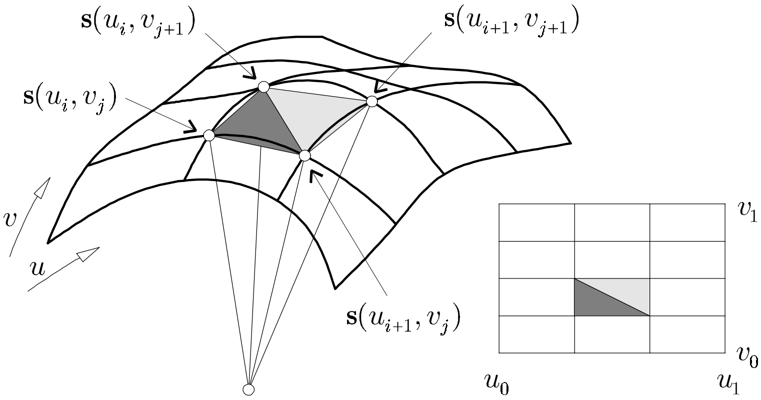

Figure 1. The approximating tetrahedra

The partial derivatives of the Bézier surface of Equation (1.1) are

$$\begin{aligned}\frac{\partial}{\partial u}\mathbf{b}\left(u,v\right) &= n\sum_{i=0}^{n}\sum_{j=0}^{m}\mathbf{b}_{ij}\left(B_{i-1}^{n-1}\left(u\right)-B_{i}^{n-1}\left(u\right)\right)B_{j}^{m}\left(v\right),\\ \frac{\partial}{\partial v}\mathbf{b}\left(u,v\right) &= m\sum_{k=0}^{n}\sum_{l=0}^{m}\mathbf{b}_{kl}B_{k}^{n}\left(u\right)\left(B_{l-1}^{m-1}\left(v\right)-B_{l}^{m-1}\left(v\right)\right)\end{aligned}$$

and the required volume is

$$\begin{aligned}V = \frac{nm}{3}\int_{0}^{1}\int_{0}^{1}&\left(\sum_{i=0}^{n}\sum_{j=0}^{m}\mathbf{b}_{ij}\left(B_{i-1}^{n-1}\left(u\right)-B_{i}^{n-1}\left(u\right)\right)B_{j}^{m}\left(v\right),\right.\\ &\left.\sum_{k=0}^{n}\sum_{l=0}^{m}\mathbf{b}_{kl}B_{k}^{n}\left(u\right)\left(B_{l-1}^{m-1}\left(v\right)-B_{l}^{m-1}\left(v\right)\right),\sum_{p=0}^{n}\sum_{q=0}^{m}\mathbf{b}_{pq}B_{p}^{n}\left(u\right)B_{q}^{m}\left(v\right)\right)dudv\\ = \frac{nm}{3}&\sum_{i=0}^{n}\sum_{j=0}^{m}\sum_{k=0}^{n}\sum_{l=0}^{m}\sum_{p=0}^{n}\sum_{q=0}^{m}\left(\mathbf{b}_{ij},\mathbf{b}_{kl},\mathbf{b}_{pq}\right)\int_{0}^{1}\int_{0}^{1}\left(B_{i-1}^{n-1}\left(u\right)-B_{i}^{n-1}\left(u\right)\right)\\ &B_{k}^{n}\left(u\right)B_{p}^{n}\left(u\right)\left(B_{l-1}^{m-1}\left(v\right)-B_{l}^{m-1}\left(v\right)\right)B_{j}^{m}\left(v\right)B_{q}^{m}\left(v\right)dudv.\end{aligned}$$

The double integrals of the sum have a closed form but the sum contains many superfluous terms, namely $((n+1)(m+1))^3$. Taking into consideration the equalities

$$\begin{aligned}(\mathbf{b}_{ij},\mathbf{b}_{kl},\mathbf{b}_{pq}) &= (\mathbf{b}_{kl},\mathbf{b}_{pq},\mathbf{b}_{ij}) = (\mathbf{b}_{pq},\mathbf{b}_{ij},\mathbf{b}_{kl}) = \\ -(\mathbf{b}_{kl},\mathbf{b}_{ij},\mathbf{b}_{pq}) &= -(\mathbf{b}_{pq},\mathbf{b}_{kl},\mathbf{b}_{ij}) = -(\mathbf{b}_{ij},\mathbf{b}_{pq},\mathbf{b}_{kl})\end{aligned}$$

and the property that a mixed product is zero if at least two of its operands are equal, we can reduce the number of terms in the sum to $\binom{(n+1)(m+1)}{3}$.

This sum consists of the terms

$$(\mathbf{b}_{ij},\mathbf{b}_{kl},\mathbf{b}_{pq})\int_0^1\int_0^1 (A_1(u,v)+A_2(u,v)+A_3(u,v)- A_4(u,v)-A_5(u,v)-A_6(u,v))\,dudv$$

where

$$\begin{aligned}
A_1(u,v) &= \left(B_{i-1}^{n-1}(u)-B_i^{n-1}(u)\right)B_j^m(v)\left(B_{l-1}^{m-1}(v)-B_l^{m-1}(v)\right) B_k^n(u)\,B_p^n(u)\,B_q^m(v),\\
A_2(u,v) &= \left(B_{k-1}^{n-1}(u)-B_k^{n-1}(u)\right)B_l^m(v)\left(B_{q-1}^{m-1}(v)-B_q^{m-1}(v)\right) B_p^n(u)\,B_i^n(u)\,B_j^m(v),\\
A_3(u,v) &= \left(B_{p-1}^{n-1}(u)-B_p^{n-1}(u)\right)B_q^m(v)\left(B_{j-1}^{m-1}(v)-B_j^{m-1}(v)\right) B_i^n(u)\,B_k^n(u)\,B_l^m(v),\\
A_4(u,v) &= \left(B_{k-1}^{n-1}(u)-B_k^{n-1}(u)\right)B_l^m(v)\left(B_{j-1}^{m-1}(v)-B_j^{m-1}(v)\right) B_i^n(u)\,B_p^n(u)\,B_q^m(v),\\
A_5(u,v) &= \left(B_{p-1}^{n-1}(u)-B_p^{n-1}(u)\right)B_q^m(v)\left(B_{l-1}^{m-1}(v)-B_l^{m-1}(v)\right) B_k^n(u)\,B_i^n(u)\,B_j^m(v),\\
A_6(u,v) &= \left(B_{i-1}^{n-1}(u)-B_i^{n-1}(u)\right)B_j^m(v)\left(B_{q-1}^{m-1}(v)-B_q^{m-1}(v)\right) B_p^n(u)\,B_k^n(u)\,B_l^m(v).
\end{aligned}$$

After some simplification we gain

$$A_1(u,v)-A_6(u,v) = \left(B_{i-1}^{n-1}(u)-B_i^{n-1}(u)\right)B_k^n(u)\,B_p^n(u) \left(B_{l-1}^{m-1}(v)\,B_q^{m-1}(v)-B_{q-1}^{m-1}(v)\,B_l^{m-1}(v)\right)B_j^m(v) =$$

$$\binom{n}{k}\binom{n}{p}\left(\frac{\binom{n-1}{i-1}}{\binom{3n-1}{i+k+p-1}}B_{i+k+p-1}^{3n-1}(u)-\frac{\binom{n-1}{i}}{\binom{3n-1}{i+k+p}}B_{i+k+p}^{3n-1}(u)\right)$$

$$\frac{\binom{m}{j}\left(\binom{m-1}{l-1}\binom{m-1}{q}-\binom{m-1}{l}\binom{m-1}{q-1}\right)}{\binom{3m-2}{j+l+q-1}}B_{j+l+q-1}^{3m-2}(v).$$

Analogously we can obtain

$$A_2(u,v) - A_4(u,v) =$$
$$\binom{n}{i}\binom{n}{p}\left(\frac{\binom{n-1}{k-1}}{\binom{3n-1}{i+k+p-1}}B^{3n-1}_{i+k+p-1}(u) - \frac{\binom{n-1}{k}}{\binom{3n-1}{i+k+p}}B^{3n-1}_{i+k+p}(u)\right)$$
$$\frac{\binom{m}{l}\left(\binom{m-1}{q-1}\binom{m-1}{j} - \binom{m-1}{j-1}\binom{m-1}{q}\right)}{\binom{3m-2}{j+l+q-1}}B^{3m-2}_{j+l+q-1}(v)\,,$$

and

$$A_3(u,v) - A_5(u,v) =$$
$$\binom{n}{i}\binom{n}{k}\left(\frac{\binom{n-1}{p-1}}{\binom{3n-1}{i+k+p-1}}B^{3n-1}_{i+k+p-1}(u) - \frac{\binom{n-1}{p}}{\binom{3n-1}{i+k+p}}B^{3n-1}_{i+k+p}(u)\right)$$
$$\frac{\binom{m}{q}\left(\binom{m-1}{j-1}\binom{m-1}{l} - \binom{m-1}{j}\binom{m-1}{l-1}\right)}{\binom{3m-2}{j+l+q-1}}B^{3m-2}_{j+l+q-1}(v)\,.$$

Utilizing that $\int_0^1 B_i^n(u)\,du = 1/(n+1), (i = 0, 1, \ldots, n)$ cf. [2] we gain

$$\int_0^1\int_0^1 A_1(u,v) - A_6(u,v)\,du dv = \frac{1}{3n(3m-1)}\frac{1}{\binom{3m-2}{j+l+q-1}}$$
$$\binom{n}{k}\binom{n}{p}\left(\frac{\binom{n-1}{i-1}}{\binom{3n-1}{i+k+p-1}} - \frac{\binom{n-1}{i}}{\binom{3n-1}{i+k+p}}\right)$$
$$\binom{m}{j}\left(\binom{m-1}{l-1}\binom{m-1}{q} - \binom{m-1}{l}\binom{m-1}{q-1}\right).$$

Simplifying the binomials above (cf. [4]) we obtain

$$\int_0^1\int_0^1 A_1(u,v) - A_6(u,v)\,dudv = \frac{(l-q)(2i-k-p)}{3n(3m-1)m(3n-1)\binom{3m-2}{j+l+q-1}\binom{3n-2}{i+k+p-1}} \binom{n}{i}\binom{n}{k}\binom{n}{p}\binom{m}{j}\binom{m}{l}\binom{m}{q}. \tag{2.1}$$

The integral of the differences $A_2(u,v) - A_4(u,v)$ and $A_3(u,v) - A_5(u,v)$ can analogously be expressed and we obtain the equations

$$\int_0^1\int_0^1 A_2(u,v) - A_4(u,v)\,dudv = \frac{(q-j)(2k-i-p)}{3n(3m-1)m(3n-1)\binom{3m-2}{j+l+q-1}\binom{3n-2}{i+k+p-1}} \binom{n}{i}\binom{n}{k}\binom{n}{p}\binom{m}{j}\binom{m}{l}\binom{m}{q} \tag{2.2}$$

and

$$\int_0^1\int_0^1 A_3(u,v) - A_5(u,v)\,dudv = \frac{(j-l)(2p-i-k)}{3n(3m-1)m(3n-1)\binom{3m-2}{j+l+q-1}\binom{3n-2}{i+k+p-1}} \binom{n}{i}\binom{n}{k}\binom{n}{p}\binom{m}{j}\binom{m}{l}\binom{m}{q}. \tag{2.3}$$

The sum of the integrals (2.1), (2.2) and (2.3) is

$$\frac{l(i-p) + q(k-i) + j(p-k)}{n(3n-1)m(3m-1)\binom{3n-2}{i+k+p-1}\binom{3m-2}{j+l+q-1}} \binom{n}{i}\binom{n}{k}\binom{n}{p}\binom{m}{j}\binom{m}{l}\binom{m}{q}.$$

The choice of the point triplets $\mathbf{b}_{ij}, \mathbf{b}_{kl}, \mathbf{b}_{pq}$ is still left. For this we transform the matrix with elements $\mathbf{b}_{ij}$, $(i = 0, 1, \ldots, n;\ j = 0, 1, \ldots, m)$ to a vector $\mathbf{c}_r$, $(r = 0, 1, \ldots, z-1)$. This vector has $z = (n+1)(m+1)$ components. The sum of the required triplets can be gained by the triple summation

$$\sum_{r=0}^{z-3}\sum_{s=r+1}^{z-2}\sum_{t=s+1}^{z-1}(\mathbf{c}_r, \mathbf{c}_s, \mathbf{c}_t).$$

The subscripts of the original matrix can be calculated from r, s, t by means of the equalities

$$\begin{aligned} j &= r \bmod (m+1)\,, i = (r-j)\,/\,(m+1)\,, \\ l &= s \bmod (m+1)\,, k = (s-l)\,/\,(m+1)\,, \\ q &= t \bmod (m+1)\,, p = (t-q)\,/\,(m+1)\,, \end{aligned}$$

which completes the proof. ∎

Although the formula looks a bit involved, it is pleasing due to its symmetry in n and m. Moreover, it is easy to implement in a programming language, and the program based on this formula will be efficient since the number of terms in the sum is minimized.

3. Conclusions

By means of Theorem 2 one can compute the volume of any solid, the boundary of which can be decomposed to Bézier surfaces. Moreover, we can compute the signed volume of the solid bounded by a B–spline patch and the cones determined by its boundary curves and the origin. In order to do this, we generate the Bézier points of the B–spline patch, cf. [1], i.e. we decompose the B–spline patch to Bézier surfaces. The sum of the signed volumes of the solids determined by these Bézier surfaces is the signed volume of the solid determined by the B–spline patch.

REFERENCES

[1] Böhm, W.: *Generating the Bézier points of B–spline curves and surfaces*, Computer–Aided Design, **13**(6), (1981), 365–366.

[2] Farin, G.: *Curves and Surfaces for Computer Aided Geometric Design*, Academic Press, New York, 1988.

[3] Juhász, I.: *Cubic parametric curves of given tangent and curvature*, Computer-Aided Design, **30**(1), (1998), 1–9.

[4] Knuth, D.: *The Art of Computer Programming*, Addison–Wesley, Reading, 1969.

Doi: 10.18514/MMN.2000.27

Mathematical Notes, Miskolc, Vol. 1., No. 2., (2000), pp. 135–143

METHOD OF LOWER AND UPPER FUNCTIONS AND THE EXISTENCE OF SOLUTIONS TO SINGULAR PERIODIC PROBLEMS FOR NONLINEAR DIFFERENTIAL EQUATIONS OF ORDER TWO

Irena Rachůnková
Department of Mathematics, Palacký University
Tomkova 40, 779 00 Olomouc, Czech Republic
rachunko@matnw.upol.cz

Milan Tvrdý
Mathematical Institute, Academy of Sciences of the Czech Republic
Žitná 25, 115 67 Praha, Czech Republic
tvrdy@math.cas.cz

[Received May 26, 2000]

Abstract. We construct nonconstant lower and upper functions for the periodic boundary value problem $u'' = f(t,u)$, $u(0) = u(2\pi)$, $u'(0) = u'(2\pi)$ and find their estimates. By means of these results we prove existence criteria for the problems $u'' \pm g(u) = e(t)$, $u(0) = u(2\pi)$, $u'(0) = u'(2\pi)$, where $\limsup_{x\to 0+} g(x) = \infty$ is allowed and $e \in \mathbb{L}[0,2\pi]$ need not be essentially bounded.

Mathematical Subject Classification: 34B15, 34C25
Keywords: Second order nonlinear ordinary differential equation, periodic solution, singular problem, lower and upper functions, attractive and repulsive singularity, Duffing equation.

1. Introduction

In this paper we construct lower and upper functions to the periodic boundary value problem

$$u'' = f(t,u), \quad u(0) = u(2\pi), \quad u'(0) = u'(2\pi). \tag{1.1}$$

By means of these results, we prove existence criteria for the problems

$$u'' \pm g(u) = e(t), \quad u(0) = u(2\pi), \quad u'(0) = u'(2\pi),$$

where $\limsup_{x\to 0+} g(x) = \infty$ is allowed and $e \in \mathbb{L}[0,2\pi]$ need not be essentially bounded. We assume that $f : [0,2\pi] \times \mathbb{R} \mapsto \mathbb{R}$ fulfills the Carathéodory conditions on $[0,2\pi] \times \mathbb{R}$, i.e. f has the following properties: (i) for each $x \in \mathbb{R}$ the function $f(.,x)$ is measurable on $[0,2\pi]$; (ii) for almost every $t \in [0,2\pi]$ the function $f(t,.)$ is continuous on $\mathbb{R}$; (iii) for each compact set $\mathrm{K} \subset \mathbb{R}$ the function $m_{\mathrm{K}}(t) = \sup_{x\in \mathrm{K}} |f(t,x)|$ is Lebesgue integrable on $[0,2\pi]$.

For a given subinterval J of $\mathbb{R}$ (possibly unbounded), $\mathbb{C}(J)$ denotes the set of functions continuous on J. Furthermore, $\mathbb{L}[0,2\pi]$ stands for the set of functions Lebesgue integrable on $[0,2\pi]$, $\mathbb{L}_2[0,2\pi]$ is the set of functions square Lebesgue integrable on $[0,2\pi]$ and $\mathbb{AC}\,[0,2\pi]$ denotes the set of functions absolutely continuous on $[0,2\pi]$. For x bounded on $[0,2\pi]$, $y \in \mathbb{L}[0,2\pi]$ and $z \in \mathbb{L}_2[0,2\pi]$ we denote

$$\|x\|_{\mathbb{C}} = \sup_{t\in[0,2\pi]} |x(t)|, \quad \overline{y} = \frac{1}{2\pi}\int_0^{2\pi} y(s)\,\mathrm{d}s,$$
$$\|y\|_1 = \int_0^{2\pi} |y(t)|\,\mathrm{d}t \text{ and } \|z\|_2 = \Big(\int_0^{2\pi} z^2(t)\,\mathrm{d}t\Big)^{\frac{1}{2}}.$$

By a *solution of* (1.1) we mean a function $u\colon\ [0,2\pi]\mapsto\mathbb{R}$ such that $u'\in\mathbb{AC}[0,2\pi]$, $u(0)=u(2\pi)$, $u'(0)=u'(2\pi)$ and

$$u''(t) = f(t,u(t)) \quad \text{for a.e.} \quad t\in[0,2\pi].$$

Definition 1. A function σ_1 is said to be a *lower function of the problem* (1.1) if $\sigma_1'\in\mathbb{AC}\,[0,2\pi]$,

$$\begin{aligned} \sigma_1''(t) &\ge f(t,\sigma_1(t)) \quad \text{for a.e.} \quad t\in[0,2\pi],\\ \sigma_1(0) &= \sigma_1(2\pi), \quad \sigma_1'(0)\ge\sigma_1'(2\pi). \end{aligned}$$

Similarly, a function σ_2 is said to be an *upper function of the problem* (1.1) if $\sigma_2'\in\mathbb{AC}\,[0,2\pi]$,

$$\begin{aligned} \sigma_2''(t) &\le f(t,\sigma_2(t)) \quad \text{for a.e.} \quad t\in[0,2\pi],\\ \sigma_2(0) &= \sigma_2(2\pi), \quad \sigma_2'(0)\le\sigma_2'(2\pi). \end{aligned}$$

The lower and upper functions approach we will use here is based on the following theorem which is contained in [8, Theorems 4.1 and 4.2].

Theorem 2. *Let σ_1 and σ_2 be a lower and an upper function of the problem* (1.1)*, respectively.*

(I) Suppose $\sigma_1(t)\le\sigma_2(t)$ on $[0,2\pi]$. Then there is a solution u of the problem (1.1) *such that $\sigma_1(t)\le u(t)\le\sigma_2(t)$ on $[0,2\pi]$.*

(II) Suppose $\sigma_1(t)\ge\sigma_2(t)$ on $[0,2\pi]$ and there is $m\in\mathbb{L}[0,2\pi]$ such that

$$\begin{aligned} f(t,x) &\ge m(t) \quad \textit{for a.e.} \quad t\in[0,2\pi] \textit{ and all } x\in\mathbb{R}\\ (\textit{or } f(t,x) &\le m(t) \quad \textit{for a.e.} \quad t\in[0,2\pi] \textit{ and all } x\in\mathbb{R}.) \end{aligned}$$

Then there is a solution u of the problem (1.1) *such that $\|u'\|_{\mathbb{C}}\le\|m\|_1$ and*

$$\sigma_2(t_u)\le u(t_u)\le\sigma_1(t_u) \quad \textit{for some} \quad t_u\in[0,2\pi].$$

2. Construction of lower and upper functions

Proposition 1. *Assume that there are $A \in \mathbb{R}$ and $b \in \mathbb{L}[0, 2\pi]$ such that*

$$\overline{b} = 0, \tag{2.1}$$

$$f(t, x) \le b(t) \text{ for a.e. } t \in [0, 2\pi] \text{ and all } x \in [A, B], \tag{2.2}$$

where

$$B = A + \frac{\pi}{3}\,\|b\|_1. \tag{2.3}$$

Then there exists a lower function σ of the problem (1.1) such that

$$A \le \sigma(t) \le B \text{ on } [0, 2\pi]. \tag{2.4}$$

Proof. Define

$$\sigma_0(t) = c_0 + \int_0^{2\pi} g(t, s) b(s)\mathrm{d}s \text{ for } t \in [0, 2\pi],$$

where

$$g(t, s) = \begin{cases} \dfrac{t(s - 2\pi)}{2\pi} & \text{if } 0 \le t \le s \le 2\pi, \\[2ex] \dfrac{(t - 2\pi)s}{2\pi} & \text{if } 0 \le s < t \le 2\pi \end{cases}$$

is the Green function of the problem $v'' = 0$, $v(0) = v(2\pi) = 0$ and

$$c_0 = -\frac{1}{2\pi} \int_0^{2\pi} \Big(\int_0^{2\pi} g(t, s) b(s)\mathrm{d}s \Big)\mathrm{d}t.$$

Then

$$\sigma_0''(t) = b(t) \quad \text{a.e. on } [0, 2\pi] \tag{2.5}$$

and

$$\sigma_0(0) = \sigma_0(2\pi). \tag{2.6}$$

Furthermore, by virtue of (2.1) we have also

$$\sigma_0'(0) = \sigma_0'(2\pi). \tag{2.7}$$

Multiplying the relation (2.5) by σ_0, integrating it over $[0, 2\pi]$ and using the Hölder inequality we get

$$\|\sigma_0'\|_2^2 \le \|b\|_1 \|\sigma_0\|_{\mathbb{C}}.$$

Further, as $\overline{\sigma}_0 = 0$, the Sobolev inequality (see [5, Proposition 1.3]) yields

$$\|\sigma_0'\|_2^2 \le \sqrt{\frac{\pi}{6}}\,\|b\|_1 \|\sigma_0'\|_2,$$

and so

$$\|\sigma_0'\|_2 \le \sqrt{\frac{\pi}{6}}\,\|b\|_1,$$

wherefrom using again the Sobolev inequality we get

$$\|\sigma_0\|_{\mathbb{C}} \le \frac{\pi}{6}\,\|b\|_1.$$

Thus, the function σ given by

$$\sigma(t) = \frac{\pi}{6}\|b\|_1 + A + \sigma_0(t) \text{ for } t \in [0, 2\pi] \tag{2.8}$$

satisfies (2.4). Furthermore, according to (2.1),(2.2) and (2.6)-(2.7) we have

$$\sigma''(t) = \sigma_0''(t) = b(t) \ge f(t, \sigma(t)) \text{ for a.e. } t \in [0, 2\pi] \tag{2.9}$$

and

$$\sigma(0) = \sigma(2\pi), \quad \sigma'(0) = \sigma'(2\pi), \tag{2.10}$$

i.e. σ is a lower function of (1.1). □

The following assertion is dual to Proposition 1 and its proof will be omitted.

Proposition 2. *Assume that there are $A \in \mathbb{R}$ and $b \in \mathbb{L}[0, 2\pi]$ such that*

$$\bar{b} = 0$$

and

$$f(t, x) \ge b(t) \textit{ for a.e. } t \in [0, 2\pi] \textit{ and all } x \in [A, B]$$

where B is given by (2.3). *Then there exists an upper function σ of the problem* (1.1) *with the property (2.4).*

3. Applications to Lazer-Solimini singular problems

In this section we will consider possible singular problems of the attractive type

$$u'' + g(u) = e(t), \quad u(0) = u(2\pi), \quad u'(0) = u'(2\pi) \tag{3.1}$$

and of the repulsive type

$$u'' - g(u) = e(t), \quad u(0) = u(2\pi), \quad u'(0) = u'(2\pi), \tag{3.2}$$

where

$$g \in \mathbb{C}(0, \infty) \text{ and } e \in \mathbb{L}[0, 2\pi] \tag{3.3}$$

and it is allowed that $\limsup_{x \to 0+} g(x) = \infty$.

The problem (3.1) has been studied by Lazer and Solimini in [6] for $e \in \mathbb{C}[0, 2\pi]$ and g positive. In [9, Corollary 3.3], their existence result has been extended to $e \in \mathbb{L}[0, 2\pi]$ essentially bounded from above. Here, we bring conditions for the existence of solutions to (3.1) without assuming boundedness of e.

Theorem 1. *Assume*(3.3) *and let there exist* $A_1, A_2 \in (0, \infty)$ *such that*

$$g(x) \geq \overline{e} \text{ for all } x \in [A_1, B_1], \tag{3.4}$$

$$g(x) \leq \overline{e} \text{ for all } x \in [A_2, B_2], \tag{3.5}$$

where

$$B_1 - A_1 = B_2 - A_2 = \frac{\pi}{3}\|e - \overline{e}\|_1 \tag{3.6}$$

and $A_2 \geq B_1$.

Then the problem (3.1) *has a solution* u *such that* $A_1 \leq u(t) \leq B_2$ *on* $[0, 2\pi]$.

Proof. Define, for a.e. $t \in [0, 2\pi]$,

$$f(t, x) = e(t) - \begin{cases} g(A_1) & \text{if } x < A_1, \\ g(x) & \text{if } x \geq A_1. \end{cases}$$

Then f satisfies the Carathéodory conditions on $[0, 2\pi] \times \mathbb{R}$. Furthermore, by (3.4) and (3.6), f satisfies (2.1)-(2.3) with $b(t) = e(t) - \overline{e}$ a.e. on $[0, 2\pi]$ and $[A, B] = [A_1, B_1]$. Hence, by Proposition 1 there exists a lower function σ_1 of (1.1) such that $\sigma_1(t) \in [A_1, B_1]$ for all $t \in [0, 2\pi]$. Similarly, (3.5), (3.6) and Proposition 2 yield the existence of an upper function σ_2 of (1.1) such that $\sigma_2(t) \in [A_2, B_2]$ on $[0, 2\pi]$. Now, since $A_2 \geq B_1$, we have $\sigma_1(t) \leq \sigma_2(t)$ on $[0, 2\pi]$ and the assertion (I) of Theorem 2 gives the existence of the desired solution u to (1.1) which is naturaly also a solution to (3.1). $\square$

Classical Lazer and Solimini's considerations [6] of the repulsive problem (3.2) have been extended by several authors (see e.g. [1]-[4], [7] and [10]). Here we present a related result, where e need not be essentially bounded.

Theorem 2. *Assume*(3.3),

$$\lim_{x \to 0+} \int_x^1 g(\xi) \mathrm{d}\xi = \infty, \tag{3.7}$$

and

$$g_* := \inf_{x \in (0, \infty)} g(x) > -\infty. \tag{3.8}$$

Furthermore, let there exist $A_1, A_2 \in (0, \infty)$ *such that*

$$g(x) \leq -\overline{e} \text{ for all } x \in [A_1, B_1], \tag{3.9}$$

$$g(x) \geq -\overline{e} \text{ for all } x \in [A_2, B_2], \tag{3.10}$$

where (3.6) *is true and* $A_1 \geq B_2$.

Then the problem (3.2) *has a positive solution.*

Proof. Denote

$$K = \|e\|_1 + |g_*|, \quad B = B_1 + 2\pi K \text{ and } K^* = K\|e\|_1 + \int_{A_2}^{B} |g(x)|\mathrm{d}x.$$

It follows from (3.7) that $\limsup_{x\to 0+} g(x) = \infty$ and there exists $\varepsilon \in (0, A_2)$ such that

$$\int_{\varepsilon}^{A_2} g(x)\mathrm{d}x > K^* \text{ and } g(\varepsilon) > 0. \tag{3.11}$$

Define

$$\widetilde{g}(x) = \begin{cases} g(x) & \text{if } x \geq \varepsilon, \\ g(\varepsilon) & \text{if } x < \varepsilon, \end{cases}$$

and

$$f(t, x) = e(t) + \widetilde{g}(x) \text{ for a.e. } t \in [0, 2\pi] \text{ and all } x \in \mathbb{R}.$$

Now, we can argue as in the proof of Theorem 1 obtaining a lower function σ_1 and an upper function σ_2 of (1.1) such that $\sigma_1(t) \geq \sigma_2(t)$ on $[0, 2\pi]$. The assertion (II) of Theorem 2 (with $m(t) = g_* + e(t)$ a.e. on $[0, 2\pi]$) implies that (1.1) has a solution u such that $u(t_u) \in [A_2, B_1]$ for some $t_u \in [0, 2\pi]$ and $\|u'\|_{\mathbb{C}} \leq K$. It remains to show that $u(t) \geq \varepsilon$ holds on $[0, 2\pi]$.

Let t_0 and $t_1 \in [0, 2\pi]$ be such that

$$u(t_0) = \min_{t\in[0,2\pi]} u(t) \text{ and } u(t_1) = \max_{t\in[0,2\pi]} u(t).$$

Clearly, $A_2 \leq u(t_1) \leq B$. With respect to the periodic boundary conditions we have $u'(t_0) = u'(t_1) = 0$. Now, multiplying the differential relation $u''(t) = e(t) + \widetilde{g}(u(t))$ by $u'(t)$ and integrating over $[t_0, t_1]$ we get

$$0 = \int_{t_0}^{t_1} u''(t)u'(t)\mathrm{d}t = \int_{t_0}^{t_1} e(t)u'(t)\mathrm{d}t + \int_{t_0}^{t_1} \widetilde{g}(u(t))u'(t)\mathrm{d}t,$$

i.e.

$$\int_{u(t_0)}^{u(t_1)} \widetilde{g}(x)\mathrm{d}x = -\int_{t_0}^{t_1} e(t)u'(t)\mathrm{d}t \leq K\|e\|_1.$$

Further,

$$\int_{u(t_0)}^{A_2} \widetilde{g}(x)\mathrm{d}x \leq K\|e\|_1 + \int_{A_2}^{B} |\widetilde{g}(x)|\mathrm{d}x = K^*$$

which, with respect to (3.11), is possible only if $u(t_0) \geq \varepsilon$. Thus, u is a solution to (3.2). □

Example 3. Let $g(x) = \frac{1}{x^\gamma}$ on $(0, \infty)$. If $\gamma > 0$, then Theorem 1 ensures the existence of a positive solution to (3.1) for any $e \in \mathbb{L}[0, 2\pi]$ such that

$$\overline{e} > 0 \text{ and } \frac{\pi}{3}\,\overline{e}^{\frac{1}{\gamma}}\,\|e - \overline{e}\|_{\mathbb{L}} < 1. \tag{3.12}$$

The function $e(t) = c + \dfrac{1}{\sqrt{2\pi t}} - \dfrac{1}{\pi}$ with $c \in \mathbb{R}$ is not essentially bounded from above on $[0, 2\pi]$. However, it satisfies (3.12) if

$$0 < c < \left(\frac{3}{\pi}\right)^{\gamma}.$$

We should mention that provided $e \in \mathbb{C}[0, 2\pi]$ or e is essentially bounded from above, the condition $\overline{e} > 0$ is sufficient for the existence of a solution to (3.1) (cf. [6] or [9], respectively).

Example 4. Let $e \in \mathbb{L}[0, 2\pi]$ be essentially unbounded from below and let

$$g(x) = \frac{1 + \sin\left(\frac{\pi}{x}\right)}{x} - \arctan(x), \quad x \in (0, \infty).$$

Then g verifies the assumptions (3.3), (3.7) and (3.8) of Theorem 2. Let us suppose that $\overline{e} = -5$. Then the equation $g(x) = 5$ has exactly 5 roots in the interval $[0.125, \infty)$. In particular, we have (see Figures 1 and 2)

$$x_1 \approx 0.126804,\ x_2 \approx 0.141071,\ x_3 \approx 0.167853,\ x_4 \approx 0.200541,\ x_5 \approx 0.244461,$$
$$g(x) > 5 \text{ on } (x_2, x_3) \cup (x_4, x_5) \text{ and } g(x) < 5 \text{ on } (x_1, x_2) \cup (x_3, x_4) \cup (x_5, \infty).$$

Therefore, by Theorem 2, if

$$\|e - \overline{e}\|_{\mathbb{L}} \le \frac{3}{\pi}(x_5 - x_4) \approx 0.0419392,$$

the problem

$$u'' = \frac{1 + \sin\left(\frac{\pi}{u}\right)}{u} - \arctan(u) + e(t), \quad u(0) = u(2\pi), \quad u'(0) = u'(2\pi) \tag{3.13}$$

has a solution u_1 such that $u_1(t^*) \in [x_4, x_5 + d_1]$ for some $t^* \in [0, 2\pi]$, where $d_1 = x_5 - x_4$ (see Figure 3).

Similarly, by Theorems 1 and 2, if

$$\|e - \overline{e}\|_{\mathbb{L}} < \frac{3}{2\pi}(x_5 - x_4) \approx 0.0209699,$$

the problem (3.13) has at least 2 different solutions u_1 and u_2, where $u_1(t^*) \in (x_5 - d_2, x_5 + d_2)$ for some $t^* \in [0, 2\pi]$ and $u_2(t) \in (x_4 - d_2, x_4 + d_2)$ for all $t \in [0, 2\pi]$, where $d_2 = \frac{1}{2}(x_5 - x_4)$ (see Figure 4).

Finally, if

$$\|e - \overline{e}\|_{\mathbb{L}} \le \frac{3}{\pi}(x_2 - x_1) \approx 0.0136238,$$

the problem (3.13) has at least 3 different solutions u_1, u_2 and u_3, where $u_1(t^*) \in [x_5 - d_3, x_5 + d_3]$ for some $t^* \in [0, 2\pi]$, $u_2(t) \in [x_4 - d_3, x_4 + d_3]$ for all $t \in [0, 2\pi]$ and $u_3(t) \in [x_1, x_2]$ for all $t \in [0, 2\pi]$, where $d_3 = x_2 - x_1$ (see Figure 5).

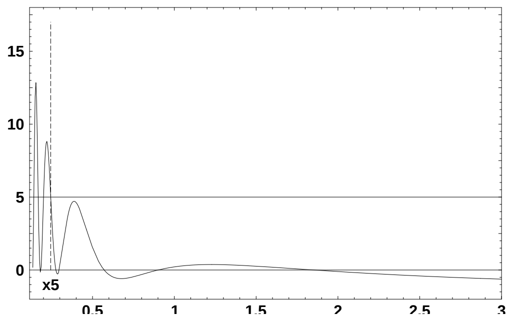

Figure 1.

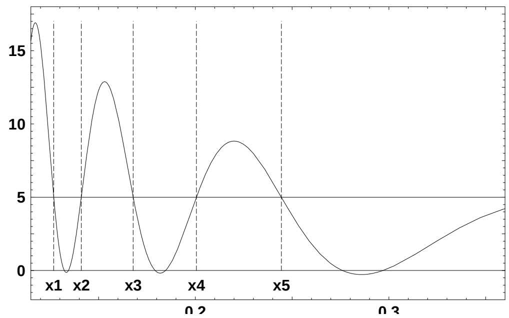

Figure 2.

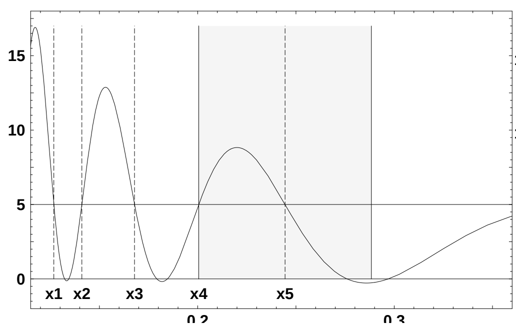

Figure 3.

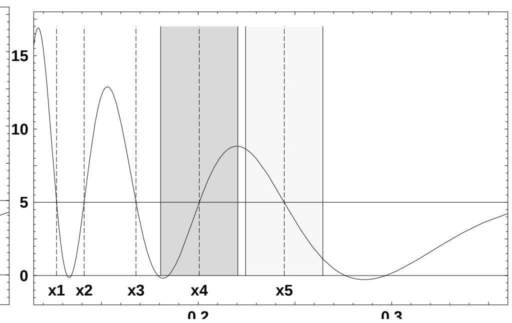

Figure 4.

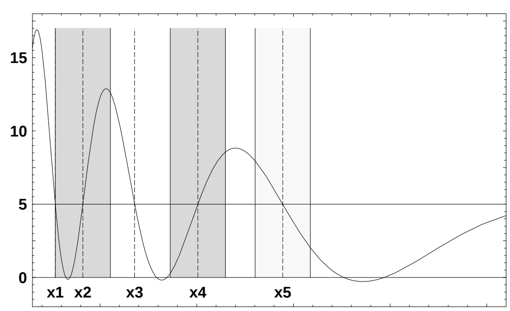

Figure 5.

Acknowledgement: The support provided to I. Rachůnková by the grant No. 201/98/0318 of the Grant Agency of the Czech Republic and by the Council of Czech Government J14/98:153100011 is gratefully acknowledged.

REFERENCES

[1] del Pino, M., Manásevich, R. and Montero, A.: *T-periodic solutions for some second order differential equations with singularities*, Proc. Royal Soc. Edinburgh, **120A**, (1992), 231–244.

[2] Fonda, A., Manásevich, R. and Zanolin, F.: *Subharmonic solutions for some second-order differential equations with singularities*, SIAM J. Math. Anal. **24**, (1993), 1294–1311.

[3] Habets, P. and Sanchez, L.: *Periodic solutions of some Liénard equations with singularities*, Proc. Amer. Math. Soc. **109**, (1990), 1035–1044.

[4] Mawhin, J.: *Topological degree and boundary value problems for nonlinear differential equations*, M. Furi (ed.) et al., *Topological methods for ordinary differential equations.* Lectures given at the 1st session of the Centro Internazionale Matematico Estivo (C.I.M.E.) held in Montecatini Terme, Italy, June 24-July 2, 1991. Berlin: Springer-Verlag, Lect. Notes Math. **1537**, (1993), 74–142.

[5] Mawhin, J and Willem, M.: *Critical Point Theory and Hamiltonian Systems*, Applied Mathematical Sciences 74, Springer-Verlag, Berlin (1989).

[6] Lazer,A. C. and Solimini, S.: *On periodic solutions of nonlinear differential equations with singularities*, Proc. Amer. Math. Soc. **99**, (1987), 109–114.

[7] Omari, P. and Ye, W.: *Necessary and sufficient conditions for the existence of periodic solutions of second order ordinary differential equations with singular nonlinearities*, Differential and Integral Equations **8**, (1995), 1843–1858.

[8] Rachůnková, I. and Tvrdý, M.: *Nonlinear systems of differential inequalities and solvability of certain nonlinear second order boundary value problems*, J. Inequal. Appl., to appear.

[9] Rachůnková, I., Tvrdý, M. and Vrkoč, I.: *Existence of nonnegative and nonpositive solutions for second order periodic boundary value problems*, Preprint. Math. Inst. Acad. Sci. Czech Rep., **134/1999**.

[10] Zhang, M.: *A relationship between the periodic and the Dirichlet BVP's of singular differential equations*, Proc. Royal Soc. Edinburgh, **128A**, (1998), 1099–1114.

Doi: 10.18514/MMN.2000.28

Mathematical Notes, Miskolc, Vol.1., No.1., (2000), pp. 145–156

CLASSIFICATION SYSTEMS AND THE DECOMPOSITIONS OF A LATTICE INTO DIRECT PRODUCTS

Sándor Radeleczki
Institute of Mathematics, University of Miskolc
3515 Miskolc – Egyetemváros, Hungary
matradi@gold.uni-miskolc.hu

[Received September 12, 2000]

Abstract. We introduce the notion of a classification system for an arbitrary complete lattice. We give a characterization of those pseudocomplemented lattices which have the property that any of their classification systems yields a decomposition of the lattice into a direct product. By applying these results we prove new stucture theorems for particular classes of complete pseudocomplemented lattices.

Mathematical Subject Classification: 08B05, 08B10
Keywords: Classification system, independent set, semicentral element

1. Introduction

The notion of the classification system has its origin in an application of Concept Lattices to one of the main problems of Group Technology, namely, to classify some technical objects on the basis of their properties.

Given a set G of (technical) objects and a set M of (possible) properties, a binary relation $I \subseteqq G \times M$ is defined as follows:

$(g, m) \in I$ if and only if the object $g \in G$ has the property $m \in M$.

The triple (G, M, I) is called *formal context* in mathematical literature and (by using the basic construction of Formal Concept Analysis) a complete lattice $\mathcal{L}(G, M, I)$ is associated with it, which is called the *concept lattice of the context* (G, M, I). (For details see [3].) If the objects with the same properties are considered identical - and this is the case in our technical application - then the context (G, M, I) is called *row-reduced.* A *classification of the elements of* G means a partition $\pi = \{G_i, i \in I\}$ of the set G, where any block G_i of π is characterized by the common properties of its objects. The observation (see [7]) which led us to our investigations was the following:

"If (G, M, I) is a row-reduced context, then to any classification $\pi = \{G_i, i \in I\}$ of the elements of G corresponds a system $\{a_i, i \in I\}$ of nonzero elements of the concept lattice $\mathcal{L}(G, M, I)$ satisfying the conditions (1) and (2) of Definition 2.1. Conversely, any set of elements of $\mathcal{L}(G, M, I)$ which satisfies conditions (1) and (2) induces a classification of the elements of G."

In the next Section, we define and study classification systems in an arbitrary complete lattice presenting some of their basic properties, making abstraction from the original problem and the theory of Concept Lattices. In Section 3 we give a characterization of those complete pseudocomplemented lattices which have the property that any of their classification systems yields a decomposition of the given lattice into a direct product. In Section 4 we apply these results to describe direct product decompositions of certain pseudocomplemented lattices.

2. Basic notions

Let 0 and 1 stand for the least and the greatest element of a bounded lattice L, respectively. $(x]$ is our notation for the principal ideal generated by an $x \in L$. The supremum of a set $A \subseteqq L$ (if it exists) is denoted by $\vee A$. We set $\vee\varnothing = 0$, as usual.

Definition 2.1. Let L be a complete lattice. A set $S = \{a_i \mid i \in I\}$, $I \neq \varnothing$ of nonzero elements of L is called a *classification system* of L if the following conditions are satisfied:

(1) $\quad a_i \wedge a_j = 0$, for all $i \neq j$,

(2) $\quad x = \bigvee_{i \in I} (x \wedge a_i)$, for all $x \in L$.

If $S = \{1\}$, then we say that S is *trivial.* S is called *weakly independent* if $a_j \wedge (\bigvee_{i \in I \setminus \{j\}} a_j) = 0$ holds for all $j \in I$. We say that S is *strongly independent* if the relation $(\bigvee_{i \in J} a_i) \wedge (\bigvee_{i \in K} a_i) = (\bigvee_{i \in J \cap K} a_j)$ holds for all $K, J \subseteqq I$.

The following lemma contains some simple properties of the classification systems.

Lemma 2.2. *Let $S = \{a_i \mid i \in I\}$ be a classification system of a complete lattice L. Then the following statements hold:*

(i) *We have $\bigvee_{i \in I} a_i = 1$ and for any $b \in L$, $b \neq 0$ the nonzero elements of the set $\{b \wedge a_i \mid i \in I\}$ form a classification system of the sublattice $(b]$.*

(ii) *If $T = \{b_j \mid j \in J\}$ is a classification system of $(a_{i_0}]$ $(i_0 \in I)$, then $S' = \{a_i \mid i \in I \setminus \{i_0\}\} \cup \{b_j \mid j \in J\}$ is also a classification system for L.*

(iii) *Let $K \subseteqq I$, $K \neq \varnothing$ arbitrary and $b = \bigvee_{i \in K} a_i$. If S is weakly independent then $S^* = \{a_i \mid i \in I \setminus K\} \cup \{b\}$ is also a classification system of L.*

Proof. (i) Substituting $x = 1$ in relation (2) we obtain $1 = \bigvee_{i \in I} a_i$. Let $V = \{b \wedge a_i \mid i \in K\}$ be the subset of nonzero elements of $\{b \wedge a_i \mid i \in I\}$ (here $K \subseteqq I$). Since by (2) we have $b = \bigvee_{i \in I} (b \wedge a_i)$, V cannot be empty. As we have $(b \wedge a_i) \wedge (b \wedge a_j) = b \wedge (a_i \wedge a_j) = 0$ for $i \neq j$, relation (1) is satisfied by V. Now take an $x \in (b]$, then

$x \wedge (b \wedge a_i) = (x \wedge b) \wedge a_i = x \wedge a_i$, $\forall i \in I$. Therefore, we can write:

$$x = \bigvee_{i \in I} (x \wedge a_i) = \bigvee_{i \in I} (x \wedge (b \wedge a_i)) = \bigvee_{i \in K} (x \wedge (b \wedge a_i)).$$

Thus V satisfies (2).

(ii) Relation (1) can be easily checked for S'. Since $\{b_j \mid j \in J\}$ is a classification system of $(a_{i_0}]$ and since $x \wedge a_{i_0} \in (a_{i_0}]$, we get

$$x \wedge a_{i_0} = \bigvee_{j \in J} ((x \wedge a_{i_0}) \wedge b_j) = \bigvee_{j \in J} (x \wedge (a_{i_0} \wedge b_j)) = \bigvee_{j \in J} (x \wedge b_j), forall x \in L.$$

Thus we can write:

$$x = \bigvee_{i \in I} (x \wedge a_i) = (\ x \wedge a_{i_0}) \vee (\bigvee_{i \in I \setminus \{i_0\}} (x \wedge a_i)) = (\bigvee_{j \in J} (x \wedge b_j)) \vee (\bigvee_{i \in I \setminus \{i_0\}} (x \wedge a_i)),$$

and this proves that relation (2) holds for S'.

(iii) We have $a_i \wedge b = a_i \wedge (\bigvee_{i \in K} a_i) \leqq a_i \wedge (\bigvee_{i \in I \setminus \{i\}} a_i)$. Since S is weakly independent $a_i \wedge (\bigvee_{i \in I \setminus \{i\}} a_i) = 0$. Thus we obtain $a_i \wedge b = 0$ for all $i \in I \setminus K$, and this ensures that relation (1) holds for S^*. Now for an $x \in L$ we have

$$x = \bigvee_{i \in I} (x \wedge a_i) \leqq \ (\bigvee_{i \in I \setminus K} (x \wedge a_i)) \vee (\bigvee_{i \in K} (x \wedge a_i)) \leqq \ (\bigvee_{i \in I \setminus K} (x \wedge a_i)) \vee (x \wedge b) \leqq x.$$

Hence we get $(\bigvee_{i \in I \setminus K} (x \wedge a_i)) \vee (x \wedge b) = x$, thus condition (2) is satisfied by S^*. □

Definition 2.3. An element $a \in L$ is called a *central element* of the lattice L if for all $x, y \in L$ the sublattice generated by $\{a, x, y\}$ is distributive and a is complemented.

A complement of an element $a \in L$ (if it exists) is denoted by $\overline{a}$. We note that the complement of a central element is unique. The central elements of L form a Boolean sublattice of L denoted by Cen$L$. We say that an ordered pair (a, b) of elements of a lattice L is a *modular pair* and we write $(a, b) \in M$ if, for all $x \in L$ $x \leqq b$ implies $x \vee (a \wedge b) = (x \vee a) \wedge b$ (see [6] or [8]). Clearly, we have $(a, b) \in M$ for any $a \in$Cen$L$ and $b \in L$. The following lemma from [6] will be useful in our proofs.

Lemma 2.4. *Let a be an element of a bounded lattice L. Then the following statements are equivalent:*

(i) $a \in$Cen$L$,

(ii) *There exists an element $a' \in L$ such that*

$x = (x \wedge a) \vee (x \wedge a') = (x \vee a) \wedge (x \vee a')$,

(iii) *There exists an element $a' \in L$ such that $a \wedge a' = 0$, $(a, a') \in M$, $(a', a) \in M$ and $x = (x \wedge a) \vee (x \wedge a')$ for every $x \in L$.*

Remark 2.5. It is easy to check that if the element a' in (ii) or (iii) exists, then $a' = \overline{a}$. Consequently, for any $a \in$Cen$L$, $\{a, \overline{a}\}$ is a classification system both for L and its dual L^d.

Definition 2.6. Let L be a complete lattice. A classification system $S = \{a_i \mid i \in I\}$ is called a *decomposition system* of L if $a_i \in \mathrm{Cen}L$, for all $i \in I$.

The following proposition clarifies the meaning of the former definition.

Proposition 2.7. *Let L be a complete lattice. Then the following assertions are true:*

(i) *If $\{a_i \mid i \in I\}$ is a (nontrivial) decomposition system of L, then $L \cong \prod_{i\in I}(a_i]$.*

(ii) *For any (nontrivial) direct decomposition $L = \prod_{i\in I} L_i$ there exist elements $a_i \in \mathrm{Cen}L$, $i \in I$ such that $L_i \cong (a_i]$ and such that $\{a_i \mid i \in I\}$ is a (nontrivial) decomposition system of L.*

Proof. Since the above statements are more or less known in the literature, we outline only the principal steps of the proof.

(i) The isomorphism is given by the map $h : L \longrightarrow \prod_{i\in I}(a_i]$, $h(x) = (x \wedge a_i)_{i\in I}$.

Indeed, it is easy to check that h is a homomorphism, since for any $a_i \in \mathrm{Cen}L$ and $x, y \in L$ we have $(x \vee y) \wedge a_i = (x \wedge a_i) \vee (y \wedge a_i)$ (and of course, $(x \wedge y) \wedge a_i = (x \wedge a_i) \wedge (y \wedge a_i)$). Since by the assumption of (i) $x = \bigvee_{i\in I}(x \wedge a_i)$ holds for all $x \in L$, $h(x_1) = h(x_2)$ implies $x_1 = x_2$, therefore h is injective. Finally, the surjectivity of h can be shown proving the equality

$$h\left(\bigvee_{i\in I} x_i\right) = (x_i)_{i\in I}, \quad \text{for all} \quad x_i \leqq a_i, i \in I.$$

(ii) Let 0_i and 1_i stand for the least and the greatest element of L_i, respectively. We define the elements $a_i \in \prod_{i\in I} L_i = L$ as follows: $(a_i)_j = 0_j$, for $j \in I$, $j \neq i$ and $(a_i)_i = 1_i$. Then obviously we have $L_i \cong (a_i]$ and $a_i \wedge a_j = 0$ for all $i \neq j$. It is also easy to check that any a_i is a central element of $\prod_{i\in I} L_i$, i.e. of L (see [4] or [6]). Clearly, for all $x = (x_i)_{i\in I} \in \prod_{i\in I} L_i$ we can write $x = \bigvee_{i\in I}(x \wedge a_i)$, thus $S = \{a_i \mid i \in I\}$ is a decomposition system of $L = \prod_{i\in I} L_i$. Evidently, whenever the set I contains more than one element, S is not trivial. □

Proposition 2.8. *Any decomposition system $S = \{c_i \mid i \in I\}$ of a complete lattice L is a strongly independent classification system of L. Moreover, we have $\bigvee_{i\in K} c_i \in \mathrm{Cen}L$, for any $K \subseteqq I$.*

Proof. In view of Proposition 2.7 we have $L \cong \prod_{i\in I}(c_i]$. For each $M \subseteqq I$ we define the elements $c^M \in \prod_{i\in I}(c_i]$ as follows: $(c^M)_i = c_i$ for all $i \in M$, otherwise $(c^M)_i = 0_i$. It is easy to check that for any $J, K \subseteqq I$ we have $c^J \wedge c^K = c^{J\cap K}$. Let $h : L \longrightarrow \prod_{i\in I}(c_i]$, $h(x) = (x \wedge c_i)_{i\in I}$ be the isomorphism as in the proof of Proposition 2.7.(i). Since we have $h\left(\bigvee_{i\in I} x_i\right) = (x_i)_{i\in I}$ for all $x_i \leqq a_i$ (as was indicated), h^{-1} maps any $x = (x_i)_{i\in I} \in \prod_{i\in I}(c_i]$ to $\bigvee_{i\in I} x_i$. Thus we get $h^{-1}(c^M) = \bigvee_{i\in M} c_i$, for all $M \subseteqq I$. Since h^{-1} is an isomorphism, now we can write:

$$\left(\bigvee_{i\in J} c_i\right) \wedge \left(\bigvee_{i\in K} c_i\right) = h^{-1}(c^J) \wedge h^{-1}(c^K) = h^{-1}(c^J \wedge c^K) = h^{-1}(c^{J\cap K}) = \bigvee_{i\in J\cap K} c_i.$$

Thus $\{c_i \mid i \in I\}$ is a strongly independent system.

In order to prove the second assertion, let us define for any $K \subseteqq I$ the direct products $A_K = \prod_{i\in K}(c_i]$ and $B_K = \prod_{i\in I\setminus K}(c_i]$, and denote by 0_{A_K},1_{A_K} and 0_{B_K},1_{B_K} the least and the greatest elements of the lattices A_K and B_K, respectively. Clearly, we have $L \cong \prod_{i\in I}(c_i] \cong A_K \times B_K$. By [4], $(1_{A_K}, 0_{B_K})$ and $(0_{A_K}, 1_{B_K})$ are central elements of the lattice $A_K \times B_K$ and correspond by isomorphism to the elements $c^K \in \prod_{i\in I}(c_i]$ and $c^{I\setminus K} \in \prod_{i\in I}(c_i]$, respectively. Since $h^{-1}(c^K) = \bigvee_{i\in K} c_i$, and since the central elements of a lattice are preserved by isomorphisms, we get $\bigvee_{i\in K} c_i \in \mathrm{Cen}L$. □

3. Classification systems in particular lattices

A lattice L with 0 is said to be *0-modular* if $a, b \in L, a \leqq c$ and $b \wedge c = 0$ imply $(a \vee b) \wedge c = a$. By J.C. Varlet's result (see e.g. [8]), this definition is equivalent to the fact that there is no N_5 sublattice in L including the element 0.

Proposition 3.1. *If L is a 0-modular complete lattice, then any of its weakly independent classification systems of it is a decomposition system.*

Proof. Let $S = \{a_i \mid i \in I\}$ be a weakly independent classification system of L and take $b_i = \bigvee_{j\in I\setminus\{i\}} a_j$, $i \in I$. Since S is weakly independent, we have $a_i \wedge b_i = 0$. In view of Lemma 2.2(iii) the set $\{a_i, b_i\}$ is a classification system of L, thus we get $x = (x \wedge a_i) \vee (x \wedge b_i)$ for all $x \in L$. Since L is 0-modular, we have $(x \vee a_i) \wedge b_i = x$ for all $x \leqq a_i$. Thus we obtain $x \vee (a_i \wedge b_i) = x = (x \vee a_i) \wedge b_i$ and this relation means that $(a_i, b_i) \in M$. Similarly we can prove $(b_i, a_i) \in M$. Now, by applying Lemma 2.4 we get $a_i \in \mathrm{Cen}L$, $i \in I$, therefore S is a decomposition system of L. □

A lattice L with 0 element is called a *pseudocomplemented lattice* if any element $x \in L$ has a *pseudocomplement* x^*, that is, for any $x \in L$ there exists an element $x^* \in L$ such that $y \wedge x = 0$ iff $y \leqq x^*$.

Definition 3.2. (R. Beazer, [2]) An element a of a pseudocomplemented lattice L is called a *semicentral element* if

$$x = (x \wedge a) \vee (x \wedge a^*) \text{ holds for all } x \in L. \qquad (*)$$

Remark 3.3. The above definition implies $a^* = \overline{a}$ and $a^{**} = a$ for any semicentral element $a \in L$. From here it follows that the pseudocomplement of a, i.e. a^*, is a semicentral element too. We note that in view of relation $(*)$ the set $\{a, a^*\}$ is a classification system of L. It is also clear that any $c \in \mathrm{Cen}L$ is a semicentral element.

Proposition 3.4. *Any classification system $S = \{a_i \mid i \in I\}$ of a complete pseudocomplemented lattice L is weakly independent and it consists of semicentral elements of L. Moreover, for any $K \subseteqq I$, $K \neq \emptyset$ $\bigvee_{i \in K} a_i$ is a semicentral element of L.*

Proof. As $a_i \wedge a_j = 0$ for $i \neq j$, we have $a_j \leqq a_i^*$ for all $j \neq i$. Thus we can write $\bigvee_{j \in I \setminus \{i\}} a_j \leqq a_i^*$. As a consequence, we get $a_i \wedge \left(\bigvee_{j \in I \setminus \{i\}} a_j \right) = 0$, $i \in I$, i.e. S is weakly independent. Furthermore, we can write $a_i^* = \bigvee_{j \in I} (a_i^* \wedge a_j) = (a_i^* \wedge a_i) \vee \left(\bigvee_{j \in I \setminus \{i\}} (a_i^* \wedge a_j) \right) \leqq \bigvee_{j \in I \setminus \{i\}} a_j$. Hence we obtain $a_i^* = \bigvee_{j \in I \setminus \{i\}} a_j$. Since S is a weakly independent classification system, by applying Lemma 2.2(iii) with $K = I \setminus \{i\}$ we get that $\{a_i, a_i^*\}$ is a classification system, too. Therefore $x = (x \vee a_i) \vee (x \wedge a_i^*)$ holds for all $x \in L$ (and $i \in I$). Thus each a_i is a semicentral element.

Finally, set $b = \bigvee_{i \in K} a_i$, for a $K \subseteqq I$, $K \neq \emptyset$. Then $S^* = \{a_i \mid i \in I \setminus K\} \cup \{b\}$ is also a classification system of L, by Lemma 2.2(iii) again. Thus, as we have already proved above, $b = \bigvee_{i \in K} a_i$ is a semicentral element. □

Corollary 3.5. *If L is a complete 0-modular pseudocomplemented lattice, then any classification system of L is its decomposition system.*

Proof. We apply Proposition 3.4 and Proposition 3.1. □

Example 3.6. Let Tol$L$ stand for the tolerance lattice of a lattice L. By [1], Tol$L$ is a pseudocomplemented and 0-modular complete lattice. Thus, according to Corollary 3.5, any classification system of Tol$L$ is its decomposition system.

The last two results naturally raise the question: "Under what conditions do the classification systems of a complete lattice coincide with its decomposition systems?" In the case of pseudocomplemented lattices the following is an answer.

Theorem 3.7. *Let L be a complete pseudocomplemented lattice. Then the following assertions are equivalent:*

(i) *Any semicentral element of L is its central element,*

(ii) *Any classification system of L is its decomposition system,*

(iii) *For any semicentral element $a \in L$ we have $(a, a^*) \in M$.*

(iv) *The complemented congruences of the algebra $(L, \wedge, ^*)$ and the factor congruences of the lattice L are the same.*

Proof. (i)⇒(ii) is obvious, since any classification system of a pseudocomplemented lattice consists of semicentral elements.

(ii)⇒(iii). Since for any semicentral element $a \in L$, the set $\{a, a^*\}$ is a classification system, by assumption of (ii) it follows that $a, a^* \in$Cen$L$. Hence $(a, a^*) \in M$.

(iii)$\Rightarrow$(i). Let $a \in L$ be a semicentral element. Then, according to Remark 3.3, a^* is a semicentral element, too, and $a^{**} = a$. Thus by assumption we have $(a, a^*) \in M$ and $(a^*, a) \in M$ and $x = (x \wedge a) \vee (x \wedge a^*)$, $\forall x \in L$. As $a \wedge a^* = 0$, by applying Lemma 2.4, we get that $a \in$Cen$L$.

(i)$\Leftrightarrow$(iv). Let $\theta_a = \{(x, y) \in L^2 \mid x \wedge a = y \wedge a\}$. In view of [2], θ_a is a complemented congruence of the algebra $(L, \wedge, ^*)$ if and only if a is a semicentral element of L. However $a \in$Cen$L$ if and only if θ_a is a factor congruence of the lattice L (see [4]), whence the required equivalence follows. $\square$

4. Classification systems in CJ-generated pseudocomplemented lattices

An element p of a complete lattice L is called *completely join-irreducible* if for any system of elements $x_i \in L$, $i \in I$, the equality $p = \bigvee\{x_i \mid i \in I\}$ implies $p = x_{i_0}$ for some $i_0 \in I$. If any element of L is a join of the completely join-irreducible elements, then L is called a *CJ-generated lattice.* The set of completely join-irreducible elements of L is denoted by $J(L)$. For an $a \in L$ let $J(a) = \{p \in J(L) \mid p \leq a\}$. It is clear that for any system of elements $a_i \in L$, $i \in I$ we have $J\left(\bigwedge\limits_{i \in I} a_i\right) = \bigcap\limits_{i \in I} J(a_i)$ and $\bigcup\limits_{i \in I} J(a_i) \subseteqq J\left(\bigvee\limits_{i \in I} a_i\right)$, moreover, if L is CJ-generated, then for any $x, y \in L$ $J(x) \subseteqq J(y) \Longleftrightarrow x \leqq y$.

For an arbitrary relation $\rho \subseteqq X \times X$ and for a set $B \subseteqq X$ we define $\rho(B) = \{x \in X \mid \exists b \in B \text{ such that } (b, x) \in \rho\}$. We say that the set B is *closed relative to the relation* ρ if $\rho(B) \subseteqq B$, i.e. for any $b \in B$ and $x \in X$ from $(b, x) \in \rho$ it follows $x \in B$. If ρ is an equivalence relation, then the equivalence class of an element $x \in X$ is denoted by $[x]_\rho$.

Let L be a CJ-generated (complete) pseudocomplemented lattice. On the set $J(L) \setminus \{0\}$ we define the relation R as follows: $pRq \Leftrightarrow p \wedge q \neq 0$ $(p, q \in J(L) \setminus \{0\})$. A partition $\pi = \{A_i \mid i \in I\}$ of $J(L) \setminus \{0\}$ is called *R-closed* if every block A_i is closed relative to the relation R.

Now we can formulate the following

Proposition 4.1.

(i) *If* $\{a_i \mid i \in I\}$ *is a classification system of* L*, then the sets* $J(a_i) \setminus \{0\}$*,* $i \in I$ *form an R-closed partition of* $J(L) \setminus \{0\}$*.*

(ii) *If* $\pi = \{A_i \mid i \in I\}$ *is an R-closed partition of the set* $J(L) \setminus \{0\}$ *then the elements* $a_i = \bigvee A_i$*,* $i \in I$ *form a classification system of* L *with* $J(a_i) \setminus \{0\} = A_i$*.*

Proof. (i) Since $a_i \neq 0$, we have $J(a_i) \setminus \{0\} \neq \emptyset$, and for $i \neq j$, $a_i \wedge a_j = 0$ gives $(J(a_i) \setminus \{0\}) \cap (J(a_j) \setminus \{0\}) = \emptyset$. We prove $\bigcup\limits_{i \in I} (J(a_i) \setminus \{0\}) = J(L) \setminus \{0\}$. The inclusion

$\bigcup_{i\in I}(J(a_i)\setminus\{0\})\subseteqq J(L)\setminus\{0\}$ is obvious. To prove the converse inclusion take an arbitrary $p\in J(L)\setminus\{0\}$. Then $p=\bigvee_{i\in I}(p\wedge a_i)$, because $\{a_i\mid i\in I\}$ is a classification system of L. As p is completely join-irreducible, there exists an $i_0\in I$ such that $p=p\wedge a_{i_0}$, i.e. such that $p\in J(a_{i_0})\subseteqq\bigcup_{i\in I}(J(a_i)\setminus\{0\})$. Thus $\{J(a_i)\setminus\{0\}\mid i\in I\}$ is a partition of $J(L)\setminus\{0\}$.

Take now any $p\in J(a_i)\setminus\{0\}$ and assume that pRq holds for some $q\in J(L)\setminus\{0\}$. Since $\{J(a_i)\setminus\{0\}\mid i\in I\}$ is a partition, we have $q\in J(a_j)\setminus\{0\}$ for some $j\in I$. Then $p\leqq a_i$ and $q\leqq a_j$ implies $p\wedge q\leqq a_i\wedge a_j$. As $a_i\wedge a_j=0$ for $i\neq j$, the assumption $p\wedge q\neq 0$ gives $i=j$. Therefore $q\in J(a_i)\setminus\{0\}$ and this means that $J(a_i)\setminus\{0\}$ is R-closed.

(ii) Let $\pi=\{A_i\mid i\in I\}$ be an R-closed partition of $J(L)\setminus\{0\}$ and put $a_i=\bigvee A_i$, $i\in I$. Since $A_i\neq\emptyset$, we have $a_i\neq 0$. First, we show that $J(a_i)\setminus\{0\}=A_i$.

Let $p\in J(a_i)\setminus\{0\}$. We claim that $p\wedge q\neq 0$ for some $q\in A_i$. Indeed, $p\wedge q=0,\forall q\in A_i$ would imply that $a_i=\bigvee\{q\mid q\in A_i\}\leqq p^*$, thus we would get $p=p\wedge a_i\leqq p\wedge p^*=0$ - a contradiction. Thus we must have pRq for some $q\in A_i$. Since A_i is R-closed by assumption, we get $p\in A_i$ and this proves $J(a_i)\setminus\{0\}\subseteqq A_i$. The inclusion $A_i\subseteqq J(a_i)\setminus\{0\}$ is obvious.

If $i\neq j$ then $J(a_i\wedge a_j)=J(a_i)\cap J(a_j)=(A_i\cup\{0\})\cap(A_j\cup\{0\})=\{0\}$, and since L is CJ-generated, from here it follows $a_i\wedge a_j=0$.

Finally, observe that, in order to prove that $\{a_i\mid i\in I\}$ is a classification system of L, it is enough to show the inequality $x\leqq\bigvee_{i\in I}(x\wedge a_i)$ for any $x\in L$. For this purpose take any $p\in J(x)\setminus\{0\}$. Since $\{A_i\mid i\in I\}$ is a partition of $J(L)\setminus\{0\}$, there exists an $i_p\in I$ such that $p\in A_{i_p}$. Then we get $p\leqq x\wedge a_{i_p}$, therefore $p\leqq\bigvee_{i\in I}(x\wedge a_i)$. Thus we obtain $x=\bigvee\{p\mid p\in J(x)\}\leqq\bigvee_{i\in I}(x\wedge a_i)$. $\square$

Let $\overline{R}$ denote the transitive hull of the relation R. Since R is reflexive and symmetric by its definition, $\overline{R}$ is an equivalence on $J(L)\setminus\{0\}$. Clearly, the equivalence classes $[x]_{\overline{R}}$, $x\in J(L)\setminus\{0\}$ form an R-closed partition of $J(L)\setminus\{0\}$. Let $D=\{d\mid d=\bigvee[x]_{\overline{R}}$, $x\in J(L)\setminus\{0\}\}$. Evidently the elements of D can be indexed by the elements of the factor set $\mathcal{C}_R=(J(L)\setminus\{0\})/\overline{R}$, i.e. we can write: $D=\{d_i\mid i\in\mathcal{C}_R\}$. Now, by applying Proposition 4.1(ii) we obtain the following

Corollary 4.2. *For any CJ-generated complete pseudocomplemented lattice L, the set $D=\{d_i\mid i\in C_R\}$ is a classification system of L.*

We call a CJ-generated lattice L *connected*, if $\overline{R}$ is the total relation on the set $J(L)\setminus\{0\}$, i.e. if for any $p,q\in J(L)\setminus\{0\}$ there exists a sequence of elements $p_0,p_1,...,p_n\in J(L)\setminus\{0\}$, such that $p_0=p$, $p_n=q$ and $p_{i-1}\wedge p_i\neq 0$, for all $1\leqq i\leqq n$. Now we have the following

Corollary 4.3 *Let L be a CJ-generated complete pseudocomplemented lattice. Then*

(i) *L admits only the trivial classification system if and only if L is connected.*

(ii) *If L is connected, then it is directly irreducible.*

Proof. (i) If L is not a connected lattice, then the equivalence $\overline{R}$ has more than one class, thus applying Corollary 4.2 we get that $D = \{d_i \mid i \in \mathcal{C}_R\}$ is a nontrivial classification system of L.

Conversely, suppose that L is connected, and let $\{a_i \mid i \in I\}$ be a classification system of L. In view of Proposition 4.1(i) the sets $J(a_i) \setminus \{0\}$, $i \in I$ form an R-closed partition of $J(L)\setminus\{0\}$. Thus for every a_i and any $x \in J(a_i)\setminus\{0\}$ we have $[x]_{\overline{R}} \subseteqq J(a_i)$. As L is connected, we have $[x]_{\overline{R}} = J(L) \setminus \{0\}$ for all $x \in J(L) \setminus \{0\}$, whence we get that $J(a_i) = J(L)$, i.e. that $a_i = 1$, for all $i \in I$. Hence any classification system of L is trivial.

(ii) On the contrary, suppose that the lattice L is directly reducible. Then, by Proposition 2.7, L admits a nontrivial decomposition system S. Since S at the same time is a classification system, the above (i) gives that L is not connected, which is a contradiction. □

Now we are able to formulate the main result of this section, which is the following

Theorem 4.4. *Let L be a CJ-generated complete pseudocomplemented lattice. Then the following assertions are equivalent.*

(i) *Any classification system of L is a decomposition system,*

(ii) *L is a direct product of connected lattices,*

(iii) *Any semicentral element of L is central.*

Proof. The equivalence (i)⇔(iii) was proved by Theorem 3.7.

(i)⇒(ii). Let $D = \{d_i \mid i \in \mathcal{C}_R\}$ be the classification system induced by the equivalence $\overline{R}$. Since by assumption D is a decomposition system of L, Proposition 2.7 gives $L \cong \prod_{i\in\mathcal{C}_R} (d_i]$. Clearly, any $(d_i]$ as a principal ideal of L is a pseudocomplemented CJ-generated complete lattice and the set of its completely join-irreducible elements is the same as $J(d_i) \subseteqq J(L)$. Since any d_i is of the form $d_i = \bigvee[x]_{\overline{R}}$ for some $x \in J(L) \setminus \{0\}$, and since $J(d_i) \setminus \{0\} = [x]_{\overline{R}}$ in view of Proposition 4.1(ii), we have $p\overline{R}q$ for all $p, q \in J(d_i) \setminus \{0\}$. Thus any lattice $(d_i]$ is connected.

(ii)⇒(iii). By assumption of (ii) we have $L \cong \prod_{i\in I} L_i$, where all L_i are connected. Then, in view of Proposition 2.7, there are $c_i \in$Cen$L$ such that $L_i \cong (c_i]$, $i \in I$ and $S = \{c_i \mid i \in I\}$ is a decomposition system of L. Now, let a be a semicentral element of L. Since $\{a, a^*\}$ is a classification system of L, if $c_i \wedge a \neq 0$ and $c_i \wedge a^* \neq 0$ for some $i \in I$, then by applying Lemma 2.2(i) we get that $\{c_i \wedge a, c_i \wedge a^*\}$ is a nontrivial classification system of the lattice $(c_i]$. (The system is nontrivial, since $c_i \wedge a = c_i$ would imply $c_i \leqq a$, i.e. $c_i \wedge a^* = 0$, - a contradiction.) Since any $(c_i]$ is a connected lattice, it admits only the trivial classification system. Thus for each $i \in I$ we must have either $c_i \wedge a = 0$ or $c_i \wedge a^* = 0$, i.e. either $c_i \leqq a^*$ or $c_i \leqq a^{**} = a$. Let $K = \{i \in I \mid c_i \leqq a\}$. Then we have $c_i \leqq a^*$ for all $i \in I \setminus K$. Set $b = \bigvee_{i\in K} c_i$ and $c = \bigvee_{i\in I\setminus K} c_i$. Then Proposition 2.8 gives $b, c \in$Cen$L$, and since S is also a classification system, we have $b \vee c = \bigvee_{i \in I} c_i = 1$.

On the other hand, we have $b = \bigvee_{i \in K} c_i \leqq a$ and $c = \bigvee_{i \in I \setminus K} c_i \leqq a^*$, so we get $c \wedge a \leqq a^* \wedge a = 0$ and $c \vee a \geqq c \vee b = 1$, whence $a = \overline{c}$. As the complement of a central element is also central, we obtain that $a \in \mathrm{Cen} L$. □

Corollary 4.5. *Any finite 0-modular pseudocomplemented lattice L is a direct product of connected directly irreducible lattices.*

Proof. In view of Corollary 3.5 any classification system of L is a decomposition system. Since any finite lattice is a CJ-generated complete lattice, we can apply Theorem 4.4 and this gives that L is a direct product of connected lattices, i.e. that $L \cong \prod_{i \in I} L_i$, where any L_i is connected and of course, finite and pseudocomplemented, too. Applying now Corollary 4.3(ii) we get that each L_i is directly irreducible. □

Finally, we present an application of our former results to the theory of Stone lattices.

A bounded pseudocomplemented lattice L is called a *Stone lattice* if $x^* \vee x^{**} = 1$ holds for all $x \in L$. We say that a lattice L with 0 is *handled*, if it contains an atom $\omega \in L$ such that $\omega \leqq x$ hold for all $x \in L \setminus \{0\}$. Clearly, any bounded handled lattice L is a pseudocomplemented lattice where $x^* = 0$ and $x^{**} = 1$, for all $x \in L \setminus \{0\}$. Thus any bounded handled lattice is a Stone lattice. It is also obvious that any CJ-generated handled lattice is connected. In [5], G. Grätzer and E.T. Schmidt gave the following fine characterization of finite distributive Stone lattices:

"A finite distributive lattice is a Stone lattice if and only if it is a direct product of finite distributive handled lattices."

Remark 4.6. It is easy to check that in a distributive Stone lattice L we have $x^* \in \mathrm{Cen} L$ for all $x \in L$. Conversely, if $x^* \in \mathrm{Cen} L$ holds for all $x \in L$ in a bounded lattice L, then L is a Stone lattice. (We note that a lattice with this property is not distributive in general.). In view of these observations, the next theorem can be considered to be a generalization of the above cited result of G. Grätzer and E.T. Schmidt.

Theorem 4.7. *Let L be a CJ-generated complete lattice. Then the following statements are equivalent:*

(i) *L is a pseudocomplemented atomic lattice, and for all $x \in L$ we have $x^* \in \mathrm{Cen} L$.*

(ii) *L is a direct product of handled lattices.*

Proof. (i)⇒(ii). Since any semicentral element $a \in L$ satisfies $a = a^{**}$, the assumption of (i) implies that any semicentral element of L is central. Thus, by applying Theorem 4.4 we get that L is of the form $L \cong \prod_{i \in I} L_i$, where all L_i are connected CJ-generated lattices. Evidently, any lattice L_i, as a factor of L, is a complete pseudocomplemented atomic lattice.

Now we prove that any L_i $(i \in I)$ is a handled lattice. Denote by 0_i and 1_i the smallest and the largest element of L_i, respectively. Let α be an atom of L_i, we have to show that $\alpha \leqq x$ for all $x \in L_i$, $x \neq 0_i$. On the contrary, let us assume that $b \wedge \alpha = 0$, for some $b \in L_i$, $b \neq 0_i$. Then we have $b^* \neq 1_i$ and $b^* \geqq \alpha > 0_i$. The latter relation implies $b^{**} \neq 1_i$, and since $b^{**} \geqq b$, we have $b^{**} \neq 0_i$. Thus we get $b^*, b^{**} \notin \{0_i, 1_i\}$.

Set now an element $z = (z_k)_{k\in I} \in \prod\limits_{k\in I} L_k$, with $z_i = b$. Since by assumption $z^*, z^{**} \in$Cen$L$, and since $z^{**} = \overline{(z^*)}$ the set $\{z^*, z^{**}\}$ is a decomposition system of L. Thus we have $x = (x \wedge z^*) \vee (x \wedge z^{**})$ for all $x \in \prod\limits_{i\in I} L_i$. Clearly, we get $x_i = (x_i \wedge (z^*)_i) \vee (x_i \wedge (z^{**})_i)$ for all $x_i \in L_i$. Observe now, that for all $x = (x_k)_{k\in I} \in \prod\limits_{k\in I} L_k$ we have $x^* = (x_k^*)_{k\in I}$. Hence $(z^*)_i = b^*$ and $(z^{**})_i = b^{**}$. Summarizing, we get that $\{b^*, b^{**}\}$ is a classification system of L_i. Since $b^* \neq 1_i$, this system is nontrivial. As L_i is a connected lattice, this result is in contradiction with Corollary 4.3(i).

(ii)$\Rightarrow$(i). Let $L = \prod\limits_{i\in I} L_i$, with all L_i handled. As L is a bounded lattice, each L_i must be bounded, and consequently, it must be pseudocomplemented. Then it is not hard to check that L itself, as a direct product of pseudocomplemented atomic lattices L_i, is a pseudocomplemented atomic lattice, too. Take now an arbitrary $x = (x_i)_{i\in I} \in L$. Then we have $x^* = (x_i^*)_{i\in I}$. Since any L_i is a bounded handled lattice, we must have $x_i^* = 0_i$, for all $x_i \neq 0_i$, and $x_i^* = 1_i$ otherwise. Thus for all $i \in I$ we have either $(x^*)_i = 0_i$ or $(x^*)_i = 1_i$, and this fact proves that x^* is a central element of the product $\prod\limits_{i\in I} L_i$. Hence $x^* \in$Cen$L$. $\square$

Acknowledgement: The support provided by the Bolyai János Grant of the Hungarian Academy of Sciences, the Hungarian National Foundation for Scientific Research (Grant No. T030243 and No. T029525) and the Higher-Educational Found for Research and Development in Hungary (Grant No. 0486/99) is gratefully acknowledged. The author wishes to express his thanks to professors E.T. Schmidt, J. Szigeti and T. Tóth for their advice and encouragement.

REFERENCES

[1] Bandelt, H.-J.: *Tolerance relations on lattices*, Bull. Austral. Math. Soc., **23**, (1981), 367–381.

[2] Beazer, R.: *Pseudocomplemented algebras with Boolean congruence lattices*, J. Austral. Math. Soc., **26**, (1978), 163–168.

[3] Ganter, B., Wille R.: *Formal Concept Analysis, Mathematical Foundations*, Springer-Verlag, Berlin, Heidelberg, New York, 1999.

[4] Grätzer, G.: *General Lattice Theory*, Birkhäuser-Verlag, Bassel, Stuttgart, 1978.

[5] Grätzer, G. and Schmidt, E. T.: *On a problem of M. H. Stone*, Acta Math. Acad. Sci. Hungar., **8**, (1957), 455-460.

[6] Maeda, F. and Maeda, S.: *Theory of Symmetric Lattices,* Springer-Verlag, 1970.

[7] Radeleczki, S.: *Concept lattices and their application in Group Technology* (in Hun-

garian), Proceedings of International Computer Science Conference: microCAD '98, Miskolc, 1999.

[8] Stern, M.: *Semimodular Lattices, Theory and Applications*, Cambridge University Press, Cambridge, New York, Melbourne, 1999.

Doi: 10.18514/MMN.2000.29

Mathematical Notes, Miskolc, Vol. 1., No. 2., (2000), pp. 157–166

ON THE INVESTIGATION OF SOME NON-LINEAR BOUNDARY VALUE PROBLEMS WITH PARAMETERS

Miklós Rontó
Institute of Mathematics, University of Miskolc
3515 Miskolc – Egyetemváros, Hungary
matronto@gold.uni-miskolc.hu

Tamara Csikos-Marinets
25 Servan Str.
1006 Lausanne, Switzerland
matronto@gold.uni-miskolc.hu

[Received June 20, 2000]

Abstract. A scheme of the numerical-analytic method based upon successive approximations for the investigation of non-linear two-point boundary value problems containing parameters both in the differential equation and in the boundary condition is given.

Mathematical Subject Classification: 34B15
Keywords: Parametrized boundary value problem, numerical-analytic method, determining equation

1. Introduction

The so-called numerical-analytic method (shortly NAM) based upon successive approximations was introduced by the first author jointly with Professor A. Samoilenko [1,2] for the purpose of studying the existence of solutions of non-linear boundary value problems (BVP) and finding approximations to them. For a survey of the further application and development of the NAM to various types of BVPs, including periodic, two-point, multipoint, impulsive, and parametrised ones, one can consult our series of papers in the *Ukrainian Mathematical Journal* joint with Samoilenko and Trofimchuk. The most recently published paper [3] from this series contains the seventh part of the survey. Extentions of NAM to some types of parametrised boundary value problems (PBVPS) can be found in [4,5].

2. Main results

We consider the following two-point non-linear boundary value problem containing parameters both in the given differential equation and in the boundary condition:

$$dx/dt = f(t, x, \lambda_1), \tag{2.1}$$

$$Ax(0) + C(\lambda_1)x(\lambda_2) = d(\lambda_1, \lambda_2), \tag{2.2}$$

$$x_1(0) = x_{10}, \quad x_2(0) = x_{20}. \tag{2.3}$$

Here, we suppose that $x : [0, T] \to \mathbb{R}^n$, $T > 0$ is fixed, the functions $f : \Omega := [0, T] \times D \times [a_1, b_1] \to \mathbb{R}^n$ and $d : I_1 \times I_2 \to \mathbb{R}^n$ are continuous in their domains of definition, $D \subset \mathbb{R}^n (n \geq 3)$ is a closed, connected, and bounded domain, and $\lambda_1 \in I_1 := [a_1, b_1], \lambda_2 \in I_2 := (0, T]$ are unknown scalar parameters. The $n \times n$ matrices A and $C(\lambda_1)$ are supposed to be such that $\det h(\lambda_1) \neq 0$ and $\operatorname{rank} \left[r_{11}(\lambda_1),\ r_{12}(\lambda_1) \right] = 2$ for some real k_1 and k_2 ($k_1 \neq k_2$) and all $\lambda_1 \in I_1$, where $h(\lambda_1) := k_1 A + k_2 C(\lambda_1)$, $H(\lambda_1) := h(\lambda_1)^{-1}$,

$$\begin{bmatrix} r_{11}(\lambda_1) & r_{12}(\lambda_1) \\ r_{21}(\lambda_1) & r_{22}(\lambda_1) \end{bmatrix} = E - k_1 H(\lambda_1) \left[A + C(\lambda_1) \right].$$

(In the equality above, the matrices $r_{11}(\lambda_1)$, $r_{12}(\lambda_1)$, $r_{21}(\lambda_1)$, and $r_{22}(\lambda_1)$ have dimension 2×2, $2 \times (n-2)$, $(n-2) \times 2$, $(n-2) \times (n-2)$, respectively.)

We aim at obtaining the values $\lambda_1^* \in I_1$ and $\lambda_2^* \in I_2$ for which the BVP (2.1), (2.2) has a solution x^* satisfying the additional condition (2.3) for its first and second components. By a solution of (2.1)–(2.3), we thus mean the pair (λ, x), where $\lambda = (\lambda_1,\ \lambda_2)$.

It is obvious that the right-hand side boundary in BVP (2.1)–(2.3) should also be regarded as a parameter.

Let us denote by $|f|$ the column $(|f_1|, |f_2|, \ldots, |f_n|)$. The inequalities between the vectors will be understood component-wise.

With this conventions adopted, we set

$$D_\beta := \{x \in \mathbb{R}^n : B(x, \beta(x)) \subset D\},$$

where $\beta : \mathbb{R}^n \to \mathbb{R}^n$, and $B(x, \beta(x))$ is the $\beta(x)$–neighbourhood of an $x \in \mathbb{R}^n$.

We also assume that the following three conditions hold for the BVP (2.1)–(2.3):

(i) f is continuous on Ω and bounded by some vector $M \in \mathbb{R}^n_+$:

$$|f(t, x, \lambda_1)| \leq M \quad \text{for all } (t, x, \lambda_1) \in \Omega,$$

and is Lipschitzian in the last two variables, i.e.,

$$|f(t, x', \lambda_1') - f(t, x'', \lambda_1'')| \leq K|x' - x''| + |\lambda_1' - \lambda_1''| M_1,$$

where K and M_1 are non-negative matrices of dimension $n \times n$ and $n \times 1$, respectively;

(ii) The set D_β, where

$$\beta(x, \lambda) := \frac{1}{2} T M' + \beta_1(x, \lambda),$$

$$d_1(x,\lambda) := d(\lambda) - [A + C(\lambda_1)]x,$$

$$\beta_1(x,\lambda) := |(k_1 - k_2)H(\lambda_1)\,[d(\lambda_1,\lambda_2) - (A + C(\lambda_1))x]\,| + |k_1 H(\lambda_1) d_1(x,\lambda)|,$$

and

$$M' := \frac{1}{2}\left[\max_{(t,x,\lambda_1)\in\Omega} f(t,x,\lambda_1) - \min_{(t,x,\lambda_1)\in\Omega} f(t,x,\lambda_1)\right],$$

is not empty:

$$D_\beta \neq \emptyset;$$

(iii) The greatest eigen-value $\lambda_{\max}(K)$ of the matrix K satisfies the inequality

$$\lambda_{\max}(K) < \frac{q}{T},$$

where $q = \frac{3}{10}$.

Let us introduce the sequence of functions

$$\begin{aligned} x_{m+1}(t,y,\lambda) &= z(y) + k_1 H(\lambda_1) d_1(z(y),\lambda) + \int_0^t f(s, x_m(s,y,\lambda), \lambda_1)ds \\ &- \frac{t}{\lambda_2}\int_0^{\lambda_2} f(\tau, x_m(\tau,y,\lambda),\lambda_1)d\tau \\ &+ \frac{t}{\lambda_2}(k_2 - k_1)H(\lambda_1)d_1(z(y),\lambda), \end{aligned} \tag{2.4}$$

where

$$\begin{aligned} z = \mathrm{col}(z_1, z_2, \dots, z_i, \dots, z_j, \dots, z_n) \\ = \mathrm{col}(y_1, y_2, \dots, y_i(y), \dots, y_j(y), \dots, y_{n-2}) = z(y), \end{aligned}$$

$y = \mathrm{col}(y_1, y_2, \dots, y_{n-2})$, and $y_i(y)$, $y_j(y)$ are solutions of the first two equations in the system

$$x_{m+1}(0,y,\lambda) = \mathrm{col}(x_{10}, x_{20}, x_3(0), \dots, x_n(0)),$$

i.e., the system

$$[E - k_1 H(\lambda_1)\{A + C(\lambda_1)\}]z = \mathrm{col}(x_{10}, x_{20}, x_3(0), \dots, x_n(0)) - d(\lambda_1,\lambda_2). \tag{2.5}$$

(Here and above, i and j denote the numbers of components of the vector z with respect to which system (2.5) is solvable.)

We set $G = \{y \in \mathbb{R}^{n-2} : z(y) \in D_\beta\}$. One can verify by direct computation that sequence (2.4) depending on the parameters λ_1, λ_2 and on the additional $(n-2)$–dimensional vector y, satisfies the boundary conditions (2.2), (2.3) for arbitrary $\lambda_1 \in I_1$, $\lambda_2 \in I_2$, and $y \in G$.

Theorem 1 *Assume that the conditions (i)–(iii) hold.*

Then:

1. *The sequence (2.4) converges to the function* $x^* = x^*(t, y, \lambda)$ *as* $m \to \infty$ *uniformly in* $(t, y, \lambda) \in [0, T] \times G \times I_1 \times I_2$;

2. *The limit function* x^* *is a solution of the "perturbed" BVP (2.6), (2.2), (2.3),*

$$dx/dt = f(t, x, \lambda_1) + \Delta(y, \lambda), \tag{2.6}$$

with the initial value $x^*(0, y, \lambda) = z(y) + k_1 H(\lambda_1) d_1(z(y), \lambda)$*, where*

$$\Delta(y, \lambda) := \frac{1}{\lambda_2}(k_2 - k_1) H(\lambda_1) d_1(z(y), \lambda) - \frac{1}{\lambda_2} \int_0^{\lambda_2} f(t, x^*(t, y, \lambda), \lambda_1) dt;$$

3. *The following error estimation holds:*

$$\begin{aligned} |x_m(t, y, \lambda) - x^*(t, y, \lambda)| \leq \overline{\alpha}_1(t, \lambda_2)[&Q^m(\lambda_2)(E - Q(\lambda_2))^{-1} M' \\ &+ K Q(\lambda_2)^{m-1}(E - Q(\lambda_2))^{-1} \beta_1(z(y), \lambda)], \end{aligned} \tag{2.7}$$

where $\overline{\alpha}_1(t, \lambda_2) := \frac{10}{9}\alpha_1(t, \lambda_2) \leq \frac{5}{9}\lambda_2$, $\alpha_1(t, \lambda_2) := 2t\left(1 - t\lambda_2^{-1}\right)$, $Q(\lambda_2) := \frac{3\lambda_2}{10} K$.

The *proof* of Theorem 1 can be carried out by using the techniques from [2] (Theorems 16.1, 18.1, and 20.1) and Theorem 1 of [4].

The following statement establishes the relation of the limit function x^* to the solution of the original BVP (2.1)–(2.3).

Theorem 2 *Under the conditions of Theorem 1, the pair* $(x^*(\cdot, y^*, \lambda^*), \lambda^*)$ *is a solution of the BVP (2.1)–(2.3) if, and only if* (y^*, λ^*) *satisfies the determining equation*

$$\Delta(y, \lambda) = \frac{1}{\lambda_2}(k_2 - k_1) H(\lambda_1) d_1(z(y), \lambda) - \frac{1}{\lambda_2} \int_0^{\lambda_2} f(t, x^*(t, y, \lambda), \lambda_1) dt = 0. \tag{2.8}$$

The proof of Theorem 2 is analogous to the corresponding statements from [2] (Theorems 16.3 and 18.3).

3. Sufficient existence conditions

In what follows, we need to consider the mth approximation to the determining equation (2.8):

$$\Delta_m(y, \lambda) := \frac{1}{\lambda_2}(k_2 - k_1) H(\lambda_1) d_1(z(y), \lambda) - \frac{1}{\lambda_2} \int_0^{\lambda_2} f(t, x_m(t, y, \lambda), \lambda_1) dt = 0. \tag{3.1}$$

Theorem 3 *Suppose that, for PBVP* (2.1)–(2.3)*, conditions (i)–(iii) hold and, furthermore,* *(iv) There exists a closed, convex subset*

$$\Omega_1 = G_1 \times I_1' \times I_2' \subset G \times I_1 \times I_2,$$

where, for some $m \geq 1$ *fixed, the approximate determining equation* (3.1) *has only one solution* $(\tilde{y}, \tilde{\lambda})$*, which has non-zero topological index;*

(v) The inequality

$$\inf_{(y,\lambda)\in\partial\Omega} |\Delta_m(y,\lambda)| > \frac{10}{27} \sup_{\lambda\in I_1'\times I_2'} \{\lambda_2 K W(y,\lambda)\} \tag{3.2}$$

is satisfied on the boundary $\partial\Omega_1$ *of the subset* Ω_1*, where*

$$W(x,y) := Q^m(\lambda_2)(E - Q(\lambda_2))^{-1} M' + K Q(\lambda_2)^{m-1}(E - Q(\lambda_2))^{-1}\beta_1(z(y),\lambda).$$

Then, there exists a solution (x^*,λ^*) *of PBVP* (2.1)–(2.3)*, and the initial value* $x^*(0)$ *of this solution at* $t = 0$ *is equal to*

$$z(y^*) + k_1 H d_1(z(y^*),\lambda^*),$$

where $y^* \in G_1$*,* $\lambda_1^* \in I_1'$*, and* $\lambda_2^* \in I_2'$*.*

Proof. Based on inequalities (2.7) and (3.2), similarly to Theorems 3.1 and 17.1 of [2], one can show that the vector fields $\Delta(\cdot,\lambda)$ and $\Delta_m(\cdot,\lambda)$ are homotopic for all λ, which, by the well-known result of degree theory, immediately implies the assertion of Theorem 3. ■

4. Necessary Existence Conditions

The following subsidiary statements will be used in the sequel.

Lemma 4 *Under conditions (i)–(iii), for an arbitrary pair*

$$\{(z',\lambda'),\ (z'',\lambda'')\} \subset D_\beta \times I_1 \times I_2, \tag{4.1}$$

the inequality

$$\begin{aligned} |x^*(t,y',\lambda') - x^*(t,y'',\lambda'')| \leq {} & \left[E + \overline{\alpha}_1(t,\gamma_2) K [E - Q(\gamma_2)]^{-1}\right] \big\{|z(y') - z(y'')| \\ & + b_1(y',y'',\lambda',\lambda'')\big\} \\ & + \overline{\alpha}_1(t,\gamma_2) K [E - Q(\gamma_2)]^{-1} |\lambda_1' - \lambda_1''| M_1 \end{aligned} \tag{4.2}$$

holds, where

$$\begin{aligned} b_1(y',y'',\lambda',\lambda'') := {} & |k_1[H(\lambda_1')d_1(z(y'),\lambda') - H(\lambda_1'')d_1(z(y''),\lambda'')]| \\ & + T|k_2 - k_1| \left| \frac{1}{\lambda_2'} H(\lambda_1')d_1(z(y'),\lambda') - \frac{1}{\lambda_2'} H(\lambda_1'')d_1(z(y''),\lambda'')\right| + 2TM, \end{aligned}$$

$$z(y') = \mathrm{col}(y'_1, y'_2, \dots, y'_{i-1}, y'_i(y'), \dots, y'_j(y'), \dots, y'_{n-2}),$$

$$z(y'') = \mathrm{col}(y''_1, y''_2, \dots, y''_{i-1}, y''_i(y''), \dots, y''_j(y''), \dots, y''_{n-2}),$$

and $\gamma_2 = \max\{\lambda'_2, \lambda''_2\}$.

Proof. By virtue of (2.4), we have

$$\begin{aligned}
x_1(t, y', \lambda') - x_1(t, y'', \lambda'') &= z(y') - z(y'') \\
&+ k_1\left[H(\lambda'_1) d_1(z(y'), \lambda') - H(\lambda''{}_1) d_1(z(y''), \lambda'')\right] \\
&+ \int_0^t \left[f(s, z(y'), \lambda'_1) - f(s, z(y''), \lambda''_1)\right] ds \\
&- \frac{t}{\lambda'_2}\int_0^{\lambda'_2} f(\tau, z(y'), \lambda'_1) d\tau + \int_0^{\lambda''_2} f(\tau, z(y''), \lambda''_1) d\tau \\
&+ \frac{t}{\lambda'_2}(k_2 - k_1) H(\lambda'_1) d_1(z(y'), \lambda') \\
&- \frac{t}{\lambda''_2}(k_2 - k_1) H(\lambda''_1) d_1(z(y''), \lambda'') \\
&= z(y') - z(y'') \\
&+ k_1\left[H(\lambda'_1) d_1(z(y'), \lambda') - H(\lambda''{}_1) d_1(z(y''), \lambda'')\right] \\
&+ \int_0^t \Bigg\{ f(s, z(y'), \lambda'_1) - f(s, z(y''), \lambda''_1) \\
&- \frac{1}{\lambda'_2}\int_0^{\lambda'_2} \left[f(\tau, z(y'), \lambda'_1) - f(\tau, z(y''), \lambda''_1)\right] d\tau \Bigg\} ds \\
&+ \frac{t}{\lambda''_2}\int_0^{\lambda''_2} f(\tau, z(y''), \lambda''_1) d\tau - + \frac{t}{\lambda'_2}\int_0^{\lambda'_2} f(\tau, z(y''), \lambda''_1) d\tau \\
&- t(k_2 - k_1)\Bigg[\frac{1}{\lambda'_2} H(\lambda'_1) d_1(z(y'), \lambda') - \\
&\frac{1}{\lambda''_2} H(\lambda''_1) d_1(z(y'), \lambda'')\Bigg].
\end{aligned}$$

By using the Lipschitz condition on f, similarly to Lemma 19.1 from [2, p. 154], we obtain

$$\begin{aligned}
|x_1(t, y', \lambda') - x_1(t, y'', \lambda'')| &\le [E + \alpha_1(t, \gamma_2) K] |z(y') - z(y'')| \\
&+ \alpha_1(t, \gamma_2) |\lambda'_1 - \lambda''_1| M_1 \\
&+ b_1(y', y'', \lambda', \lambda'').
\end{aligned}$$

One can prove by induction that

$$|x_m(t,y',\lambda') - x_m(t,y'',\lambda'')| \le \sum_{i=0}^{m} \alpha_i(t,\gamma_2) K^i |z(y') - z(y'')| + \sum_{i=0}^{m} \alpha_i(t,\gamma_2) K^{i-1} |\lambda_1' - \lambda_1''| M_1 + \sum_{i=0}^{m-1} \alpha_i(t,\gamma_2) K^i b_1(y',y'',\lambda',\lambda''), \tag{4.3}$$

where (*see*, e.g., [2, p. 148] or [5])

$$\alpha_{m+1}(t,\gamma) := \left(1 - \frac{t}{\gamma}\right) \int_0^t \alpha_m(s,\gamma) ds + \frac{t}{\gamma} \int_t^\gamma \alpha_m(s,\gamma) ds,$$

and $\alpha_0(t,\gamma) \equiv 1$.

Taking into account estimate (*see* Lemma 4 in [6])

$$\alpha_{m+1}(t,\gamma) \le \left(\frac{3}{10}\gamma\right)^m \overline{\alpha}_1(t,\gamma),$$
$$\alpha_1(t,\gamma) = \frac{10}{9}\overline{\alpha}_1(t,\gamma) \le \frac{5}{9}\gamma$$

and passing to the limit as $m \to \infty$ in (4.3), we obtain the required inequality (4.2). ∎

Lemma 5 *Let us suppose that BVP* (2.1)–(2.3) *satisfies conditions (i)–(iii).*

Then the determining function Δ is continuous in the domain $G \times I_1 \times I_2$ and, for arbitrary pairs (4.1)*, the following relation holds:*

$$|\Delta(y',\lambda') - \Delta(y'',\lambda'')| \le b_2(y',y'',\lambda',\lambda'') + \frac{\gamma_2}{\gamma_1}|\lambda_1' - \lambda_1''| M_1 + \frac{\gamma_2}{\gamma_1} K \left[E + \frac{10}{27}\gamma_2 K (E - Q(\gamma_2))^{-1}\right] \Big(|z(y') - z(y'')| + b_1(y',y'',\lambda',\lambda'')\Big) =: \epsilon(\Delta(y',\lambda'), \Delta(y'',\lambda'')), \tag{4.4}$$

where

$$b_2(y',y'',\lambda',\lambda'') := |k_2 - k_1| \left| \frac{1}{\lambda_2'} H(\lambda_1') d_1(z(y'),\lambda') - \frac{1}{\lambda_2'} H(\lambda_1') d_1(z(y'),\lambda') \right| + 2M$$

and $\gamma_1 := \min\{\lambda_2', \lambda_2''\}$.

Proof. For every $\{y', y''\} \subset G$ such that $\{z(y'), z(y'')\} \subset D_\beta$, there exists a continuous limit function of the uniformly convergent function sequence (2.4). The determining function is thus also continous and bounded in the domain $G \times I_1 \times I_2$. Due to the form of the function Δ in (2.8), we have

$$\begin{aligned}\Delta(y',\lambda') - \Delta(y'',\lambda'') &= \frac{1}{\lambda_2'}(k_2-k_1)H(\lambda_1)d_1(z(y'),\lambda') \\ &\quad - \frac{1}{\lambda_2''}(k_2-k_1)H(\lambda_1)d_1(z(y''),\lambda'') \\ &\quad - \frac{1}{\lambda_2'}\int_0^{\lambda_2'} f(t,x^*(t,y',\lambda'),\lambda_1')dt + \frac{1}{\lambda_2''}\int_0^{\lambda_2''} f(t,x^*(t,y'',\lambda''),\lambda_1'')dt.\end{aligned}$$

By direct computation, using the Lipschitz condition on f and estimate (4.2), we obtain

$$\begin{aligned}|\Delta(y',\lambda') - \Delta(y'',\lambda'')| &\le b_2(y',y'',\lambda',\lambda'') \\ &\quad + \left|\frac{1}{\lambda_2'}\int_0^{\lambda_2'}\left[f(t,x^*(t,y',\lambda'),\lambda_1') - f(t,x^*(t,y'',\lambda''),\lambda_1'')\right]dt\right| \\ &\le b_2(y',y'',\lambda',\lambda'') + \frac{1}{\lambda_2'}\int_0^{\lambda_2'}\left\{K|x^*(t,y',\lambda') - x^*(t,y'',\lambda'')| + |\lambda_1'-\lambda_1''|M_1\right\}dt \\ &\le b_2(y',y'',\lambda',\lambda'') + \frac{1}{\lambda_2'}\int_0^{\lambda_2'}\Big\{K\Big[\left(E+\overline{\alpha}_1(t,\gamma_2)K(E-Q(\gamma_2))^{-1}\right)|z(y')-z(y'')| \\ &\quad + \left(E+\overline{\alpha}_1(t,\gamma_2)K(E-Q(\gamma_2))^{-1}\right)b_1(y',y'',\lambda',\lambda'')\Big] + |\lambda_1'-\lambda_1''|M_1\Big\}dt \\ &\le b_2(y',y'',\lambda',\lambda'') \\ &\quad + \frac{\gamma_2}{\gamma_1}K\left[E+\frac{10}{27}\gamma_2K(E-Q(\gamma_2))^{-1}\right]\Big(|z(y')-z(y'')| + b_1(y',y'',\lambda',\lambda'')\Big) \\ &\quad + \frac{\gamma_2}{\gamma_1}|\lambda_1'-\lambda_1''|M_1 \le \epsilon(\Delta(y',\lambda'),\Delta(y'',\lambda'')),\end{aligned}$$

as required. ■

The following statement gives a necessary condition for the existence of solutions of PBVP (2.1)–(2.3).

Theorem 6 *Assume that conditions (i)–(iii) hold. Then the subset*

$$\Omega_2 = G_2 \times I_1'' \times I_2'' \subset G \times I_1 \times I_2$$

may contain a pair (y^*,λ^*) *generating a solution*

$$x^*(t,y^*,\lambda^*) = \lim_{m\to\infty} x_m(t,y^*,\lambda^*)$$

of PBVP (2.1)–(2.3) *only if, for every* $m \ge 1$ *and every pair* $(\tilde{y},\tilde{\lambda})$*, the following* *relation holds true:*

$$\Delta_m(\tilde{y},\tilde{\lambda}) \leq \sup_{(y,\lambda)\in\Omega_2} \Bigg\{ b_2(\tilde{y},y,\tilde{\lambda},\lambda) + \frac{\gamma_2}{\gamma_1}|\tilde{\lambda}_1 - \lambda_1|M_1 + \frac{\gamma_2}{\gamma_1}K\left[E + \frac{10}{27}\gamma_2 K(E-Q(\gamma_2))^{-1}\right]\Big(|z(\tilde{y}) - z(y)| + b_1(\tilde{y},y,\tilde{\lambda},\lambda)\Big)\Bigg\} + \epsilon(\Delta(\tilde{y},\tilde{\lambda}),\Delta_m(\tilde{y},\tilde{\lambda})). \quad (4.5)$$

Proof. Let the determining function Δ vanish at $y = y^*$, $\lambda = \lambda^*$, i.e., that $x^*(\cdot, y^*, \lambda^*)$ is a solution of the PBVP (2.1)–(2.3). Rewriting inequality (4.4) for the pairs $(y', \lambda') = (\tilde{y}, \tilde{\lambda})$ and $(y'', \lambda'') = (y^*, \lambda^*)$, we obtain

$$|\Delta(\tilde{y},\tilde{\lambda})| \leq b_2(\tilde{y},y^*,\tilde{\lambda},\lambda^*) + \frac{\gamma_2}{\gamma_1}|\tilde{\lambda}_1 - \lambda_1^*|M_1 + \frac{\gamma_2}{\gamma_1}K\left[E + \frac{10}{27}\gamma_2 K(E-Q(\gamma_2))^{-1}\right]\Big(|z(\tilde{y}) - z(y^*)| + b_1(\tilde{y},y^*,\tilde{\lambda},\lambda^*)\Big).$$

Relations (2.8) and (3.1) yield

$$|\Delta(y,\lambda) - \Delta_m(y,\lambda)| = \left|\frac{1}{\lambda_2}\int_0^{\lambda_2} [f(t,x^*(t,y,\lambda),\lambda_1) - f(t,x_m(t,y,\lambda),\lambda_1)]\,dt\right|$$

$$\leq \frac{1}{\lambda_2}KW(y,\lambda)\int_0^{\lambda_2}\overline{\alpha}_1(t,\lambda_2)dt = \frac{10}{27}\lambda_2 KW(y,\lambda) = \epsilon(\Delta_m(y,\lambda),\Delta_m(y,\lambda)). \quad (4.6)$$

Relation (4.6) with $(y,\lambda) = (\tilde{y},\tilde{\lambda})$ implies

$$|\Delta_m(\tilde{y},\tilde{\lambda})| \leq |\Delta(\tilde{y},\tilde{\lambda})| + \epsilon(\Delta_m(\tilde{y},\tilde{\lambda}),\Delta_m(\tilde{y},\tilde{\lambda})). \quad (4.7)$$

Combining (4.6) and (4.7), we obtain the desired necessary condition (4.5). ■

Acknowledgement: The research was partially supported by the Hungarian-Ukrainian Intergovernmental Scientific and Technical Cooperation Programme, Grant UK–3/1999 and the Hungarian National Research Foundation, Grant T031961.

REFERENCES

[1] Samoilenko, A. M. and Ronto, N. I.: *Numerical-Analytic Methods of Investigating Periodic Solutions,* Mir Publishers, Moscow, 1980.

[2] Samoilenko, A. M. and Ronto, N.I.: *Numerical-Analytic Methods in the Theory of Boundary-Value Problems*, World Scientific, Singapore, 2000.

[3] Ronto, N. I., Samoilenko, A. M. and Trofimchuk, S. I.: *Theory of numerical-analytic method: Achievements and new trends of development. VII.*, Ukr. Mat. Zh., **51**(9), (1999), 1244–1261.

[4] Ronto, N. I.: *On numerical-analytic method for boundary value problems with parameters*, Publ. Univ. Miskolc, Ser. D. Natur. Sci. Math., **36**(2), (1996), 125–132.

[5] Ronto, N. I.: *On some existence results for parametrized boundary value problems*, Publ. Univ. Miskolc, Ser. D. Natur. Sci. Math., **37**(2), (1997), 95–103.

[6] Rontó, M. and Mészáros, J.: *Some remarks on the convergence analysis of the numerical-analytic method based upon successive approximations*, Ukr. Mat. Zh., **48**(1), (1996), 90–95.

Doi: 10.18514/MMN.2000.30

Mathematical Notes, Miskolc, Vol. 1., No. 2., (2000), pp. 167–170

SOLUTIONS OF A SYSTEM OF SECOND-ORDER ORDINARY DIFFERENTIAL EQUATIONS

Kurt W. Tomantschger
Department of Mathematics, Technical University of Graz
Steyrergasse 30, A-8010 Graz, Austria
toman@weyl.math.tu-graz.ac.at

[Received March 24, 2000]

Abstract. A general solution of a coupled system of second-order ordinary differential equations is obtained. The solution is represented by modified Bessel functions of the first and second kind. The corresponding boundary conditions are also given. The boundary value problem describes the motion of a micropolar suspension between two coaxial cylinders, and the two unknown functions are, respectively, the velocity and the velocity of microrotation of the micropolar theory.

Mathematical Subject Classification: 34B05, 33C10
Keywords: Coupled linear ordinary differential equations, modified Bessel functions, MacDonald functions

1. Introduction

Classical continuum mechanics is not sufficient to describe the behaviour of certain materials (granular materials, fluid suspensions, liquid crystals, blood flow, polymeric substances, composite materials, etc.). For this reason the continuum with microstructure was intruduced [1]. The new model is called the micropolar continuum, and it possesses two independent kinematic quantities: the velocity vector $\vec{v}$ (macromotion) and the spin, or microrotation vector $\vec{\nu}$.

Since 1965, the micropolar theory has attracted a great deal of attention.

Paper [2] considers the stationary motion of a suspension between two coaxial cylinders. The inner cylinder with the radius R_1 is immobile, while the outer one with the radius R_2 ($R_2 > R_1$) rotates with angular velocity Ω. The axis of rotation is horizontal.

Applying the basic relations and balance laws of micropolar theory, the following result was obtained: The behaviour of the micropolar suspension between the two coaxial cylinders is described by the system of the following two coupled differential equations [2]:

$$(\mu + k)[r^2 v'' + rv' - v] - kr^2 \nu' = 0, \tag{1}$$

$$\gamma\,[r\nu'' + \nu'] + k\,(vr)' - 2kr\nu \;=\; 0, \qquad R_1 \le r \le R_2\,. \tag{2}$$

Hence, $v(r)$ represents the velocity of the suspension (macromotion); $\nu(r)$ is the microrotational velocity r is one of the coordinates of the cylindrical system; γ, μ, and k denote viscosity coefficients of the micropolar continuum (they are positive constants).

We treat this kind of motion of suspension when the suspension flows down the walls of the cylinders, so that the boundary conditions for the velocity $v(r)$ and the microrotation velocity $\nu(r)$ are

$$v(R_1) = 0, \quad v(R_2) = \Omega R_2 \qquad \text{and} \qquad \nu(R_1) = 0, \quad \nu(R_2) = 0\,. \tag{3}$$

In this paper, the general solution $v(r)$ and $\nu(r)$ of the differential equations (1) and (2) are determined. The arbitrary constants of this solution are calculated on the basis of the boundary conditions (3). Finally, we show, that in a special case, we obtain from our result the same velocity $v(r)$ which we would have got if we solved this technical problem taking into account the laws of classical physics.

2. Solution of the coupled system

After multiplying (2) by r, the terms in the brackets of (1) and (2) are brought to the form $r^n y^{(n)}$, $y = y(r)$. These expressions may be reduced to derivatives with constant coefficients if we substitute

$$r = R_2 e^{-t}, \qquad \text{resp.} \qquad t = \ln R_2 - \ln r. \tag{4}$$

If the change of variable (4) is applied to $v(r)$ and $\nu(r)$, we determine the resulting functions

$$V(t) = v(R_2 e^{-t}) = v(r) \qquad \text{and} \qquad U(t) = \nu(R_2 e^{-t}) = \nu(r). \tag{5}$$

Substituting (4), (5), $rv' = -\dot V$, $r^2 v'' = \ddot V + \dot V$ and the analogous ν derivatives into equations (1) and (2) then multiplying by r, we obtain

$$\ddot V - V = -kR_2(\mu + k)^{-1}\, e^{-t}\, \dot U, \tag{6}$$

$$\ddot U - 2k\gamma^{-1} R_2^2 e^{-2t} U = k\gamma^{-1} R_2 e^{-t}\, (\dot V - V). \tag{7}$$

After multyplying the above equation by $\gamma e^t/(kR_2)$ and differentiatiing it, we have

$$\dot V - V = \frac{\gamma}{kR_2}\, e^t \ddot U - 2R_2 e^{-t} U, \quad \ddot V - \dot V = \frac{\gamma}{kR_2} \frac{d(e^t \ddot U)}{dt} - 2R_2 \frac{d(e^{-t} U)}{dt}\,. \tag{8}$$

These two equations and equation (6) are a system of three independent differential equations. Consequently, V, $\dot V$ and $\ddot V$ can be eliminated. In this way we obtain

$$\dddot U \,+\, 2\,\ddot U \,-\, M^2 e^{-2t}\, \dot U \;=\; 0, \qquad M = \left[\frac{k}{\gamma}\left(\frac{2\mu + k}{\mu + k}\right)\right]^{1/2} R_2\,. \tag{9}$$

Suppose the Laplace transform of $U(t)$ is denoted by $\Upsilon(s)$. Then we take the Laplace transform of both sides of equation (9). Dividing the transformed equation by $s^2(s+2)$, then using the method of partial fraction and taking into account the boundary conditions $U(0) = \nu(R_2) = 0$, we get

$$\Upsilon(s) - \frac{M^2}{s^2}\,\Upsilon(s+2) = \frac{1}{s^2}\,\dot{U}(0) + \left(-\frac{1}{4s} + \frac{1}{2s^2} + \frac{1}{4(s+2)}\right)\ddot{U}(0)\,.$$

Applying to this relation the inverse Laplace transform on this relation, we obtain the integral equation

$$U(t) - M^2 \int_0^t \int_0^{t_1} e^{-2t_2}\, U(t_2)\, dt_2\, dt_1 = t\,\dot{U}(0) + [-1 + 2t + e^{-2t}]\frac{\ddot{U}(0)}{4}\,.$$

Introducing the new variables

$$t_1 = \ln(M/x_1)\,, \qquad t_2 = \ln(M/x_2)\,, \qquad t = \ln(M/x) \tag{10}$$

and setting

$$w(\xi) = U[\ln(M/\xi)]$$

we bring the integral equation to the form

$$w(x) - \int_M^x \int_M^{x_1} x_2\, w(x_2)dx_2\, \frac{dx_1}{x_1} = \ln(\frac{M}{x})\,\dot{U}(0) + [-1 + 2\ln\frac{M}{x} + \frac{x^2}{M^2}]\frac{\ddot{U}(0)}{4}\,.$$

If we apply to the transformed integral equation the differential operator $d/dx\,(x\,d/dx)$, we obtain the modified inhomogeneous Bessel equation with zero order

$$xw'' + w' - xw = xM^{-2}\ddot{U}(0), \qquad w = w(x).$$

The general solution of this differential equation is defined by

$$w(x) = C_1 I_0(x) + C_2 K_0(x) - M^{-2}\ddot{U}(0). \tag{11}$$

I_n (resp. K_n) is called the *modified Bessel function of the first* (resp., *second kind) of order* n. K_n is also known as *MacDonald function*.

Substituting $x = Me^{-t}$ (see (10)) and (11) into the first equation (8), we get

$$x\tilde{V}'(x) + \tilde{V}(x) = \frac{kR_2}{\mu + k}\,\frac{x}{M}\,(C_1 I_0(x) + C_2 K_0(x)) - 2R_2\,\frac{x}{M^3}\,\ddot{U}(0)\,. \tag{12}$$

By variation of constants we obtain a particular solution of the form $\tilde{V}_p = C(x)x^{-1}$. The general solution of equation (7) (resp. (12)) is now explicitely expressed:

$$\tilde{V} = -\frac{R_2}{M^3}\,\ddot{U}(0)\,x + C_3\frac{1}{x} + \frac{kR_2}{\mu + k}\,\frac{1}{M}\,(C_1 I_1(x) - C_2 K_1(x))\,. \tag{13}$$

Substituting $e^{-t} = x/M$ into (4), using (5), $x = Me^{-t}$, (11) and (13), we find

$$v = -\frac{\ddot{U}(0)}{M^2}\, r + \tilde{C}_3 \frac{1}{r} + \frac{k}{\mu + k}\frac{1}{N}\, \left(C_1 I_1(Nr) - C_2 K_1(Nr)\right)\,, \tag{14}$$

$$\nu = C_1 I_0(Nr) + C_2 K_0(Nr) - \frac{\ddot{U}(0)}{M^2}\,, \quad N = M/R_2\,. \tag{15}$$

These two functions satisfy equations (1) and (2) identically. To determine the arbitrary constants, we apply the boundary conditions (3) to (14) and (15). This yields

$$\begin{aligned}
C_1 &= \frac{\Omega R_2^2}{H}[K_0(a_2) - K_0(a_1)]\,, \qquad C_2 = \frac{\Omega R_2^2}{H}[I_0(a_2) - I_0(a_1)]\,,\\
\frac{\ddot{U}(0)}{M^2} &= \frac{\Omega R_2^2}{H}[I_0(a_1)K_0(a_2) - I_0(a_2)K_0(a_1)], \qquad a_j = NR_j, \quad j = 1, 2,\\
\tilde{C}_3 &= \frac{\Omega R_1 R_2^2}{H}\,(I_1(a_1)[K_0(a_1) - K_0(a_2)] + K_1(a_1)[I_0(a_1) - I_0(a_2)] + \\
&\qquad + R_1[I_0(a_1)K_0(a_2) - I_0(a_2)K_0(a_1)])\,,\\
H &= (R_1^2 - R_2^2)[I_0(a_1)K_0(a_2) - I_0(a_2)K_0(a_1)] + [K_0(a_1) - K_0(a_2)] * \\
&\quad *[R_1 I_1(a_1) - R_2 I_1(a_2)] + [I_0(a_1) - I_0(a_2)][R_1 K_1(a_1) - R_2 K_1(a_2)].
\end{aligned}$$

3. Conclusion

The classic case of suspension motion is obtained if the material constant k of the micropolar suspension is equal to zero. In that case, the relation for the velocity of suspension motion is reduced to the well-known form

$$v = \frac{\Omega}{R_2^2 - R_1^2}\left[R_2^2\, r - \frac{R_1^2 R_2^2}{r}\right]; \qquad \nu = 0.$$

REFERENCES

[1] Eringen, A.C. and Suhubi, E.S.: *Nonlinear theory of simple microelastic solids*, Int. J. Engng. Sci., **2** (1964), 189-202.

[2] Cvetković, P., Tomantschger, K.W., Kuzmanović, D. and Golubović, Z.: *The motion of a suspension between two coaxial cylinders*, Facta Universitatis, in press.

[3] Eringen, A.C.: *Simple microfluids*, Int. J. Engng. Sci., **2,** (1964), 205-217.

[4] Tomantschger, K.W., Cvetković, P., Kuzmanović, D. and Golubović, Z.: *Solving the problem of motion of a micropolar suspension*, *ZAMM*, in press.

Twenty offprints of each paper will be provided free of charge and mailed to the corresponding author.

The Mathematical Notes of the University of Miskolc (Series D of the Publications of the University of Miskolc) is abstracted in Zentralblatt für Mathematik, Mathematical Reviews, and Referativnyij Zhurnal

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: 200
Put to the Press on December 15, 2000
Number of permission: TU 00-40-ME **HU e-ISSN 1787-2413**