Issue 2 JULY–DECEMBER 2001 ISSN 1083-0464

SOUTHWEST JOURNAL
OF
PURE AND APPLIED MATHEMATICS

CAMERON UNIVERSITY
DEPARTMENT OF MATHEMATICAL SCIENCES
LAWTON, OK 73505, U.S.A

Southwest Journal
of
Pure and Applied Mathematics

Editor in Chief: Ioannis Argyros
Managing Editor: Mohammad Tabatabai
Technical Editor: Gary Huckabay

Editorial Office

SWJPAM
Cameron University
Department of Mathematical Sciences
Lawton, OK 73505, U.S.A.

Publication Program

2 issues per year
July, December

Information for Authors

Scope. SWJPAM is an electronic journal devoted to all aspects of Pure and Applied mathematics, and related topics. Authoritative expository and survey articles on subjects of special interest are also welcomed.

Since SWJPAM aims to serve not only the specialists, but also the user of mathematics (promoting, in such a way, a strong interaction between them), particular emphasis is placed on applications. Thus, SWJPAM publishes papers which, although not directly contributing to mathematics, contain, however, significant results (from any field) obtained by making an essential and well stressed use of the methods and results of mathematics.

Submission of Manuscripts. All manuscripts must be submitted in AMSTEXformat. Papers for publication should be sent as e-mail attachments to swj@rattler.cameron.edu. The articles must be placed in the 'SWJPAM_incoming' directory. The submission should include the title of the paper, the author's name and affiliation, and a short abstract which summarizes the paper; a list of keywords and a classification according to the Mathematical Reviews' Subject Classification (1991). Authors are informed by e-mail of referee communication.

Key Words. Key words should be comprehensive and subject specific. It is not necessary to list the subject area of the journal's coverage as a key word. Four key words should be sufficient to cover the major subjects off a given paper, although more can always be supplied if the author deems it necessary.

(continued on inside back cover)

SOUTHWEST JOURNAL
OF
PURE AND APPLIED MATHEMATICS
ISSUE 2, DECEMBER–JULY 2001

CONTENTS

ISSN 1083-0464

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 2 December 2001, pp. 1 – 6
Submitted: April 4, 2001. Published: December 31, 2001

On σ-Chebyshev Subspaces of Banach Spaces

H. Mazaheri and A. Dehghan Nezhad

Abstract

σ-Chebyshev subspaces of a Banach space X are defined as those in which the set of best approximants of any vector x in X is non-empty and σ-compact. Moreover, it is shown that there exists a proximinal subspace G of X which is σ-Chebyshev, but is not quasi-Chebyshev in X. Also, some other related results are presented.

(1991) A.M.S Subject Classification Codes. 46B25, 46A20, 46A25, 41A65.
Keywords and Phrases. σ-Chebyshev subspace, Quasi-Chebyshev subspace, Reflexive Banach space.

1 Introduction

Let X be a (complex or real) Banach space and let G be a linear subspace of X. A point $y_0 \in G$ is said to be a best approximation for $x \in X$ if

$$\|x - y_0\| = d(x, G) = \inf\{\|x - y\| : y \in G\}.$$

Departments of Mathematics, Yazd University, Yazd, Iran
E-mail : hmazaheri@yazduni.net, anezhad@yazduni.net

If each $x \in X$ has at least one best approximation in G, then G is called a proximinal subspace of X. If each $x \in X$ has a unique best approximation in G, then G is called a Chebyshev subspace of X (see[7,8]). For $x \in X$, put

$$P_G(x) = \{y \in G : \|x - y\| = d(x, G)\},$$

and

$$\hat{G} = P_G^{-1}(0) = \{x \in X : \|x\| = d(x, G)\}.$$

It is clear that $P_G(x)$ is a closed, bounded and convex subset of X. We say that G is a quasi-Chebyshev subspace of X if $P_G(x)$ is a non-empty and compact set in X for every $x \in X$.

In the following we give a definition which extends the definition of quasi-Chebyshev subspace.

1.1. Definition. Let X be a Banach space. A linear subspace G of X is called σ-Chebyshev if $P_G(x)$ is non-empty and σ-compact set in X for every $x \in X$ (see[3,4,5,6]).

It is clear that every quasi-Chebyshev subspace is σ-Chebyshev. In the following we shall give an example in which the converse is not true.

Let X^* be the dual space of the Banach space X. For $0 \neq f \in X^*$, put

$$M_f = \{x \in X : f(x) = \|f\|, \|x\| = 1\}.$$

We say that a Banach space X is quasi-strictly convex if for every $0 \neq f \in X^*$ the set M_f is compact (see [7]). Every k-strictly convex Banach space is quasi-strictly convex $(k = 1, 2, \ldots)$.

1.2. Lemma([1]). Let $C_R([0, 1])$ be class of all continuous real functions on $[0, 1]$ with the supremum norm. Let $c_0, c_1 > 0$ and let S consist of all differentiable functions $f : [0, 1] \longrightarrow R$ such that $\|f\| \leq c_0$ and $\|f^{'}\| \leq c_1$ then S is a compact subset of $C_R([0, 1])$.

2 Introduction

Now, we are ready to state and prove our main results.

2.1. lemma. *Let X be a Banach space and let G is a proximinal subspace of X. If $\hat{G}$ is a σ-compact set, then G is a σ-Chebyshev subspace.*

Proof. Let $x \in X \backslash G$. We have

$$P_G(x) = \{y \in G : \|x - y\| = d(x, G)\}$$
$$= \{y \in G : x - y \in \hat{G}\} = G \bigcap (x - \hat{G})$$

.
Since $\hat{G}$ is a σ-Chebyshev set, then $x - \hat{G}$ and hence $P_G(x)$ is σ-compact set.$\square$

2.2.Example. Let $X = G^1_R([0, 1])$ be the Banach space of all real-valued differentiable continuously functions on $[0, 1]$, and equipped with the supremum norm. Suppose $G = \{g \in G^1_R([0, 1]) : g(0) = 0\}$. Then for every $f \in X$, we have $g = f - f(0) \in G$ is a best approximation for f in X. In fact

$\|f - g\| = |f(0)| = |f(0) - h(0)| \leq \|f - h\|$, for all $h \in G$.

Thus G is proximinal subspace of X. Now it is sufficent that prove $\hat{G}$ is a σ-Chebyshev set.

We have

$$\hat{G} = \{f \in C^1([0, 1]) : \|f\| = d(f, G)\}$$
$$= \{f \in C^1([0, 1]) : \|f\| = |f(0)|\}$$
$$= \bigcup_{n=1}^{\infty} \{f \in C^1([0, 1]) : \|f\| = |f(0)|, \|f^{'}\| \leq n\}$$

.

By lemma 1.3, the set $\{f \in C^1([0, 1]) : \|f\| = |f(0)|, \|f^{'}\| \leq n\}$ is a compact subset of $C^1([0, 1])$. Therefore $\hat{G}$ is σ-Chebyshev subspace of X. Now, we show that G is not quasi-Chebyshev subspace of X.

For that, let $f = 1$ and $g_n(x) = 1 - cosnx(n = 1, 2, ..., x \in [0, 1])$. It is clear that $g_n \in P_G(f)$ and $g_n(1/n) = 1 - cos1$ for every $n = 1, 2, ...,$. Since $g_n(0) = 1(n = 1, 2, ...)$ and $g(0) = 1$ for all $g \in G$, it follows that $\{g_n\}_{n \geq 1}$ has not any uniformly convergent subsequence in $G_G(f)$. Hence G is not a quasi-Chebyshev subspace of $C^1_R([0,1])$.

2.3. Theorem. *Let X be a Banach space and let G is a proximinal subspace of X. If M_f is σ-compact for every $0 \neq f \in G^{\perp}$, then G is σ-Chebyshev in X.*

Proof. Let M_f is σ-compact for every $0 \neq f \in G^{\perp}$. It is clear that for each $0 \neq f \in G^{\perp}$ and each $x_0 \in M_f$

$$x_0 - P_G(x_0) = M_f \bigcap (x_0 + G)$$

.

Now, suppose that $x_0 \in X - G$, put $y_0 = x_0/\|x_0\|$. Then by an consequence of the Hahn-Banach theorem, there exists $0 \neq f \in G^{\perp}$ such that $\|f_0\| = f_0(y_0)$, therefore $y_0 \in M_f$. Since M_{f_0} is σ-compact, then $y_0 - P_G(y_0)$ is σ-compact and hence $P_G(x_0) = \|x_0\| P_G(y_0)$ is σ-compact. $\square$

The following example show that the inverse of Theorem 2.3 does not hold in general.

Example 2.4. Let ℓ^1 with the basis $\{e_n\}_{n\geq 1}$. Put $X = \ell^1 \bigoplus_1 \Re$and $G = \{0\} \times \Re$. First, we show that G is a σ-Chebyshev subspace of X. Let $z = (x, y) \in X$ be arbitrary. let $g = (0, y)$. Then $g \in G$and we have

$$\|z - g\|_1 = \|(x, y) - (0, y)\|_1$$

$$= \|(x, 0)\|_1 = \|x\|_1 \leq \|x\|_1 + |y - u|$$

$$\leq \|(x, y) - (0, u)\|_1$$

for all $(o, u) \in G$. It follows that $g \in P_G(z)$ and hence $P_G(z)$ is non-empty for every $z \in X$. Therefore, G is a proximinal subspace of X.

It is not difficult to show that for every $z = (x, y) \in X$ we have $P_G(z) = \{(o, y)\}$, hence $P_G(z)$ is σ-compact for each $(x, y) \in X$.

Finally, define $f_0 : X \longrightarrow \mathcal{C}$ by $f_0(x,y) = \sum_{n=1}^{\infty} \mathcal{C}_n$ for all $(x,y) \in X$, where $x = \sum_{n=1}^{\infty} e_n$ and $\mathcal{C}$ is the field of complex numbers. Then we have

$$|f_0(x,y)| = |\sum_{n=1}^{\infty} \mathcal{C}_n| \leq \sum_{n=1}^{\infty} |\mathcal{C}_n| = \|x\|_1 \leq \|x\|_1 + |y| = \|(x,y)\|_1$$

for all $(x,y) \in X$. It follows that $f_0 \in X^*$, $f_0|_G = 0$ and $\|f_0\| = 1$. But, we have

$$\begin{aligned} M_{f_0} &= \{(x,y) \in X : f_0(x,y) = \|f_0\|, \|(x,y)\|_1 = 1\} \\ &= \{(x,y) \in X : \sum_{n=1}^{\infty} \mathcal{C}_n = 1, \|x\|_1 + |y| = 1\} \\ &= \{(x,y) \in X : \|x\|_1 = 1, y = 0\} = \{(x,0) \in X : \|x\|_1 = 1\} \\ &= S_{\ell^1} \times \{0\} \end{aligned}$$

where $S_{\ell^1} = \{x \in \ell^1 : \|x\|_1 = 1\}$.

Hence M_{f_0} is not σ-compact.

References

1. Bridges, D.S., Foundations of Real and Abstract Analysis, Springer-Verlag, 1997.

2. Conway, J.B., A Coures in Functional Analysis, Springer-Verlag, 1985.

3. Mohebi, H., On Quasi-Chebyshev Subspaces of Banach Spaces, to appear in Journal of Approximation Theory (2000).

4. Mohebi, H., Pseudo-Chebyshev Subspaces in L^1, The Korean Journal of Computational and Applied Mathematics, 7(2000), 465-475.

5. Mohebi, H., Pseudo-Chebyshev Subspaces in Dual Spaces, Journal of Natural Geometry, 17(2000), 93-104.

6. Mohebi, H., Radjavi, H., On Compactness of the Best Approximant Set, Submitted.

7. Singer, I., The Theory of Best Approximation and Functional Analysis, Regional conference series in Applied Mathematics, 13(1974).

8. Singer, I., Best Approximation in Normal Linear Spaces by Elements of Linear Subspaces, Springer-Verlag, 1970.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2001, pp. 7–18.
Submitted: July 18, 2001. Published: December 31 2001.

STABILITY FOR NONLINEAR 2-D MULTIRESOLUTION: THE APPROACH OF A.HARTEN

S. AMAT AND S. BUSQUIER

Abstract. In this paper a class of multiresolution representation based on nonlinear prediction in 2-D is presented. In contrast with the wavelet framework these representations cannot be consider as a change of basis, and the stability theory requires different considerations. Some error-control algorithms in order to obtain stability are proposed.

1991 A.M.S. (MOS) Subject Classification Codes. 41A05, 41A15

Key Words and Phrases. Multiscale decomposition, nonlinearity, stability

1. Introduction

Multiresolution representations are a powerful tool for several applications like data compression, numerical simulation of PDE's and statistical estimation. Their best property is to obtain high accuracy approximations of functions by a very small set of coefficients. In wavelet basis decompositions the accuracy is limited by the presence of singularities. Nonlinear multiscale representations are a possible track for such improvements. We will consider the multiresolution framework introduced by Harten (see e.g. [9] [10]). Formally speaking, the multiresolution of Harten has a similar form than the original wavelet representations: a sequence $\bar{f}^L$ of data sampled at the finest resolution level is encoded into $(\bar{f}^0, e^1, \dots, e^L)$, where the $\bar{f}^0$ corresponds to a sampling at the coarsest resolution and each sequence e^k represents the details which are necessary to recover $\bar{f}^k$ from $\bar{f}^{k-1}$. Nevertheless, in the case of Harten's multiresolution the process which connects two scales can be nonlinear. The nonlinear process allows a better adapted treatment of singularities, in the sense that they do not generate so many large detail coefficients as in standard wavelet transform.

On the other hand, nonlinear process are more difficult to analyze. In particular the stability of the multiresolution is not clear. In practical applications, as we

Departamento de Matemática Aplicada y Estadística, Cartagena, Murcia, Spain
E-mail Address: sergio.amat@upct.es
Australian National University, Canberra ACT 2601, Australia
E-mail Address: sonia.busquier@uv.es

said before, a discrete sequence $\bar{f}^L$ is encoded to produce a multiscale representation of its information contents, $(\bar{f}^0, e^1, e^2, \dots, e^L)$; this representation is then processed and the end result of this step is a modified multiscale representation $(\hat{f}^0, \hat{e}^1, \hat{e}^2, \dots, \hat{e}^L)$ which is close to the original one, i.e. such that (in some norm)

$$||\hat{f}^0 - \bar{f}^0|| \leq \epsilon_0 \qquad ||\hat{e}^k - e^k|| \leq \epsilon_k \quad 1 \leq k \leq L,$$

where the truncation parameters $\epsilon_0, \epsilon_1, \dots, \epsilon_L$ are chosen according to some criteria specified by the user. After decoding the processed representation, we obtain a discrete set $\hat{f}^L$ which is expected to be close to the original discrete set $\bar{f}^L$. In order for this to be true, some form of stability is needed, i.e. we must require that

$$||\hat{f}^L - \bar{f}^L|| \leq \sigma(\epsilon_0, \epsilon_1, \dots, \epsilon_L)$$

where $\sigma(\cdot, \dots, \cdot)$ satisfies

$$\lim_{\epsilon_l \to 0,\ 0 \leq l \leq L} \sigma(\epsilon_0, \epsilon_1, \dots, \epsilon_L) = 0.$$

The stability analysis for linear prediction processes can be carried out using tools coming from wavelet theory, subdivision schemes and functional analysis (see [10], [4]), however none of these techniques is applicable in general when the prediction process is nonlinear.

In the nonlinear case, stability can be ensured by modifying the encoding algorithm. The idea of a modified-encoding to deal with nonlinear multiresolution schemes is due to Harten; one dimensional algorithms in several settings can be found in [9], [5], [3]. The goal of a modified-encoding procedure is to keep track of the accumulation error in processing the values in the multiscale representation. The aim of this paper is to present two-dimensional multiresolution algorithms that ensure stability in the case of nonlinear prediction processes. We will consider two frameworks: point-values and cell-average. In the case of linear reconstructions the first corresponds with the interpolatory wavelets and the second with some biorthogonal wavelets.

The paper is organized as follows: we recall in section 2 the discrete framework for multiresolution introduced by Harten. We present the point-values and the cell-average (2-D) multiresolutions in sections 3 and 4 respectively. The error-control algorithms are discussed in section 5. Finally we give stability results in section 6.

2. Harten's framework for multiresolution

Harten's framework is based on two fundamental tools: discretization and reconstruction. The discretization operator D_k obtains discrete information from a (non-discrete) signal ($f \in F$ functional space) at a particular resolution level k ($\bar{f}^k \in V^K$). The reconstruction operator R_k, on the other hand, produces an approximation to a signal from its discrete values. This reconstruction can be nonlinear. Then we can connect the different resolution levels by $D_k^{k-1} := D_{k-1} R_k$ (decimation operator) and $P_{k-1}^K := R_{k-1} D_k$ (prediction operator). The basic consistency property that these two operators have to satisfy is

$$D_k^{k-1} P_{k-1}^k = I_{V^{k-1}}$$

where $I_{V^{k-1}}$ is the identity operator in V^{k-1}. In this case, $D_k^{k-1} e^k = 0$ where e^k denotes the prediction error, $e^k = \bar{f}^k - P_{k-1}^k \bar{f}^{k-1}$.

3. Interpolatory multiresolution analysis in 2-D

Let $X^k = \{x_i^k, y_j^k\}_{i,j=0}^{J_k}$, $J_k = 2^k J_0$, J_0 some integer, $x_{2i}^k = x_i^{k-1}$, $y_{2j}^k = y_j^{k-1}$ and $x_{2i-1}^k = (x_i^{k-1} + x_{i-1}^{k-1})/2$, $y_{2j-1}^k = (y_j^{k-1} + y_{j-1}^{k-1})/2$.

We consider

$$D_k : C([0,1] \times [0,1]) \to V^k \qquad \bar{f}_{i,j}^k = (D_k f)_{i,j} = f(x_i^k, y_j^k,), \quad 0 \le i, j \le J_k.$$

In this case, $\dim V^k = (J_k+1) \times (J_k+1)$ and the decimation operators D_k^{k-1} are

$$\bar{f}_{i,j}^{k-1} = (D_k^{k-1} \bar{f}^k)_{i,j} = \bar{f}_{2i,2j}^k, \quad i, j = 0, 1, \dots, J_{k-1}.$$

Since $\bar{f}_{i,j}^{k-1} = \bar{f}_{2i,2j}^k$, we obtain $N(D_k^{k-1}) = \{v^k \in V^k | v_{2i,2j}^k = 0\}$. Thus, if we denote by e^k the prediction errors, we will need to keep $e_{2i-1,2j-1}^k, e_{2i-1,2j}^k, e_{2i,2j-1}^k$ only.

A reconstruction procedure for this discretization is given by any operator R_k such that

$$R_k : V^k \to C([0,1] \times [0,1]); \qquad D_k R_k \bar{f}^k = \bar{f}^k,$$

which means

$$(R_k \bar{f}^k)(x_i^k, y_j^k) = \bar{f}_{i,j}^k = f(x_i^k, y_j^k).$$

Therefore, R_k should be a continuous function that interpolates the data $\bar{f}^k$ on at the grid points of X^k. Finally the prediction operators are

$$P_{k-1}^k := D_k R_{k-1}$$

Many different approaches have been used to derive the prediction operators, see [2], [6], [7], [8], [11].

The encoding and decoding algorithms take the following form:

$\mu(\bar{f}^L) = M \bar{f}^L$ (Encoding)

$for\ k = L, \dots, 1$
$\quad for\ i, j = 0, \dots, J_{k-1}$
$\quad\quad \bar{f}_{i,j}^{k-1} = \bar{f}_{2i,2j}^k$
$\quad end$
$\quad for\ i = 1, \dots, J_{k-1}$
$\quad for\ j = 0, \dots, J_{k-1}$
$\quad\quad f_{2i-1,2j}^P = (P_{k-1}^k \bar{f}^{k-1})_{2i-1,2j}$
$\quad\quad e_{2i-1,2j}^k = \bar{f}_{2i-1,2j}^k - f_{2i-1,2j}^P$
$\quad end$
$\quad end$
$\quad for\ i = 0, \dots, J_{k-1}$
$\quad for\ j = 1, \dots, J_{k-1}$
$\quad\quad f_{2i,2j-1}^P = (P_{k-1}^k \bar{f}^{k-1})_{2i,2j-1}$

$$
\begin{array}{l}
\qquad\qquad\qquad e^k_{2i,2j-1} = \bar{f}^k_{2i,2j-1} - f^P_{2i,2j-1} \\
\qquad\qquad end \\
\qquad\qquad end \\
\qquad\qquad for\ i,j = 1,\dots,J_{k-1} \\
\qquad\qquad\qquad f^P_{2i-1,2j-1} = (P^k_{k-1}\bar{f}^{k-1})_{2i-1,2j-1} \\
\qquad\qquad\qquad e^k_{2i-1,2j-1} = \bar{f}^k_{2i-1,2j-1} - f^P_{2i-1,2j-1} \\
\qquad\qquad end \\
\qquad end \\
M^M \bar{f}^L = \{\bar{f}^0, e^1, \dots, e^L\}
\end{array}
$$

$\bar{f}^L = M^{-1}\mu(\bar{f}^L)$ (Decoding)

$$
\begin{aligned}
&for\ k = 1, \dots, L\\
&\quad for\ i,j = J_{k-1}, \dots, 0\\
&\qquad \bar{f}^k_{2i,2j} = \bar{f}^{k-1}_{i,j}\\
&\quad end\\
&\quad for\ i = J_{k-1}, \dots, 1\\
&\quad for\ j = J_{k-1}, \dots, 0\\
&\qquad f^P_{2i-1,2j} = (P^k_{k-1}\bar{f}^{k-1})_{2i-1,2j}\\
&\qquad \bar{f}^k_{2i-1,2j} = e^k_{2i-1,2j} + f^P_{2i-1,2j}\\
&\quad end\\
&\quad end\\
&\quad for\ i = J_{k-1}, \dots, 0\\
&\quad for\ j = J_{k-1}, \dots, 1\\
&\qquad f^P_{2i,2j-1} = (P^k_{k-1}\bar{f}^{k-1})_{2i,2j-1}\\
&\qquad \bar{f}^k_{2i,2j-1} = e^k_{2i,2j-1} + f^P_{2i,2j-1}\\
&\quad end\\
&\quad end\\
&\quad for\ i,j = J_{k-1}, \dots, 1\\
&\qquad f^P_{2i-1,2j-1} = (P^k_{k-1}\bar{f}^{k-1})_{2i-1,2j-1}\\
&\qquad \bar{f}^k_{2i-1,2j-1} = e^k_{2i-1,2j-1} + f^P_{2i-1,2j-1}\\
&\quad end\\
&end
\end{aligned}
$$

4. Cell-Averages multiresolution analysis in 2-D

In this case we have

$$D_k : L^1([0,1]\times[0,1]) \to V^k,$$

$$\bar{f}^k_{i,j} = (D_k f)_{i,j} = \frac{1}{h_k^2}\int_{x^k_{i-1}}^{x^k_i}\int_{y^k_{j-1}}^{y^k_j} f(x,y)dydx, \quad 1 \le i,j \le J_k;$$

where $L^1([0,1]\times[0,1])$ is the space of absolutely integrable functions in $[0,1]\times[0,1]$ and the set of nested grids X^k is as in above section. This analysis turns out to be appropriate for data compression of discontinuous, piecewise smooth signals.

It is sufficient to consider weighted averages $\bar{f}^k_{i,j}$ for $1 \le i,j \le J_k$ since these contain information on f over $[0,1]\times[0,1]$. Thus, V^k is the space of sequences with $J_k \times J_k$ components.

Moreover

$$\bar{f}^{k-1}_{i,j} = (D^{k-1}_k \bar{f}^k)_{i,j} = \frac{1}{4}(\bar{f}^k_{2i-1,2j-1} + \bar{f}^k_{2i-1,2j} + \bar{f}^k_{2i,2j-1} + \bar{f}^k_{2i,2j})$$

$i,j = 1, 2, \dots, J_{k-1}$.

On the other hand, since

$$0 = (D^{k-1}_k e^k)_{i,j} = (e^k_{2i-1,2j-1} + e^k_{2i-1,2j} + e^k_{2i,2j-1} + e^k_{2i,2j})/4$$

we will keep $e^k_{2i-1,2j-1}, e^k_{2i-1,2j}, e^k_{2i,2j-1}$ only.

A reconstruction operator for this discretization is any operator R_k satisfying

$$R_k : V_k \to L^1([0,1]\times[0,1]),$$

$$(D_k R_k \bar{f}^k)_{i,j} = \frac{1}{h_k^2}\int_{x_{i-1}^k}^{x_i^k}\int_{y_{j-1}^k}^{y_j^k} (R_k \bar{f}^k)(x,y)dxdy = \bar{f}^k_{i,j}.$$

That is, $R_k \bar{f}^k(x,y)$ has to be a function in $L^1([0,1]\times[0,1])$ whose mean value on the (i,j)-th cell coincides with $\bar{f}^k_{i,j}$, $\forall i,j$, and $P^k_{k-1} := D_k R_{k-1}$.

The multiresolution transform and its inverse are now

$\mu(\bar{f}^L) = M\bar{f}^L$ (Encoding)

$$
\begin{array}{l}
for\ k = L, \dots, 1 \\
\quad for\ i,j = 1,\dots, J_{k-1} \\
\quad\quad \bar{f}^{k-1}_{i,j} = \frac{1}{4}(\bar{f}^k_{2i-1,2j-1} + \bar{f}^k_{2i-1,2j} + \bar{f}^k_{2i,2j-1} + \bar{f}^k_{2i,2j}) \\
\quad end \\
\quad for i,j = 1,\dots, J_{k-1} \\
\quad\quad f^P_{2i-1,2j-1} = (P^k_{k-1}\bar{f}^{k-1})_{2i-1,2j-1} \\
\quad\quad e^k_{2i-1,2j-1} = \bar{f}^k_{2i-1,2j-1} - f^P_{2i-1,2j-1} \\
\quad\quad f^P_{2i-1,2j} = (P^k_{k-1}\bar{f}^{k-1})_{2i-1,2j} \\
\quad\quad e^k_{2i-1,2j} = \bar{f}^k_{2i-1,2j} - f^P_{2i-1,2j} \\
\quad\quad f^P_{2i,2j-1} = (P^k_{k-1}\bar{f}^{k-1})_{2i,2j-1} \\
\quad\quad e^k_{2i,2j-1} = \bar{f}^k_{2i,2j-1} - f^P_{2i,2j-1} \\
\quad end \\
end
\end{array}
$$

$M^M\bar{f}^L = \{\bar{f}^0, e^1, \dots, e^L\}$

$\bar{f}^L = M^{-1}\mu(\bar{f}^L)$ (Decoding)

$$
\begin{array}{l}
for\ k = 1, \dots, L \\
\quad for\ i,j = J_{k-1},\dots, 1 \\
\quad\quad f^P_{2i-1,2j-1} = (P^k_{k-1}\bar{f}^{k-1})_{2i-1,2j-1} \\
\quad\quad \bar{f}^k_{2i-1,2j-1} = f^P_{2i-1,2j-1} + e^k_{2i-1,2j-1} \\
\quad\quad f^P_{2i-1,2j} = (P^k_{k-1}\bar{f}^{k-1})_{2i-1,2j} \\
\quad\quad \bar{f}^k_{2i-1,2j} = f^P_{2i-1,2j} + e^k_{2i-1,2j} \\
\quad\quad f^P_{2i,2j-1} = (P^k_{k-1}\bar{f}^{k-1})_{2i,2j-1} \\
\quad\quad \bar{f}^k_{2i,2j-1} = f^P_{2i,2j-1} + e^k_{2i,2j-1} \\
\quad\quad \bar{f}^k_{2i,2j} = 4\bar{f}^{k-1}_{i,j} - \bar{f}^k_{2i-1,2j-1} - \bar{f}^k_{2i-1,2j} - \bar{f}^k_{2i,2j-1} \\
\quad end \\
end
\end{array}
$$

5. Multiresolution-based compression schemes with error control

Multiresolution representations lead naturally to data-compression algorithms. The simplest data compression procedure is obtained by setting to zero all scale coefficients which fall below a prescribed tolerance. Let us denote

$$(\hat{e}^k)_{i,j} = tr(e^k_{i,j}; \epsilon_k)$$

and refer to this operation as truncation. This type of data compression is used primarily to reduce the "dimensionality" of the data. A different strategy, which is used to reduce the digital representation of the data is "quantization", which can be modeled by

$$(\hat{e}^k)_{i,j} = qu(e^k_{i,j}; \epsilon_k) = 2\epsilon_k round\left[\frac{e^k_{i,j}}{2\epsilon_k}\right],$$

where $round[\cdot]$ denotes the integer obtained by rounding. For example, if $|e^k_{i,j}| \leq 256$ and $\epsilon_k = 4$ then we can represent $e^k_{i,j}$ by an integer which is not larger than 32 and commit a maximal error of 4. Observe that if $|e^k_{i,j}| < \epsilon_k \Rightarrow qu(e^k_{i,j}; \epsilon_k) = 0$ and that in both cases

$$|e^k_{i,j} - \hat{e}^k_{i,j}| \leq \epsilon_k.$$

By applying the inverse multiresolution transform to the compressed representation, we obtain $\hat{f}^L = M^{-1}\{\bar{f}^0, \hat{e}^1, \ldots, \hat{e}^L\}$, an approximation to the original signal $\bar{f}^L$. We expect the information contents of $\hat{f}^L$ to be very close to those of the original signal $\bar{f}^L$, and in order for this to be true, the stability of the multiresolution scheme with respect to perturbations is essential. Studying the effect of using $\hat{e}^k_{i,j}$ instead of $e^k_{i,j}$ in the input of M^{-1} is equivalent to studying the effect of a perturbation in the scale coefficients in the outcome of the inverse multiresolution transform.

Given a discrete sequence $\bar{f}^L$ and a tolerance level ϵ for accuracy, our task is to come up with a compressed representation

$$\{\bar{f}^0, \hat{e}^1, \ldots, \hat{e}^L\} \tag{10}$$

such that if $\hat{f}^L = M^{-1}\{\bar{f}^0, \hat{e}^1, \ldots, \hat{e}^L\}$, we have

$$||\bar{f}^L - \hat{f}^L|| \leq C\epsilon \tag{11}$$

for an appropriate norm.

As observed by Harten [9], one possible way to accomplish this goal is to modify the encoding procedure in such a way that the modification allows us to keep track of the cumulative error and truncate accordingly.

In what follows we present a two-dimensional extension of the one dimensional algorithms in [9], [3] and the two dimensional tensor product in [1]. Given a tolerance level ϵ, the outcome of the modified encoding procedure is a compressed representation (10) satisfying (11). This enables us to specify the desired level of accuracy in the decompressed signal. A modified encoding procedure is designed keeping in mind the particular decoding procedure to be used.

We will consider truncation, but the algorithms are identical for another compression process.

Both point-values and cell-averages have been analyzed.

In next section, we will obtain explicit error bound corresponding to these algorithms, and in particular we will obtain their stability.

(Encoding for point-values)

$$
\begin{array}{l}
for\ k = L, \dots, 1 \\
\quad for\ i, j = 0, \dots, J_{k-1} \\
\quad\quad \bar{f}^{k-1}_{i,j} = \bar{f}^k_{2i,2j} \\
\quad end \\
end \\
\\
Set\ \hat{f}^0 = \bar{f}^0 \\
\\
for\ k = 1, \dots, L \\
\\
\quad for\ i = 1, \dots, J_{k-1} \\
\quad for\ j = 0, \dots, J_{k-1} \\
\quad\quad \hat{e}^k_{2i-1,2j} = tr(\bar{f}^k_{2i-1,2j} - (P^k_{k-1}\hat{f}^{k-1})_{2i-1,2j}, \epsilon_k) \\
\quad\quad \hat{f}^k_{2i-1,2j} = (P^k_{k-1}\hat{f}^{k-1})_{2i-1,2j}, + \hat{e}^k_{2i-1,2j} \\
\quad end \\
\quad end \\
\\
\quad for\ i = 0, \dots, J_{k-1} \\
\quad for\ j = 1, \dots, J_{k-1} \\
\quad\quad \hat{e}^k_{2i,2j-1} = tr(\bar{f}^k_{2i,2j-1} - (P^k_{k-1}\hat{f}^{k-1})_{2i,2j-1}, \epsilon_k) \\
\quad\quad \hat{f}^k_{2i,2j-1} = (P^k_{k-1}\hat{f}^{k-1})_{2i,2j-1} + \hat{e}^k_{2i,2j-1} \\
\quad end \\
\quad end \\
\\
\quad for\ i = 1, \dots, J_{k-1} \\
\quad for\ j = 1, \dots, J_{k-1} \\
\quad\quad \hat{e}^k_{2i-1,2j-1} = tr(\bar{f}^k_{2i-1,2j-1} - (P^k_{k-1}\hat{f}^{k-1})_{2i-1,2j-1}, \epsilon_k) \\
\quad\quad \hat{f}^k_{2i-1,2j-1} = (P^k_{k-1}\hat{f}^{k-1})_{2i-1,2j-1} + \hat{e}^k_{2i-1,2j-1} \\
\quad end \\
\quad end \\
\\
\quad for\ i = 0, \dots, J_{k-1} \\
\quad for\ j = 0, \dots, J_{k-1} \\
\quad\quad \hat{f}^k_{2i,2j} = \hat{f}^{k-1}_{i,j} \\
\quad end \\
\quad end \\
\\
end
\end{array}
$$

(Encoding for cell-average)

$for\ k = L, \ldots, 1$
  $for\ i, j = 1, \ldots, J_{k-1}$
    $\bar{f}^{k-1}_{i,j} = \frac{1}{4}(\bar{f}^k_{2i-1,2j-1} + \bar{f}^k_{2i-1,2j} + \bar{f}^k_{2i,2j-1} + \bar{f}^k_{2i,2j})$
  end
end

$Set\ \hat{f}^0 = \bar{f}^0$

$for\ k = 1, \ldots, L$

  $for\ i = 1, \ldots, J_{k-1}$
  $for\ j = 1, \ldots, J_{k-1}$
    $f^P_{2i-1,2j-1} = (P^k_{k-1}\hat{f}^{k-1})_{2i-1,2j-1}$
    $\hat{e}^k_{2i-1,2j-1} = tr([\bar{f}^k_{2i-1,2j-1} - f^P_{2i-1,2j-1}] - [\bar{f}^{k-1}_{i,j} - \hat{f}^{k-1}_{i,j}], \epsilon_k)$
    $\hat{f}^k_{2i-1,2j-1} = f^P_{2i-1,2j-1} + \hat{e}^k_{2i-1,2j-1}$
    $f^P_{2i-1,2j} = (P^k_{k-1} hatf^{k-1})_{2i-1,2j}$
    $\hat{e}^k_{2i-1,2j} = tr([\bar{f}^k_{2i-1,2j} - f^P_{2i-1,2j}] - [\bar{f}^{k-1}_{i,j} - \hat{f}^{k-1}_{i,j}], \epsilon_k)$
    $\hat{f}^k_{2i-1,2j} = f^P_{2i-1,2j} + \hat{e}^k_{2i-1,2j}$
    $f^P_{2i,2j-1} = (P^k_{k-1}\hat{f}^{k-1})_{2i,2j-1}$
    $\hat{e}^k_{2i,2j-1} = tr([\bar{f}^k_{2i,2j-1} - f^P_{2i,2j-1}] - [\bar{f}^{k-1}_{i,j} - \hat{f}^{k-1}_{i,j}], \epsilon_k)$
    $\hat{f}^k_{2i,2j-1} = f^P_{2i,2j-1} + \hat{e}^k_{2i,2j-1}$
    $\hat{f}^k_{2i,2j} = 4\hat{f}^{k-1}_{i,j} - \hat{f}^k_{2i-1,2j-1} - \hat{f}^k_{2i-1,2j} - \hat{f}^k_{2i,2j-1}$
  end
  end

end

6. Stability Analysis

We use the following matrix norms:

$$||A||_1 = max_{||x||_1} ||Ax||_1$$
$$||A||_\infty = max_{||x||_\infty} ||Ax||_\infty$$
$$||A||_2 = max_{||x||_2} ||Ax||_2$$

The proofs are similar with norm 1 and ∞. Moreover, to derive the l^2 bound, we simply note that for a vector, x, one has $||x||_2 \le ||x||_1$. Thus we are proving the propositions with $p = \infty$.

Proposition 6.1. *Given a discrete sequence $\bar{f}^L$ and a tolerance level ϵ, if the truncation parameters ϵ_k in the modified encoding algorithm (point values) are chosen so that*

$$\epsilon_k = \epsilon$$

then the sequence $\hat{f}^L = M^{-1}\{\bar{f}^0, \hat{e}^1, \ldots, \hat{e}^L\}$ satisfies

$$||\bar{f}^L - \hat{f}^L||_p \le \epsilon \qquad (12)$$

for $p = \infty, 1$ *and* 2.

Proof

Let us define the cumulative compression error at the k-th level by $E^k_{i,j}$,

$$E^k_{i,j} = \bar{f}^k_{i,j} - \hat{f}^k_{i,j}$$

Notice that

$$|E^k_{2i,2j}| = |\bar{f}^k_{2i,2j} - \hat{f}^k_{2i,2j}| = |E^{k-1}_{i,j}| \leq ||E^{k-1}||_\infty$$

$$\begin{aligned} |E^k_{2i,2j-1}| &= |\bar{f}^k_{2i,2j-1} - \hat{f}^k_{2i,2j-1}| \\ &= |\bar{f}^k_{2i,2j-1} - f^P_{2i,2j-1} - tr(\bar{f}^k_{2i,2j-1} - f^P_{2i,2j-1}, \epsilon_k)| \leq \epsilon_k \\ |E^k_{2i-1,2j}| &= |\bar{f}^k_{2i-1,2j} - \hat{f}^k_{2i-1,2j}| \\ &= |\bar{f}^k_{2i-1,2j} - f^P_{2i-1,2j} - tr(\bar{f}^k_{2i-1,2j} - f^P_{2i-1,2j}, \epsilon_k)| \leq \epsilon_k \end{aligned}$$

$$\begin{aligned} |E^k_{2i-1,2j-1}| &= |\bar{f}^k_{2i-1,2j-1} - \hat{f}^k_{2i-1,2j-1}| \\ &= |\bar{f}^k_{2i-1,2j-1} - f^P_{2i-1,2j-1} - tr(\bar{f}^k_{2i-1,2j-1} - f^P_{2i-1,2j-1}, \epsilon_k)| \leq \epsilon_k \end{aligned}$$

then

$$||E^k||_\infty \leq \max\left(\frac{1}{J_k+1}\sum_{i=0}^{J_k}|\epsilon_k|, \frac{1}{J_k+1}(||E^{k-1}||_\infty + \sum_{i=1}^{J_{k-1}}|\epsilon_k|)\right) \qquad k = 1, \ldots, l$$

$||E^0||_\infty = 0$ which implies

$$||E^L||_\infty \leq \max(\epsilon_L, \frac{1}{J_L+1}||E^{L-1}||_\infty + \frac{J_{L-1}}{J_L+1}\epsilon_L)$$

taking $\epsilon_k = \epsilon$ we obtain $||E^L||_\infty \leq \epsilon$, and hence (12).

□

Proposition 6.2. *Given a discrete sequence* $\bar{f}^L$ *and a tolerance level* ϵ, *if the truncation parameters* ϵ_k *in the modified encoding algorithm (cell-average) are chosen so that*

$$\epsilon_k = \frac{\epsilon}{2} \cdot q^{L-k} \qquad 0 < q < 1$$

then the sequence $\hat{f}^L = M^{-1}\{\bar{f}^0, \hat{d}^1, \ldots, \hat{d}^L\}$ *satisfies*

$$||\bar{f}^L - \hat{f}^L||_p \leq (1-q)^{-1}\epsilon \tag{13}$$

for $p = \infty, 1$ *and* 2.

Proof Let us define the cumulative compression error at the k-th level by $E^k_{i,j}$,

$$E^k_{i,j} = \bar{f}^k_{i,j} - \hat{f}^k_{i,j}$$

and the modified prediction error at the k-th level by $\tilde{e}^k_{i,j}$,

$$\tilde{e}^k_{i,j} = \bar{f}^k_{i,j} - (P^k_{k-1}\hat{f}^{k-1})_{i,j}$$

With this notation we get

$$\begin{aligned} E^k_{2i-1,2j-1} &= \bar{f}^k_{2i-1,2j-1} - \hat{f}^k_{2i-1,2j-1} \\ &= \bar{f}^k_{2i-1,2j-1} - f^P_{2i-1,2j-1} - \hat{e}^k_{2i-1,2j-1} \\ &= \tilde{e}^k_{2i-1,2j-1} - tr(\tilde{e}^k_{2i-1,2j-1} - E^{k-1}_{i,j}, \epsilon_k), \end{aligned} \tag{14}$$

$$\begin{aligned} E^k_{2i-1,2j} &= \bar{f}^k_{2i-1,2j} - \hat{f}^k_{2i-1,2j} \\ &= \bar{f}^k_{2i-1,2j} - f^P_{2i-1,2j} - \hat{e}^k_{2i-1,2j} \\ &= \tilde{e}^k_{2i-1,2j} - tr(\tilde{e}^k_{2i-1,2j} - E^{k-1}_{i,j}, \epsilon_k), \end{aligned} \tag{15}$$

$$\begin{aligned} E^k_{2i,2j-1} &= \bar{f}^k_{2i,2j-1} - \hat{f}^k_{2i,2j-1} \\ &= \bar{f}^k_{2i,2j-1} - f^P_{2i,2j-1} - \hat{e}^k_{2i,2j-1} \\ &= \tilde{e}^k_{2i,2j-1} - tr(\tilde{e}^k_{2i,2j-1} - E^{k-1}_{i,j}, \epsilon_k), \end{aligned} \tag{16}$$

$$\frac{1}{4}(E^k_{2i-1,2j-1} + E^k_{2i-1,2j} + E^k_{2i,2j-1} + E^k_{2i,2j}) = E^{k-1}_{i,j}$$

Let us now examine the two possibilities in (14), (15) and (16),

$$|\tilde{e}^k_{2i-1,2j-1} - E^{k-1}_{i,j}| > \epsilon_k \Rightarrow E^k_{2i-1,2j-1} = E^{k-1}_{i,j},$$

$$|\tilde{e}^k_{2i-1,2j-1} - E^{k-1}_{i,j}| \leq \epsilon_k \Rightarrow E^k_{2i-1,2j-1} = \tilde{e}^k_{2i-1,2j-1}$$

$$|\tilde{e}^k_{2i-1,2j} - E^{k-1}_{i,j}| > \epsilon_k \Rightarrow E^k_{2i-1,2j} = E^{k-1}_{i,j},$$

$$|\tilde{e}^k_{2i-1,2j} - E^{k-1}_{i,j}| \leq \epsilon_k \Rightarrow E^k_{2i-1,2j} = \tilde{e}^k_{2i-1,2j}$$

$$|\tilde{e}^k_{2i,2j-1} - E^{k-1}_{i,j}| > \epsilon_k \Rightarrow E^k_{2i,2j-1} = E^{k-1}_{i,j},$$

$$|\tilde{e}^k_{2i,2j-1} - E^{k-1}_{i,j}| \leq \epsilon_k \Rightarrow E^k_{2i,2j-1} = \tilde{e}^k_{2i,2j-1}$$

We have:

$$\begin{aligned} ||E^k||_\infty = \max_i(\max(\frac{1}{J_k}(\sum_{j=1}^{\frac{J_k}{2}}(|E^k_{2i-1,2j-1}| + |E^k_{2i-1,2j}|), \\ \frac{1}{J_k}\sum_{j=1}^{\frac{J_k}{2}}(|E^k_{2i,2j-1}| + |E^k_{2i,2j}|))) \end{aligned}$$

Moreover:

$$|E^k_{2i-1,2j-1}| + |E^k_{2i-1,2j}| = \max(|E^k_{2i-1,2j-1} + E^k_{2i-1,2j}|, |E^k_{2i-1,2j-1} - E^k_{2i-1,2j}|)$$

$$|E^k_{2i,2j-1}| + |E^k_{2i,2j}| = \max(|E^k_{2i,2j-1} + E^k_{2i,2j}|, |E^k_{2i,2j-1} - E^k_{2i,2j}|)$$

And after straightforward algebraic manipulations we obtain:

$$\begin{aligned} ||E^k||_\infty &\leq \max(\frac{1}{J_k}(\sum_{j=1}^{\frac{J_k}{2}}(2|E_{i,j}^{k-1}| + 4\epsilon_k))) \\ &\leq ||E^{k-1}||_\infty + 2\epsilon_k \end{aligned}$$

then

$$||E^L||_\infty \leq ||E^{L-1}||_\infty + 2\epsilon_L \leq \ldots \leq 2\sum_{l=1}^{L}\epsilon_l$$

taking $\epsilon_k = \frac{\epsilon}{2}q^{L-k}$ where $0 < q < 1$ we obtain $\sum_{l=1}^{L}\epsilon_l \leq \epsilon\frac{1}{1-q}$ and hence (13) for $p = \infty$.

□

References

1. Amat S., Aràndiga F., Cohen A. and Donat R.:, *Tensor product multiresolution analysis with error control for compact image representation.*, To appear in Signal processing. (2001).
2. Amat S., Aràndiga F., Cohen A., Donat R., García G. and von Oehsen M.:, *Data compresion with ENO schemes: A Case Study.*, To appear in ACHA. (2001).
3. Aràndiga F. and Donat R.:, *Nonlinear Multi-scale Decompositions: The Approach of A.Harten.*, Numerical Algorithms. **23** (2000), 175-216.
4. Aràndiga F., Donat R. and Harten A.:, *Multiresolution Based on Weighted Averages of the Hat Function I: Linear Reconstruction Operators.*, SIAM J. Numer. Anal. **36 (1)** (1999), 160-203.
5. Aràndiga F., Donat R. and Harten A.:, *Multiresolution Based on Weighted Averages of the Hat Function II: Nonlinear Reconstruction Operators.*, SIAM J. Sci. Comput. **20 (3)** (1999), 1053-1093.
6. Bamberger R.H.:, *A Method for Image Interpolation Based on a Novel Multirate Filter Bank Structure and Properties of the Human Visual System.*, Proc. of SPIE **1657 (Image Processing Algorithms and Techniques)** (1992), 351-362.
7. Chan A.K., Chui C.K., Zha J. and Liu Q.:, *Local cardinal spline interpolation and its application to image proccessing.*, Proc. of SPIE **1610 (Curves and Surfaces in Computer Vision and Graphics)** (1991), 272-283.
8. Chen G. and de Figueiredo Rui J.P.:, *A Unified Approach to Optimal Image Interpolation Problems Based on Linear Partial Differential Equations.*, IEEE Trans. on Image Processing **2 (1)** (1993), 41-49.
9. Harten A.:, *Discrete Multiresolution Analysis and Generalized Wavelets.*, J. Appl. Numer. Math. **12** (1993), 153-192.
10. Harten A.:, *Multiresolution Representation of Data II: General Framework.*, SIAM J. Numer. Anal. **33 (3)** (1996), 1205-1256.
11. Ho Y.S. and Gersho A.:, *Variable-rate contour-based interpolative vector quantization for image coding.*, Proc. IEEE Intl. Conf. on ASSP (1998), 750-754.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2001, pp. 19–23.
Submitted: October 25, 2001. Published: December 31 2001.

CONVERGENCE THEOREMS FOR NEWTON-LIKE METHODS USING DATA FROM A SET OR A SINGLE POINT AND OUTER INVERSES

Ioannis K. Argyros

Abstract. Semilocal convergence theorems for Newton-like methods are provided using information from a set or a single point and outer inverses.

A.M.S. (MOS) Subject Classification Codes. 65H10, 47H17, 49M15.

Key Words and Phrases. Newton-like methods, Banach space, semilocal convergence theorem, outer inverse.

1. Introduction

In this study we are concerned with the problem of approximating a locally unique solution x^* of the equation

$$LF(x) = 0, \tag{1}$$

where F is a Fréchet-differentiable operator defined on an open convex subset D of a Banach space X with values in a Banach space Y, $L \in L(X, Y)$ the space of bounded linear operators from X into Y.

We consider Newton-like methods of the form

$$x_{n+1} = x_n - A(x_n)^{\#} F(x_n) \quad (n \geq 0) \tag{2}$$

for generating a sequence converging to x^*. Here $A(x_n) \in L(X, Y)$ $(n \geq 0)$ is an approximation of the Fréchet-derivative $F'(x_n)$ $(n \geq 0)$ and $A(x_n)^{\#}$ denotes an outer inverse of $A(x_n)$, i.e., $A(x_n)^{\#} A(x_n) A(x_n)^{\#} = A(x_n)^{\#}$. This setting includes generalized Newton methods for undetermined systems [1], [3].

Semilocal convergence theorems for Newton-like methods of the form (2) are provided using information from a set or a single point.

Department of Mathematics, Cameron University, Lawton, OK 73505, U.S.A.
E-mail Address: ioannisa@cameron.edu

2. Convergence Analysis

Define parameters r_i, $i=0,1,\dots,5$, h_0 and γ by

$$\begin{aligned}
&r_0 = \frac{1}{\gamma}\left(1-\sqrt[3]{\frac{2\gamma}{M+L}}\right), \quad M+L\neq 0, \quad r_1=\frac{1-\ell}{L}, \quad L\neq 0,\\
&r_2=\frac{1}{\gamma}, \quad \gamma\neq 0, \quad r_3=\min\{r_0,r_1,r_2\},\\
&r_3=r_0 \text{ if } L=\gamma=0, \quad r_3=\min\{r_0,r_2\} \text{ if } L=0,\\
&r_3=\min\{r_0,r_1\} \text{ if } \gamma=0, \quad r_4=\min\{r_3,r^{**}\},\\
&r_5=\min\{r_1,r_2\}, \quad r_5=r_1 \text{ if } \gamma=0,\\
&r_5=r_2 \text{ if } L=0, \quad r_5=\infty \text{ if } L=\gamma=0,\\
&h_0=(M+L)\eta-\frac{1}{2}(1-b)^2, \text{ for } L\geq 0,\ M\geq 0\\
&\gamma=\sup_{k>1}\left\|\frac{1}{k!}A^{\#}F^{(k)}(x_0)\right\|, \quad x_0\in X \quad \text{(if it exists)}
\end{aligned}$$

where $A^{\#}$ is a bounded outer inverse of A $(=A(x_0))$ and functions f, g, f_1, f_2 by

$$\begin{aligned}
f(r)&=\frac{\sigma}{2}r^2-(1-b)r+\eta, \quad \sigma=\max\{K,M+L\},\\
g(r)&=Lr+\ell-1,\\
f_1(r)&=\frac{M+L}{2}r^2-(1-b)r+\eta,\\
f_2(r)&=\frac{\gamma r^2}{(1-\gamma r)^3}-(1-b)r+\eta,
\end{aligned}$$

where r^*, r^{**} are the zeros of function f_1.

Assume:

(C) $r_0\geq 0$, $h_0\leq 0$ and $r^*\leq r_3$

Note that under condition (C), $0\leq r^*\leq r^{**}$, $f_1(r)\geq 0$, $g(r)<0$ and $\frac{2\gamma}{(1-\gamma r)^3}\equiv K_0\leq M+L$ on $[0,r^*]$.

or

(C_1) $r_0\leq 0$ (or $M+L=0$) and there exists $r_6\in(0,r_5)$ or $r_6\in[r_0,r_5))$ such that $f_2(r_6)\leq 0$.

Denote by r_7 the zero of function f_2 in $(0,r_6)$. (Such a zero exists by the intermediate value theorem.) Note that under condition (C_1), $f_2(r)\geq 0$, $g(r)<0$ and $M+L\leq K_0$ on $[0,r_6]$.

We can state and prove the following semilocal convergence theorem for Newton-like methods using information from a set and outer inverses.

Theorem 1. *Let $F\colon D\subseteq X\to Y$ be twice-Fréchet-differentiable and let $A(x)\in L(X,Y)$ $(x\in D)$ be an approximation of $F'(x)$. Assume:*

(a) *there exists an open convex subset D_0 of D, $x_0\in D_0$, a bounded outer inverse $A^{\#}$ of A $(=A(x_0))$ and constants η, $K>0$, M, L, μ, $\ell\geq 0$ such that for al $x\in D_0$*

the following conditions hold:

$$\|A^{\#}F(x_0)\| \le \eta, \tag{3}$$

$$\|A^{\#}F''(x)\| \le K, \tag{4}$$

$$\|A^{\#}(F'(x) - A(x))\| \le M\|x - x_0\| + \mu \tag{5}$$

$$\|A^{\#}(A(x) - A)\| \le L\|x - x_0\| + \ell, \tag{6}$$

$$b = \mu + \ell < 1, \tag{7}$$

$$h = \sigma\eta \le \frac{1}{2}(1 - b)^2 \tag{8}$$

and

$$\overline{U}(x_0, t^*) \subseteq D_0, \tag{9}$$

where, t^* is the small zero of function f.

Then,

(1) *Newton-like method $\{x_n\}$ $(n \ge 0)$ generated by (2) is well defined, remains in $U(x_0, t^*)$ for all $n \ge 0$ and converges to a solution $x^* \in \overline{U}(x_0, t^*)$ of equation $A^{\#}F(x) = 0$, with*

$$A(x_n)^{\#} = (I + A^{\#}(A(x_n) - A))^{-1}A^{\#};$$

(2) *equation $A^{\#}F(x) = 0$ has a unique solution x^* in $\widetilde{U} \cap \{R(A^{\#}) + x_0\}$, where*

$$\widetilde{U} = \begin{cases} \overline{U}(x_0, t^*) \cap D_0, & \text{if } h = \frac{1}{2}(1 - b)^2 \\ U(x_0, t^{**}) \cap D_0, & \text{if } h < \frac{1}{2}(1 - b)^2, \end{cases}$$

where t^{**} is the large zero of function f and

$$R(A^{\#}) + x_0 = \{x + x_0; x \in R(A^{\#})\};$$

(3) *the following error bounds hold for all $n \ge 0$*

$$\|x_{n+1} - x_n\| \le t_{n+1} - t_n$$

and

$$\|x_n - x^*\| \le t^* - t_n,$$

where,

$$t_0 = 0, \quad t_{n+1} = t_n - \frac{f(t_n)}{g(t_n)}.$$

Proof. Lipschitz condition (3.2) in [3] follows from (4). The rest of the theorem follows exactly as Theorem 3.1 in [3].

That completes the proof of the theorem.

We can show the following semilocal result for Newton-like methods using information from a set of a single point and outer inverses.

Theorem 2. *Let $F\colon X \to Y$ be analytic on some open convex subset D of X and consider $A(x) \in L(X,Y)$ $(x \in D)$ to be an approximation of $F'(x)$. Assume*:

(a) *there exists an open convex subset D_0 of D, $x_0 \in D_0$, a bounded outer inverse $A\#$ of A $(= A(x_0))$ and constants $\eta > 0$, M, L, μ, $\ell \geq 0$ such that for all $x \in D_0$ conditions (3), (5), (6), and (7) hold*;

(b) *conditions (C) (or (C_1)) hold*;

(c) $\overline{U}(x_0, r^*) \subseteq D_0$ *(or* $\overline{U}(x_0, r_1) \subseteq D_0$*)*.

Then,

(1) *Newton's method $\{x_n\}$ $(n \geq 0)$ generated by (2) is well defined, remains in $U(x_0, r^*)$ (or $U(x_0, r_7)$) for all $n \geq 0$ and converges to a solution $x^* \in \overline{U}(x_0, r^*)$ (or $\overline{U}(x_0, r_7)$) of equation $A^{\#}F(x) = 0$ with $A(x_n)^{\#} = (I + A^{\#}(A(x_n) - A))^{-1}A^{\#}$;*

(2) *the equation $A^{\#}F(x) = 0$ has a unique solution in $\widetilde{U}_1 \cap \{R(A^{\#}) + x_0\}$ (or $\widetilde{U}_2 \cap \{R(A^{\#} + x_0\})$, where*

$$\widetilde{U}_1 = \begin{cases} \overline{U}(x_0, r^*) \cap D_0 & \text{if } h_0 = 0 \\ U(x_0, r_4) \cap D_0 & \text{if } h_0 < 0, \end{cases}$$

$$\widetilde{U}_2 = \begin{cases} \overline{U}(x_0, r_5) \cap D_0 & r_6 < r_5 \\ U(x_0, r_6) \cap D_0 & r_5 = r_6; \end{cases}$$

(3) *the following error bound hold for all $n \geq 0$*:

$$\|x_{n+1} - x_n\| \leq r_{n+1} - r_n \quad (\text{or } r^1_{n+1} - r^1_n), \tag{10}$$

$$\|x_n - x^*\| \leq r^* - r_n \quad (\text{or } r_6 - r^1_n), \tag{11}$$

where,

$$r_0 = 0, \quad r_{n+1} = r_n - \frac{f_1(r_n)}{g(r_n)}$$

and

$$r^1_0 = 0, \quad r^1_{n+1} = r^1_n - \frac{f_2(r^1_n)}{g(r^1_n)}.$$

Proof. Since $\gamma r^* < 1$ under (C) (or $\gamma r_6 < 1$ under (C_1)), we find for all $x \in U(x_0, r^*)$ (or r_6) that

$$\begin{aligned} \|A^{\#}F''(x)\| &= \left\| \sum_{n=0}^{\infty} \frac{1}{k!} A^{\#}F^{(k+2)}(x_0)(x - x_0)^k \right\| \\ &\leq \gamma \sum_{k=0}^{\infty} (k+2)(k+2)(\|x - x_0\|)^k \leq K_0. \end{aligned}$$

Estimates (10) and (11) are shown exactly as in Theorem 1 since under condition (C) $r^* \leq r^{**}$, $\frac{2\gamma}{(1-\gamma r)^3} = K_0 \leq M + L$, $f_1(r) \geq 0$, $g(r) < 0$ on $[0, r^*]$ (or under (C_1): $f_2(r) \geq 0$, $g(r) < 0$ and $M + L < K_0$ on $[0, r_6]$).

The rest follows exactly as in Theorem 1.

That completes the proof of the theorem.

REFERENCES

[1] Argyros, I.K., Convergence domains for some iterative processes in Banach space using outer and generalized inverses, *Journ. Computational Anal. and Applic.* **1**, 1 (1999), 87–104.

[2] Argyros, I.K. and Szidarovszky, F., *The Theory and Application of Iteration Methods*, CRC Press, Inc., Boca Raton, Florida, U.S.A., 1993.

[3] Chen, X. and Nashed, M.Z., Convergence of Newton-like methods for singular operator equations using outer inverses, *Numer. Math.* **66** (1993), 235–257.

[4] Kantorovich, L.V. and Akilov, G.P., *Functional Analysis*, Pergamon Press, Oxford, 1982.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2001, pp. 24–29.
Submitted: October 5, 2001. Published: December 31 2001.

A GENERALIZATION OF BEATTY'S THEOREM

Arthur Holshouser and Harold Reiter

Abstract. In 1926 Sam Beatty made the following discovery, which he posed as a problem in [1]. If a is a positive irrational number, the sequences $m(1+a), m = 1, 2, \ldots$ and $n(1+a^{-1}), n = 1, 2, \ldots$ together contain exactly one number from each of the intervals $(k, k+1), k = 1, 2, 3, \ldots$. The problem was solved by Ostrowski and Aitken[4] and generalized to a larger class of sequences by Lambeck and Moser[3]. The authors are grateful for UNC Charlotte undergraduate James Rudzinski for suggesting the combinatorial game that lead to this paper.

1991 A.M.S. (MOS) Subject Classification. 11B05, 11B75, 11K36

Keywoards and Phrases:. Beatty's theorem, combinatorial game, Ostrowski, Aitken

The purpose of this paper is to extend this theorem from sequences to continuous functions. We first establish some notation. Let $P = (a_0 = 0, a_1, a_2, \ldots)$ be a strictly increasing unbounded sequence of real numbers. For any nonnegative number x, let $\lfloor x \rfloor_P$ be the largest member of P that does not exceed x, and for positive x,

$$\rfloor x \lfloor_P = \begin{cases} \lfloor x \rfloor_P & \text{if } x \text{ is not in } P \\ a_{i-1} & \text{if } x \in P \text{ and } \lfloor x \rfloor_P = a_i. \end{cases}$$

Also, if $t > 0$, we define $N_t = \{0, t, 2t, 3t, 4t, \cdots\}$ and $N_t^+ = N_t \backslash \{0\}$.

Main Theorem. *Let F and G be real, continuous, strictly increasing, functions with domains $[0, \infty)$ satisfying $F(0) = G(0) = 0$ and $\lim_{x\to\infty}(F(x) + G(x)) = \infty$. For all $t > 0$, let $P_t = \{0, (F+G)^{-1}(t), (F+G)^{-1}(2t), (F+G)^{-1}(3t), \ldots\}$, and $P_t^+ = P_t \backslash \{0\}$. Then the two sequences A_t and B_t defined by $A_t = \{\rfloor(F^{-1})(n)\lfloor_{P_t}:$*

Department of Mathematics, UNC Charlotte, Charlotte, NC
E-mail Address: hbreiter@email.uncc.edu

$n \in N_t^+\}$ *and* $B_t = \{\lfloor (G^{-1})(n) \rfloor_{P_t} : n \in N_t^+\}$ *partition* P_t^+ *for all* $t > 0$*. Also, the elements of* A_t *are distinct and the elements of* B_t *are distinct.*

We conjecture that the theorem remains true even when $F+G$ is bounded. Also, it is easy to see that an analogous version exists for functions $F : [a, b) \to [0, \infty)$ and $G : [a, b) \to [0, \infty)$, where F and G are continuous, strictly increasing, and satisfy $F(a) = G(a) = 0$ and $\lim_{x \to b}(F(x) + G(x)) = \infty$.

Note that the two functions F and G in the theorem are independent, unlike those in the Lambek and Moser paper where each one determines the other. This shows that using continuous functions provides a much richer theory. Note also that our conclusion is true for all $t > 0$. Thus we have an uncountable number of partitions. This forces rigid restrictions on the sequences A_t and B_t.

The following is an example illustrating the theorem. We can choose F and G independently. However, we have chosen F and G so that we can compute $F^{-1}, G^{-1}, (F+G)^{-1}$ explicitly. Of course, if we cannot compute the inverse of one of our functions explicitly, then we just leave that inverse in implicit form just as we would for the function sin^{-1}. Let $F(x) = ax^2 + bx, G(x) = cx^2 + dx, a, b, c, d > 0$. Then $(F + G)(x) = (a + c)x^2 + (b + d)x, F^{-1}(n) = \frac{-b+\sqrt{b^2+4an}}{2a}, G^{-1}(n) = \frac{-d+\sqrt{d^2+4cn}}{2c}$ and $(F + G)^{-1}(n) = \frac{-(b+d)+\sqrt{(b+d)^2+4(a+c)n}}{2(a+c)}$. Then

$$P_t = \left\{ \frac{-(b+d) + \sqrt{(b+d)^2 + 4(a+c)n}}{2(a+c)} : n = 0, t, 2t, 3t, \ldots \right\},$$

$$A_t = \left\{ \left\lfloor \frac{-b + \sqrt{b^2 + 4an}}{2a} \right\rfloor_{P_t} : n = t, 2t, 3t, \ldots \right\},$$

and

$$B_t = \left\{ \left\lfloor \frac{-d + \sqrt{d^2 + 4cn}}{2c} \right\rfloor_{P_t} : n = t, 2t, 3t, \ldots \right\}.$$

The theorem guarantees that for all $t > 0$, A_t and B_t partition P_t^+ and the members of A_t and B_t are distinct. The functions $F(x) = \frac{x}{1+a}$ and $G(x) = \frac{ax}{1+a}$, where a is a positive irrational number, can be used to prove Beatty's theorem. The reader might also like to use $F(x) = 3(x^2 + x), G(x) = x^3$ to give an example for himself.

We next prove the main theorem by first proving a more specialized result, for which we need more notation. Let $N = \{0, 1, 2, 3, \ldots\}, Z = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$, and for any $t > 0$, $Z_t = \{\ldots, -3t, -2t, -t, 0, t, 2t, 3t, \ldots\}$ and $N_t = \{0, t, 2t, 3t, \ldots\}$. Thus $N = N_1$ and $Z = Z_1$. Suppose $x = nt + \underline{\epsilon}$ where $n \in Z$ and $0 \leq \underline{\epsilon} < t$. Then $\lfloor x \rfloor_t = nt$. Suppose $= x = mt + \bar{\epsilon}$ where $m \in Z$ and $0 < \bar{\epsilon} \leq t$. Then $\lceil x \rceil_t = mt + t$. Thus $\lfloor x \rfloor_t$ and $\lceil x \rceil_t$ are the *floor* and *ceiling* functions with respect to the grid Z_t. Of course, $\lfloor x \rfloor_1 = \lfloor x \rfloor$ and $\lceil x \rceil_1 = \lceil x \rceil$ where $\lfloor x \rfloor$ and $\lceil x \rceil$ are the usual floor and ceiling functions. If $F : A \to B$ is a function on A where A is a countable linearly ordered set, then we consider $S = \{F(x) : x \in A\}$ to be both a set and a sequence. The reason for this is that we wish to discuss the set S of functional values and we also need to discuss the fact that the members of the *sequence* are distinct.

<u>Definition</u>: Suppose f and g are any real functions with domain $[0, \infty)$. We say that f and g are complementary if f and g satisfy $f(0) = g(0) = 0$, f and g are strictly increasing on $[0, \infty)$, and $f(x) + g(x) = x$ for all $x \geq 0$.

It is easy to see that for any $\epsilon > 0$, $\quad f(x+\epsilon)-f(x) < \epsilon$ and $0 < g(x+\epsilon)-g(x) < \epsilon$ for all non-negative x, so complimentary functions are continuous. Also, it is easy to see that $0 < f(x) < x$ and $0 < g(x) < x$ for all positive x.

Theorem 1. *Suppose $f : [0, \infty) \to R,\ g : [0, \infty) \to R$ are complementary. Then*

a. *$\forall t > 0$, the following two sets partition the set $N_t^+ = \{t, 2t, 3t, 4t, \cdots\}$: $A = \left\{\left\lceil f^{-1}(n)\right\rceil_t : n \in N_t^+\right\}$ and $B = \left\{\left\lfloor g^{-1}(n) + t\right\rfloor_t : n \in N_t\right\}$. Also, the elements of the sequence A are distinct and the elements of the sequence B are distinct.*

b. *Also, for every $t > 0$, the following two sets partition N_t^+ : $C = \left\{\left\lceil f^{-1}(n) - t\right\rceil_t : n \in N_t^+\right\}$ and $D = \left\{\left\lfloor g^{-1}(n)\right\rfloor_t : n \in N_t^+\right\}$. Also, the elements of the sequence C are distinct and the elements of the sequence D are distinct.*

   We note that if one of f, g is bounded above, then in both a. and b., one of the two sets is finite.

We will now prove theorem 1 through a series of lemmas. It is convenient for notational reasons to assume that $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = \infty$. Although we will assume this, a slightly more clumsy version holds without the assumption. In the following lemmas, we consider $t > 0$ to be arbitrary but fixed.

Lemma 1. *Suppose $h : N_t \to N_t$ satisfies $h(0) = 0$, $h(n + t) - h(n) \in \{0, t\}$ for all $n \in N_t$, and $\lim_{n\to\infty} h(n) = \lim_{n\to\infty}(n - h(n)) = \infty$.*

Define $k : N_t \to N_t$ as follows: $\forall n \in N_t, k(n) = n - h(n)$. It is easy to see that $k(0) = 0, \lim_{n\to\infty} k(n) = \infty$ and $\forall n \in N_t, k(n + t) - k(n) \in \{0, t\}$. Also both h and k are non-decreasing on N_t.

Next, $\forall m \in N_t$, define $H(m)$ to be the smallest member of N_t such that $h(H(m)) = m$ and $K(m)$ to be the smallest member of N_t such that $k(K(m)) = m$. Then

a. *$H(0) = K(0) = 0$.*
b. *H and K are strictly increasing.*
c. *the two sets $\left\{H(n) : n \in N_t^+\right\}$ and $\left\{K(n) : n \in N_t^+\right\}$ partition the set N_t^+.*

Proof. Parts a and b are easy to see. To prove c, first observe that $\forall n \in N_t^+$, either $h(n) - h(n - t) = t$ and $k(n) - k(n - t) = 0$ or $h(n) - h(n - t) = 0$ and $k(n) - k(n - t) = t$. This is because $h(n) + k(n) = n$. From this it follows that $\forall x \in N_t^+, x$ is in the range set of exactly one of H and K. $\square$

Definitions: Let f, g be the two complementary functions used in the hypothesis of the Theorem 1. For convenience we again assume that $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = \infty$.

Let us define for all positive numbers t, $h : N_t \to N_t$ by $h(n) = \lfloor f(n)\rfloor_t$ and define $k : N_t \to N_t$ by $k(n) = n - h(n)$. It follows that $k(n) = n - \lfloor f(n)\rfloor_t = n + \lceil -f(n)\rceil_t = \lceil n - f(n)\rceil_t = \lceil g(n)\rceil_t$. i.e., $k(n) = \lceil g(n)\rceil_t$. Note that k is being defined from h the same way as it was in lemma 1. These functions h and k are used in the remainder of our proof of the theorem.

Lemma 2.. *The function h defined above satisfies the hypothesis given in lemma 1, namely*

1. *$h(0) = 0$,*
2. *$\forall n \in N_t, h(n + t) - h(n) \in \{0, t\}$,*
3. *$\lim_{n\to\infty} h(n) = \lim_{n\to\infty}(n - h(n)) = \infty$.*

Proof. Since $h(n) = \lfloor f(n) \rfloor_t$, $n - h(n) = \lceil g(n) \rceil_t$, $f(0) = 0$ and $\lim_{n\to\infty} f(x) = \lim_{n\to\infty} g(x) = \infty$, conclusions 1 and 3 are easy to see. We now prove conclusion 2.

Now $0 < f(n+t) - f(n) < t$ since $\forall \epsilon > 0, \forall x \in [0, \infty), 0 < f(x+\epsilon) - f(x) < \epsilon$.

Therefore $\lfloor f(n+t) \rfloor_t - \lfloor f(n) \rfloor_t \in \{0, t\}$. It follows that $h(n+t) - h(n) \in \{0, t\}$ since $h(n+t) = \lfloor f(n+t) \rfloor_t$ and $h(n) = \lfloor f(n) \rfloor_t$. This completes the proof of lemma 2. $\square$

Since $k(n) = n - h(n)$, as in lemma 1, it follows that $k(0) = 0$, $\lim_{n\to\infty} k(n) = \infty$ and $\forall n \in N_t$, $k(n+t) - k(n) \in \{0, t\}$. Now $\forall m \in N_t$, $H(m), K(m)$ are defined in lemma 1 since h and k satisfy the conditions specified in lemma 1. That is, $\forall m \in N_t$, $H(m)$ is the smallest member of N_t such that $h(H(m)) = \lfloor f(H(m)) \rfloor_t = m$ and $\forall m \in N_t$, $K(m)$ is the smallest member of N_t such that $= k(K(m)) = \lceil g(K(m)) \rceil_t = m$. The following two lemmas will now add considerable power to lemma 1 when we use $h(n) = \lfloor f(n) \rfloor_t$ and $= k(n) = n - h(n) = \lceil g(n) \rceil_t$ in lemma 1.

Since $f : [0, \infty) \to R$ and $g : [0, \infty) \to R$ are complementary, we recall that $f(0) = g(0) = 0$, f, g are strictly increasing and continuous on $[0, \infty)$ and also $\forall \epsilon > 0, \forall x \in [0, \infty), 0 < f(x+\epsilon) - f(x) < \epsilon, 0 < g(x+\epsilon) - g(x) < \epsilon$. As always, for convenience, we are assuming that $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = \infty$.

Lemma 3. $\forall n \in N_t, H(n) = \lceil f^{-1}(n) \rceil_t$.

Proof. Let us abbreviate $f^{-1}(n) = x$. Therefore $f(x) = n$.

Define $\overline{x} = \lceil x \rceil_t$. Of course, $= \overline{x} \in N_t$. Now by lemma 1, $H(0) = 0$. Also, $f^{-1}(0) = 0$. So $H(0) = \lceil f^{-1}(0) \rceil_t = \lceil 0 \rceil_t = 0$ is obvious. Therefore, we now assume that $n \in \{t, 2t, 3t, 4t, \cdots\}$. Then note that $f^{-1}(n) = x > t$ since $\forall x \in (0, \infty), 0 < f(x) < x$. Now since by lemma 2, $\forall \overline{n} \in= N_t, h(\overline{n}+t) - h(\overline{n}) \in \{0, t\}$, lemma 3 is proved once we prove the following:

a. $h(\overline{x}) = \lfloor f(\overline{x}) \rfloor_t = n$

b. $h(\overline{x} - t) = \lfloor f(\overline{x} - t) \rfloor_t = n - t$.

<u>Proof of a</u>: Now $\overline{x} - x = \overline{\epsilon}$ where $0 \leq \overline{\epsilon} < t$. Now $\lfloor f(\overline{x}) \rfloor_t = \lfloor f(x + \overline{\epsilon}) \rfloor_t = \lfloor f(x+\overline{\epsilon}) - f(x) + f(x) \rfloor_t = \lfloor f(x+\overline{\epsilon}) - f(x) \rfloor_t + n$. Now $0 \leq f(x+\overline{\epsilon}) - f(x) \leq \overline{\epsilon} < t$ since $0 \leq \overline{\epsilon}$. It follows that $\lfloor f(x+\overline{\epsilon}) - f(x) \rfloor_t = 0$. Therefore $\lfloor f(\overline{x}) \rfloor_t = \lfloor f(x+\overline{\epsilon}) - f(x) \rfloor_t + n = 0 + n = n$.

<u>Proof of b</u>: Define $\underline{x} = \overline{x} - t = \lceil x - t \rceil_t$. Let $x - \underline{x} = \underline{\epsilon}$, where $0 < \underline{\epsilon} \leq t$. Now $\lfloor f(\overline{x} - t) \rfloor_t = \lfloor f(\underline{x}) \rfloor_t = \lfloor f(x - \underline{\epsilon}) \rfloor_t = \lfloor f(x-\underline{\epsilon}) - f(x) + f(x) \rfloor =_t = \lfloor f(x-\underline{\epsilon}) - f(x) \rfloor_t + n$. Since $0 < \underline{\epsilon}$, we see that $0 < f(x) - f(x-\underline{\epsilon}) < \underline{\epsilon} \leq t$. Therefore $-t < f(x-\underline{\epsilon}) - f(x) < 0$. It follows that $\lfloor f(x-\underline{\epsilon}) - f(x) \rfloor_t = -t$. Hence $\lfloor f(\overline{x} - t) \rfloor_t = \lfloor f(x-\underline{\epsilon}) - f(x) \rfloor_t + n = n - t$. $\square$

Lemma 4. $K(0) = 0$ *and* $\forall n \in N_t$, $K(n+t) = \lfloor g^{-1}(n) + t \rfloor_t$.

Proof. Recall that K was defined at the end of the proof of lemma 2. Note that $K(0) = 0$ from lemma 1. We now show that for all $n \in N_t$, $K(n+t) = \lfloor g^{-1}(n) + t \rfloor_t$. Let us first consider $n = 0$. We will show that $K(0+t) = \lfloor g^{-1}(0) + t \rfloor_t$. Now $g^{-1}(0) = 0$. Therefore, we need to show that $K(t) = \lfloor t \rfloor_t = t$. Now $0 < g(t) < t$ is true. Therefore, $k(t) = \lceil g(t) \rceil_t = t$. Since $k(0) = 0$, $= k(t) = t$, we see that $K(t) = t$ must be true by the definition of K. We will now suppose $n \in \{t, 2t, 3t, \cdots\}$. We must show that $K(n+t) = \lfloor g^{-1}(n) + t \rfloor_t$. Let us abbreviate $g^{-1}(n) = x$. Therefore, $g(x) = n$. Since $n \in \{t, 2t, \cdots\}$ and for all

$x \in (0,\infty)$, $0 < g(x) < x$, it follows that $x > t$. Define $\overline{x} = \lfloor x+t \rfloor_t$. Of course, $\overline{x} \in N_t$. Since $\forall \overline{n} \in N_t$, $k(\overline{n}+t) - k(\overline{n}) \in \{0,t\}$, the lemma is proved once we prove the following:

a. $k(\overline{x}) = \lceil g(\overline{x}) \rceil_t = n+t$.

b. $k(\overline{x}-t) = \lceil g(\overline{x}-t) \rceil_t = n$.

Proof of a: Define $\overline{x} - x = \overline{\epsilon}$, where $0 < \overline{\epsilon} \le t$. Now $\lceil g(\overline{x}) \rceil_t = \lceil g(x+\overline{\epsilon}) \rceil_t = \lceil g(x+\overline{\epsilon}) - g(x) + g(x) \rceil_t = \lceil g(x+\overline{\epsilon}) - g(x) \rceil_t + n$. Now $0 < g(x+\overline{\epsilon}) - g(x) < \overline{\epsilon} \le t$, since $0 < \overline{\epsilon}$. Therefore $\lceil g(x+\overline{\epsilon}) - g(x) \rceil_t = t$. Therefore $\lceil g(\overline{x}) \rceil_t = \lceil = g(x+\overline{\epsilon}) - g(x) \rceil_t + n = n+t$.

Proof of b: Remember $x > t$. Let $\underline{x} = \overline{x} - t = \lfloor x \rfloor_t$. We must show $\lceil = g(\underline{x}) \rceil_t = n$. Define $x - \underline{x} = \underline{\epsilon}$, where $0 \le \underline{\epsilon} < t$. Now $\lceil g(\underline{x}) \rceil_t = \lceil g(x-\underline{\epsilon}) \rceil_t = \lceil g(x-\underline{\epsilon}) - g(x) + g(x) \rceil_t = \lceil g(x-\underline{\epsilon}) - g(x) \rceil_t + n$. Now $0 \le g(x) - g(x-\underline{\epsilon}) \le \underline{\epsilon} < t$, since $0 \le \underline{\epsilon} < t$. Therefore $-t < g(x-\underline{\epsilon}) - g(x) \le 0$. Therefore $\lceil g(\underline{x}) \rceil_t = \lceil g(x-\underline{\epsilon}) - g(x) \rceil_t + n = 0 + n = n$. $\square$

We now have enough information to prove the theorem 1.

Proof. We know that f and g are complementary functions. Also, we know that f and g are continuous and bounded above by the identity function. We also recall that we defined $h : N_t \to N_t$ and $k : N_t \to N_t$ as follows: $\forall n \in N_t, h(n) = \lfloor f(n) \rfloor =_t, k(n) = n - h(n) = \lceil g(n) \rceil_t$. In lemma 2 we showed that this h satisfies the hypothesis of lemma 1. Also this k is defined from this h in the same way as in lemma 1.

We also defined the functions $H(m)$ and $K(m)$ in lemma 1. We also know from lemma 3, 4 that $\forall n \in N_t, H(n) = \lceil f^{-1}(n) \rceil_t$ and $\forall n \in N_t, K(n+t) = \lfloor g^{-1}(n) + t \rfloor_t$.

¿From lemma 1, we know that $\forall t > 0$, the following two sets partition the set N_t^+ : $\{H(n) : n \in N_t^+\} = \{\lceil f^{-1}(n) \rceil_t : n \in N_t^+\}$ and $\{K(n) : n \in N_t^+\} = \{\lfloor g^{-1}(n) + t \rfloor_t : n \in N_t\}$.

Recall that H and K are strictly increasing functions whose image sets are disjoint. Therefore, the elements of these two sets, when viewed as a sequence, are all distinct.

Now since $g^{-1}(0) = 0$, we know that $\{\lfloor g^{-1}(n) + t \rfloor_t : n \in N_t\} = \{t\} \cup \{\lfloor g^{-1}(n) + t \rfloor_t : n \in N_t^+\}$. Therefore, $\{\lceil f^{-1}(n) \rceil_t : n \in N_t^+\}$ and $\{\lfloor g^{-1}(n) + t \rfloor_t : n \in N_t^+\}$ partition the set $\{2t, 3t, 4t, 5t, \cdots\}$. Therefore $\{\lceil f^{-1}(n) - t \rceil_t : n \in N_t^+\}$ and $\{\lfloor g^{-1}(n) \rfloor_t : n \in N_t^+\}$ partition N_t^+. Also, the elements of these two sets, when viewed as a sequence, are all distinct.

We now use theorem 1 to prove the Main Theorem.

Suppose $h : [0,\infty) \to [0,\infty)$ is continuous, strictly increasing, and satisfies both $h(0) = 0$ and $\lim_{x\to\infty} h(x) = \infty$. We define for all $t > 0$, the sets P_t and P_t^+ as follows:

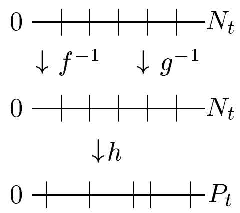
$$P_t = \{h(0) = 0, h(t), h(2t), \ldots\} \text{and} P_t^+ = \{h(t), h(2t), \ldots\}.$$

Next consider the following lemma.

Lemma 5. *Suppose f and g are complementary functions and h satisfies the conditions listed above. Then the sets A and B defined by $A = \{\rfloor (fh^{-1})^{-1}(n) \lfloor_{P_t} : n \in N_t^+\}$ and $B = \{\lfloor (gh^{-1})^{-1}(n) \rfloor_{P_t} : n \in N_t^+\}$ partition P_t^+ for all $t > 0$. Also, the elements of A (when viewed as a sequence) are distinct and the same is true for the elements of B.*

Proof. Note that $(fh^{-1})^{-1}(n) = hf^{-1}(n)$ and $(gh^{-1})^{-1}(n) = hg^{-1}(n)$. By part b of theorem 1, $\{\lceil f^{-1}(n) - t\rceil_t :\ n \in N_t^+\}$ and $\{\lfloor g^{-1}(n)\rfloor_t :\ n \in N_t^+\}$ partition N_t^+.

Also, the elements of each of the two sets, when viewed as sequences, are distinct. Since h is an order preserving change of variable and since $\rfloor x\lfloor_{P_t}$ is defined for the partition of P_t topologically analogous to the way $\lceil x - t\rceil_t$ is defined for the partition of N_t and since $\lfloor x\rfloor_{P_t}$ is defined for the partition P_t topologically analogous to the way $\lfloor x\rfloor_t$ is defined for the partition of N_t, the conclusion is easy to see. See figure 1. $\square$



h is an order preserving change of variable.

We now prove the Main Theorem.

Proof. Since both F and G are strictly increasing, so is $F + G$. Also, $F + G$ is unbounded and $(F+G)(0) = 0$. It follows that $(F+G)^{-1}$ exists for all non-negative x. Also, $((F+G)(F+G)^{-1})(x) = (F(F+G)^{-1})(x) + (G(F+G)^{-1})(x) = x$ for all nonnegative x. For convenience of notation, denote $F(F+G)^{-1}$ by f, $G(F+G)^{-1}$ by g and $(F+G)^{-1}$ by h. Note that f, g, and h satisfy the hypothesis of the last lemma and that $fh^{-1} = F$ and $gh^{-1} = G$. The lemma completes the proof. $\square$

References

[1] Sam Beatty, *Problem 3173*, American Mathematical Monthly **33** (1926), 159.

[2] Ross Honsberger, *Ingenuity in Mathematics*, New Mathematical Library, American Mathematical Association, Providence, 1989, pp. 93-110.

[3] J. Lambek and L. Moser, *Inverse and Complementary Sequences of Natural Numbers*, American Mathematical Monthly **61** (1954), 454.

[4] Ostrowski and Aitken, *Solution to Problem 3173*, American Mathematical Monthly **34** (1927), 159.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2001, pp. 30–42.
Submitted: February 1, 2000. Published: December 31 2001.

IRREDUCIBLE MAPPING AND THE LIGHTNESS OF OPEN MAPPINGS

Pavel Pyrih
Tomáš Bárta
David Opěla
Pavel Růžička
Robert Šámal

Abstract. We study continua with the property that each nonconstant open mapping is light. We define a large class of continua and show that each member of this class has this property. In this class we present an arcwise connected continuum which can be mapped under an irreducible mapping onto a continuum admitting a nonconstant non-light open mapping. Two closely related questions are answered.

1991 A.M.S. (MOS) Subject Classification. 54E40, 54F15, 54F50

Key Words and Phrases. Irreducible, atomic, light, open

1. Introduction and notation.

For the convenience of the reader we repeat the relevant material. All spaces considered in the paper are assumed to be metric and all mappings are continuous. A *continuum* means a compact connected space. Given a space X and its subset S, we denote by $\operatorname{cl} S$ the closure of S.

Let X be a continuum and $A \subset X$. Then X is said to be *irreducible about* A provided no proper subcontinuum of X contains A. A continuum X is said to be *irreducible* provided that X is irreducible about $\{p, q\}$ for some $p, q \in X$ (then X is said to be *irreducible between* p *and* q).

Department of Mathematical Analysis, Prague, Czech Republic
E-mail : pyrih@karlin.mff.cuni.cz, barta@kma14.karlin.mff.cuni.cz,opela@kma14.karlin.mff.cuni.cz
ruzicka@karlin.mff.cuni.cz, samal@atrey.karlin.mff.cuni.cz

Joint research of the Open Problem Seminar, Charles University, Prague.
Research supported by CEZ J13/98113200007 and GAUK 165/99 of Charles University.

Let X be a continuum, and let A and B be nonempty compact subsets of X. We say that a subcontinuum C of X is *irreducible from A to B*, written $C = \operatorname{irr}(A, B)$, provided that $C \cap A \neq \emptyset$, $C \cap B \neq \emptyset$, and no proper subcontinuum of C intersects both A and B. If A or B is a singleton set, we omit the set notation; thus, e.g. we write $C = \operatorname{irr}(p, B)$, $C = \operatorname{irr}(p, q)$.

Let X and Y be continua. A mapping $f : X \to f(X) = Y$ is said to be:
– *atomic* provided that for each subcontinuum $K \subset X$ with nondegenerate image $f(K)$ the equality $f^{-1}(f(K)) = K$ holds;
– *irreducible* provided that for each nondegenerate irreducible subcontinuum $I \subset Y$ the inverse image $f^{-1}(I)$ is irreducible;
– *monotone* provided that for each subcontinuum Q of Y the inverse image $f^{-1}(Q)$ is connected;
– *light* provided that for each point $y \in Y$ each component of the inverse image $f^{-1}(y)$ is a singleton;
– *interior at a point* $p \in X$ provided that for each open neighborhood U of p in X the point $f(p)$ is an interior point of the image $f(U)$ in Y;
– *open* provided that for each open subset U of X its image $f(U)$ is an open subset of Y.

Obviously a mapping is open if and only if it is interior at each point of its domain.

Given continua X and Y, we will denote by

$$\mathsf{Atomic}, \quad \mathsf{Irreducible}, \quad \text{and} \quad \mathsf{Monotone}$$

the family of all atomic, irreducible and monotone mappings $f : X \to f(X) = Y$, respectively. We recall from [15, Corollary 2.9, p. 36] (see also [12, 3.7, p. 13], [8, Theorem 1, p. 49]) the following

1.1. Statement. $\mathsf{Atomic} \subsetneq \mathsf{Irreducible} \subsetneq \mathsf{Monotone}$.

Let L denote the class of all continua X such that each nonconstant open mapping defined on X is light, and let $\mathfrak{M}$ be an arbitrary class of mappings between continua that contains the class of homeomorphisms. Consider classes $\mathfrak{M}$ for which the implication holds :

$$(\heartsuit) \qquad \text{if} \quad f \in \mathfrak{M} \quad \text{and} \quad X \in \mathsf{L}, \quad \text{then} \quad f(X) \in \mathsf{L}.$$

Recall the following assertions (see [13, Propositions 2.7-2.9, p. 783], [7, Example 2.10]).

(1.2) *The implication* $(\heartsuit)$ *holds if* $\mathfrak{M}$ *is the class of open mappings.*
(1.3) *The implication* $(\heartsuit)$ *does not hold if* $\mathfrak{M}$ *is the class of atomic mappings.*
(1.4) *The implication* $(\heartsuit)$ *holds if* $\mathfrak{M}$ *is the class of atomic mappings and the continuum* X *is arcwise connected.*

Notice that each atomic mapping defined on an arcwise connected continuum is already a homeomorphism (see [12, 6.3, p. 51]). A natural question arises (see [16, Problem 35]) whether the implication $(\heartsuit)$ is true if the continuum X is arcwise connected and $\mathfrak{M}$ is the class of irreducible mappings. Answering this question in the negative, we construct in Section 3 of the present paper a large family $\mathcal{S}_c$ of continua in L and in Example 3.3 an arcwise connected continuum in the family $\mathcal{S}_c$ as a counterexample.

Let us recall two (still open) problems (see [5, Problem 3, p. 214 and Problem 11, p. 217]).

1.5. Problem. What topological spaces X and Y have the property that each open mapping from X onto Y is light?

1.6. Problem. Characterize all continua X being in the class L.

Note that the family of continua in L that is defined in the present paper gives a partial answer to Problems 1.5 and 1.6.

2. Spirals.

A *ray* means a one-to-one image of the closed half-line $[0, +\infty)$, and the image of 0 is called the *end point of the ray.*

If S is a dense subspace of a compact space Z, then Z is called a *compactification* of S, and $Z \setminus S$ is called the *remainder* of S in Z (see e.g. [1, p. 34]). It is known that if S is a locally compact, noncompact, separable metric space, then each continuum is a remainder of S in some compactification of S, [1, Theorem, p. 35]. Taking as S a one-to-one image of the real half-line $[0, \infty)$ we conclude the following statement.

2.1. Statement. *Each nondegenerate continuum B is a remainder of a ray S in some compactification of S, and then $Y = B \cup S$ is a continuum having S as an arc-component with $B = \operatorname{cl} S \setminus S$.*

Let S^- and S^+ are two disjoint subcontinua of a continuum X. An open set $S \subset X$ is called a *spiral from S^- to S^+* if

- (S1) there is a one-to-one surjective mapping $g : (-\infty, \infty) \to S$ of the open interval $(-\infty, \infty)$ onto S;
- (S2) $S^+ \cup g([0, \infty))$ is a compactification of $g([0, \infty))$ having S^+ as the remainder of $g([0, \infty))$ in $S^+ \cup g([0, \infty))$;
- (S3) $S^- \cup g((-\infty, 0])$ is a compactification of $g((-\infty, 0])$ having S^- as the remainder of $g((-\infty, 0])$ in $S^- \cup g((-\infty, 0])$;
- (S4) $(S^- \cup S^+) \cap S = \emptyset$.

If conditions (S1)-(S4) are satisfied, then we say that the above mentioned mapping $g : (-\infty, \infty) \to S$ *describes the spiral S from S^- to S^+*, and the union

$$X = S^- \cup S \cup S^+$$

is named a *spiral-arc from S^- to S^+*. Obviously each arc is a spiral-arc (with its ends as degenerate continua S^- and S^+); and the $\sin(1/x)$-curve $\operatorname{cl}\{(x, \sin(1/x)) : x \in (0, 1]\}$ is a spiral-arc from a singleton $\{(1, \sin 1)\}$ to the closed segment $\{(0, y) : y \in [-1, 1]\}$.

If S is a spiral from a continuum A to a continuum B, we denote by S^- and S^+ the continua A and B, respectively.

We say that S is a *spiral* in X if S is a spiral from some subcontinuum (S^-) to some subcontinuum (S^+) in X.

For $S = g((-\infty, +\infty)) \subset X$ a spiral, the above notation means that

$$S^+ = \{x \in X, \text{ there exists } \{t_n\}_{n=1}^{\infty} \subset \mathbb{R} \text{ such that } \lim_{n\to\infty} t_n = +\infty \;\&\; \lim_{n\to\infty} g(t_n) = x\}$$
$$S^- = \{x \in X, \text{ there exists } \{t_n\}_{n=1}^{\infty} \subset \mathbb{R} \text{ such that } \lim_{n\to\infty} t_n = -\infty \;\&\; \lim_{n\to\infty} g(t_n) = x\}.$$

A spiral S is said to be *crooked near* a nondegenerate continuum K if for each point $a \in K$ there exists an arbitrarily small neighborhood U of a such that for each $b \in K \cap U$ there exists in $U \cap S$ a sequence of arcs J_n with end points α_n, β_n satisfying
(CS1) $\lim \alpha_n = \lim \beta_n \in \{a, b\}$;
(CS2) $\{a, b\} \subset \liminf J_n$.

A spiral S from a continuum S^- to a nondegenerate continuum S^+ is said to be *crooked (from S^- to S^+)* if S is crooked near S^+.

We will use the notion of *order of a point* in the sense of Menger-Urysohn (see e.g. [11, §51, I, p. 254]). In particular, a point p of a subset S of a space X is called an *end point* of S provided that it is of order one in S, i.e., for each $\varepsilon > 0$ there is a neighborhood U of p such that $\operatorname{diam} U < \varepsilon$ and $\operatorname{card}(S \cap \operatorname{bd} U) = 1$.

We start with recalling a well known result, see [17, Chapter 8, (7.31), p. 147].

2.2. Statement. *The order of a point is never increased under an open mapping.*

2.3. Proposition. *If $S = g(\mathbb{R}) \subset X$ is a spiral and $f : X \to f(X) = Y$ is a nonconstant open mapping, then $f(S)$ is homeomorphic to one of the following four sets:*
(a) *the open interval* $(0, 1)$*;*
(b) *the half-open interval* $[0, 1)$*;*
(c) *the closed interval* $[0, 1]$*;*
(d) *the circle.*

Moreover, if (a) holds, then $f|S$ is injective and if (b) holds, then exists $c \in \mathbb{R}$ such that $f|g((-\infty, c])$ and $f|g([c, \infty))$ are injective.

Proof. The first part is obvious, because the open mapping does not increase the order of the point due to Statement 2.2. The first part of 'moreover' statement is also obvious, it is sufficient to prove the last one. To see this, let $h : f(S) \to [0, 1)$ be the above mentioned homeomorphism. Denote $p = h \circ f \circ g$. Choose $c \in \mathbb{R}$ such that $p(c) = 0$. Suppose on the contrary that $p|[c, \infty)$ is not injective. This means that there are $a, b \in \mathbb{R}$ with $c \le a < b$ such that $p(a) = p(b) = y$ for some $y \in [0, 1)$. Then $p([a, b])$ is an arc J in $[0, 1)$.

If $0 \in J$, then fix some $d \in (a, b]$ such that $p(d) = 0$; now $p([c, d))$ is an arc $0z$ for some $z \in (0, 1)$ and f is not interior at $f^{-1}(h^{-1}(z))$, a contradiction.

If $0 \notin J$, then fix an end point z distinct from y; now f is not interior at $f^{-1}(h^{-1}(z)) \in g([c, b])$, a contradiction.

We have shown that f is injective on $g([c, \infty))$. Similarly we can show that f is injective on $g((-\infty, c])$. Note further that the point c is uniquely determined. □

Let K be a subset of a continuum X and S be an open spiral in X. We say that a mapping f *cycles on S near K* if either

(i) $f|S \cap U \cap V$ is not injective for each open set U containing K and each open set V containing S^+

or

(ii) $f|S \cap U \cap V$ is not injective for each open set U containing K and each open set V containing S^-.

We say that a mapping f *cycles on* S if a mapping f cycles on S near S^+ or near S^-.

2.4. Proposition. *Let S be a spiral in a continuum X. If a nonconstant open mapping $f : X \to Y$ cycles on S near S^+ then $f(S^+) = f(S) = f(X)$ is homeomorphic either to the unit circle or to the unit interval.*

Proof. We know by Proposition 2.3 that $f(X)$ is homeomorphic either to the unit circle or to the unit interval. Fix first an open $U \supset S^+$.

Claim. In U there exist 3 distinct points with the same image under f.

(i) "circle case"

We find reals $s_1 < t_1 < s_2 < t_2 < s_3 < t_3$ such that $g([s_1, \infty)) \subset U$ and $f(g(s_i)) = f(g(t_i)) = y_i$ for $i \in \{1, 2, 3\}$. The image $f(g[s_i, t_i])$, for $i \in \{1, 2, 3\}$, is a closed connected subset of Y and equals Y (cannot be an arc J in Y because the mapping f must be interior at the preimage of the end point of J distinct from y_i). Then each $y \in Y$ has a preimage in $g[s_i, t_i]$, for $i \in \{1, 2, 3\}$. Hence Claim holds.

(ii) "interval case"

We find reals $s_i < t_i < s_{i+1}$ such that $g([s_1, \infty)) \subset U$ and $f(g(s_i)) = f(g(t_i)) = y_i$ for $i \in \{1, 2, 3, 4, 5\}$. Fix $i \in \{1, 2, 3, 4, 5\}$. The image $f(g[s_i, t_i])$ is an arc J_i in Y. At least one end point of J_i is distinct from y_i and belongs to $\{0, 1\}$ due to the openness of f at X. We have found $x_i \in (s_i, t_i)$ such that $f(g(x_i)) \in \{0, 1\}$. Among five such points $g(x_i)$ at least three have the same image under f. Claim is proved.

Let $a, b, c \in U$ be such that $f(a) = f(b) = f(c)$, $g^{-1}(a) < g^{-1}(b) < g^{-1}(c)$. Then $f(g([g^{-1}(a), g^{-1}(b)])) = f(g((g^{-1}(a), g^{-1}(b))))$ is both open and closed subset of Y, hence equals Y.

Finally, we conclude that $f(S^+) = Y$ due to the continuity of f. $\square$

2.5. Proposition. *Let S be a spiral in a continuum X crooked near S^+. Let $f : X \to Y$ be a nonconstant open mapping. Then $f : X \to Y$ cannot cycle on S near S^+.*

Proof. Suppose on the contrary that f cycles on S near S^+. Using Proposition 2.4 and composing f with an open projection from a circle onto a unit interval (if necessary) we may assume that $f(X) = Y = [0, 1]$. By Proposition 2.4 we see that $f(S^+)$ (which equals $f(X)$) is not a singleton. Take $a \in S^+$ such that $0 \neq f(a) \neq 1$ such that arbitrarily close to a there are points in S^+ which are not mapped to $f(a)$. Then there is an open set W containing a such that $f(W) \subset (0, 1)$.

By the crookedness of S there is a neighborhood $U \subset W$ of a such that for any fixed point $b \in U \cap S^+$ with $f(b) \neq f(a)$ there exists in $U \cap S$ a sequence of arcs J_n with end points α_n, β_n satisfying conditions (CS1) - (CS2).

Let $\varepsilon = |f(b) - f(a)|/3$. By the continuity of f there are in U open neighborhoods V_a and V_b of a and of b respectively, such that $f(V_a) \subset \{y \in (0, 1) : |y - f(a)| < \varepsilon\}$ and $f(V_b) \subset \{y \in (0, 1) : |y - f(b)| < \varepsilon\}$.

In the sequence $\{J_n\}$ of arcs in S one can find an arc $J \subset C \cap U$ such that J meets both V_a and V_b and its end points α and β are contained in one of the mentioned sets; we may assume that they are contained in V_a. Note that $f(J)$ is an arc in $Y = [0,1]$ not containing the end points of Y. Moreover, $f(J)$ meets both $f(V_a)$ and $f(V_b)$. Hence, at least one of the end points of $f(J)$, denote it by y, is distinct from both $f(\alpha)$ and $f(\beta)$. But then f is not interior at that point $x \in J \setminus \{\alpha, \beta\}$ which is mapped onto y. We have obtained a contradiction with the openness of f. $\square$

We say that a subcontinuum K of a continuum X *is controlled by* a (finite or infinite) sequence of spirals $\{S_i\}_{i\in I}$ in X if K does not meet any S_i for $i \in I$ and in K there exists a point x with positive distance to

$$X \setminus \left(K \cup \bigcup_{i \in I} \operatorname{cl} S_i \right).$$

Denote the set of these points by $K * \{S_i\}_{i\in I}$. If K is a continuum controlled by a (finite or infinite) sequence of spirals $\{S_i\}_{i\in I}$ in X, then $K * \{S_i\}_{i\in I}$ contains a nondegenerate subcontinuum due to [14, Corollary 3.5, p. 74].

2.6. Proposition. *Let K be a nondegenerate subcontinuum of a continuum X. Let K be controlled by a finite sequence $\{S_i\}_{i=1}^n$ of spirals in X. Let $f : X \to Y$ be a nonconstant open mapping. If $f(K)$ is a singleton in Y, then for some $i \in \{1, \cdots, n\}$ the mapping f cycles on S_i .*

Proof. Let a nondegenerate subcontinuum K of a continuum X be controlled by a finite sequence $\{S_i\}_{i=1}^n$ of spirals in X. Let $f : X \to Y$ be a nonconstant open mapping, $f(K)$ be a singleton.

Suppose on the contrary that for $i \in \{1, 2, \ldots, n\}$ the mapping f does not cycle on S_i. We conclude a contradiction.

If K has a nonempty interior we obtain a contradiction with the openness of f. Hence we can assume that

(2.6.1) arbitrarily close to any point of K there are points of $\bigcup_{i=1}^n S_i$.

Notice first that if for some $i, j \in \{1, 2, \ldots, n\}$ the image $f(S_j^+)$ meets $f(S_i)$ in two distinct points, then f cycles on S_j near S_j^+ and the proof is finished.

Now we can assume that they meet at most at some singleton. If the inverse image of this singleton meets S_i arbitrarily close to either S_i^+ or S_i^- we conclude that f cycles on S_i and the proof is finished. The same arguments hold for S_j^-.

In all other cases we can replace the spiral S_i by two spirals $S_{i+1/3} \subset S_i$ from $c_i^+ \in S_i$ to S_i^+ and $S_{i-1/3} \subset S_i$ from $c_i^- \in S_i$ to S_i^- in such a way that $f(S_j^+) \cap f(S_{i\pm 1/3}) = \emptyset$ and $f(S_j^-) \cap f(S_{i\pm 1/3}) = \emptyset$ for $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$. A similar replacement can be done to obtain f injective on each 'new' spiral.

Due to this discussion we can assume that K is controlled by a finite sequence $\{S_i\}_{i=1}^n$ of spirals in X with this properties :

(2.6.2) for $i \in \{1, 2, \ldots, n\}$ the mapping f is injective on S_i,
(2.6.3) $f(S_j^+) \cap f(S_i) = \emptyset$ for $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$,
(2.6.4) $f(S_j^-) \cap f(S_i) = \emptyset$ for $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$.

To obtain a contradiction suppose that $f(K) = \{y\}$ and take distinct points $x_1, x_2, \ldots, x_{n+1}$ in K with a positive distance to

$$X \setminus \left(K \cup \bigcup_{i=1}^{n} \operatorname{cl} S_i \right)$$

and their disjoint open neighbourhoods $U_1, U_2, \ldots, U_{n+1}$ such that $U_i \subset K \cup \bigcup_{j=1}^{n} \operatorname{cl} S_j$ for $i \in \{1, \ldots, n+1\}$. Then for $i \in \{1, \ldots, n+1\}$ the image $f(U_i)$ and also their intersection V is an open neighbourhood of y. Since f is a nonconstant open mapping and due to (2.6.1) we can take

$$y' \in \left(V \cap \bigcup_{i=1}^{n} f(S_i) \right) \setminus \{y\}.$$

For $i \in \{1, \ldots, n+1\}$ there exists $z_i \in U_i$ such that $f(z_i) = y'$. Now z_i cannot be in K for $i \in \{1, \ldots, n+1\}$ (because $f(K) = \{y\}$ and $y' \neq y$). Moreover, they cannot be in S_j^+ due to the construction. Indeed, the images of S_j^+ do not meet images of S_i for $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ due to (2.6.3). The same is true for S_j^- due to (2.6.4).

We conclude that $z_i \in S_{j_i}$ for some $j_i \in \{1, 2, \ldots, n\}$. There must be $j_i = j_k$ for some $i \neq k$ what is a contradiction with the injectivity of f on spirals in $\{S_i\}_{i=1}^{n}$ due to (2.6.2). $\square$

We show in Example 2.8 that the we cannot replace the finite sequence in Proposition 2.6 by an infinite one.

2.7. Lemma. *Let $\varepsilon > 0$ and $d \in [-1, 0]$ be given. Put $L = \{(x, 0) \in \mathbb{R}^2 : x \in [-1, 1]\}$, $K = \{(x, 0) \in \mathbb{R}^2 : x \in [-1, 0]\}$, $c = (\varepsilon, 0)$. For $n \in \mathbb{N}$ put $s_n = (0, \varepsilon/n)$ and $t_n = (-1, 3\varepsilon/4n)$. Put*

$$A = cs_1 \cup \bigcup_{n=1}^{\infty} s_n t_n \cup \bigcup_{n=1}^{\infty} t_n s_{n+1},$$

$B = \{(x, y) \in \mathbb{R}^2 : (x, -y) \in A\}$ and $X = A \cup B \cup L$.

Then there exists a nonconstant mapping $g : X \to [0, 1]$ which is interior at $(d, 0)$ such that $g|(X \setminus K)$ is open.

Proof. See Figure 1. Let $g|L$ be a mapping sending K to $0 \in [0, 1]$ and identity elsewhere. Let $\{A_n\}_{n=1}^{\infty}$ be any sequence of mutually disjoint arcs in A, $c \in A_1$, converging to $(d, 0)$ such that each $A_n \setminus \{c\}$ separates $\{c\}$ and A_{n+1} in A. Let $g|A$ be any homeomorphism of A onto the interval $(0, \varepsilon]$ already defined at $c \in A \cap L$. For $n \in \mathbb{N}$ let $\{B_n\}$ be any sequence of mutually disjoint arcs in $B \setminus \{c\}$ converging to $(d, 0)$ such that each B_n separates $\{c\}$ and B_{n+1} in B. Let $g|B$ be any homeomorphism of B onto the interval $(0, \varepsilon]$ already defined at $c \in B \cap L$ such that $g(B_n)$ meets both $g(A_n)$ and $g(A_{n+1})$. We see that g is interior at $(d, 0)$ and $g|(X \setminus K)$ is open. $\square$

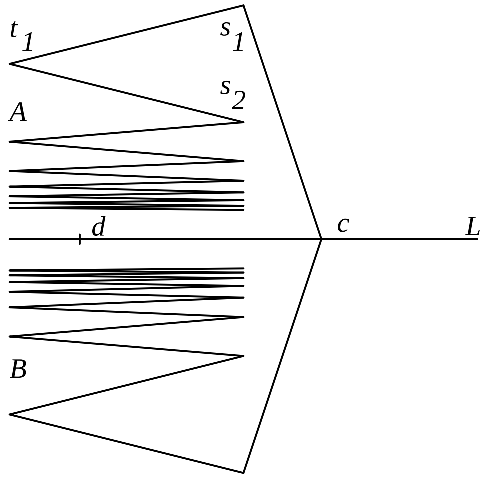


FIGURE 1 (LEMMA 2.7)

Using Lemma 2.7 we show that the finite sequence of spirals in Proposition 2.6 cannot be changed to an infinite one.

2.8. Example. *There exist a continuum X and a nondegenerate subcontinuum K controlled by an infinite sequence $\{S_i\}_{i=1}^{\infty}$ of spirals in X and a nonconstant open mapping $f : X \to Y$ such that $f(K)$ is a singleton in Y and f is injective on all S_i for $i \in \mathbb{N}$.*

Proof. Put $L = \{(x, 0, 0) \in \mathbb{R}^3 : x \in [-1, 1]\}$ and let $\{(d_n, 0, 0)\}_{n=1}^{\infty}$ be a dense sequence in $K = \{(x, 0, 0) \in \mathbb{R}^3 : x \in [-1, 0]\}$. Fix $n \in \mathbb{N}$.

We use Lemma 2.7 with $\varepsilon = 1/n$ and $d = d_n$. We obtain $A_n = A$, $B_n = B$ and $g_n = g$. We put $S_n = \pi_n(A_n \cup B_n)$ where π_n is the orthogonal projection from the plane $P_0 = \{(x, y, z) \in \mathbb{R}^3 : z = 0\}$ onto the plane $P_n = \{(x, y, z) \in \mathbb{R}^3 : z = ny\}$.

Now let

$$X = L \cup \bigcup_{n=1}^{\infty} S_n.$$

We see that X is a continuum. Let $Y = [0, 1]$. We define a mapping $f : X \to Y$ in the following way.

Let $f|L$ be a mapping sending $(x, 0, 0) \in L$ to $\max(0, x) \in Y$. For $n \in \mathbb{N}$ define $f|S_n = g_n \circ \pi_n^{-1}$.

We see that in the continuum X there is a nondegenerate subcontinuum K controlled by an infinite sequence $\{S_i\}_{i=1}^{\infty}$ of spirals in X and that the nonconstant mapping $f : X \to Y$ satisfies

(2.8.1) f is open;
(2.8.2) $f(K)$ is a singleton in Y;
(2.8.3) f is injective on all S_i for $i \in \mathbb{N}$.

To see (2.8.1) we distinguish points in K, L and S_n. The openness of f follows from the definition of f in X and the density of points $(d_n, 0, 0)$ in K, because each g_n guarantees the interiority of f at $(d_n, 0, 0)$. Properties (2.8.2) and (2.8.3) are easy observations. This finishes the proof. □

2.9 Remark. In [7, Remark 3.5] a question is posed, whether the openness of a spiral S from a continuum to a nondegenerate continuum S^+ guarantees that the image of S^+ under any nonconstant open mapping f is nondegenerate. We have seen in Example 2.8 that the answer is negative. Indeed, there can be infinitely many open spirals which ensures the interiority of f at S^+.

In [7, Remark 3.5] a question is posed, whether the openness of a spiral S crooked from a subcontinuum S^- to a nondegenerate subcontinuum S^+ in a given continuum X guarantees, that if the image $f(X)$ is either an arc or a simple closed curve, then f is not open. We will see in the following example that the answer to this question is negative.

2.10 Example. *There exist a continuum X, an open mapping $f : X \to [-1,1]$ and an open spiral S in X crooked near a nondegenerate subcontinuum S^+ of X.*

Proof. Put $L = [-1,1]$. For $n \in \mathbb{N}$ let $r_{2n-1} = -1/5$ and $\{r_n\}_{n=1}^{\infty}$ be a sequence containing all rational numbers in the interval $[-1/5, 1/5]$. Put

$$A = \bigcup_{n=1}^{\infty} a_n a_{n+1},$$

where $a_n = (-1/n, r_n)$ for $n \in \mathbb{N}$. Put

$$B = \bigcup_{n=1}^{\infty} b_n b_{n+1},$$

where $b_n = (1/n, (-1)^n)$ for $n \in \mathbb{N}$. Now we set $X = A \cup B \cup L$. The continuum is pictured on Figure 2.

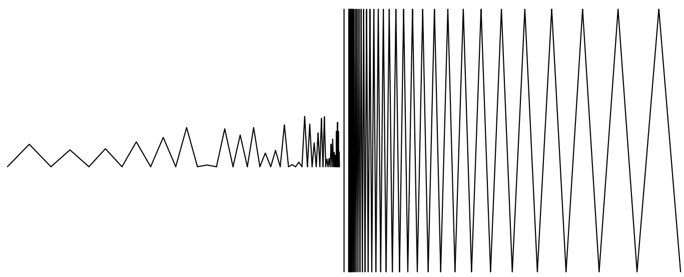


FIGURE 2 (EXAMPLE 2.10, CONTINUUM X)

We see that $S = A \setminus \{a_1\}$ is a spiral crooked near $S^+ = \{0\} \times [-1/5, 1/5]$. We define the mapping $f : X \to [-1,1]$ in the following way

(2.10.1) $f|A$ is any homeomorphism of A onto $(0,1]$ such that $f(a_1) = 1$;

(2.10.2) $f|(B \cup L) : B \cup L \to [-1,1]$ is a modification of the open non-light mapping from the $\sin(1/x)$-curve onto $[-1,1]$ sending $\{0\} \times [-1/3, 1/3]$ to $0 \in [-1,1]$ (see [6, Remark 18, p. 242]).

Then we observe that f is an open mapping of the continuum X onto $[-1,1]$. □

3. Classes $\mathcal{S}$ and $\mathcal{S}_c$.

We have seen in Example 2.8 that we can guarantee the interiority of a mapping at n points of a subcontinuum with $2n$ spirals and that if there is an infinite number of spirals approximating a nondegenerate continuum, then there is an open mapping from this space which is not light. This motivates the definition of suitable classes of continua using only (locally) finitely many spirals in the following paragraph.

The class $\mathcal{S}$ denotes the class of all continua at which all subcontinua are controlled by a finite sequence of spirals.

The class $\mathcal{S}_c$ denotes the class of all continua in $\mathcal{S}$ which have a nondegenerate subcontinuum K such that K is controlled by a finite sequence $\{S_i\}$ of crooked spirals and $K \subset \cup S_i^+$.

Notice that continua pictured Figures 1-4 are in $\mathcal{S}$ and continuum pictured in Figure 3 is in $\mathcal{S}_c$. See also continua in [6, Figures 1-9] which are in $\mathcal{S}$ and continua pictured in [7, Figures 1-2] which are in $\mathcal{S}_c$.

3.1. Proposition. *Let $X \in \mathcal{S}$ and $f : X \to Y$ be a nonconstant open non-light mapping. Then*

(i) f cycles on some spiral in X;

(ii) $f(X)$ is homeomorphic either to the unit circle or to the unit interval.

Proof. We have (i) due to Proposition 2.6 and (ii) due to Proposition 2.4. $\square$

3.2. Theorem. $\mathcal{S}_c \subset \mathsf{L}$.

Proof. Let $X \in \mathcal{S}_c$ and $f : X \to Y$ be a nonconstant open non-light mapping. We conclude a contradiction.

According to Proposition 3.1., $f(X)$ is homeomorphic to the unit circle or to the unit interval. Composing f with the open projection from the unit circle onto the unit interval (if necessary) we may assume that $f(X)$ is homeomorphic to the unit interval. Let K be a subcontinuum of X controlled by a finite sequence $\{S_i\}_{i=1}^n$ of crooked spirals and $K \subset \bigcup_{i=1}^n S_i^+$.

For $i \in \{1, \cdots, n\}$ the mapping f cannot cycle on S_i near S_i^+ according to Proposition 2.5., and so we can assume that f is injective on S_i due to Proposition 2.3 (we can replace S_i by a spiral from some $s_i \in S_i$ to S_i^+). Hence $f(S^+)$ is a singleton because $f(X)$ is homeomorphic to the unit interval. We conclude that $f(K)$ is a singleton. By Proposition 2.6. we get that for some $i \in \{1, \cdots, n\}$ the mapping f cycles on the spiral S_i, what is the contradiction due to Proposition 2.5, because spirals in $\{S_j\}_{j=1}^n$ are crooked. $\square$

Notice that in [7, Theorem 2.6] only continua of the form $X = S^- \cup S \cup S^+$ are considered, where S is a crooked spiral. Theorem 3.2 extends these results to a very large family.

3.3. Example. *There exist an arcwise connected continuum X in* L and an irreducible mapping f such that $f(X)$ is not in L .

Proof. For t real we denote $a_t = (0, t)$ and $b_t = (8, t)$. Set

$$E = b_3b_0 \cup b_0a_0 \cup a_0a_8 \cup a_8b_8 \cup b_8b_5.$$

For $n \in \mathbb{N}$ let $r_{2n-1} = 5$ and $\{r_n\}_{n=1}^{\infty}$ be a sequence containing all rational numbers in the interval $[5, 7]$. Put

$$C = \bigcup_{n=1}^{\infty} c_n c_{n+1},$$

where $c_n = (8/n, r_n)$ for $n \in \mathbb{N}$. Put

$$D = \bigcup_{n=1}^{\infty} d_n d_{n+1},$$

where $d_n = (8/n, 2-(-1)^n)$ for $n \in \mathbb{N}$. Now we set $X = C \cup D \cup E$. The continuum is pictured on Figure 3.

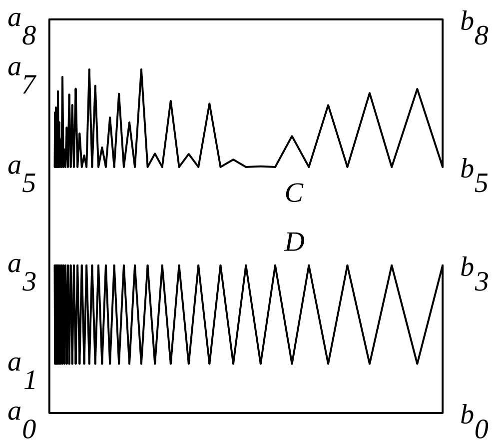


FIGURE 3 (EXAMPLE 3.3, CONTINUUM X)

We see that $X \in \mathcal{S}_c$. Now we construct an irreducible mapping $f : X \to Y$. We define f in the following way

(i) $f(a_5a_7) = \{a_5\}$;

(ii) for $(0, y) \in a_7a_8$ we define $f(0, y) = (0, 3y - 16)$;

(iii) for $(x, y) \in C$ we define $f(x, y) = (x, 5)$;

(iv) f is the identity elsewhere.

The mapping f is continuous and $Y = f(X)$ is a continuum. See Figure 4.

We show that f is an irreducible mapping and that and $Y \notin \mathsf{L}$.

We start with the proof that f is irreducible. Let $I \subset Y$ be a nondegenerate subcontiuum irreducible about $\{p, q\}$. Put $J = f^{-1}(I)$. If $\{p, q\} \cap \{a_5\} = \emptyset$, then J is irreducible about $\{f^{-1}(p), f^{-1}(q)\}$. If $p = a_5$ and J is an arc then it is irreducible about its end points. If J is not an arc then it contains either a spiral $C' \subset C$ from some $c \in C$ to a_5a_7 or a spiral $D' \subset D$ from some $d \in D$ to a_1a_3. In both cases J is irreducible about $\{f^{-1}(q), a_5\}$.

To show that $Y \notin \mathsf{L}$ we define an open non-light mapping $g : Y \to a_1a_3$ in the following way :

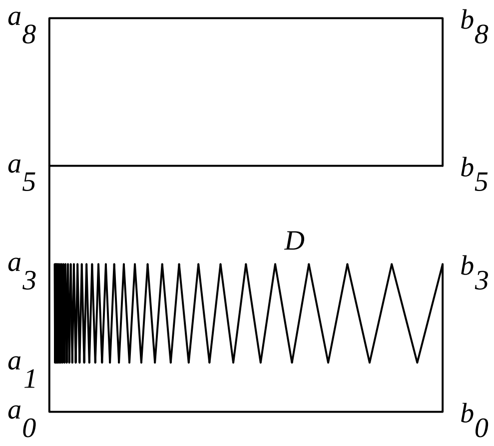


FIGURE 4 (EXAMPLE 3.3, CONTINUUM Y)

(α) We adapt the open non-light mapping f from the $\sin(1/x)$-curve (see [6, Remark 18, p. 242]) to get an open non-light surjective mapping $g : (D \cup a_1a_3) \to a_1a_3$, $g(b_3) = a_3$. This gives $g(a_{5/3}a_{7/3}) = \{a_2\}$.

(β) We define $g|(a_1a_0 \cup a_0b_0 \cup b_0b_3)$ to be any homeomorphism onto a_1a_3 (already defined in a_1 and b_3).

(γ) We define $g|(a_3a_8 \cup a_8b_8)$ to be any homeomorphism onto a_1a_3 (already defined in a_3).

(δ) We define $g|(a_5b_5 \cup b_5b_8)$ to be any homeomorphism onto $a_1g(a_5)$ (already defined in a_5 and b_8).

The continuity and the openness of g and the fact that g is not light follows from the construction of g. □

REFERENCES

1. J. M. Aarts, P. van Emde Boas, *Continua as remainders in compact extensions*, Nieuw Arch. Wisk. **15** (1967), 34–37.
2. D. E. Bennet, J. B. Fugate, *Continua and their non-separating subcontinua*, Dissertationes Math. (Rozprawy Mat.) **149** (1977), 1 – 50.
3. J. J. Charatonik, *Absolutely terminal continua and confluent mappings*, Comment. Math. Univ. Carol. **32,2** (1991), 377 – 382.
4. J. J. Charatonik, W. J. Charatonik, S. Miklos, *Confluent mappings on fans*, Dissertationes Math. (Rozprawy Mat.) **301** (1990), 1–86.
5. J. J. Charatonik, K. Omiljanowski, *On light open mappings*, Baku International Topological Conference Proceedings, ELM, Baku, 1989, pp. 211–219.
6. J. J. Charatonik, P. Pyrih, *Continua with light open mappings*, Questions Answers Gen. Topology **17** (1999), 237 – 256.
7. J. J. Charatonik, P. Pyrih, *Atomic mappings can spoil lightness of open mappings*, Tsukuba J. of Math. (to appear).
8. A. Emeryk, Z. Horbanowicz, *On atomic mappings*, Colloq. Math. **27** (1973), 49–55.
9. J. B. Fugate, *Decomposable chainable continua*, Trans. Amer. Math. Soc. **123** (1966), 466 – 468.
10. G. R. Gordh Jr., *Terminal subcontinua of hereditarily unicoherent continua*, Pacific J. Math. **47** (1973), 457 – 464.
11. K. Kuratowski, *Topology, Vol. II*, Academic Press, New York, N.Y., 1968.
12. T. Maćkowiak, *Continuous mappings on continua*, Dissertationes Math. (Rozprawy Mat.) **158** (1979), 1–95.
13. W. Makuchowski, *On open light mappings*, Comment. Math. Univ. Carol **35,4** (1994), 779–788.

14. S. B. Nadler Jr., *Continuum theory: an introduction*, M. Dekker, Inc., New York, Basel and Hong Kong, 1992.
15. P. Pyrih, T. Bárta, D. Opěla, P. Růžička, R. Šámal, *Irreducible mappings and HU-terminal continua*, South-West Journal of Pure and Applied Mathematics **5** (1999), 31–41.
16. P. Pyrih, *Problem Book*, Open Problem Seminar, Department of Math. Analysis, Charles University, Prague.
17. G. T. Whyburn, *Analytic topology*, Amer. Math. Soc. Colloq. Publ. 28, Providence, 1942, reprinted with corrections 1971.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2001, pp. 43–52.
Submitted: April 2, 2001. Published: December 31 2001.

SOME FORMULAS FOR CONJUGATION

Ismet Karaca and I. Yaslan Karaca

Abstract. In this paper we prove a formula involving the canonical anti-automorphism χ of the mod-p Steenrod algebra and its dual.

1991 A.M.S. (MOS) Subject Classification Codes. 55S10, 55S05

Key Words and Phrases. Steenrod algebra, anti-automorphism, Milnor basis

1. INTRODUCTION AND MAIN RESULTS

It is well known that the mod-p Steenrod algebra $\mathcal{A}$, being a cocommutative Hopf algebra, has a canonical conjugation χ. This linear map was first introduced by Thom [8] and later Milnor [5] discovered most of what we know today about χ. In 1974 Davis [1] proved nice formulae involving canonical anti-automorphism.

We recall the structure of dual algebra $\mathcal{A}^*$ of $\mathcal{A}$ which was computed by Milnor [5]. The algebra $\mathcal{A}^*$ is the tensor product of a polynomial algebra on generators ξ_i $i \geq 1$ of degree $2(p^i - 1)$ by an exterior algebra on generators τ_i $i \geq 0$ of degree $2p^i - 1$. So the monomial basis in $\mathcal{A}^*$ can be written in the form

$$\xi^I = \tau_0^{\epsilon_0} \xi_1^{r_1} \tau_1^{\epsilon_1} \cdots$$

Then dual to the the monomial basis ξ^I is a basis for the mod-p Steenrod algebra known as Milnor basis. So $Q_0^{\epsilon_0} Q_1^{\epsilon_1} \cdots P(r_1, r_2, \dots)$ will denote the dual of ξ^I.

The structure of the mod-p Steenrod algebra $\mathcal{A}$ is given as a graded $\mathbb{Z}_p$-module on the Milnor basis with product and coproduct defined in terms of that basis. If $R = (r_1, r_2, \dots, r_m)$ is a sequence of nonnegative integers, we will write $\mathcal{P}(R)$ for the corresponding Milnor basis element. The degree of $\mathcal{P}(R)$ is $2\,|R|$ where $|R| = \sum_{i=1}^{m} \left(p^i - 1\right) r_i$ and the excess of $\mathcal{P}(R)$ is $2e(R)$ where $e(R) = \sum_{i=1}^{m} r_i$. Let B be the vector space of $\mathcal{A}$ with the basis the set of all $\mathcal{P}(R)$. For $\mathcal{P}(a) \in B$, write $\hat{\mathcal{P}}(a)$

Department of Mathematics Ege University, IZMIR 35100 TURKEY
E-mail Addresss : karaca@fenfak.ege.edu.tr, yaslan@fenfak.ege.edu.tr

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX for $(-1)^a\chi(\mathcal{P}(a))$ where χ denotes the canonical conjugation. Given pair (k, n) of nonnegative integers, we define

$$X(k,n) = \mathcal{P}(p^k(p^{n+1}-1))\mathcal{P}(p^{k-1}(p^{n+1}-1))\cdots\mathcal{P}(p(p^{n+1}-1))\mathcal{P}((p^{n+1}-1)).$$

First main result which is a conjugation formula for a certain Steenrod opreations $X(k,n)$ is mod-p analogue of Theorem 3.1 in [6]

Theorem 1.1. *For all positive integers i and j, we have*

$$\widehat{X}(j,i) = X(i,j).$$

Our second main result is also a conjugation formula for the certain elements $\mathcal{W}_I(k)$ in the dual $\mathcal{A}^*$ of Steenrod algebra which will be defined in Section 3. This formula generalizes the Proposition 5.5 in [7].

Theorem 1.2. *For any integers $k \geq 1$ and $i_0 > 0$ we have*

$$\widehat{\mathcal{W}_I}(k) \equiv \xi_k^{\frac{p^{i_0}-1}{p-1}}\widehat{\mathcal{R}}_I(k)$$

modulo elements of length $> k$, where $\mathcal{R}_I(k)$is defined in Section 3.

2. PRELIMINARIES

We will introduce the useful notation: each natural number a has a unique p-adic expansion

$$a = \sum_{i=0}^{\infty} \alpha_i(a)p^i$$

with $0 \leq \alpha_i(a) < p$. It is fact that

$$\binom{a}{b} \equiv \prod_{i=0}^{\infty} \binom{\alpha_i(a)}{\alpha_i(b)} \mod p. \tag{1}$$

We have following definition;

Definition 2.1. We say that m and n have no carries if $\alpha_i(m) + \alpha_i(n) \leq p-1$ for all i. This is equivalent to the condition that the binomial coefficient $\binom{m+n}{m}$ is nonzero mod p . If m and n have no carries, we will write $m \asymp n$.

Using Davis' method [1] we can derive the formulae,

Proposition 2.2.

$$\mathcal{P}(u)\cdot\widehat{\mathcal{P}}(v) = \sum_R \binom{|R|+e(R)}{pu}\mathcal{P}(R) \tag{2}$$

$$\widehat{\mathcal{P}}(u)\cdot\mathcal{P}(v) = \sum_R \binom{e(R)}{v}\mathcal{P}(R) \tag{3}$$

where the sum is taken over all R for which $|R| = (p-1)(u+v)$.

Proof. Corollary $1a$ in [2] is exactly formula (2) when $t = 1$. So we will prove formula (3). Using Gallant's Proposition 1 that $\chi(\mathcal{P}(a))$ is $(-1)^a$ times the sum of all $\mathcal{P}(R)$ in the appropriate dimension, we see that the product $\widehat{\mathcal{P}}(u)\cdot\mathcal{P}(v)$ contains a term

$$\prod_i \binom{r_i}{s_i}\mathcal{P}(r_1, r_2, \dots)$$

for each Milnor matrix,

$$\begin{pmatrix} * & s_1 \\ r_1 - s_1 & s_2 \\ r_2 - s_2 & s_3 \\ \vdots & \vdots \\ r_{n-1} - s_{n-1} & s_n \\ r_n - s_n & s_{n+1} \\ \vdots & \vdots \end{pmatrix}$$

such that $v = \sum_i s_i$. Thus

$$\widehat{\mathcal{P}}(u)\cdot\mathcal{P}(v) = \sum_R \sum_S \prod_i \binom{r_i}{s_i}\mathcal{P}(R)$$

where S ranges over sequences $(s_1, s_2, \dots)$ having $\sum\limits_i s_i = v$. The equation $\sum\limits_S \prod\limits_i \binom{r_i}{s_i} = \binom{\sum_i r_i}{v}$ follows immediately from considering the coefficient of x^v in the expansion of

$$(1+x)^{r_1}\cdot(1+x)^{r_2}\cdots(1+x)^{r_n}.$$

Hence we have

$$\widehat{\mathcal{P}}(u)\cdot\mathcal{P}(v) = \sum_R \binom{\sum_i r_i}{v}\mathcal{P}(R) = \sum_R \binom{e(R)}{v}\mathcal{P}(R).$$

□

Using formulae above, we can prove the following proposition.

Proposition 2.3. *For nonnegative integers k, l, m, n and $k > l$, suppose that*

(1) $m + n = p^k - p^l$
(2) $m < p^{k-1}$
(3) $m \equiv 0 \mod p^l$.

Then

(i) *When $l = 0$, we have*

$$\mathcal{P}(m)\cdot\widehat{\mathcal{P}}(n) = \widehat{\mathcal{P}}(n-(p-1)m-1)\mathcal{P}(pm+1)$$

(ii) *When $l > 0$, we have*

$$\mathcal{P}(m)\cdot\widehat{\mathcal{P}}(n) = \widehat{\mathcal{P}}(n-(p-1)(m-p^l))\mathcal{P}(pm+(p-1)p^l) - \sum_{w=1}^{p-1}\binom{p-1}{w}\mathcal{P}(m+wp^{l-1})\cdot\widehat{\mathcal{P}}(n-wp^{l-1})$$

Proof.
(i) Let $l = 0$. Using Proposition 2.2, we have

$$\mathcal{P}(m) \cdot \widehat{\mathcal{P}}(n) = \sum_R \binom{|R| + e(R)}{pm} \mathcal{P}(R),$$

and

$$\widehat{\mathcal{P}}(n - (p-1)m - 1) \cdot \mathcal{P}(pm+1) = \sum_R \binom{e(R)}{pm+1} \mathcal{P}(R)$$

where $|R| = (p-1)(p^k - 1)$ and $1 \le e(R) \le p^k - 1$. In order to prove these sums are equivalent in mod-p, we need to show that their binomial coefficients are equivalent in mod-p i.e.

$$\binom{|R| + e(R)}{pm} \equiv \binom{e(R)}{pm+1} \mod p.$$

We know that $|R| = \sum\limits_{i=1}^{\infty} r_i(p^i - 1)$ and $e(R) = \sum\limits_{i=1}^{\infty} r_i$. Using these facts, we have

$$\begin{aligned}
p^k - 1 = (p-1)(p^{k-1} + p^{k-2} + \cdots + p + 1) &= \frac{|R|}{p-1} \\
&= \sum_{i=1}^{\infty} r_i(p^{i-1} + p^{i-2} + \cdots + p + 1) \\
&= r_1 + \sum_{i=2}^{\infty} r_i(p^{i-1} + p^{i-2} + \cdots + p + 1) \\
&= e(R) - \sum_{i=2}^{\infty} r_i + \sum_{i=2}^{\infty} r_i(p^{i-1} + p^{i-2} + \cdots + p + 1) \\
&= e(R) + \sum_{i=2}^{\infty} r_i p(p^{i-2} + p^{i-3} + \cdots + p + 1).
\end{aligned}$$

Since $\sum\limits_{i=2}^{\infty} r_i p(p^{i-2} + p^{i-3} + \cdots + p + 1) \equiv 0 \mod p$, $e(R) \equiv p - 1 \mod p$. Using Equation 1, and the upper bounds of $e(R)$ and m, we have

$$\binom{|R| + e(R)}{pm} = \binom{(p-1)(p^k - 1) + e(R)}{pm} \equiv \binom{e(R)}{pm+1} \mod p.$$

This completes the proof of part (i).
(ii) Let $l > 0$. Again using Proposition 2.2, we have

$$\mathcal{P}(m) \cdot \widehat{\mathcal{P}}(n) = \sum_R \binom{|R| + e(R)}{pm} \mathcal{P}(R),$$

$$\mathcal{P}(m + wp^{l-1}) \cdot \widehat{\mathcal{P}}(n - wp^{l-1}) = \sum_R \binom{|R| + e(R)}{pm + wp^l} \mathcal{P}(R),$$

and

$$\widehat{\mathcal{P}}(n-(p-1)(m-p^l))\cdot\mathcal{P}(pm+(p-1)p^l)=\sum_R\binom{e(R)}{pm+(p-1)p^l}\mathcal{P}(R)$$

where $|R|=(p-1)(p^k-p^l)$ and $1\le e(R)\le p^k-p^l$. In order to prove the sums in part (ii) are equivalent, we need to show that

$$\binom{|R|+e(R)}{pm}+\sum_{w=1}^{p-1}\binom{p-1}{w}\binom{|R|+e(R)}{pm+wp^l}\equiv\binom{e(R)}{pm+(p-1)p^l}\mod p.$$

Case 1: $0\le\alpha_l(e(R))<p-1$. Then $\binom{e(R)}{pm+(p-1)p^l}$ are equivalent to zero in mod-p. So it is enough to show that

$$\binom{|R|+e(R)}{pm}+\sum_{w=1}^{p-1}\binom{p-1}{w}\binom{|R|+e(R)}{pm+wp^l}\equiv 0\mod p,$$

i.e.

$$1+\sum_{w=1}^{p-1}\binom{p-1}{w}\binom{1+\alpha_l(e(R))}{w}\equiv 0\mod p$$

for $0\le\alpha_l(e(R))<p-1$. The following equivalent holds

$$\sum_{w=0}^{p-1}\binom{p-1}{w}\binom{1+\alpha_l(e(R))}{w}\equiv\binom{p+\alpha_l(e(R))}{p-1}\mod p$$

by considering coefficient of x^{p-1} in the binomial expansion of

$$(x+1)^{p+\alpha_l(e(R))}=(x+1)^{p-1}(x+1)^{1+\alpha_l(e(R))}.$$

Since $0\le\alpha_l(e(R))<p-1$, $\binom{p+\alpha_l(e(R))}{p-1}$ is equivalent to zero in mod-p. Hence

$$\sum_{w=0}^{p-1}\binom{p-1}{w}\binom{1+\alpha_l(e(R))}{w}\equiv 0\mod p,$$

i.e.

$$1+\sum_{w=1}^{p-1}\binom{p-1}{w}\binom{1+\alpha_l(e(R))}{w}\equiv 0\mod p$$

So the result holds.

Case 2: $\alpha_l(e(R))=p-1$. Then

$$\sum_{w=1}^{p-1}\binom{p-1}{w}\binom{|R|+e(R)}{pm+wp^l}\equiv 0\mod p$$

and

$$\binom{|R|+e(R)}{pm}\equiv\binom{e(R)}{pm+(p-1)p^l}\mod p.$$

Therefore the result holds. $\square$

3. Conjugation in Dual of Steenrod algebra

Let $\mathcal{S}$ be the monoid of sequences of nonnegative integers almost all of which are zero; write $0_{\mathcal{S}}$ for the trivial element and $B(j)$ for the sequence whose i-th coordinate is δ_{ij}. If $S = (s_1, s_2, s_3, \dots)$ has $s_i = 0$ for $i > N$, we identify S with $S = (s_1, s_2, \dots, s_N)$. Suppose that $S = (s_1, s_2, \dots, s_N)$ with $s_N \neq 0$. The length of S is $len(S) = N$. We will write $\mathcal{S}^*$ for $\mathcal{S} \cup \{*\}$ where $*$ satisfies $x + * = *$ for all $x \in \mathcal{S}^*$.

Let $\mathcal{S}(k)$ be the symmetric group with identity Id^k acting on $\{0, 1, 2, 3, \dots, k-1\}$. For $\tau \in \mathcal{S}(k)$ and $i \geq 0$. Define

$$Z_i(k;\tau) = \begin{cases} \sum_{j=0}^{k-1} B(i + \tau(j) - j)p^j & i + \tau(j) - j \geq 0 \text{ for all } j \\ * & \text{else} \end{cases}$$

and

$$T_i(k;\tau) = \begin{cases} \prod_{j=0}^{k-1} \xi_{i+\tau(j)-j}^{p^j} & i + \tau(j) - j \geq 0 \text{ for all } j \\ 0 & \text{else.} \end{cases}$$

¿From these definitions we can observe that

(1) $Z_i(k; Id^k) = \frac{p^k-1}{p-1} B(i)$.

(2) $T_i(k; Id^k) = \xi_i^{\frac{p^k-1}{p-1}}$.

So define

$$\mathcal{W}_i(k) = \sum_{\tau \in \mathcal{S}(k)} T_i(k;\tau).$$

Then $\mathcal{W}_i(1) = T_i(1; Id^1) = \xi_i$. So we have following lemma;

Lemma 3.1. *For any integer $k \geq 1$, we have $\widehat{\xi}_k = \mathcal{W}_1(k)$.*

Proof:. We prove this by induction on k. Let $k = 1$. Then

$$\mathcal{W}_1(1) = \xi_1 = \widehat{\xi}_1.$$

Let $k = 2$.

$$\mathcal{W}_1(2) = T_1(2; Id^2) + T_1(2; (0,1)) = \xi_1^{p+1} + \xi_2^{p^0} \xi_0^{p^1} = \widehat{\xi}_2.$$

Suppose that the statement is true for $k' < k$. Note that if $T_1(k;\tau) \neq 0$, we must have $\tau(j) \geq j - 1$ for all j. So if $T_1(k;\tau) \neq 0$ and l satisfies $\tau(l) = k - 1$, we must have $\tau(k-1) = k-2$, $\tau(k-2) = k-3$, $\dots$, $\tau(l+1) = l$ so that τ has cycle-decomposition of the form $(k-1, k-2, \dots, l)\sigma$ for some $\sigma \in \mathcal{S}(l)$. In this case we have

$$T_1(k;\tau) = \xi_{k-1}^{p^l} T_1(k;\sigma).$$

So let $\mathcal{S}_l(k) = \{\tau \in \mathcal{S}(k) : \tau(l) = k-1\}$. Then

$$\mathcal{W}_1(k) = \sum_{l=0}^{k-1} \sum_{\tau \in \mathcal{S}_l(k)} T_1(k;\tau) = \sum_{l=0}^{k-1} \xi_{k-l}^{p^l} \sum_{\sigma \in \mathcal{S}(l)} T_1(k;\sigma)$$

$$\sum_{l=0}^{k-1} \xi_{k-l}^{p^l} \widehat{\xi}_l = \widehat{\xi}_k.$$

□

Definition 3.2. For $k \geq 1$, let $\mathcal{I}(k)$ be the set of nondecreasing sequences $(i_0, i_1, \ldots, i_{k-1})$ of positive integers. For $\tau \in \mathcal{S}(k)$ and $I \in \mathcal{I}(k)$, define

$$Z_I(k;\tau) = \begin{cases} \sum_{j=0}^{k-1} B(i_{\tau(j)} + \tau(j) - j)p^j & i_{\tau(j)} + \tau(j) - j \geq 0 \text{ for all } j \\ * & \text{else} \end{cases}$$

$$T_I(k;\tau) = \begin{cases} \prod_{j=0}^{k-1} \xi_{i_{\tau(j)}+\tau(j)-j}^{p^j} & i_{\tau(j)} + \tau(j) - j \geq 0 \text{ for all } j \\ 0 & \text{else.} \end{cases}$$

$$\mathcal{W}_I(k) = \sum_{\tau \in \mathcal{S}(k)} T_I(k;\tau).$$

We further define

$$P_I(k;\tau) = \begin{cases} \sum_{j=0}^{k-1} B(i_{\tau(j)} + \tau(j) - j - i_0)p^j & i_{\tau(j)} + \tau(j) - j - i_0 \geq 0 \text{ for all } j \\ * & \text{else} \end{cases}$$

$$R_I(k;\tau) = \begin{cases} \prod_{j=0}^{k-1} \xi_{i_{\tau(j)}+\tau(j)-j-i_0)}^{p^{i_0+j}} & i_{\tau(j)} + \tau(j) - j - i_0 \geq 0 \text{ for all } j \\ 0 & \text{else.} \end{cases}$$

$$\mathcal{R}_I(k) = \sum_{\tau \in \mathcal{S}(k)} R_I(k;\tau).$$

¿From this definition, we observe that

(1) If $I = (i, i, i, \ldots, i) \in I(k)$ is a constant sequence, then $Z_I(k;\tau) = Z_i(k;\tau)$. Moreover for $\tau \neq Id^k$, $P_I(k;\tau) = *$ and consequently $\mathcal{R}_I(k) = R_I(k; Id^k) = 1$.

(2) For general I, $Z_I(k; Id^k) = \sum_{j=0}^{k-1} B(i_{\tau(j)})p^j$.

(3) If $I = (i_0, i_1, \ldots, i_{k-1}) \in \mathcal{I}(k)$ and $i_0 > 1$, write $I[-1]$ for the sequence $(i_0 - 1, i_1 - 1, \ldots, i_{k-1} - 1)$. Then $\mathcal{R}_I(k) = [\mathcal{R}_{I[-1]}(k)]^p$.

4. PROOFS OF MAIN RESULTS

Proof of Theorem 1.1. We are going to prove the theorem by induction on i under the assumption that $i \leq j$. For $i = 0$ and all j , $\widehat{X}(j, 0) = \mathcal{P}(p^{j+1} - 1) = X(0, j)$ by Davis' formula in [1]. Assume that for all $\hat{i} \leq i - 1$ and all j, and for $\hat{i} = i$ and $\hat{j} \leq j - 1$,

$$\widehat{X}(\hat{i}, \hat{j}) = X(\hat{j}, \hat{i}).$$

The inductive proof will draw on the following remark: under the above assumptions,

$$\widehat{\mathcal{P}}(cp^{l-1}) \cdot X(l-1, i) = 0 \tag{4}$$

where c is a unit in mod-p and $1 \leq l \leq i$. Indeed by induction on l, we have

$$\begin{aligned}
\widehat{\mathcal{P}}(cp^{l-1}) \cdot X(l-1,i) &= \widehat{[\widehat{X}(l-1,i) \cdot \mathcal{P}(cp^{l-1})]} \\
&= \widehat{[X(i,l-1) \cdot \mathcal{P}(cp^{l-1})]} \\
&= \widehat{[\mathcal{P}(p^i(p^l-1)) \cdots \mathcal{P}(p(p^l-1))\mathcal{P}((p^l-1)) \cdot \mathcal{P}(cp^{l-1})]}.
\end{aligned}$$

By Adem relations, $\mathcal{P}(p^l-1) \cdot \mathcal{P}(cp^{l-1}) = 0$. Therefore this verifies equation (4).

We claim that for $0 \leq l \leq j$

(5)
$$X(l,i-1) \cdot \widehat{\mathcal{P}}(p^i(p^{j+1}-1)) = \widehat{\mathcal{P}}(p^i(p^{j+1}-p^{l+1})) \cdot X(l,i)$$

The case $l = 0$ follows from Proposition 2.3 (i). Suppose that the statement is true for $l-1$. Then by induction and Proposition 2.3 (ii), we have

$$\begin{aligned}
X(l,&i-1) \cdot \widehat{\mathcal{P}}(p^i(p^{j+1}-1)) = \mathcal{P}(p^l(p^i-1)) \cdot \widehat{\mathcal{P}}(p^i(p^{j+1}-p^l)) \cdot X(l-1,i) \\
&= [\widehat{\mathcal{P}}(p^i(p^{j+1}-p^{l+1})) \cdot \mathcal{P}(p^l(p^{i+1}-1))] \cdot X(l-1,i) \\
&- [\sum_{w=1}^{p-1} \binom{p-1}{w} \mathcal{P}(p^l(p^i-1)+wp^{l-1}) \cdot \widehat{\mathcal{P}}(p^{i+j+1}-p^{l+i}-wp^{l-1})] \cdot X(l-1,i)
\end{aligned}$$

From Equation (4),

$$\widehat{\mathcal{P}}(p^{i+j+1}-p^{l+i}-wp^{l-1}) \cdot X(l-1,i) = 0$$

for every $w = 1, 2, 3, \dots, p-1$.

Hence we have

$$X(l,i-1) \cdot \widehat{\mathcal{P}}(p^i(p^{j+1}-1)) = \widehat{\mathcal{P}}(p^i(p^{j+1}-p^{l+1})) \cdot X(l,i).$$

This proves our claim. Finally, taking $l = j$, we find that

$$\begin{aligned}
\widehat{X}(i,j) &= \widehat{X}(i-1,j)) \cdot \widehat{\mathcal{P}}(p^i(p^{j+1}-1)) \\
&= X(j,i-1) \cdot \widehat{\mathcal{P}}(p^i(p^{j+1}-1)) \\
&= \widehat{\mathcal{P}}(0) \cdot X(j,i).
\end{aligned}$$

This completes the proof. $\square$

Proof of Theorem 1.2. Fix k. We prove by induction on i_0. Let $i_0 = 1$.

(6)
$$\xi_k \widehat{\mathcal{R}}_I(k) = \sum_{\tau} \xi_k \widehat{\xi}^{p^k}_{i_{\tau(k-1)}+\tau(k-1)-k} \prod_{j \neq k-1} \widehat{\xi}^{p^{j+1}}_{i_{\tau(j)}+\tau(j)-(j+1)}.$$

On the other hand,

$$\widehat{\mathcal{W}}_I(k) = \sum_{\rho} \widehat{\xi}^{p^0}_{i_{\rho(0)}+\rho(0)} \prod_{j \neq 0} \widehat{\xi}^{p^j}_{i_{\rho(j)}+\rho(j)-j}.$$

For each $\rho \in \mathcal{S}(k)$, define ρ' by

$$(6) \qquad \rho'(l) = \begin{cases} \rho(0) & l = k-1 \\ \rho(l+1) & 0 \le l \le k-2. \end{cases}$$

Then

$$(8) \qquad \widehat{\mathcal{W}}_I(k) = \sum_{\rho'} \widehat{\xi}_{i_{\rho'(k-1)}+\rho'(k-1)} \prod_{l \neq k-1} \widehat{\xi}^{p^{l+1}}_{i_{\rho'(l)}+\rho'(l)-(l+1)}.$$

By Milnor formula,

$$(9) \qquad \widehat{\xi}_{i_{\rho'(k-1)}+\rho'(k-1)} \equiv \sum_{n=1}^{k} \xi_n \widehat{\xi}^{p^n}_{i_{\rho'(k-1)}+\rho'(k-1)-n}$$

modulo terms of length $> k$. So we have

$$(10) \qquad \widehat{\mathcal{W}}_I(k) \equiv \sum_{\rho'} (\sum_{n=1}^{k} \xi_n \widehat{\xi}^{p^n}_{i_{\rho'(k-1)}+\rho'(k-1)-n}) \prod_{l \neq k-1} \widehat{\xi}^{p^{l+1}}_{i_{\rho'(l)}+\rho'(l)-(l+1)}.$$

Claim that (6) and (10) agree. Indeed their difference is

$$(11) \qquad \begin{aligned} &\sum_{\rho'} (\sum_{n=1}^{k-1} \xi_n \widehat{\xi}^{p^n}_{i_{\rho'(k-1)}+\rho'(k-1)-n}) \prod_{l \neq k-1} \widehat{\xi}^{p^{l+1}}_{i_{\rho'(l)}+\rho'(l)-(l+1))} \\ &= \sum_{n=1}^{k-1} \sum_{\rho'} \xi_n \widehat{\xi}^{p^n}_{i_{\rho'(k-1)}+\rho'(k-1)-n} \widehat{\xi}^{p^{(n-1)+1}}_{i_{\rho'(n-1)}+\rho'(n-1)-n} \prod_{l \neq n-1, k-1} \widehat{\xi}^{p^{l+1}}_{i_{\rho'(l)}+\rho'(l)-(l+1)}. \end{aligned}$$

It can be easily verified that summand in (11) associated to n and ρ' equals to the term associated to n and ρ'' where

$$\rho''(l) = \begin{cases} \rho'(l) & l \neq n-1, k-1 \\ \rho'(n-1) & l = k-1 \\ \rho(k-1) & l = n-1. \end{cases}$$

Each term in (11) appears p-times. So sums in (6) and (10) are equal. Hence $\xi_k \widehat{\mathcal{R}}_I(k)$ and $\widehat{\mathcal{W}}_I(k)$ are congruent modulo elements of length $> k$. So theorem holds for $i_0 = 1$.

The proof for general I is similar. Suppose that the statement is true for non-decreasing sequences $(i_0 - 1, i_1 - 1, \dots i_k - 1) = I[-1]$. Let $i = i_0$. From Lemma 3.1 and induction hypothesis, we have

$$\begin{aligned} \xi_k^{\frac{p^i-1}{p-1}} \widehat{\mathcal{R}}_I(k) &= [\xi_k^{\frac{p^{i-1}-1}{p-1}} \widehat{\mathcal{R}}_{I[-1]}(k)]^p \xi_k \\ &\equiv (\widehat{\mathcal{W}}_{I[-1]}(k))^p \xi_k \\ &= \xi_k \sum_{\tau} \prod_{j=0}^{k-1} \widehat{\xi}^{p^{j+1}}_{(i_{\tau(j)}-1)+\tau(j)-j} \\ &= \sum_{\tau} \xi_k \widehat{\xi}^{p^k}_{(i_{\tau(k-1)}-1)+\tau(k-1)-(k-1)} \prod_{j \neq k-1} \widehat{\xi}^{p^{j+1}}_{(i_{\tau(j)}-1)+\tau(j)-j} \\ (12) \qquad &= \sum_{\tau} \xi_k \widehat{\xi}^{p^k}_{i_{\tau(k-1)}+\tau(k-1)-k} \prod_{j \neq k-1} \widehat{\xi}^{p^{j+1}}_{i_{\tau(j)}+\tau(j)-(j+1)}. \end{aligned}$$

On the other hand, we have

$$\begin{aligned}\widehat{\mathcal{W}}_I(k) &= \sum_{\rho} \prod_{j=0}^{k-1} \widehat{\xi}^{p^j}_{i_{\rho(j)}+\rho(j)-j} \\ &= \sum_{\rho} \widehat{\xi}^{p^0}_{i_{\rho(0)}+\rho(0)-0} \prod_{j\neq 0} \widehat{\xi}^{p^j}_{i_{\rho(j)}+\rho(j)-j}.\end{aligned} \tag{13}$$

So one can define ρ' as in (7) and use (9) as in the case $i_0 = 1$ to establish the congruence of the sums in (12) and (13) modulo elements of length $> k$. This establishes the inductive step and hence completes the proof. $\square$

Define

$$\widehat{\mathcal{W}}'_I(k) = \sum_{Id^k \neq \tau \in \mathcal{S}(k)} \widehat{T}_i(k;\tau).$$

Then we have following corollary;

Corollary 4.1. *For any integers $k \geq 1$ and $i_0 > 0$, we have*

$$\widehat{T}_I(k; Id^k) \equiv \widehat{\mathcal{W}}'_I(k) + \xi_k^{\frac{p^{i_0}-1}{p-1}} \widehat{\mathcal{R}}_I(k)$$

modulo elements of length $> k$.

REFERENCES

[**1**] D.M. Davis, *The Antiautomorphism of the Steenrod Algebra*, Proc. Amer. Math. Soc. **44** (1974), 235–236.

[**2**] A.M. Gallant, *Excess and Conjugation in the Steenrod Algebra*, Proc. Amer. Math. Soc. **76** (1979), 161–166.

[**3**] I. Karaca, *On the action of Steenrod algebra on Polynomial algebra*, Turkish Journal of Math. **22** (1998), 163–170.

[**4**] I. Karaca, *The nilpotence height of P^s_t for odd primes*, Trans. Amer. Math. Soc. **351** (1999), 547–558.

[**5**] J. Milnor, *The Steenrod Algebra and Its Dual*, Ann. of Math. **67** (1958), 150–171.

[**6**] J.H. Silverman, *Conjugation and Excess in the Steenrod Algebra*, Proc. Amer. Math. Soc. **119** (1993), 657–661.

[**7**] J.H. Silverman, *Hit polynomials and conjugation in the dual Steenrod Algebra*, Math. Proc. Camb. Philos. Soc. **124** (1998), 531–547.

[**8**] R. Thom, *Quelques proprietes globales des varietes differentiables*, Comment. Math. Helv. **28** (1954), 17–86.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2001, pp. 53–60.
Submitted: April, 9, 2001. Published: December 31 2001.

DYNAMIC ANALYSIS OF OLIGOPSONY UNDER ADAPTIVE EXPECTATIONS

Ferenc Szidarovszky[1] and Koji Okuguchi[2]

Abstract. This paper analyzes the global stability of equilibrium for oligopsony facing perfectly competitive product price under adaptive expectations, assuming each firm's production function to be quadratic and the factor price function to be linear. The general stability conditions are simplified for some special cases, and a formal similarity is noted between the stability for oligopoly and that for oligopsony.

AMS Subject Classification Codes. 90

Key words. Oligopsony, oligopoly, adaptive expectations, stability

1. Introduction

In spite of many papers on the exitence and stability of Cournot oligopoly and other oligopolies under various forms of firms' expectations, few economists have analyzed the existence and stability of equilibrium in oligopsony and oligopoly with oligopsony (oligopoly–oligopsony). Okuguchi (1998) and Szidarovszky and Okuguchi (1997) have examined the existence of a unique equilibrium for Cournot oligopoly–oligopsony. Chiarella and Okuguchi (1997) have analyzed the existence and stability of Cournot duopoly–duopsony equilibrium under the assumption of two identical firms, one factor of production and a linear inverse demand function for the product. Earlier, Okuguchi (1970) had analyzed the stability of the equilibrium of oligopoly with adaptive expectations and without product differentiation, while Okuguchi and Szidarovszky (1999) have investigated the stability of the equilibrium of oligopoly with multi-product firms with adaptive expectations, and given a comprehensive summary of the earlier works on oligopoly–oligopsony.

[1]Systems and Industrial Engineering Department, University of Arizona, Tucson, AZ 85721-0020, U.S.A.

[2]Department of Economics and Information, Gifu-Shotokugakuen University, Nakauzura, Gifu-ken 500-8288, Japan.

In this paper we will introduce adaptive expectations into oligopoly facing perfect competition in product market and with several factors of production, and analyze the global stability of the oligopsony equilibrium. We will assume quadratic production functions and linear factor price functions to make our analysis technically tractable. We will derive a general stability condition, which is then interpreted for some special cases including Cournot oligopsony. We will find a remarkable similarity between the stability condition for Cournot oligopoly and that for Cournot oligopsony.

2. Model

Similarly to the model examined in Okuguchi (1998), consider an n-firm oligopsony where only a single product is produced using m factors of production. We assume perfectly competitive product market and an oligopsonistic factor market. The profit of firm i is given as

$$\Pi_i = pf_i(\mathbf{l}_i) - \mathbf{l}_i^T w(\mathbf{L}) \quad (i = 1, 2, \ldots, n) \tag{1}$$

where p is the competitive product price, f_i is the production of firm i, $\mathbf{l}_i$ is its m-element production factor, w is the m-element factor price vector and $\mathbf{L}$ is a m-element vector of total production factor amounts used by the industry: $\mathbf{L} = \sum_{i=1}^{n} \mathbf{l}_i$. For the sake of simplicity assume that

$$\text{(A)} \qquad f_i(\mathbf{l}_i) = \mathbf{l}_i^T \mathbf{C}_i \mathbf{l}_i + \mathbf{c}_i^T \mathbf{l}_i + \gamma_i \text{ and } w(\mathbf{L}) = \mathbf{A}\mathbf{L} + \mathbf{a}, \tag{2}$$

where $\mathbf{c}_i$ and $\mathbf{a}$ are m-element vectors, $\mathbf{C}_i$ and $\mathbf{A}$ are $m \times m$ matrices, and γ_i is a scalar.

Assume that at the initial time period $t = 0$, each firm selects an initial $\mathbf{l}_i(0)$, and at each later time period $t \geq 1$, each firm first predicts the production factor usage of the rest of the industry adaptively:

$$\mathbf{s}_i^E(t) = \mathbf{M}_i \sum_{j \neq i} \mathbf{l}_j(t-1) + (\mathbf{I} - \mathbf{M}_i)\mathbf{s}_i^E(t-1), \tag{3}$$

where $\mathbf{M}_i$ is an $m \times m$ given constant matrix, usually $\alpha_i \mathbf{I}$ with some $\alpha_i \in (0, 1]$. Then the expected profit of firm i at time period t can be expressed as

$$\Pi_i(t) = p\big(\mathbf{l}_i^T \mathbf{C}_i \mathbf{l}_i + \mathbf{c}_i^T \mathbf{l}_i + \gamma_i\big) - \mathbf{l}_i^T \big(\mathbf{A}\mathbf{l}_i + \mathbf{A}\mathbf{s}_i^E(t) + \mathbf{a}\big). \tag{4}$$

Excluding corner optimum, the expected profit maximizing production factors of firm i at time period t are the solutions of equation

$$p\big((\mathbf{C}_i + \mathbf{C}_i^T)\mathbf{l}_i + \mathbf{c}_i\big) - (\mathbf{A} + \mathbf{A}^T)\mathbf{l}_i - \mathbf{A}\mathbf{s}_i^T(t) - \mathbf{a} = \mathbf{0}. \tag{5}$$

Here we assume that the second-order optimality condition holds, that is,

(B) matrix $p(\mathbf{C}_i + \mathbf{C}_i^T) - (\mathbf{A} + \mathbf{A}^T)$ is negative definite for all i.

Combining equations (3) and (5), a discrete dynamic system is obtained with state variables $\mathbf{l}_i(t)$ and $\mathbf{s}_i^E(t)$ $(i = 1, 2, \dots, n)$. The state transition equation for variables $\mathbf{l}_i$ is given as follows:

(6)

$$\mathbf{l}_i(t) = -\big(\mathbf{A} + \mathbf{A}^T - p(\mathbf{C}_i + \mathbf{C}_i^T)\big)^{-1}\mathbf{A}\left(\mathbf{M}_i \sum_{j \neq i} \mathbf{l}_j(t-1) + (\mathbf{I} - \mathbf{M}_i)\mathbf{s}_i^E(t-1)\right)$$
$$+ \ \big(\mathbf{A} + \mathbf{A}^T - p(\mathbf{C}_i + \mathbf{C}_i^T)\big)^{-1}(p\mathbf{c}_i - \mathbf{a}).$$

The state transition equations for variables $\mathbf{s}_i^E$ are given by (3). The asymptotical stability of this system will be analyzed in the next section of the paper.

3. Stability Analysis

Introduce the notation

$$\mathbf{B} = \big(\mathbf{A} + \mathbf{A}^T - p(\mathbf{C}_i + \mathbf{C}_i^T)\big)^{-1}\mathbf{A}, \tag{7}$$

then the coefficient matrix of the system can be written as

$$\mathbf{H} = \tag{8}$$

$$\left\{\begin{array}{cccccccc}
\mathbf{O} & -\mathbf{BM}_1 & \cdots & -\mathbf{BM}_1 & -\mathbf{B}(\mathbf{I}-\mathbf{M}_1) & & & \\
-\mathbf{BM}_2 & \mathbf{O} & \cdots & -\mathbf{BM}_2 & & -\mathbf{B}(\mathbf{I}-\mathbf{M}_2) & & \\
\vdots & \vdots & \vdots & & & & \ddots & \\
-\mathbf{BM}_n & -\mathbf{BM}_n & \cdots & \mathbf{O} & & & & -\mathbf{B}(\mathbf{I}-\mathbf{M}_n) \\
\mathbf{O} & \mathbf{M}_1 & \cdots & \mathbf{M}_1 & \mathbf{I}-\mathbf{M}_1 & & & \\
\mathbf{M}_2 & \mathbf{O} & \cdots & \mathbf{M}_2 & & \mathbf{I}-\mathbf{M}_2 & & \\
\vdots & \vdots & \vdots & & & & \ddots & \\
\mathbf{M}_n & \mathbf{M}_n & \cdots & \mathbf{O} & & & & \mathbf{I}-\mathbf{M}_n
\end{array}\right\}.$$

It is well known (see, for example, Szidarovszky and Bahill, 1998) that the equilibrium of the system is asymptotically stable if and only if all eigenvalues of matrix $\mathbf{H}$ are inside the unit circle. The eigenvalue equation can be written as

$$\left.\begin{array}{r} -\mathbf{BM}_i \sum_{j \neq i} \mathbf{u}_j - \mathbf{B}(\mathbf{I} - \mathbf{M}_i)\mathbf{v}_i = \lambda \mathbf{u}_i \\ \mathbf{M}_i \sum_{j \neq i} \mathbf{u}_j - (\mathbf{I} - \mathbf{M}_i)\mathbf{v}_i = \lambda \mathbf{v}_i \end{array}\right\} \quad (i = 1, 2, \dots, n). \tag{9}$$

Adding the $\mathbf{B}$-multiple of the second equation to the first equation yields a relation

$$\lambda(\mathbf{u}_i + \mathbf{B}\mathbf{v}_i) = \mathbf{0}. \tag{10}$$

Since zero eigenvalues do not disturb asymptotical stability, we may assume that $\lambda \neq 0$. Then

$$\mathbf{u}_i = -\mathbf{B}\mathbf{v}_i, \tag{11}$$

and by substituting this relation into the second equation of (9) we have

$$-\mathbf{M}_i\mathbf{B}\sum_{j\neq i}\mathbf{v}_j + (\mathbf{I}-\mathbf{M}_i)\mathbf{v}_i = \lambda\mathbf{v}_i \quad (i=1,2,\dots,n). \tag{12}$$

That is, the nonzero eigenvalues of $\mathbf{H}$ coincide with the eigenvalues of matrix

$$\mathbf{H}^{(1)} = \begin{pmatrix} \mathbf{I}-\mathbf{M}_1 & -\mathbf{M}_1\mathbf{B} & \cdots & -\mathbf{M}_1\mathbf{B} \\ -\mathbf{M}_2\mathbf{B} & \mathbf{I}-\mathbf{M}_2 & \cdots & -\mathbf{M}_2\mathbf{B} \\ \vdots & \vdots & & \vdots \\ -\mathbf{M}_n\mathbf{B} & -\mathbf{M}_n\mathbf{B} & \cdots & \mathbf{I}-\mathbf{M}_n \end{pmatrix}. \tag{13}$$

Hence the equilibrium is asymptotically stable if and only if all eigenvalues of $\mathbf{H}^{(1)}$ are inside the unit circle.

We will next further reduce the size of the eigenvalue problem that has to be examined in order to establish the asymptotical stability of the equilibrium. With the new variable $\mathbf{v} = \sum_{j=1}^{n}\mathbf{v}_j$, we see that (12) is rewritten as

$$-\mathbf{M}_i\mathbf{B}\mathbf{v} + (\mathbf{I}-\mathbf{M}_i+\mathbf{M}_i\mathbf{B}-\lambda\mathbf{I})\mathbf{v}_i = 0. \tag{14}$$

Assuming that matrix $\mathbf{I}-\mathbf{M}_i+\mathbf{M}_i\mathbf{B}$ is non-singular,

$$\mathbf{v}_i = (\mathbf{I}-\mathbf{M}_i+\mathbf{M}_i\mathbf{B}-\lambda\mathbf{I})^{-1}\mathbf{M}_i\mathbf{B}\mathbf{v}, \tag{15}$$

and adding these equations for $i=1,2,\dots,n$, we have

$$\left(\sum_{i=1}^{n}(\mathbf{I}-\mathbf{M}_i+\mathbf{M}_i\mathbf{B}-\lambda\mathbf{I})^{-1}\mathbf{M}_i\mathbf{B}-\mathbf{I}\right)\mathbf{v} = \mathbf{0}. \tag{16}$$

Notice that $\mathbf{v}=\mathbf{0}$ implies that for all i, $\mathbf{v}_i=\mathbf{0}$ which is impossible, since eigenvectors must differ from zero. Therefore equation (16) has nonzero solutions if and only if

$$\det\left(\sum_{i=1}^{n}(\mathbf{I}-\mathbf{M}_i+\mathbf{M}_i\mathbf{B}-\lambda\mathbf{I})^{-1}\mathbf{M}_i\mathbf{B}-\mathbf{I}\right) = 0. \tag{17}$$

Hence, we have proved that under the assumption that matrices $\mathbf{I}-\mathbf{M}_i+\mathbf{M}_i\mathbf{B}$ are non-singular, the equilibrium is asymptotically stable if and only if all roots of equation (17) are inside the unit circle.

Simple sufficient conditions can be given for the asymptotical stability of the equilibrium by requiring that some norm of matrix $\mathbf{H}^{(1)}$ is less than one. Selecting the block-row norm, the condition is the following:

$$\|\mathbf{I}-\mathbf{M}_i\| + \sum_{j\neq i}\|\mathbf{M}_i\mathbf{B}\| < 1 \quad (i=1,2,\dots,n) \tag{18}$$

which holds if

$$\|\mathbf{I}-\mathbf{M}_i\| + (n-1)\|\mathbf{M}_i\|\cdot\|\mathbf{B}\| < 1 \tag{19}$$

which is equivalent to the assumption that

$$\|\mathbf{B}\| < \min_i \left\{ \frac{1 - \|\mathbf{I} - \mathbf{M}_i\|}{(n-1)\|\mathbf{M}_i\|} \right\}. \tag{20}$$

Selecting the block-column norm, the condition is the following:

$$\|\mathbf{I} - \mathbf{M}_i\| + \sum_{j \neq i} \|\mathbf{M}_j \mathbf{B}\| < 1 \quad (i = 1, 2, \dots, n) \tag{21}$$

which holds if

$$\|\mathbf{I} - \mathbf{M}_i\| + \|\mathbf{B}\| \cdot \sum_{j \neq i} \|\mathbf{M}_j\| < 1 \tag{22}$$

which is equivalent to the condition

$$\|\mathbf{B}\| < \min_i \left\{ \frac{1 - \|\mathbf{I} - \mathbf{M}_i\|}{\sum\limits_{j \neq i} \|\mathbf{M}_j\|} \right\}. \tag{23}$$

4. Special Cases

Assume first that $\mathbf{M}_i \equiv \mathbf{M}$, that is, the firms have identical expectations. In this case, matrix $\mathbf{H}^{(1)}$ can be rewritten as

$$\mathbf{H}^{(1)}_{spec1} = \begin{pmatrix} \mathbf{I} - \mathbf{M} & -\mathbf{MB} & \cdots & -\mathbf{MB} \\ -\mathbf{MB} & \mathbf{I} - \mathbf{M} & \cdots & -\mathbf{MB} \\ \vdots & \vdots & & \vdots \\ -\mathbf{MB} & -\mathbf{MB} & \cdots & \mathbf{I} - \mathbf{M} \end{pmatrix}, \tag{24}$$

and equation (17) simplifies to the following:

$$\det\bigl(n(\mathbf{I} - \mathbf{M} + \mathbf{MB} - \lambda \mathbf{I})^{-1}\mathbf{MB} - \mathbf{I}\bigr) = 0. \tag{25}$$

Notice that inequalities (20) and (23) reduce to the following:

$$\|\mathbf{B}\| < \frac{1 - \|\mathbf{I} - \mathbf{M}\|}{(n-1)\|\mathbf{M}\|}. \tag{26}$$

In the further special case of Cournot expectations, $\mathbf{M} = \mathbf{I}$. Then $\mathbf{H}^{(1)}$ becomes

$$\mathbf{H}^{(1)}_{spec2} = \begin{pmatrix} \mathbf{O} & -\mathbf{B} & \cdots & -\mathbf{B} \\ -\mathbf{B} & \mathbf{O} & \cdots & -\mathbf{B} \\ \vdots & \vdots & & \vdots \\ -\mathbf{B} & -\mathbf{B} & \cdots & \mathbf{O} \end{pmatrix}, \tag{27}$$

which is the Kronecker product of matrices

$$\mathbf{E} = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix} \tag{28}$$

and $-\mathbf{B}$. It is known (see, for example, Okuguchi and Szidarovszky, 1999) that the eigenvalues of $\mathbf{E}$ are -1 and $(n-1)$, therefore the eigenvalues of $\mathbf{H}^{(1)}_{spec2}$ are λ_i and $-(n-1)\lambda_i$ where $\lambda_i (1 \le i \le m)$ are the eigenvalues of $\mathbf{B}$. Hence the equilibrium is asymptotically stable if and only if for $i = 1, 2, \ldots, m$,

$$|\lambda_i| < \frac{1}{n-1}\,. \tag{29}$$

Assume next that for all i, $\mathbf{C}_i = \mathbf{O}$ and $\mathbf{A}$ is an invertible symmetric matrix. Then $\mathbf{B} = \frac{1}{2}\mathbf{I}$, and therefore $\mathbf{H}^{(1)}$ has the special form

$$\mathbf{H}^{(1)}_{spec3} = \begin{pmatrix} \mathbf{I}-\mathbf{M}_1 & -\frac{1}{2}\mathbf{M}_1 & \cdots & -\frac{1}{2}\mathbf{M}_1 \\ -\frac{1}{2}\mathbf{M}_2 & \mathbf{I}-\mathbf{M}_2 & \cdots & -\frac{1}{2}\mathbf{M}_2 \\ \vdots & \vdots & & \vdots \\ -\frac{1}{2}\mathbf{M}_n & -\frac{1}{2}\mathbf{M}_n & \cdots & \mathbf{I}-\mathbf{M}_n \end{pmatrix}. \tag{30}$$

In order to apply equation (17), we have to assume that the matrix

$$\mathbf{I} - \mathbf{M}_i + \mathbf{M}_i\mathbf{B} = \mathbf{I} - \frac{1}{2}\mathbf{M}_i \tag{31}$$

is non-singular. In this case, equation (17) specializes as

$$\det\left(\sum_{i=1}^{n} \frac{1}{2}\left(\mathbf{I} - \frac{1}{2}\mathbf{M}_i - \lambda\mathbf{I}\right)^{-1}\mathbf{M}_i - \mathbf{I}\right) = 0. \tag{32}$$

In addition to the above conditions assume that for all i, $\mathbf{M}_i = \alpha_i\mathbf{I}$ with some $\alpha_i \in (0, 1]$. Then matrix (30) can be further reduced as follows:

$$\mathbf{H}^{(1)}_{spec4} = \begin{pmatrix} \mathbf{I}(1-\alpha_1) & -\frac{\alpha_1}{2}\mathbf{I} & \cdots & -\frac{\alpha_1}{2}\mathbf{I} \\ -\frac{\alpha_2}{2}\mathbf{I} & \mathbf{I}(1-\alpha_2) & \cdots & -\frac{\alpha_2}{2}\mathbf{I} \\ \vdots & \vdots & & \vdots \\ -\frac{\alpha_n}{2}\mathbf{I} & -\frac{\alpha_n}{2}\mathbf{I} & \cdots & \mathbf{I}(1-\alpha_n) \end{pmatrix}. \tag{33}$$

Since matrix (31) is $\mathbf{I}\left(1 - \frac{\alpha_i}{2}\right)$ and non-singular, we can use equation (32), which can be significantly simplied:

$$\sum_{i=1}^{n} \frac{\frac{\alpha_i}{2}}{\left(1 - \frac{\alpha_i}{2}\right) - \lambda} = 1. \tag{34}$$

We will next find sufficient and necessary conditions that the roots of this equation are inside the unit circle. Let $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_r \le 1$ denote the distinct α_i values with multipliticies $m_1, m_2, \ldots, m_r$. Then (34) can be rewritten as

$$g(\lambda) = \sum_{i=1}^{r} \frac{\frac{m_i\alpha_i}{2}}{\left(1 - \frac{\alpha_i}{2}\right) - \lambda} = 1. \tag{35}$$

Notice first that this is equivalent to a polynomial equation of degree r, therefore there are r real or complex roots. Furthermore, for all i,

$$\lim_{\lambda \to 1-\frac{\alpha_i}{2}+0} g(\lambda) = -\infty, \tag{36}$$

$$\lim_{\lambda \to 1-\frac{\alpha_1}{2}-0} g(\lambda) = \infty, \tag{37}$$

and

$$\lim_{\lambda \to \pm\infty} g(\lambda) = 0. \tag{38}$$

Since $g'(\lambda) > 0$, function g is locally strictly increasing everywhere. There are $r-1$ real roots between $1-\frac{\alpha_{i+1}}{2}$ and $1-\frac{\alpha_i}{2}$ $(i = 1, 2, \ldots, r-1)$, and there is an additional real root below $1-\frac{\alpha_r}{2}$. Hence all roots are real, and they are between -1 and $+1$ if and only if $g(-1) < 1$. This condition can be rewritten as

$$\sum_{i=1}^{n} \frac{\alpha_i}{4-\alpha_i} < 1. \tag{39}$$

Assume first that $n = 2$. Then

$$\sum_{i=1}^{2} \frac{\alpha_i}{4-\alpha_i} \leq 2 \cdot \frac{1}{4-1} < 1, \tag{40}$$

therefore dupolies are always asymptotically stable under the above simplifying assumptions. If $n \geq 3$, then (39) is not satisfied necessarily. It holds if the α_i values are sufficiently small. For example, if $\alpha_i \equiv \alpha$, then (39) holds if and only if

$$\frac{n\alpha}{4-\alpha} < 1, \tag{41}$$

that is, when

$$\alpha < \frac{4}{n+1}. \tag{42}$$

Notice that this condition is equivalent to the asymptotical stability of dynamic oligopolies under similar simplifying conditions (see, Okuguchi and Szidarovszky, 1999, Theorem 4.2.4).

Assume finally that the factor market is also perfectly competitive. Then matrix $\mathbf{A}$ is zero, and therefore at each time period the profit (4) of firm i $(i = 1, 2, \ldots, n)$ does not depend on the activity of the competitors:

$$\Pi_i(t) = p\big(\mathbf{l}_i^T \mathbf{C}_i \mathbf{l}_i + \mathbf{c}_i^T \mathbf{l}_i + \gamma_i\big) - \mathbf{l}_i^T \mathbf{a}. \tag{43}$$

Therefore at each time period the firms maximize their own payoffs independently of each other:

$$\mathbf{l}_i(t) = \big[p(\mathbf{C}_i + \mathbf{C}_i^T)\big]^{-1} (\mathbf{a} - p\mathbf{c}_i). \tag{44}$$

5. Conclusions

Special dynamic oligopsonies were examined with linear price and quadratic production functions. We proved that the equilibrium is asymptotically stable if and only if all eigenvalues of matrix (8) or (13) are inside the unit circle. We showed that under special conditions it is sufficient to examine the more simple determinant equation (17). Sufficient conditions were also derived (relations (20) and (23)) which are very easy to check in practical cases. Some special cases were discussed such as identical expectations, and in further special cases Cournot expectations, linear production functions, and special speeds of adjustment were analyzed. The general stability conditions were simplified in these special cases and the analogy between these conditions and those for oligopolies known from the literature were observed.

References

[1] Chiarella, C. and Okuguchi, K. (1997). A dynamic analysis of Cournot duopoly in imperfectly competitive product and factor markets, *Keio Economic Studies*, **34**, 21–33.

[2] Okuguchi, K. (1970). Adaptive expectations in an oligopoly model, *Review of Economic Studies*, **36**, 233–237.

[3] Okuguchi, K. (1998), Existence of equilibrium for Cournot oligopoly-oligopsony, *Keio Economic Studies*, **35**, 45–53.

[4] Okuguchu, K. and Szidarovszky, F. (1999). *The Theory of Oligopoly with Multi-Product Firms* (2nd edition), Springer-Verlag, Berlin/Heidelberg/New York.

[5] Szidarovszky, F. and Bahill, A.T. (1998). *Linear Systems Theory* (2nd edition), CRC Press, Boca Raton/London.

[6] Szidarovszky, F. and Okuguchi, K. (1997). Oligopoly-oligopsony equilibrium as a solution of a complementarity problem, *Mimeo*, University of Arizona.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 2 December 2001, pp. 61 – 72
Submitted: April 6, 2001. Published: December 31, 2001

Geometry and Convergence of some third-order methods

Sergio Amat and Sonia Busquier

Abstract

We study the global convergence of two third-order methods by using their geometric interpretation. A comparison of these methods with the classic ones is discussed also.

A.M.S. Subject Classification Codes 65F15, 65H05.
Key Words and Phrases. Non-linear equations, third order iterative methods, global convergence

1 Introduction

We consider the basic problem, finding the zeros of a given function f, that is, finding the values α for which $f(\alpha) = 0$. Determining the zeros of a function has attracted the attention of pure and applied mathematicians for centuries. Solving equations numerically is a classical problem. General problems may be formulated in terms of finding zeros. The roots of a nonlinear equation $f(x) = 0$ cannot in general be expressed in closed form. Thus, in order to solve nonlinear equations, we have to use approximate methods.

Departamento de Matemática Aplicada y Estadística. Universidad Politécnica de Cartagena, Spain
Research supported in part by DGICYT PB97-1402, and in part by a Grant FPI from Generalitat Valenciana,FPI-97-CN-10-0204-G.
Departament de Matemàtica Aplicada. Universitat de València, Spain
E-mail : sergio.amat@upct.es, sonia.busquier@uv.es

One of the most important techniques to study these equations is the use of iterative processes, starting from an initial approximation x_0, called pivot, successive approaches (until some predetermined convergence criterion is satisfied) x_i are computed , $i = 1, 2, \ldots$, with the help of certain iteration function $\Phi : E \to E$:

$$x_{i+1} := \Phi(x_i), \quad i = 0, 1, 2 \ldots \tag{1}$$

One of the most studied methods is Newton method:

$$x_{k+1} = \Phi(x_k) := x_k - \frac{f(x_k)}{f'(x_k)}. \tag{2}$$

The advance of computational techniques has allowed the development of some more complicated iterative methods in order to obtain greater order of convergence. When the evaluation of first and second derivatives is not very complicated, third order methods may be useful.

Definition 1 *A method with iteration function Φ is third order if $\Phi(\alpha) = \alpha$, $\Phi'(\alpha) = \Phi''(\alpha) = 0$ and $\Phi'''(\alpha) \neq 0$.*

On the other hand, unconstrained optimization deals with the problem of minimizing or maximizing a function in absence of any restrictions. Even though most practical optimization problems have side restrictions that must be satisfied, the study of techniques for unconstrained optimization is important for several reasons. In particular, many algorithms solve a constrained problem by converting it into a sequence of unconstrained problems see [3]. Usually Newton method is used, see [17]. But in some cases, the functions do not verify the hypothesis of the theorems, and the convergence is not clear. In order to model this type of problems we consider the simple equation $f(x) = \sqrt{|x|} = 0$. In this case $f(x)$ is not derivable and we can not apply the theory. Moreover, the function does not change its sign and we cannot apply typical first order methods. If we consider Newton type methods we will not have convergence, since in this case $x_{k+1} = -x_k, \forall k$. Thus, we consider third order methods. In table 4 and figure 4 (see section 4) we analize third order methods. We obtain linear converge. Let us remark that Chebyshev method (third order method) is exact for every initial data, but in general an order reduction takes place when we consider these type of equations.

In this paper, we present two third-order methods. As we shall see later on, these methods are based on an interpolation by an exponential and a logarithmic function, respectively. We derive sufficient conditions for their convergence. Finally, we compare these methods with the classical ones, both from an analytical and a numerical point of view.

2 Geometric and convegence of some iterative methods

The following well-known theorem, giving enough conditions for the convergence of Newton method, follows easily from its geometric interpretation.

Theorem 1 *Let f'' be continuous on an interval J, containing a root x^* of f; let $f' \neq 0$ and $f'' \geq 0$ or $f'' \leq 0$ on J. Then Newton method converges monotonically to x^* from any point $x_0 \in J$ such that $f(x_0)f''(x_0) \geq 0$.*

Melman (see [12]) proposes similar results for one of the most famous third order method, the Halley method

$$x_{k+1} = x_k - \frac{2f(x_k)f'(x_k)}{2f'^2(x_k) - f(x_k)f''(x_k)}. \tag{3}$$

Theorem 2 *Let f''' be continuous, $f' \neq 0$, and $((\eta f')^{-\frac{1}{2}})'' \geq 0$ on an interval J containing a root x^* of f with $\eta = sgn(f')$. Then Halley method converges monotonically to x^* from any point of the interval.*

Halley method uses a rational function $\frac{x+a}{bx+c}$, which interpolates the function f and its two first derivatives at the point x_k. And the zero of the rational function is the new approach.

Now, we introduce new third order methods (see definition 1), being based on same geometric arguments, previously mentioned. The idea is to consider new interpolation functions to obtain more alternatives.

2.1 Exponential method

It is considered the exponential function

$$g(x) = a + ke^{bx} \tag{4}$$

such that:

$$f(\overline{x}) = g(\overline{x}), \tag{5}$$

$$f'(\overline{x}) = g'(\overline{x}), \tag{6}$$

$$f''(\overline{x}) = g''(\overline{x}). \tag{7}$$

This third order interpolation leads to the iterative method (see [4])

$$x_{k+1} = x_k + \frac{f'(x_k)}{f''(x_k)} \log(1 - \frac{f(x_k)f''(x_k)}{f'(x_k)^2}). \tag{8}$$

As for the classical methods, we can obtain the following covergence theorem.

Theorem 3 *Let f''' be continuous on an interval J containing a root x^* of f, let $f' \neq 0$, and sgn $(f')\left(\log|f'|\right)'' \leq 0$ in J. Then the exponential method converges monotonically to x^* from any point of the interval.*

Proof

We suppose $f' > 0$ (for $f' < 0$ similar).

As $f' > 0$ then for hypothesis $\left(\log|f'(x)|\right)'' \leq 0$, therefore

$$kb \exp(bx) \geq f'(x) > 0.$$

For $x \geq \overline{x}$

$$\int_{\overline{x}}^{x} kb \exp(bt)\, dt \geq \int_{\overline{x}}^{x} f'(t)\, dt$$

so

$$a + k \exp(bx) - a - k \exp(b\overline{x}) \geq f(x) - f(\overline{x})$$

then

$$a + k \exp(bx) \geq f(x)$$

Therefore, if we begin from a point on the left of x^*, we will obtain a monotonic increasing sequence, bounded from above by x^*, then it converges at the limit $\gamma \leq x^*$. So, for the construction of the method and the continuity of f the convergence is obtained, $\gamma = x^*$. If we begin from a point on the right of root, we will obtain $a + k \exp(bx) \leq f(x)$ and the convergence will be monotonic from the right. $\square$

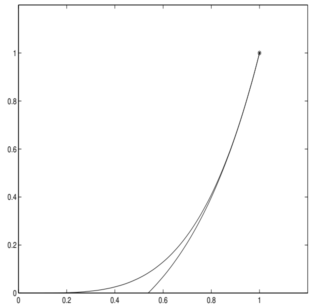


Figure 1: ‘ function, exponential interpolation, $x_0 = 1$’

2.2 Logarithmic method

We consider the logarithmic function

$$g(x) = a + k\log(x+b)$$

and the conditions (5), (6) and (7).

The iteration formula, based on logarithmic interpolation (see [4]) is given by

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}(\exp(\frac{f(x_k)f''(x_k)}{f'(x_k)^2}) - 1). \tag{9}$$

This leads to the following theorem.

Theorem 4 *Let f''' be continuous on an interval J containing a root x^* of f, let $f' \neq 0$, and $(\frac{1}{f'})(\frac{1}{f'})'' \geq 0$ (if and only if $2(f'')^2 \geq (f'''f')$) on J. Then the logarithmic method converges monotonically to x^* from any point of the interval .*

Proof

We suppose $f' > 0$ (for $f' < 0$ similar).

As $f' > 0$ then for hypothesis $\left(\frac{1}{f'}\right)'' \geq 0$, and therefore $\frac{k}{x+b} \geq f'(x) > 0$. For $x \geq \overline{x}$

$$\int_{\overline{x}}^{x} \frac{k}{t+b}dt \geq \int_{\overline{x}}^{x} f'(t)\,dt$$

so

$$a + k\log(x+b) - a - k\log(\overline{x}+b) \geq f(x) - f(\overline{x})$$

then

$$a + k\log(x+b) \geq f(x).$$

Therefore, if we begin from a point on the left of x^*, we will obtain a monotonic increasing sequence, bounded from above by x^*, then it converges to a limit $\gamma \leq x^*$. So, for the construction of the methods and the continuity of f the convergence is obtained, $\gamma = x^*$. If we begin from a point on the right of the root, we will obtain

$$a + k \log(x + b) \leq f(x)$$

and the convergence will be monotonic from the right. □

Notice that these methods can be written in the form

$$x_{k+1} = x_k + (1 + \theta_k + O(\theta_k^2)) \frac{f(x_k)}{f'(x_k)}$$

where $\theta_k := \frac{1}{2} \frac{f(x_k) f''(x_k)}{(f'(x_k))^2}$ see [1].

3 A comparison of sufficient conditions for global convergence

In the situation where the sufficient conditions of all four theorems hold, Newton method will always produce a smaller step than third-order methods. The reason for this is that when f is convex (concave), the tangent to the function at the approximation point always lie below (above) those three approximations (see [12] and figure 3).

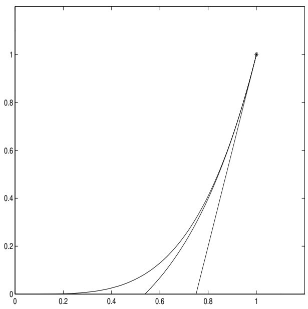


Figure 2: ‘ function, exponential interpolation, Newton interpolation (tangent), $x_0 = 1$ ’

In certain cases, it is possible to compare the third order methods. Considering, for instance, the case $f' > 0$ the conditions of the theorems (2), (3) and (4) are equivalent to:

$$\frac{3}{2}(f'')^2 \geq f'f''' \tag{10}$$

$$(f'')^2 \geq f'f''' \tag{11}$$

$$2(f'')^2 \geq f'f''' \tag{12}$$

respectively (see [2]).

These convergence conditions could also have been written in terms of the degree of logarithmic convexity of f' at x_0. The degree of logarithmic is a local functional defined by Garay and Hernández in [9]. If $g \in C^2(I)$, $x_0 \in I$, the degree of logarithmic convexity of g at x_0 is defined by

$$L_g(x_0) = \frac{g(x_0)g''(x_0)}{g'(x_0)^2}$$

Indeed, one can write the conditions as:

$$\frac{3}{2} \geq L_{f'}(t) \tag{13}$$

$$1 \geq L_{f'}(t) \tag{14}$$

$$2 \geq L_{f'}(t) \tag{15}$$

respectively.

Moreover, in [1], we can see the following interesting property:

Lemma 1 (Altman)

Let be f smooth enough such that $f(t) \geq 0$, $f' \neq 0$ in $t \in [0, t^*]$ and $f''' \geq 0$ in $t \in [0, t^*]$ and $f(t^*) = 0$. Then

$$L_f(t) \leq \frac{1}{2}. \tag{16}$$

The degree of logarithmic convexity is related to the speed of convergence and relation (16) is a typical hipothesis in convergence theorems, (for more details see [8],[9],[10]).

Remark 1 *It is possible to find (with similar ideas) a general global convergence theorem (we will assume $f^{'} > 0$). If $g \in C^2$ such that conditions (5), (6) and (7) hold.*

Theorem 5 *Let $f^{'''}$ be continuous on an interval J containing a root x^* of f, let $f^{'} > 0$, and $((f^{'})^{-1})^{''} \cdot ((f^{'})^{-1})^{'} \geq 0$. Then the iterative method associated to g converges monotonically to x^* from any point of the interval.*

4 Numerical experiments and conclusions

In order to show the performance of the different iterative methods, we have tested them on the results obtained for the zeros of some equations.

In table 1 we study the equation $\sin(x) - x = 0$, for different initial data we compute the number of iterations (until the convergence criterion is satisfied), the exponential method needs less iterations. With the same equation, in table 2 we compute the error for different number of iterations, the best results are obtained with exponential method too.

In table 3 we consider $x^5 - 1 = 0$, in this case, $f(x)$ verifies the hypothesis of all the theorems. We obtain the same conclusions as before (the convergences are uniform).

Finally we consider in $[0, 1]$ the Cauchy problem:

$$\begin{aligned} y' &= -4 \cdot 10^4\, y^3, \\ y(0) &= 1. \end{aligned} \tag{17}$$

We approximate the solution of the stiff problem (17) at $t = 1$ by applying implicit Euler method,

$$y_{n+1} = y_n + h \cdot f(t_{n+1}, y_{n+1}).$$

The different iterative processes are used for the non linear equation in the implicit method. We compute the number of iterations, in each step, until obtaining an error less than 10^{-4}. Exponential method needs less number of iterations ($j = jmax$) than the others, see table 5.

Remark 2 *Exponential method can decrease the error to* $3.00\ e - 04$ *with* $jmax = 15$ *(number of iterations implicit method=* 20*) and to* $9.29\ e - 06$ *with* $jmax = 6$ *(number of iterations implicit method=1000), but the accuracy is limited by the other methods.*

We have analyzed two new iterative methods. We have studied their convergence and we have tested their competitivity with respect the classical methods. Exponential method seemed to work very well in our preliminary numerical results. We have presented several reasons in order to use third order methods. The computational cost is not considered, since in 1-D all methods studied have similar properties. Nevertheless, 1-D problems are not trivial, since in some cases n-D problem are reduced to several 1-D problems based on local characteristic fields. In further papers we will establish these methods on Banach spaces, their convergence conditions and the computational evaluation.

Input Data	Exponential	Logarithmic	Halley
$x_0 = 0.001$	3	4	4
$x_0 = 0.01$	6	8	7
$x_0 = 0.1$	9	11	10
$x_0 = 0.3$	11	13	13
$x_0 = 0.5$	11	14	13
$x_0 = 0.9$	12	15	14
$x_0 = 1$	12	15	14
$x_0 = 2$	13	16	15

Table 1: Number of iterations untill to have an error $< 10^{-4}$, $f(x) = sin(x) - x$

Iterations	Exponential	Logarithmic	Halley
1	$4.50e-04$	$5.26e-04$	$4.99e-04$
2	$2.03e-04$	$2.76e-04$	$2.49e-04$
3	$9.15e-05$	$1.45e-04$	$1.24e-04$
4	$4.12e-05$	$7.66e-05$	$6.24e-05$
5	$1.85e-05$	$4.03e-05$	$3.12e-05$
6	$8.38e-06$	$2.12e-05$	$1.56e-05$
7	$3.77e-06$	$1.11e-05$	$4.81e-06$
8	$1.70e-06$	$5.87e-06$	$3.90e-06$
9	$7.67e-07$	$3.08e-06$	$1.95e-06$
10	$3.45e-07$	$1.62e-06$	$9.76e-07$

Table 2: Error, $x_0 = 0.001,\ f(x) = sin(x) - x$

Iterations	Exponential	Logarithmic	Halley
1	$5.51e-02$	$1.23e-01$	$1.00e-01$
2	$1.08e-04$	$4.59e-03$	$1.73e-03$
3	$8.50e-13$	$3.20e-07$	$1.04e-08$

Table 3: Error, $x_0 = 1.5,\ f(x) = x^5 - 1$

Iterations	Exponential	Logarithmic	Halley
1	$-3.86e-01$	$-2.64e-01$	$-3.33e-01$
2	$+1.49e-01$	$+6.98e-02$	$+1.11e-02$
3	$-5.76e-02$	$-1.84e-02$	$-3.70e-02$
4	$+2.22e-02$	$+4.87e-03$	$+1.23e-03$
5	$-8.60e-03$	$-1.28e-03$	$-4.11e-03$
6	$+3.32e-03$	$+3.40e-04$	$+1.37e-03$
7	$-1.28e-03$	$-8.99e-05$	$-4.57e-04$
10	$+7.39e-05$	$+1.65e-06$	$+1.69e-05$
20	$+5.47e-09$	$+3.94e-11$	$+2.86e-10$

Table 4: Iterations x_k, $x_0 = 1$, $f(x) = |x|^{\frac{1}{2}}$

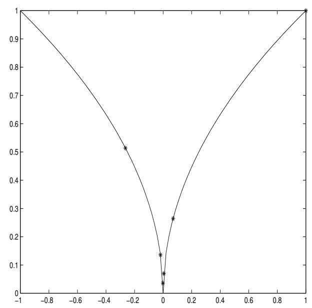


Figure 3: ' - ' $\sqrt{|x|}$, ' * ' log-iterations

Im.Met.	Expon.	Logarit.	Halley
20	$4.00e-04,\ j=12$	$4.00e-04,\ j=22$	$4.00e-04,\ j=25$
100	$7.89e-05,\ j=10$	$9.79e-05,\ j=20$	$9.79.e-04,\ j=25$
1000	$1.15e-05,\ j=5$	$1.51e-05,\ j=6$	$1.27e-05,\ j=10$

Table 5: Error and j =number iterations, stop when error$< 10^{-4}$

References

[1] Altman G.: Iterative methods of higher orde, Bull. Acad. Pollon. Sci. (Série des sci. math., astr., phys.). IX, (1961), 62-68.

[2] Amat S., Busquier S. and Candela V.: Métodos numéricos de alto orden para la resolución de ecuaciones escalares no lineales, Actas del XVI-CEDYA y VI-CMA (Canarias). II, (1999), 869-874.

[3] Bazaraa M.S. and Shetty C.M.: Nonlinear programming. Theory and Algorithms, John Wiley and Sons, Inc, 1979.

[4] Busquier S.: Métodos numéricos de alto orden para la resolucion de ecuaciones no lineales, Valencia GrAN Report, 2000.

[5] Candela V.F. and Marquina A.: Recurrence Relations for Rational Cubic Methods I: The Halley Method, Computing. 44, (1990), 169-184.

[6] Candela V.F. and Marquina A.: Recurrence Relations for Rational Cubic Methods II: The Chebyshev Method, Computing. 45, (1990), 355-367.

[7] Dahlquist G. and Björk A.: Numerical Methods, Prentice-Hall, inc., New Yersey, 1974.

[8] Ezquerro J.A.: Contrucción de procesos iterativos mediante acelaración del método de Newton, Ph. D. Thesis, Universidad de la Rioja, 1996.

[9] Garay J. and Hernández M.A.: Degree of logarithmic convexity, Publicaciones del Sem. Mat. García Galdeano, serie II, 26, Zaragoza, 1988.

[10] Gutiérrez J.M.: El método de Newton en espacios de Banach, Ph. D. Thesis, Universidad de la Rioja, 1995.

[11] Halley E.: Methodus nova, accurata ε facilis inveniendi radices aequationum quarumcumque generaliter, sine praevia reductione, Pholos. Trans. Roy. Soc. London. 18, (1694), 139-148.

[12] Melman A.: Geometry and Convergence of Euler's and Halley's Methods, SIAM Rev. 39, 4, (1997), 728-735.

[13] Salehov G.S.: On the convergence of the process of tangent hyperbolas, Dokl. Akad. Nauk SSSR. 82, (1952), 525-528, (in Russian).

[14] Šafiev R.A.: The method of tangent hyperbolas, Dokl. Akad. Nauk SSSR. 149, (1963), 788-791, (in Russian). Translation in Soviet Math. Dokl., 4, (1963), 482-485.

[15] Scavo T.R. and Thoo J.B.: On the geometry of Halley's method, Amer. Math. Monthly. 102, (1995), 417-426.

[16] Stoer J. and Burlirsch R.: Introduction to Numerical Analysis, Springer-Verlag, New York, 1993.

[17] Tapia R.A.: Newton's Method for optimization problems with Equality Constraints, SIAM J. Numerical Analysis. 11, (1974), 874-886.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 2, December 2001, pp. 73–76.
Submitted: May 3, 2001. Published: December 31 2001.

DYNAMIC OLIGOPOLIES WITH PRODUCTION FACTORS

László Kapolyi[1] and Ferenc Szidarovszky[2]

Abstract. A dynamic single-product oligopoly is introduced with continuous time scales that includes production factors used by the firms. The existence of the Nash–Cournot equilibrium is first established, and then the local and global asymptotical stability of interior equilibria are proven under realistic conditions.

1. Introduction

Since the pioneering work of Cournot (1838) has been published a large number of researchers have investigated the different aspects of oligopoly markets. The existence and uniqueness of the Nash–Cournot equilibrium was the first area that resulted in a series of work. Okuguchi (1976) gave a comprehensive summary of relevant results on single-product models without and with product differentiation. Multi-product models and related results are summarized in Okuguchi and Szidarovszky (1999), and also several model variants are introduced and examined including models with market saturation, labor-managed oligopolies, rent-seeking games, oligopsonies among others. Most of the models consider N-firms competing on the market in which the sale price is a function of the total output of the industry, and the boundedness of the available resources is taken into account by bounded feasible output sets. Oligopsonies include the competition of the firms in the labor and capital markets as well. A slightly more general model has been introduced and examined in Szidarovszky and Okuguchi (2001) in which multiple production factors were considered in a single-product oligopoly. The asymptotical stability of the equilibrium was examined under discrete time scales. In this paper the continuous time scale counterpart of the same model will be examined. The paper develops as follows. The next section introduces the basic model and the dynamic system, and the asymptotical behavior of the state trajectory will then be investigated.

---

[1]System Consulting Rt., Budapest 114, Pf. 709, H-1535, Hungary.

[2]Systems and Industrial Engineering Department, University of Arizona, Tucson, AZ 85721-0020, U.S.A.

2. The Basic Model and Dynamic System

Consider an N-firm oligopoly with single-product firms without product differentiation, and assume that the firms use M factors of production. Perfectly competitive product market and an oligopsonistic factor market are assumed. The profit of firm k is given as

$$\varphi_k(\mathbf{l}_1, \dots, \mathbf{l}_n) = pf_k(\mathbf{l}_k) - \mathbf{l}_k^T \mathbf{w}(\mathbf{L}) \tag{1}$$

where p is the competitive product price, f_k is the production of firm k, $\mathbf{l}_k$ is its M-element production factor, $\mathbf{w}$ is the M-element factor price vector, and $\mathbf{L} = \sum_{i=1}^N \mathbf{l}_i$ is the total production factor usage of the industry. For the sake of mathematical simplicity we assume that

(i) $f_k(\mathbf{l}_k) = \mathbf{l}_k^T \mathbf{C}_k \mathbf{l}_k + \mathbf{c}_k^T \mathbf{l}_k + \gamma_k$ for all k,

and

(ii) $\mathbf{w}(\mathbf{L}) = \mathbf{A}L + \mathbf{a}$,

where $\mathbf{c}_k$ and $\mathbf{a}$ are M-element vectors, $\mathbf{C}_k$ is an $M \times M$ symmetric matrix, $\mathbf{A}$ is an $M \times M$ matrix, and γ_k is a scalar. The available amount of the production factors is usually limited; therefore we assume that for all k,

(iii) $\mathbf{l}_k \in S_k$,

where S_k is a convex, bounded, and closed subset of R_+^M.

The Cournot–Nash equilibrium of this N-person noncooperative game is a vector $(\mathbf{l}_1^*, \dots, \mathbf{l}_N^*)$ such that

(a) $\mathbf{l}_k^* \in S_k$ for all k,

and

(b) for all k and $\mathbf{l}_k \in S_k$,

$$\varphi_k(\mathbf{l}_1^*, \dots, \mathbf{l}_N^*) \geq \varphi_k(\mathbf{l}_1^*, \dots, \mathbf{l}_{k-1}^*, \mathbf{l}_k, \mathbf{l}_{k+1}^*, \dots, \mathbf{l}_N^*). \tag{2}$$

Theorem 1. *Assume that $\mathbf{C}_k$ is negative definite for all k, and $\mathbf{A} + \mathbf{A}^T$ is positive semidefinite. Then there is at least one Cournot–Nash equilibrium.*

Proof. Since $\mathbf{C}_k$ is negative definite, f_k is concave for all k. The assumption that $\mathbf{A} + \mathbf{A}^T$ is positive semidefinite implies that w is monotonic in $\mathbf{L}$ in the sense that for all feasible $\mathbf{L}$ and $\mathbf{L}'$,

$$(\mathbf{L} - \mathbf{L}')^T \big(w(\mathbf{L}) - w(\mathbf{L}')\big) \geq 0$$

(see for example, Ortega and Rheinboldt, 1970). Therefore Lemma 3.1.1 of Okuguchi and Szidarovszky (1999, pp. 23–24) implies that $\mathbf{l}_k^T \mathbf{w}(\mathbf{L})$ is convex in $\mathbf{l}_k$. Therefore φ_k is concave in $\mathbf{l}_k$. Since it is continuous, and for all k, S_k is a convex, bounded, and closed subset of a finite dimensional Euclidean space, the Nikaido–Isoda theorem (see Nikaido and Isoda, 1955) implies the existence of an equilibrium. □

If the production factor usages are at the equilibrium, then it is the interest of all firms to keep the equilibrium, since none of the firms can increase its profit by unilaterally changing its production factor use. Otherwise it is usually assumed that the firms change their usages proportionally to their marginal profits. Notice that the gradient of φ_k with respect to $\mathbf{l}_k$ is the following:

$$2\mathbf{C}_k p\mathbf{l}_k + p\mathbf{c}_k - (\mathbf{A} + \mathbf{A}^T)\mathbf{l}_k - \mathbf{A}\sum_{i\neq k}\mathbf{l}_i - \mathbf{a},$$

therefore the dynamic model can be formulated as

$$\dot{\mathbf{l}}_k = \mathbf{K}_k\left(2\mathbf{C}_k p\mathbf{l}_k + p\mathbf{c}_k - (\mathbf{A} + \mathbf{A}^T)\mathbf{l}_k - \mathbf{A}\sum_{i\neq k}\mathbf{l}_i - a\right), \tag{3}$$

where $\mathbf{K}_k$ is a diagonal matrix with positive diagonal elements for all k. Under the assumptions of Theorem 1, a vector $(\mathbf{l}_1^*, \dots, \mathbf{l}_N^*)$ is an equilibrium of dynamic system (3) if and only if it is an interior Nash–Cournot equilibrium of the static N-person game.

3. Stability Analysis

The main result of this section can be stated as follows.

Theorem 2. *Under the conditions of Theorem 1, the interior Nash–Cournot equilibrium is asymptotically stable.*

Proof. The coefficient matrix of system (3) has the special structure $\mathbf{K}H$ with

$$\mathbf{H} = \begin{pmatrix} 2\mathbf{C}_1 p - (\mathbf{A} + \mathbf{A}^T) & -\mathbf{A} & \cdots & -\mathbf{A} \\ -\mathbf{A} & 2\mathbf{C}_2 p - (\mathbf{A} + \mathbf{A}^T) & \cdots & -\mathbf{A} \\ \vdots & \vdots & \ddots & \\ -\mathbf{A} & -\mathbf{A} & \cdots & 2\mathbf{C}_N p - (\mathbf{A} + \mathbf{A}^T) \end{pmatrix}$$

and

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_1 & & & \\ & \mathbf{K}_2 & & \\ & & \ddots & \\ & & & \mathbf{K}_N \end{pmatrix}.$$

Since $\mathbf{K}$ is positive definite, it is sufficient to show that $\mathbf{H} + \mathbf{H}^T$ is a negative definite matrix, and then Theorem 4.9 of Szidarovszky and Bahill (1998, pp. 213–214) implies that all eigenvalues of $\mathbf{K}H$ has negative real parts.

Notice that

$$\mathbf{H} + \mathbf{H}^T = \begin{pmatrix} 4\mathbf{C}_1 p & & & \\ & 4\mathbf{C}_2 p & & \\ & & \ddots & \\ & & & 4\mathbf{C}_N p \end{pmatrix} \tag{4}$$

$$+ \begin{pmatrix} -2(\mathbf{A} + \mathbf{A}^T) & -(\mathbf{A} + \mathbf{A}^T) & \cdots & -(\mathbf{A} + \mathbf{A}^T) \\ -(\mathbf{A} + \mathbf{A}^T) & -2(\mathbf{A} + \mathbf{A}^T) & \cdots & -(\mathbf{A} + \mathbf{A}^T) \\ & & \ddots & \\ -(\mathbf{A} + \mathbf{A}^T) & -(\mathbf{A} + \mathbf{A}^T) & \cdots & -2(\mathbf{A} + \mathbf{A}^T) \end{pmatrix}.$$

The first term is negative definite, since $\mathbf{C}_k$ is negative definite for all k, and $p > 0$. The second term is the Kronecker product of $\mathbf{A} + \mathbf{A}^T$ and $-(\mathbf{I} + \mathbf{E})$, where $\mathbf{I}$ is the $M \times M$ identity matrix and $\mathbf{E}$ is the $M \times M$ matrix, all elements of which are equal to unity. It is well known (see, for example Okuguchi and Szidarovszky, 1999, p. 69) that the eigenvalues of $-(\mathbf{I} + \mathbf{E})$ are -1 and $-M - 1$, both are negative. The eigenvalues of $\mathbf{A} + \mathbf{A}^T$ are nonnegative, since $\mathbf{A} + \mathbf{A}^T$ is positive semidefinite. It is also known (see for example, Ròzsa, 1974) that the eigenvalues of the Kronecker product of two matrices equal the products of the eigenvalues of the two factors. Since this matrix is symmetric, it is negative semidefinite. Hence $\mathbf{H} + \mathbf{H}^T$ is the sum of a negative definite matrix and a negative semidefinite matrix, which is negative definite with real and negative eigenvalues. $\square$

In the case of time invariant linear systems local and global asymptotical stability are equivalent to each other. Therefore the interior equilibrium of system (3) is globally asymptotically stable implying its uniqueness.

References

[1] Cournot, A. (1838). *Recherches sur les Principles Mathématiques de la Théorie des Richesses*, Hachette, Paris. (English translation (1960) *Researches into the Mathematical Principles of the Theory of Wealth*, Kelley, New York.)

[2] Nikaido, H. and Isoda, K. (1995). Note on noncooperative convex games, *Pac. J. of Math.*, Vol. 5, pp. 807–815.

[3] Okuguchi, K. (1976). *Expectations and Stability in Oligopoly Models*, Springer-Verlag, Berlin/Heidelberg/New York.

[4] Okuguchu, K. and Szidarovszky, F. (1999). *The Theory of Oligopoly with Multi-Product Firms* (2nd edition), Springer-Verlag, Berlin/Heidelberg/New York.

[5] Ortega, J.M. and Rheinboldt, W.C. (1970). *Iterative Solutions of Nonlinear Equations in Several Variables*, Academic Press, New York/London.

[6] Ròzsa, P. (1974). *Linear Algebra with Applications* (in Hungarian), Müszaki Könyvkiadò, Budapest.

[7] Szidarovszky, F. and Bahill, A.T. (1998). *Linear Systems Theory* (2nd edition), CRC Press, Boca Raton/London.

[8] Szidarovszky, F. and Okuguchi, K. (2001). Dynamic analysis of oligopsony under adaptive expectations, *Southwest Journal of Pure and Applied Mathematics*, to appear.

Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISBN 1083-0464
Issue 2 December 2001, pp. 77 – 85
Submitted: January 15, 2001. Published: December 31, 2001

A note on Chebyshev series solution of the Crank-Gupta equation

Dal, F.

Abstract

A moving boundary problem arising in biomechanical diffusion theory which has previously been investigated by Crank and Gupta (1972) is one of the simplest examples of a moving boundary problem in a partial differential equation; it has been used as a test problem by number of different authors.In this paper ; we show an accurate solution may be obtained by Chebyshev series solution.

A.M.S. Mathematics Subject Classification Codes 35K57,35C10.
Key Words and Phrases. Moving Boundary Problem, Chebyshev series solution

1 INTRODUCTION

The one-dimensional problem of the oxygen diffusion in an absorbing medium is one of the simplest examples of parabolic equation involving moving boundary.So, it has been used as a test problem in development of several new methods for such problems. A recent summary of the methods may be found in Furzeland (1977) , Rubinstein (1977) and comprehensive

Ege University, Department of Mathematics
E-mail: dal@fenfak.ege.edu.tr

book of Crank (1984) and related papers. The mathematical features of such problem can be stated as follows: The frontal velocity is explicitly related to the first spatial derivative of the solution value at the front through a mass or energy balance across the front. Thus, we wish to determine $C(x, t)$; the solution of

$$\frac{\partial C}{\partial t} = \frac{\partial^2 C}{\partial x^2} - 1 \tag{1}$$

x in $0 \leq x \leq S(t)$, $t \geq 0$ with the initial conditions

$$C(x, 0) = \frac{1}{2}(1 - x)^2, \quad 0 < x < 1 \tag{2}$$

$$S(0) = 1 \tag{3}$$

and the boundary conditions

$$\frac{\partial C}{\partial x} = 0 \quad on \quad x = 0 \tag{4}$$

$$C = 0 \quad \text{and} \quad \frac{\partial \mathrm{C}}{\partial \mathrm{x}} = 0 \quad \text{on} \quad \mathrm{x} = \mathrm{S(t)} \tag{5}$$

2 SOLUTION IN CHEBYSHEV SERIES

If the moving boundary conditions were replaced by a fixed boundary at $x = 1$ with the condition $\frac{\partial C}{\partial t} = 0$ on it , we may write the Chebyshev series as a solution of equation (1) as

$$C(x, t) = \sum_{r=0}^{n} a_r(t) T_r^*(x) \tag{6}$$

In order to determine the coefficients , we insert the eq. (6) into eq. (1) to obtain ;

$$\sum_{r=0}^{n} a_r^{\bullet}(t) T_r^*(x) = \sum_{r=0}^{n} a_r^{(2)} T_r^*(x) - 1 \tag{7}$$

where

$$\frac{\partial C}{\partial t} = \sum_{r=0}^{n} a_r^{\bullet}(t) T_r^*(x) \tag{8}$$

$$a_r^{\bullet}(t) = \frac{da_r}{dt} \tag{9}$$

$$\frac{\partial^2 C}{\partial x^2} = \sum_{r=0}^{n} a_r^{(2)}(t) T_r^*(x) \tag{10}$$

$$a_r^{(2)}(t) = 16 \sum_{s=1}^{n} s(r+s)(r+2s) a_{(r+2s)}(t) r = 0,1,...,n \tag{11}$$

and $T_r^*(x)$ is the Chebyshev polinomial defined in the interval $0 \leq x \leq 1$. More clearly, the equation eq. (7) can be written as;

$$\frac{1}{2} a_0^{\bullet}(t) T_0^*(x) + a_1^{\bullet}(t) T_1^3 * (x) + \quad ... \quad + a_n^{\bullet}(t) T_n^*(x) = \tag{12}$$
$$\frac{1}{2} a_0^{(2)}(t) T_0^*(x) + a_1^{(2)}(t) T_1^*(x) + \quad ... \quad + a_n^{(2)}(t) T_n^*(x) - T_0^*(x) \tag{13}$$

and using undefined coefficients for $T_r^*(x)$ $r = 0,1,...,n$ Namely ;

$$\begin{aligned} a_0^{\bullet}(t) &= a_0^{(2)}(t) - 2 \\ a_1^{\bullet}(t) &= a_1^{(2)}(t) \\ &\vdots \\ a_n^{\bullet}(t) &= a_n^{(2)}(t) \end{aligned} \tag{14}$$

Using the second terms of eq. (9) , eq. (10) and eq. (11) we get the system for $a_r^{\bullet}(t)$; $r = 0,1,...,n$ i.e:

$$\begin{aligned} \frac{da_0(t)}{dt} &= 16 \sum_{s=1}^{\frac{n-r}{2}} (s)(s)(2s) a_{2s-2} \\ \frac{da_1(t)}{dt} &= 16 \sum_{s=1}^{\frac{n-r-1}{2}} (s)(1+s)(1+2s) a_{1+2s} \\ &\vdots \\ \frac{da_r(t)}{dt} &= 16 \sum_{s=1}^{\frac{n-r}{2}} (s)(r+s)(r+2s) a_{r+2s} \end{aligned} \tag{15}$$

where $r = 0,1,...,n$; $r + 2s = n$ and $s = \frac{n-r-1}{2}$ if r is odd $s = \frac{n-r}{2}$ if r is even. This function clearly satisfies the boundary condition

$$\frac{\partial C}{\partial x} = 0 \quad \text{on} \quad \mathrm{x = S(t)} \quad \text{for} \quad \mathrm{t \geq 0}$$

Hence

$$\frac{\partial C}{\partial x} = \sum_{r=1}^{n} \frac{d}{dx} T_r^*(x) a_r(t) \tag{16}$$

$$\frac{d}{dx} T_r^*(x) = (-1)^{(r-1)} 2r^2, x = 0 \tag{17}$$

Inserting eq. (17) into eq. (16) and using the boundary condition ;

$$\sum_{r=1}^{n} (-1)^{(r-1)} 2r^2 a_r(t) = 0 \tag{18}$$

is obtained and also

$$a_n(t) = \frac{2a_1(t)}{2n^2} - \frac{\sum_{r=2}^{(n-1)} (-1)^r 2r^2 a_r(t)}{2n^2} \tag{19}$$

is easily written which is the extra condition which we need. Solving system eq. (15) with eq. (19) one can easily find the coefficients $a_0(t), a_1(t), ..., a_n(t)$. Also the function will satisfy the initial condition eq. (2) on $t = 0$ provided that

$$a_0(0) = \frac{3}{8}, a_1(0) = -\frac{1}{4}, a_2(0) = \frac{1}{16}, a_4(0) = 0, ..., a_n(0) = 0. \tag{20}$$

3 POSITION OF MOVING BOUNDARY

In order to determine the location of the moving boundary , $S(t)$, we use transformation $x = \frac{x}{S(t)}$ on to fix the boundary and hence $T_m^*(x)$ can be given as

$$T_m^*(x) = \frac{1}{2} \sum_{k=0}^{m} (-1)^k A 2^{2m-2k} \left(\frac{x}{S(t)} \right)^{m-k} \tag{21}$$

where

$$A = \frac{2(2m-k)!}{(2m-2k)!k!} - \frac{(2m-(k+1))!}{(2m-2k-1)!k!} \tag{22}$$

(see, Fox and Parker,1968) Inserting eq. (21) into the series, we find the solution on the moving boundary by

$$C(x,t) = \sum_{r=0}^{n}{}' a_r(t) \frac{1}{2} \sum_{k=0}^{m} (-1)^k A 2^{2m-2k} (\frac{x}{S(t)})^{(m-k)} \tag{23}$$

where the coefficients $a_r(t)$'s are the coefficients which are found previously in section 2. Now to find the position of moving boundary , we integrate both side of the differential equation eq. (1) with respect to x over the range for which it is valid.i.e. $0 \leq x \leq S(t)$. Thus we obtain

$$\frac{\partial}{\partial t}\int_0^{S(t)} C dx = -S(t) \tag{24}$$

Inserting equation (23) into eq. (24) we get an ordinary differential equation for the position of moving boundary $S(t)$, with t as the independent variable, i.e.

$$\frac{dS(t)}{dt} = \frac{-S(t)}{f} \tag{25}$$

where $f = \int_0^{S(t)} C dx$

4 NUMERICAL RESULTS AND DISCUSSION

In order to evaluted the numerical solution, in some sense of "good approximation" one requires a means of selecting number of terms to be used in the solution series. To do this we made a algorithm in seek a way that the difference in n th step solution and $n-1$ th step should be less than the required tolerance value i.e. $|C_n(x,t) - C_{n-1}(x,t)| < \varepsilon$ then we stop the calculating higher steps of the solution in series and last stop is taken as "good approximation" solution to the problem. In calculation we have used only eight step of our series to get "good approximation" with in the tolerance of 10^{-4}. The results are shown in the Table 1 and Table 2.There is comparison of concentrations at different times in table 1. As shown in the table 2 there is a small discrepancy in the position of the boundary which shows increment as the critical point is reached. Our calculation give the critical value as $t_c = 0.1998$ and is belived to be correct to the of figures quoted. However the Fourier solution given by Dahmardah and Mayers (1983) obtains useful accuracy with n about 20 and they quotes that the Fourier series convergences rather slowly, for small values of t and is overcomed by using the numerical integration from the positive to, using as a starting condition the approximation given Crank and Gupta (1972) and this manipulation gives very good accuracy for to about 0.01 where the Fourier series now converges rapidly. Also they quot as t increases the boundary moves left so that s decreases the boundary reaches $s = 0$ at a time t_c shortly before $t = 0.2$ and at this pint their solution becomes singular and they investigate the behavior of the solution near this point with a special change of variables. This also effects the accuracy and the effeciency of the Fourier method besides economy. Also the calculation of Fourier coefficients is not possible analytically , therefore the choice of numerical method is very important. Moreover their system of solution gives stiff equations which needs special attention. In our case, no such problem were encountered. We , for example avoided the singularity of the moving boundary at time t_c shortly before $t = 0.2$ with the Chebyshev form of Fourier series with the advange that the terminal discontinuities are eliminated. Besides in the theories of minimax, if the given function is continous in given interval, the minimax polynomial $P_n(x)$ is that polynomial for which the error $e_n(x) = f(x) - p_n(x)$ has equal and successively opposite values at $n+2$ points in the interval and the error gets successively smaller in absulate value with increasing n. Also known that none of the finite approximations to infinite series have had exactly the minimax property that minimum error of the resulting polynomial is smaller than of any other polynomial of the same degree. The truncation of the Chebyshev series at a point where the coefficients are decreasing rapidly will , of course , give a very close approximation to the minimax polynomial of corresponding degree and in practice the effort to improve on this is well rewarded in our case , as it can be be seen from the results. But yet, it is not clear how satisfactory the technique may be when applied to more general moving boundary problems, which is , of course the interest of further research.

References

[1] Crank,J. and R. S. Gupta, 1972, Jour Inst. Math. and Its Appl. 10, 19.

[2] Crank,J. 1984, Free and Moving Boundary problems, Clarendon Press, Oxford.

[3] Dahmardah, H. O. and Mayers, D.F. ,1983, IMA J. Numer. Analysis, 3, 81.

[4] Fox , L. and Parker, I.B. 1968, Chebyshev Polynomials in Numerical Analysis, Tec. Officer, Oxford Universty Press, London.

[5] Furzeland, R. M. 1977, Brunel Universty Math. Report TR/76.

[6] Hansen, E. and Hougaard,P. 1974, Inst. Math. and Its Appl., 13, 385.

[7] Rubinstein, L. I. 1971 , Trans. Math. Monographs 27. Am. Math. Soc., Providence , R.I..

Table 1: Comparison of consentrations at different times

x	Crank and Gupta (1972)	Hansen and Hougaard (1974)	Present
t=0.001			
0.1	0.40500	0.40460	0.40461
0.2	0.32000	0.32000	0.32002
0.3	0.24500	0.24500	0.24500
0.4	0.18000	0.18000	0.17999
0.5	0.12500	0.12500	0.12500
0.6	0.08000	0.08000	0.07999
0.7	0.04500	0.04500	0.04500
0.8	0.02000	0.02000	0.02002
0.9	0.00500	0.00500	0.00499
t=0.01			
0.1	0.36567	0.36507	0.36507
0.2	0.31072	0.30995	0.30995
0.3	0.24372	0.24327	0.24327
0.4	0.17999	0.17980	0.17980
0.5	0.12500	0.12498	0.12497
0.6	0.08000	0.07999	0.07980
0.7	0.04500	0.04500	0.04327
0.8	0.02000	0.02000	0.01990
0.9	0.00500	0.00500	0.00450
t=0.02			
0.1	0.32622	0.32588	0.32588
0.2	0.28718	0.28667	0.28667
0.3	0.23379	0.23327	0.23325
0.4	0.17696	0.17660	0.17645
0.5	0.12437	0.12419	0.12330
0.6	0.07990	0.07984	0.07960
0.7	0.04499	0.04497	0.04317
0.8	0.01999	0.01999	0.01800
0.9	0.00500	0.00500	0.00410

Table 2: Comparison of position of moving boundary

Time	Crank and Gupta (1972)	Dahmardah and Mayers (1983)	Present
0.02	1.00000	0.99999	1.00000
0.1	0.93518	0.93501	0.93020
0.12	0.87885	0.87917	0.84199
0.14	0.79756	0.79894	0.74705
0.16	0.68502	0.68337	0.58839
0.19	0.33873	0.34537	0.31107
0.195	0.16128	0.20652	0.19897
0.196	–	0.16312	0.16687
0.198	–	–	0.00050
0.1989	–	–	0.000003
0.1990	–	–	0.0000028
0.1998	–	–	0.0000004

Information for Authors cont.

Copyright. Authors will be asked, upon acceptance of an article, to transfer copyright of the article to the publisher. This will ensure the widest possible dissemination of information under copyright laws. Preprints are available via web page at http://rattler.cameron.edu/SWJPAM. Authorization to copy is granted provided that the SWJPAM copyright symbol is not removed.

Tables. Tables must be numbered consecutively with roman numerals and have titles. They should be prepared in such a manner that no break is necessary.

Figures. All graphs and diagrams should be referred to as 'figure' (not abbreviated) in the text. A detailed caption should be provided for each figure, and figures should be numbered consecutively.

Footnotes. The use of footnotes should be minimized.

Notification. The author will be notified of the posting of his/her work and the time of release for SWJPAM.

References. References should be listed at the end of each paper. They should be numbered consecutively using arabic numbers between square brackets in the text. In the list of references, the numbers should appear without square brackets. A journal reference should comprise of name(s) and initial(s) of author(s), title of article, abbreviated name of journal (following the style of 'Abbreviations of Names of Serials reviewed in Mathematical Reviews'), volume number, year of publication (within parentheses), and full page numbers, e.g.

1. Argyros, I.K. and Szidarovszky, F.: On the Monotone Convergence of Algorithmic Models, Appl. Math. Com. 48, (1992), 167–176.

A book reference should comprise of name(s) and initial(s) of author(s), title of book, publisher, town, year of publication, e.g.

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

An edited volume reference should comprise of name(s) and initial(s) of author(s), title of paper, name(s) of editor(s), title of book, publisher, town, year, inclusive page numbers, e.g.

12. Prenter, P.: On Polynomial Operators and Equations, L.B. Rall (ed), Computational Solution of nonlinear operator equations, Johy Wiley Publisher, New York, 1971, pp. 361–398.

Indexing and Abstracting services. Articles published in SWJPAM are selectively indexed or abstracted in: Mathematical Reviews of A.M.S., Zentralblatt für Mathematik, European Mathematical Society (EMIS).