



COMPUTERS and SIMULATION in MODERN SCIENCE

Volume IV

Editors:

Nikos Mastorakis,

Metin Demiralp,

Valeri M. Mladenov

COMPUTERS and SIMULATION in MODERN SCIENCE - Volume IV



Selected papers from WSEAS Conferences

University of Cambridge (UK),
University of Harvard (USA),
Massachusetts Institute of Technology -- MIT (USA),
China Jiliang University (China),
Beijing Jiaotong University (China),
Univ. Politecnica of Bucharest (Romania),
Technical University of Sofia (Bulgaria),
Tianjin University of Technology and Education (China),
Zhejiang Univ. of Technology (China),
University of Alcala, Madrid (Spain),
University of Cantabria, Santander (Spain),
Budapest Tech (Hungary),
University of La Laguna (Spain),
Politehnica University of Timisoara (Romania)

ISSN: 1792-6882

ISBN: 978-960-474-267-7



COMPUTERS and SIMULATION in MODERN SCIENCE Volume IV

"This book contains important selected papers from WSEAS Conferences organized in the University of Cambridge (UK), University of Harvard (USA), Massachusetts Institute of Technology -- MIT (USA), China Jiliang University (China), Beijing Jiaotong University (China), Univ. Politecnica of Bucharest (Romania), Technical University of Sofia (Bulgaria), Tianjin University of Technology and Education (China), Zhejiang Univ. of Technology (China), University of Alcala, Madrid (Spain), University of Cantabria, Santander (Spain), Budapest Tech (Hungary), University of La Laguna (Spain), Politehnica University of Timisoara (Romania)"

COMPUTERS and SIMULATION in MODERN SCIENCE

Volume IV

"This book contains important selected papers from WSEAS Conferences organized in the University of Cambridge (UK), University of Harvard (USA), Massachusetts Institute of Technology -- MIT (USA), China Jiliang University (China), Beijing Jiaotong University (China), Univ. Politecnica of Bucharest (Romania), Technical University of Sofia (Bulgaria), Tianjin University of Technology and Education (China), Zhejiang Univ. of Technology (China), University of Alcala, Madrid (Spain), University of Cantabria, Santander (Spain), Budapest Tech (Hungary), University of La Laguna (Spain), Politehnica University of Timisoara (Romania)"

Published by WSEAS Press

www.wseas.org

Copyright © 2010, by WSEAS Press

All the copyright of the present book belongs to the World Scientific and Engineering Academy and Society Press. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Editor of World Scientific and Engineering Academy and Society Press.

All papers of the present volume were peer reviewed by two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.

See also: <http://www.worldses.org/review/index.html>

ISSN: 1792-6882

ISBN: 978-960-474-267-7



World Scientific and Engineering Academy and Society

COMPUTERS and SIMULATION in MODERN SCIENCE

Volume IV

"This book contains important selected papers from WSEAS Conferences organized in the University of Cambridge (UK), University of Harvard (USA), Massachusetts Institute of Technology -- MIT (USA), China Jiliang University (China), Beijing Jiaotong University (China), Univ. Politecnica of Bucharest (Romania), Technical University of Sofia (Bulgaria), Tianjin University of Technology and Education (China), Zhejiang Univ. of Technology (China), University of Alcala, Madrid (Spain), University of Cantabria, Santander (Spain), Budapest Tech (Hungary), University of La Laguna (Spain), Politehnica University of Timisoara (Romania)"

Editors:

Prof. Nikos Mastorakis, Technical University of Sofia, BULGARIA / Hellenic Naval Academy, GREECE
Prof. Metin Demiralp, Istanbul Technical University, TURKEY
Prof. Valeri M. Mladenov, Technical University of Sofia, BULGARIA

Editor-in-Chief:

Prof. Nikos Mastorakis, Technical University of Sofia, BULGARIA / Hellenic Naval Academy, GREECE

Editorial Board:

Prof. Siavash H. Sohrab, Northwestern University, Evanston, Illinois, USA
Prof. Mikhael Wasfy, University of Central Florida, USA
Prof. Slobodan I. Babic, Ecole Polytechnique de Montreal, CANADA
Prof. Yiming Li, National Chiao Tung University, Hsinchu 300, TAIWAN
Prof. Petr Y. Ekel, Pontif. Cathol. Univ. of Minas Gerais, 30535-610, Belo Horizonte, BRAZIL
Prof. Tuan Pham, James Cook University, Townsville, QLD 4811, AUSTRALIA
Prof. M. Isabel Garcia-Planas, Univ. Polit. de Catalunya, C.Mineria 1, 08038, Barcelona, SPAIN
Prof. Demetrios Kazakos, Texas Southern University, 3100 Cleburne Str., Houston, Texas 77004, USA
Prof. Olga Martin, University Politehnica of Bucharest, ROMANIA
Prof. José M^a Zamanillo Sáinz de la Maza, University of Cantabria, SPAIN
Prof. Alexander Zemliak, Puebla Aut. Univ, Av. San Claudio s/n, Ciudad, Puebla, Pue, 72570, MEXICO
Prof. Jose Carlos Quadrado, Electrical Engineering Department, ISEL, Lisbon, PORTUGAL
Prof. Pablo Luis López Espí, Universidad de Alcalá, SPAIN

Preface

This book contains important selected papers from WSEAS Conferences organized in the University of Cambridge (UK), University of Harvard (USA), Massachusetts Institute of Technology -- MIT (USA), China Jiliang University (China), Beijing Jiaotong University (China), Univ. Politecnica of Bucharest (Romania), Technical University of Sofia (Bulgaria), Tianjin University of Technology and Education (China), Zhejiang Univ. of Technology (China), University of Alcalá, Madrid (Spain), University of Cantabria, Santander (Spain), Budapest Tech (Hungary), University of La Laguna (Spain), Politehnica University of Timisoara (Romania)

The contents of this Book will be sent to the WSEAS collaborating indices after the conference: **www.worldses.org/indexes**

The contents of this Book are also published in CD-ROM. Both will be sent to the WSEAS collaborating indices after the conference: **www.worldses.org/indexes**

The friendliness and openness of the WSEAS conferences, adds to their ability to grow by constantly attracting young researchers. The WSEAS Conferences attract a large number of well-established and leading researchers in various areas of Science and Engineering as you can see from **<http://www.wseas.org/reports>**. Your feedback encourages the society to go ahead as you can see in **<http://www.worldses.org/feedback.htm>**

Expanded and enhanced versions of papers published in this conference proceedings are also going to be considered for possible publication in one of the WSEAS journals that participate in the major International Scientific Indices (Elsevier, Scopus, EI, ACM, Compendex, INSPEC, CSA see: www.worldses.org/indexes) these papers must be of high-quality (break-through work) and a new round of a very strict review will follow. (No additional fee will be required for the publication of the extended version in a journal). WSEAS has also collaboration with several other international publishers and all these excellent papers of this volume could be further improved, could be extended and could be enhanced for possible additional evaluation in one of the editions of these international publishers.

Finally, we cordially thank all the people of WSEAS for their efforts to maintain the high scientific level of conferences, proceedings and journals.

Table of Contents

Chapter 1:

- The Simple Genetic Algorithm with Alpha-Selection, Uniform Crossover and Bitwise Mutation** 10
Andre Neubauer

Chapter 2:

- Safety of Masonry Arch Bridges: A Quick Scan Method** 24
A. Audenaert, G. Reniers, W. Dullaert, H. Peremans

Chapter 3:

- Moving Surface Reconstruction by Self Organization** 35
M. V. Bessmeltsev

Chapter 4:

- Highest Label Preflow Algorithm for the Parametric Minimum Flow Problem - A Linguistic Rule-Based Network Partitioning Approach** 45
L. Sangeorzan, M. N. Parpalea, M. M. Parpalea

Chapter 5:

- Information Process in the Biosensor of the Angular Motion of the Head** 56
V. V. Alexandrov, T. B. Alexandrova, G. Castillo-Quiroz, A. Ortega, R. Vega, N. Shulenina, G Sidorenko, E. Soto

Chapter 6:

- Numerical Analysis of Electromagnetic (EM) Absorption Reduction by using Material Attachment** 63
M. R. I. Faruque, M. T. Islam, N. Misran

Chapter 7:

- Comparison of a Pressure Correction based Solver VoF Method in Moving Floor Container** 73
R. Ghandriz, M. Goudarzi

Chapter 8:

- Possibilities of Air Quality Modelling based on IF-Sets Theory** 90
V. Olej, P. Hajek

Chapter 9:

- Reduction Method with Finite-Difference Approximation for the Model of Small Transverse Vibrations in Thin Elastic Plates and some Applications** 101
A. A. Kuleshov

Chapter 10:**Principal Component Analysis for Pose Estimation using Range Data 111***Galina Okouneva, Donald J. McTavish, Aradhana Choudhuri***Chapter 11:****A Parametric Approach to Matrix Root Clustering 124***V. G. Melnikov***Chapter 12:****Towards a Deterministic Model for Course Timetabling 134***P. Pozos-Parra, O. Chavez-Bosquez, J. L. Gomez Ramos***Chapter 13:****Periodic Flows Associated with Solutions of Hill Equation 145***G. Bognar, S. H. Sohrab***Chapter 14:****Simulation Model of the JPEG Sub-Image Compression based on an Upgraded Petri Net 155***P. Strbac, M. Tuba***Chapter 15:****BIOSARP – Bio-Inspired Self-Optimized Routing Algorithm using Ant Colony Optimization for Wireless Sensor Network – Experimental Performance Evaluation 165***Kashif Saleem, Norsheila Fisal, M. Ariff Baharudin, Adel Ali Ahmed, Sharifah Hafizah, Sharifah Kamilah***Chapter 16:****Characterization of all Optimal Flows in Networks 176***L. Ciupala, E. Ciurea***Chapter 17:****Parameterization to avoid the Gibbs Phenomenon 186***E. Chicurel-Uziel***Chapter 18:****Multi Party Computation using Infinite Product Series over Distributed Data 196***Abdolreza Rasouli, Mohd Aizaini Maarof, Mahboubah Shamsi***Chapter 19:****A Method of Extended Linearization for Polynomial Periodic and Autonomous Systems 207***V. G. Melnikov*

Chapter 20:

Power Transmission Line Magnetic Fields for different Phase Arrangements	216
<i>L. Neamt, L. E. Petrean, O. Chiver, Z. Erdei</i>	

Chapter 21:

An Ellipse Framework for Precise Iris Edge Detection	227
<i>Mahboubah Shamsi, Puteh Bt Saad, Subariah BT Ibrahim, Abdolreza Rasouli</i>	

Chapter 22:

Adaptive, Model-Following Chaos Suppression and Synchronization	237
<i>J. Kabzinski</i>	

Author Index	249
---------------------	------------

Chapter 1

Simple Genetic Algorithm with α -Selection, Uniform Crossover and Bitwise Mutation

A. Neubauer

Information Processing Systems Lab
Department of Electrical Engineering and Computer Science
Münster University of Applied Sciences
Stegerwaldstraße 39, D-48565 Steinfurt, Germany
E-mail: andre.neubauer@fh-muenster.de

Abstract. Genetic algorithms are random heuristic search algorithms which mimic biological evolution and molecular genetics in simplified form. These algorithms can be theoretically described by an infinite population model with the help of a deterministic dynamical system by which the stochastic population trajectory is characterized using a deterministic heuristic function and its fixed points. For practical problem sizes the determination of the fixed points is unfeasible even for the simple genetic algorithm with fitness-proportional selection, crossover and bitwise mutation. The simple genetic algorithm with α -selection allows the analytical calculation of the unique fixed point of the corresponding intrinsic system model. In this paper, an overview of the theory of the simple genetic algorithm with α -selection, uniform crossover and bitwise mutation is given and experimental results are presented showing a close agreement to the theoretical predictions.

Keywords. Simple genetic algorithm, α -selection, random heuristic search, dynamical system model, intrinsic system model

1.1 Introduction

As specific instances of random heuristic search (RHS), genetic algorithms mimic biological evolution and molecular genetics in simplified form [2,10]. These evolutionary algorithms have been successfully applied in a large variety of applications [1,3,8,9,11]. Genetic algorithms (GA) process populations of individuals which evolve according to selection and genetic operators like crossover and mutation. The algorithm's stochastic dynamics can be described with the help of a dynamical system model introduced by VOSE et al. [10,13,14]. The population trajectory is attracted by the fixed points of an underlying deterministic heuristic function which also yields the expected next population. However, even for moderate problem sizes the calculation of the fixed points is difficult.

The simple genetic algorithm (SGA) with α -selection recently introduced in [4–7] allows to explicitly derive the fixed points of the heuristic function. In this selection scheme, the best or α -individual is mated with individuals randomly chosen from the current population with uniform probability. For the simple genetic algorithm with α -selection it is further possible to formulate an intrinsic system model which is compatible with the equivalence relation imposed by schemata. The intrinsic system model provides a means to analyze the genetic algorithm's exploitation and exploration of the search space due to the mixing operation caused by crossover and mutation irrespective of the fitness function.

This paper gives an overview of the theoretical results for the simple genetic algorithm with α -selection and its intrinsic system model. In addition to the theoretical analysis experimental results are presented. The paper is organized as follows. The simple genetic algorithm with α -selection is described as a specific instance of random heuristic search in Sect. 2 based on the notion of the best individual randomly mating with other individuals in the current population. In Sect. 3 the corresponding dynamical system model is derived based on which the intrinsic system model of the simple genetic algorithm with α -selection is formulated in Sect. 4. Simulation results for the simple genetic algorithm with α -selection, uniform crossover and bitwise mutation are presented in Sect. 5 showing a close agreement between theory and experiment. A brief conclusion is given in Sect. 6.

1.2 Simple Genetic Algorithm with α -Selection

In this section the simple genetic algorithm with α -selection, uniform crossover and bitwise mutation is described following the notation and definition of the simple genetic algorithm (SGA) in [13]. It is assumed that the genetic algorithm is used for the maximization of a fitness function f which is defined over the search space $\Omega = \{0,1\}^\ell$ consisting of binary tuples $(a_0, a_1, \dots, a_{\ell-1})$.

Each binary tuple $(a_0, a_1, \dots, a_{\ell-1}) = a_0 a_1 \dots a_{\ell-1}$ will be identified with the integer $a = a_0 \cdot 2^{\ell-1} + a_1 \cdot 2^{\ell-2} + \dots + a_{\ell-1} \cdot 2^0$ leading to the search space $\Omega = \{0, 1, \dots, n-1\}$ with cardinality $|\Omega| = n = 2^\ell$. The fitness values are given by $f(a) = f_a$. Based on the binary number representation the bitwise modulo-2 addition $a \oplus b$, bitwise modulo-2 multiplication $a \otimes b$ and bitwise binary complement \bar{a} are defined. Vice versa, the integer $a \in \Omega$ is viewed as a column vector $(a_0, a_1, \dots, a_{\ell-1})^T$. The all-one tuple $\mathbf{1}$ corresponds to the integer $n-1 = 2^\ell - 1$. The indicator function is defined by $[i = j] = 1$ if $i = j$ and 0 if $i \neq j$.

1.2.1 Algorithm

The *simple genetic algorithm with α -selection* works over populations $P(t)$ defined as multisets of r individual binary tuples $a(t) \in \Omega$. For the creation of offspring individuals in each generation t genetic operators like crossover χ_Ω and mutation μ_Ω are applied to parental individuals (see Fig. 1.1).

```

 $t := 0$ ;
initialize population  $P(0)$ ;
while end of adaptation  $\neq$  true do
  select  $\alpha$ -individual  $b(t)$  as first parent;
  for the creation of  $r$  offspring do
    select second parent  $c(t)$  randomly;
    apply crossover  $\chi_\Omega$  and mutation  $\mu_\Omega$ 
       $a(t+1) := \mu_\Omega(\chi_\Omega(b(t), c(t)))$ ;
  end
  increment  $t := t + 1$ ;
end

```

Fig. 1.1. Simple genetic algorithm with α -selection [4–7]

1.2.2 α -Selection

For the α -selection scheme let

$$b(t) = \operatorname{argmax} \{f_i : i \in P(t)\} \quad (1.1)$$

be the best individual or α -individual in the current population $P(t)$. In the simple genetic algorithm with α -selection the α -individual $b(t)$ is mated with individuals randomly chosen from the current population $P(t)$ with uniform probability r^{-1} .

1.2.3 Mixing

The *crossover* operator $\chi_\Omega : \Omega \rightarrow \Omega$ randomly generates an offspring tuple $(a_0, a_1, \dots, a_{\ell-1})$ according to $a = \chi_\Omega(b, c)$ with crossover probability χ from two parental tuples $(b_0, b_1, \dots, b_{\ell-1})$ and $(c_0, c_1, \dots, c_{\ell-1})$. With the crossover mask $m \in \Omega$ the tuples

$$a = b \otimes m \oplus \bar{m} \otimes c \quad (1.2)$$

or

$$a = b \otimes \bar{m} \oplus m \otimes c \quad (1.3)$$

are generated one of which is chosen as offspring a with equal probability 2^{-1} . For *uniform crossover* the crossover mask m is randomly chosen from Ω according to the probability distribution vector $\chi = (\chi_0, \chi_1, \dots, \chi_{\ell-1})^T$ with [13]

$$\chi_m = \begin{cases} 1 - \chi + \chi \cdot 2^{-\ell}, & m = 0 \\ \chi \cdot 2^{-\ell}, & m > 0 \end{cases} \quad (1.4)$$

The bitwise *mutation* operator $\mu_\Omega : \Omega \rightarrow \Omega$, which randomly flips each bit of the tuple $(a_0, a_1, \dots, a_{\ell-1})$ with mutation probability μ , is defined with the help of the mutation mask $m \in \Omega$ according to $\mu_\Omega(a) = a \oplus m$. The mutation mask m is randomly chosen from Ω according to the probability distribution vector $\mu = (\mu_0, \mu_1, \dots, \mu_{\ell-1})^T$ with [13]

$$\mu_m = \mu^{1^T m} \cdot (1 - \mu)^{\ell - 1^T m}. \quad (1.5)$$

A typical value of the mutation probability is $\mu \propto \frac{1}{\ell}$.

1.3 Dynamical System Model

In the *dynamical system model* [13] the dynamics of the simple genetic algorithm is compactly formulated by defining the population vector $\mathbf{p} = (p_0, p_1, \dots, p_{\ell-1})^T$. Each component

$$p_i = \frac{1}{r} \sum_{j \in P} [j = i] \quad (1.6)$$

gives the proportion of the element $i \in \Omega$ in the current population P . The population vector \mathbf{p} is an element of the simplex

$$\Lambda = \left\{ \mathbf{p} \in \mathbb{R}^n : p_i \geq 0 \wedge \sum_{i \in \Omega} p_i = 1 \right\}. \quad (1.7)$$

For a population of size r the number of possible population vectors is given by $\binom{n+r-1}{r}$.

In the limit of infinite populations with $r \rightarrow \infty$ the population vectors are dense in the simplex Λ . For simplicity we will take the simplex Λ as the defining region of the population vector \mathbf{p} which is strictly valid only for large populations with $r \gg 1$ in the sense of an *infinite population model*.

The simple genetic algorithm is now described as an instance of RHS $\tau : \Lambda \rightarrow \Lambda$ according to $\mathbf{p}(t+1) = \tau(\mathbf{p}(t))$ with τ depending on the random selection and genetic operators. As outlined in [13] τ can be equivalently represented by a suitable heuristic function $\mathcal{G} : \Lambda \rightarrow \Lambda$ which for a given population vector \mathbf{p} yields the probability distribution $\mathcal{G}(\mathbf{p})$. This probability distribution

$$\mathcal{G}(\mathbf{p})_i = \Pr\{\text{individual } i \text{ is sampled from } \Omega\} \quad (1.8)$$

is used to generate the next population as illustrated in Fig. 1.2. The transition probabilities of the RHS τ are given by the formula [13]

$$\Pr\{\tau(\mathbf{p}) = \mathbf{q}\} = r! \prod_{i \in \Omega} \frac{\mathcal{G}(\mathbf{p})_i^{rq_i}}{(rq_i)!} . \quad (1.9)$$

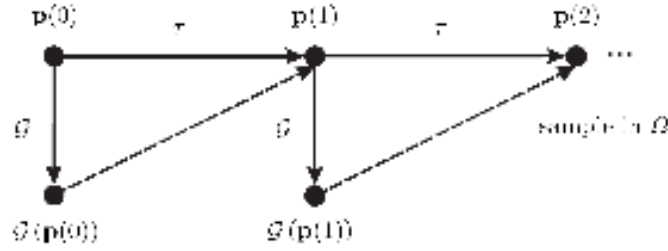


Fig. 1.2. Simple genetic algorithm as RHS τ with heuristic function \mathcal{G} [13]

The trajectory $\mathbf{p}, \tau(\mathbf{p}), \tau^2(\mathbf{p}), \dots$ approximately follows the trajectory $\mathbf{p}, \mathcal{G}(\mathbf{p}), \mathcal{G}^2(\mathbf{p}), \dots$ of the deterministic dynamical system with

$$\mathbb{E}\{\tau(\mathbf{p})\} = \mathcal{G}(\mathbf{p}) . \quad (1.10)$$

Because of the corresponding mean quadratic deviation

$$\mathbb{E}\{\|\tau(\mathbf{p}) - \mathcal{G}(\mathbf{p})\|^2\} = \frac{1}{r} \cdot (1 - \|\mathcal{G}(\mathbf{p})\|^2) \quad (1.11)$$

the RHS τ behaves like the deterministic dynamical system model in the limit of infinite populations with $r \rightarrow \infty$. As illustrated by experimental evidence the RHS τ shows *punctuated equilibria*, i.e. phases of relative stability nearby a fixed point $\mathbf{w} = \mathcal{G}(\mathbf{w})$ of the heuristic function \mathcal{G} disrupted by sudden transitions to another dynamical equilibrium near another fixed point. We call this the *fixed point hypothesis* of genetic algorithms.

1.3.1 Heuristic

In the simple genetic algorithm with α -selection the α -individual

$$b = \operatorname{argmax}\{f_i : i \in \Omega \wedge p_i > 0\} \quad (1.12)$$

is selected as the first parent for creation of a new offspring, whereas the second parent is chosen uniformly at random from the current population according to the probability distribution p_j over Ω with $j \in \Omega$. The heuristic function $\mathcal{G}(\mathbf{p})$ follows to

$$\mathcal{G}(\mathbf{p})_i = \sum_{j \in \Omega} p_j \cdot \Pr\{\mu_\Omega(\chi_\Omega(b, j)) = i\}. \quad (1.13)$$

The probability distributions for crossover χ_Ω and mutation μ_Ω lead to

$$\Pr\{\mu_\Omega(\chi_\Omega(b, j)) = i\} = \sum_{u, v \in \Omega} \mu_v \cdot \frac{\chi_u + \chi_{\bar{u}}}{2} \cdot [b \otimes u \oplus \bar{u} \otimes j = i \oplus v]. \quad (1.14)$$

By defining the $n \times n$ *mixing matrix* [13]

$$M_{i,j} = \sum_{u, v \in \Omega} \mu_v \cdot \frac{\chi_u + \chi_{\bar{u}}}{2} \cdot [i \otimes u \oplus \bar{u} \otimes j = v] \quad (1.15)$$

this yields $\Pr\{\mu_\Omega(\chi_\Omega(b, j)) = i\} = M_{i \oplus b, i \oplus j}$ and finally

$$\mathcal{G}(\mathbf{p})_i = \sum_{j \in \Omega} p_j \cdot M_{i \oplus b, i \oplus j}. \quad (1.16)$$

With the permutation matrix $(\sigma_b)_{i,j} = [i \oplus j = b]$ and the *twist* $(M^*)_{i,j} = M_{i \oplus j, i}$ of the symmetric mixing matrix $M = M^T$ the new population vector is given by

$$\boxed{\mathbf{q} = \mathcal{G}(\mathbf{p}) = \sigma_b M^* \sigma_b \cdot \mathbf{p}}. \quad (1.17)$$

This *dynamical system model* is illustrated in Fig. 1.3.

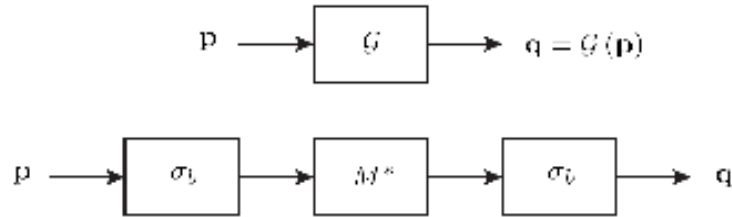


Fig. 1.3. Dynamical system model of the simple genetic algorithm with α -selection [5]

1.3.2 Mixing Matrix

The calculation of the mixing matrix M can be carried out efficiently with the help of the WALSH transform [12]. For a matrix M the WALSH transform is $\hat{M} = W \cdot M \cdot W$ with the $n \times n$ WALSH matrix $W_{i,j} = n^{-1/2} \cdot (-1)^{i^T j}$. The WALSH matrix W is symmetric and orthogonal, i.e. $W^{-1} = W^T = W$. The WALSH transform of a vector \mathbf{v} yields $\hat{\mathbf{v}} = M \cdot \mathbf{v}$. In Fig. 1.4 the WALSH matrix W is illustrated for $n = 2^6 = 64$.

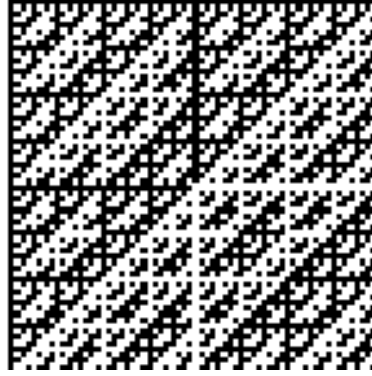


Fig. 1.4. Illustration of the WALSH matrix W for $n = 2^6 = 64$

For crossover and bitwise mutation the WALSH transform of the mixing matrix M is given by [13]

$$\hat{M}_{i,j} = [i \otimes j = 0] \cdot \frac{(1-2\mu)^{i^T(i \oplus j)}}{2} \sum_{k \in \Omega_{\bar{i} \oplus \bar{j}}} (\chi_{k \oplus i} + \chi_{k \oplus j}) \quad (1.18)$$

with

$$\Omega_k = \{i \in \Omega : i \otimes \bar{k} = 0\}. \quad (1.19)$$

Due to the factor $[i \otimes j = 0]$ the components $\hat{M}_{i,j}$ are nonzero only if $i \otimes j = 0$ or $j \in \Omega_{\bar{i}}$, respectively. The WALSH transform of the twist of the mixing matrix can be calculated from

$$(M^{*\wedge})_{i,j} = \hat{M}_{i \oplus j, j}. \quad (1.20)$$

1.4 Intrinsic System Model

The matrix $\sigma_b M^* \sigma_b$ of the dynamical system model of the simple genetic algorithm with α -selection in Eq. (1.17) depends on the mixing matrix M and the α -individual b . Because of $\sigma_b^{-1} = \sigma_b$ this yields the equivalent formulation

$$\boxed{\sigma_b \mathbf{q} = M^* \cdot \sigma_b \mathbf{p}}. \quad (1.21)$$

The permuted population vector $\sigma_b \mathbf{p}$ develops according to the matrix M^* which is independent of the α -individual b . The matrix M^* defines the *intrinsic system model* of the genetic algorithm with α -selection [5]. Because σ_0 is equal to the identity matrix the intrinsic system model of the simple genetic algorithm with α -selection corresponds to the underlying dynamical system model for the best or α -individual $b = 0$.

The fixed points of the intrinsic system model are obtained from the eigenvectors of M^* to eigenvalue $\lambda = 1$, i.e.

$$\boxed{\boldsymbol{\omega} = M^* \cdot \boldsymbol{\omega}}. \quad (1.22)$$

The fixed points of the heuristic function \mathcal{G} of the dynamical system model follow from the permutation $\sigma_b \boldsymbol{\omega}$ for a given α -individual b . For the fixed point analysis of the dynamical system model it therefore suffices to analyze the intrinsic system model shown in Fig. 1.5. To this end the WALSH transform of both sides of the equation $\mathbf{q} = M^* \cdot \mathbf{p}$ is taken yielding $\hat{\mathbf{q}} = W \cdot \mathbf{q} = W \cdot M^* \cdot W \cdot W \cdot \mathbf{p} = M^{*\wedge} \cdot \hat{\mathbf{p}}$. For an eigenvector \mathbf{v} with eigenvalue λ it follows $M^* \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$ and equivalently $M^{*\wedge} \cdot \hat{\mathbf{v}} = \lambda \cdot \hat{\mathbf{v}}$, i.e. the matrix M^* and its WALSH transform $M^{*\wedge}$ have the same eigenvalues with eigenvectors which are also related by the WALSH transform.

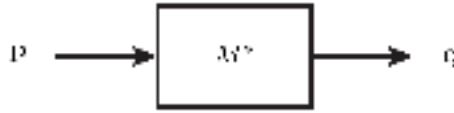


Fig. 1.5. Intrinsic system model of the genetic algorithm with α -selection [5]

For crossover and mutation the WALSH transform of the mixing matrix fulfills $\hat{M}_{i,j} \propto [i \otimes j = 0]$, i.e. \hat{M} is separative. $M^{*\wedge} = M^{*\wedge\wedge}$ is a lower

triangular matrix the spectrum of which is given by the first column of \hat{M} [13]. Since the spectrum of M^* and its WALSH transform $M^{*\wedge}$ are the same this yields the eigenvalues

$$\lambda_i = (M^{*\wedge})_{i,i} = \hat{M}_{0,i}. \quad (1.23)$$

For crossover and bitwise mutation the eigenvalues are given by

$$\lambda_i = \frac{(1-2\mu)^{1^T i}}{2} \sum_{k \in \Omega_T} (\chi_k + \chi_{k \oplus i}). \quad (1.24)$$

Because of $\lambda_0 = 1$ and $0 \leq \lambda_i \leq 1/2 - \mu < 1/2$ for $1 \leq i \leq n-1$ there exists a single eigenvector ω which is a unique fixed point of the intrinsic system model. For uniform crossover the eigenvalues are obtained from

$$\lambda_i = (1-2\mu)^{1^T i} \cdot \left(\chi \cdot 2^{-1^T i} + \frac{1-\chi}{2} \right) \quad (1.25)$$

for $1 \leq i \leq n-1$. The fixed points of the heuristic function \mathcal{G} of the simple genetic algorithm with α -selection are obtained from the permutation $\sigma_b \omega$ for a given α -individual b . According to the *fixed point hypothesis* the population will stay near this fixed point $\sigma_b \omega$ and converge to a new fixed point if a better α -individual b is found.

The unique fixed point ω of the intrinsic system model can be determined explicitly with the help of the WALSH transform. Due to the relation $\hat{\omega} = M^{*\wedge} \cdot \hat{\omega}$ and the lower triangular matrix $M^{*\wedge}$ the WALSH transform of the fixed point can be recursively calculated according to

$$\hat{\omega}_i = \frac{1}{1 - \hat{M}_{0,i}} \cdot \sum_{j=0}^{i-1} \hat{M}_{i \oplus j, j} \cdot \hat{\omega}_j \quad (1.26)$$

for $1 \leq i \leq n-1$ starting with $\hat{\omega}_0 = n^{-1/2}$ which ensures $\sum_{i \in \Omega} \omega_i = 1$. The unique fixed point ω is then obtained via the inverse WALSH transform $\omega = W \cdot \hat{\omega}$.

Under the assumption of the *fixed point hypothesis* for the intrinsic system model the permuted population vector $\sigma_b \mathbf{p}$ will stay near this fixed point ω . The population is therefore approximately sampled from the search space Ω according to the probability distribution $\sigma_b \omega$ with time-independent fixed point ω and α -individual b .

1.5 Experimental Results

In this section the ONEMAX problem with fitness function

$$f_i = \mathbf{1}^T i \quad (1.27)$$

is considered, i.e. f_i denotes the number of 1's in the binary representation of $i \in \Omega$. A simple genetic algorithm with α -selection using uniform crossover, bitwise mutation and random initial population is used with the strategy parameters listed in Tab. 1.1.

Table 1.1. Strategy parameters for the simple genetic algorithm with α -selection

ℓ	$n = 2^\ell$	χ	μ	r
10	1024	1.0	0.1	100

The intrinsic system model of the simple genetic algorithm with α -selection is defined by the twist of the mixing matrix M^* which leads to the unique fixed point ω shown in Fig. 1.6. The largest fraction of the population according to the largest component ω_i of the fixed point ω occurs at the individual $i = 0$. This is in line with the observation that the intrinsic system model of the simple genetic algorithm with α -selection corresponds to the underlying dynamical system model for the best or α -individual $b = 0$.

According to the *fixed point hypothesis* the permuted population vector $\sigma_{b(t)} \mathbf{p}(t)$ will stay near this fixed point ω . The corresponding distance between the permuted population vector $\sigma_{b(t)} \mathbf{p}(t)$ in generation t to the fixed point ω can be measured by the EUCLIDEAN distance which is defined by

$$\|\sigma_{b(t)} \mathbf{p}(t) - \omega\| = \sqrt{\sum_{i \in \Omega} \left((\sigma_{b(t)} \mathbf{p}(t))_i - \omega_i \right)^2}. \quad (1.28)$$

In Fig. 1.7 the EUCLIDEAN distance $\|\sigma_{b(t)} \mathbf{p}(t) - \omega\|$ is shown for one simulation run of the simple genetic algorithm with α -selection and strategy parameters as in Tab. 1.1 for 100 generations. The permuted population vector $\sigma_{b(t)} \mathbf{p}(t)$ stays close to the unique fixed point ω of the intrinsic system model of the simple genetic algorithm with α -selection. There is a close match between the theoretical prediction and the

experimental result, thereby confirming the *fixed point hypothesis*.

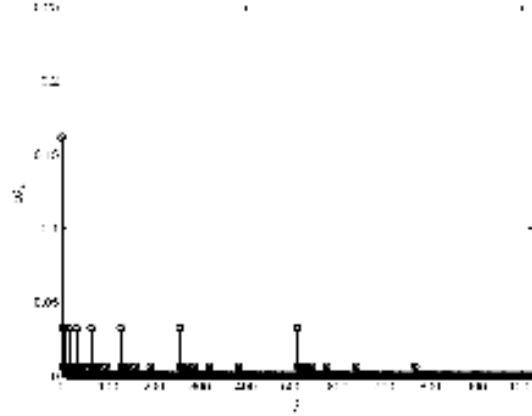


Fig. 1.6. Fixed point ω of the intrinsic system model of the simple genetic algorithm with α -selection

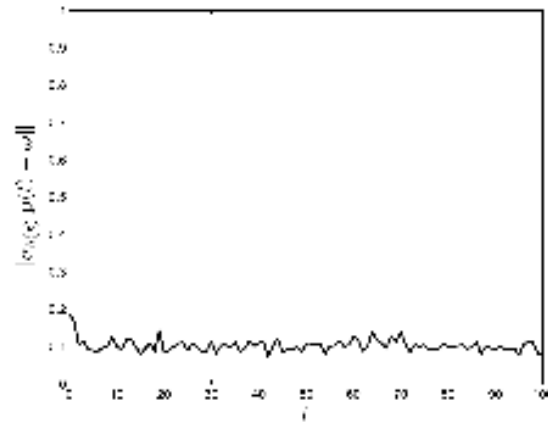


Fig. 1.7. EUCLIDEAN distance $\|\sigma_{b(t)}\mathbf{p}(t) - \omega\|$ over generation t for the simple genetic algorithm with α -selection

1.6 Conclusion

The intrinsic system model for the simple genetic algorithm with α -selection simplifies the analysis of the dynamical system model of

genetic algorithms. It is defined by the mixing matrix M and enables the explicit derivation of the unique fixed point ω . The simplifications are gained because the fitness function f is hidden from the mathematical formulation by making use of the α -individual b . Since b enters the dynamical system model via a permutation σ_b according to the system matrix $\sigma_b M^* \sigma_b$ the intrinsic system model can be formulated with the help of the twist of the mixing matrix M^* . It has to be observed, however, that the permutation σ_b depends on the α -individual b and therefore on the population vector \mathbf{p} .

The intrinsic system model provides a means to analyze the genetic algorithm's exploitation and exploration of the search space Ω irrespective of the fitness function f . This model is compatible with the equivalence relation imposed by schemata as shown in [5] by explicitly deriving the coarse-grained system model for a given schemata family. Experimental results for the simple genetic algorithm with α -selection, uniform crossover and bitwise mutation presented in this paper show a close agreement to the theoretical predictions obtained from the intrinsic system model.

References

1. BAJUELOS, A.; CANALES, S.; HERNÁNDEZ, G.; MARTINS, A.M. (2008) *Optimizing the Minimum Vertex Guard Set on Simple Polygons via a Genetic Algorithm*. In: WSEAS Transactions on Information Science and Applications. Iss. 11, Vol. 5, pp. 1584-1596
2. HOLLAND, J.H. (1992) *Adaptation in Natural and Artificial Systems – An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge: First MIT Press Edition
3. KUMSAWAT, P.; ATTAKITMONGCOL, K.; SRIKAEW, A. (2009) *An Optimal Robust Digital Image Watermarking Based on Genetic Algorithms in Multiwavelet Domain*. In: WSEAS Transactions on Signal Processing. Iss. 1, Vol. 5, pp. 42-51
4. NEUBAUER, A. (2009) *Simple Genetic Algorithm with Generalised α^* -Selection*. In: Proceedings of the International Joint Conference on Computational Intelligence – IJCCI 2009. Madeira, Portugal: pp. 204-209
5. NEUBAUER, A. (2008) *Intrinsic System Model of the Genetic Algorithm with α -Selection*. In: Rudolph, G.; Jansen, T.; Lucas, S.; Poloni, C.; Beume, N. (Eds.): Parallel Problem Solving from Nature PPSN X. Berlin: Springer-Verlag, pp. 940-949

6. NEUBAUER, A. (2008) *Theory of the Simple Genetic Algorithm with α -Selection*. In: Proceedings of the 10th Annual Genetic and Evolutionary Computation Conference – GECCO 2008. Atlanta, USA: pp. 1009-1016
7. NEUBAUER, A. (2008) *Theory of Genetic Algorithms with α -Selection*. In: Proceedings of the 1st IAPR Workshop on Cognitive Information Processing – CIP 2008. Santorini, Greece: pp. 137-141
8. PAN, C.-H. (2009) *An Efficient MIMO Detection Algorithm Employed in Imperfect Noise Estimation*. In: WSEAS Transactions on Communications. Iss. 8, Vol. 8, pp. 941-958
9. POPESCU, M.-C.; POPESCU, L.; MASTORAKIS, N. (2009) *Applications of Genetic Algorithms*. In: WSEAS Transactions on Information Science and Applications. Iss. 11, Vol. 6, pp. 1782-1791
10. REEVES, C.R.; ROWE, J.E. (2003) *Genetic Algorithms – Principles and Perspectives, A Guide to GA Theory*. Boston: Kluwer Academic Publishers
11. REIS, C.; TENREIRO MACHADO, K.A. (2009) *Crossing Genetic and Swarm Intelligence Algorithms to Generate Logic Circuits*. In: WSEAS Transactions on Computers. Iss. 9, Vol. 8, pp. 1419-1428
12. VOSE, M.D.; WRIGHT, A.H. (1998) *The Simple Genetic Algorithm and the Walsh Transform – Part I Theory, Part II The Inverse*. In: Evolutionary Computation. Vol. 6, No. 3, pp. 253-273, 275-289
13. VOSE, M.D. (1999) *The Simple Genetic Algorithm – Foundations and Theory*. Cambridge: MIT Press
14. VOSE, M.D. (1999) *Random Heuristic Search*. In: Theoretical Computer Science. Vol. 229, No. 1-2, pp. 103-142

Chapter 2

Safety of Masonry Arch Bridges: A Quick Scan Method

A. Audenaert,^{1,2} G. Reniers,^{2,3} W. Dullaert,⁴ H. Peremans²

¹University College Antwerp, Artesis, Paardenmarkt 92, B-2000 Antwerp, Belgium, E-mail: amaryllis.audenaert@artesis.be

²Department of environment, technology and technology management, University of Antwerp, Prinsstraat 13, B-2000 Antwerpen, Belgium, E-mail: genserik.reniers@ua.ac.be; herbert.peremans@ua.ac.be

³Centre for Sustainable Entrepreneurship, Hogeschool-Universiteit Brussel, Stormstraat 2, B-1000 Brussels, Belgium

⁴Institute of Transport and Maritime Management Antwerp, University of Antwerp, B-2000 Antwerpen, Belgium, E-mail: wout.dullaert@ua.ac.be

Abstract. Using a 2D decision support system based on a recent analytical model, this paper is the first to demonstrate that bridge specific characteristics require a detailed analysis to determine the collapse load. Because of the complexity of finite element modeling and the computation time requirements of 3D finite elements models for a large number of load positions, it is regularly assumed that collapse positions are situated at the crown of the arc. Empirical testing in this paper clearly demonstrates that this is not the case. The here developed approach allows for a quick scan of any type of masonry bridge, thus providing the necessary information on the critical range of load positions to be analyzed in more detail by a 3D modeling approach.

Keywords. arches, collapse load, equilibrium, safety

2.1 Introduction

The analysis of load-bearing unreinforced masonry structures such as arches, vaults and buttresses has become the subject of renewed academic

interest. The growing interest in the preservation of historical buildings and structures gives researchers the incentive, financially as well as socially, to develop methods of analysis for these structures. Moreover, arch bridges, which were designed in the days of horses and carriages are also required to function under the 21st century loadings. Nowadays, these structures are subject to heavy freight transport which they were originally not constructed for. The latter remark justifies the approach of studying the collapse behavior of these historical structures as accurately as possible. An extensive literature review on the topic is provided by Boothby [1].

The rigid block theory is the basic model for understanding the fundamental behavior of masonry arches [2, 3, 4]. This theory simplifies the masonry arch structure as a collection of rigid blocks and calculates its stability using the principle of virtual work, which is an alternative way of expressing the equations of motion and equilibrium. This principle is often used in mechanics of structures and is also the basis of FEM-models [5]. Eventually the rigid-block methodology gives rise to an upper bound for the collapse load for a given load position and thus offers a first quick insight into the arch behavior. The collapse load is an objective measure to determine the bearing capacity of a structure.

A more recent research technique uses finite element analysis. The models range from 1-dimensional [6, 7], over 2-dimensional [8, 9], up to fully 3-dimensional models [10, 11] for understanding three-dimensional effects. These 3D-models do require a precise knowledge of parameters which are usually not well-known in practice. Moreover, finite element programs are often computationally expensive and require a high level of modeling expertise.

A major shortcoming of both the rudimentary rigid block method and the more accurate 3-dimensional finite element analysis is their inability to carry out detailed sensitivity analysis. Software based on these models can only assess an arch bridge for specific load conditions and parameter settings. Neither the influence of material properties, nor the influence of different loadings can be evaluated easily. Especially for determining the bearing load on arch bridges, such a parameter analysis might prove to be very important.

To alleviate these modeling shortcomings, this paper embeds an analytical model for determining the collapse load for a single loading position in a decision support tool. The resulting software is capable of identifying the specific position (of the load) which allows the smallest collapse load, thus governing the safety of the arch.

2.2 An analytical model for the bearing capacity of an arch bridge

To determine the bearing capacity of an arch bridge, we first determine the collapse load for all possible load positions on the bridge using an analytical model.

The analytical model is based on the three fundamental equilibrium equations: horizontal equilibrium, vertical equilibrium and equilibrium of moments. The geometry of an arch is defined by an angle θ , a radius $r(\theta)$ and a height $b(\theta)$ as shown in Fig.1.

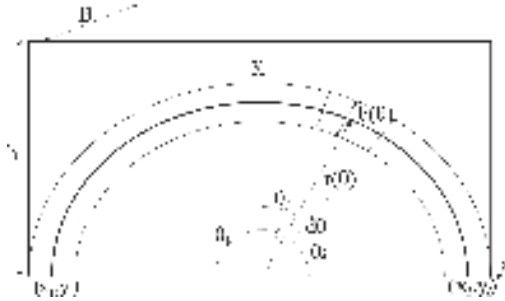


Fig 1. Definition of the geometric parameters of an arch in function of θ

All these geometrical parameters are function of the angle μ , capable of modeling any shape of arch barrel. Audenaert et al. [13] show how the equilibrium equations can be transformed to a set of ordinary differential equations (1), (2) and (3) to calculate normal forces $N(\theta)$, shear forces $V(\theta)$ and bending moments $M(\theta)$.

$$-V' - N - \gamma r^2 \eta \cos \theta + p_r - \sum_i P_i \delta(\theta - \alpha_i) \cos \theta + \sum_j H_j \delta(\theta - \alpha_j) \sin \theta = 0 \quad (1)$$

$$N' - V + \gamma r^2 \eta \sin \theta + p_\theta + \sum_i P_i \delta(\theta - \alpha_i) \sin \theta + \sum_j H_j \delta(\theta - \alpha_j) \cos \theta = 0 \quad (2)$$

$$M' + Nr' + Vr + \gamma r^3 \frac{\eta^3}{12} \sin \theta + p_\theta \frac{b}{2} + \sum_i P_i \delta(\theta - \alpha_i) |r(\alpha_i) \sin \alpha_i - r(\theta) \sin \theta| + \sum_j H_j \delta(\theta - \alpha_j) |r(\alpha_j) \cos \alpha_j - r(\theta) \cos \theta| \quad (3)$$

In equations (1), (2) and (3) a prime denotes the derivative with respect to θ ; p_r = the radial distributed force; p_θ = the tangential distributed force; α_i = the position of the vertical point load P_i ; α_j = the position of the horizontal point load H_j ; γ = the specific weight of the arch and $\eta = b/r$.

The general solution of this set of differential equations includes three constants. To find the unique solution for the internal forces and the internal moments, additional constraints need to be imposed. Boundary conditions on the horizontal u , vertical v and angular deflections φ at the abutments were introduced, based on the so-called Bresse equations [12]. This procedure yields the deflections at the right fixed support (u_2 ; v_2 ; φ_2), given their values at the left fixed support (u_1 ; v_1 ; φ_1).

with φ = the rotation of the elastic line; u = the horizontal deflection; v = the vertical deflection; A = the area of the cross-section; I = the rotational inertia of the cross-section; E = the modulus of elasticity; x = the horizontal position coordinate; y = the vertical position coordinate; the subscripts 1 and 2 denote the left and right side supports of the arch bridge respectively, see Fig. 1.

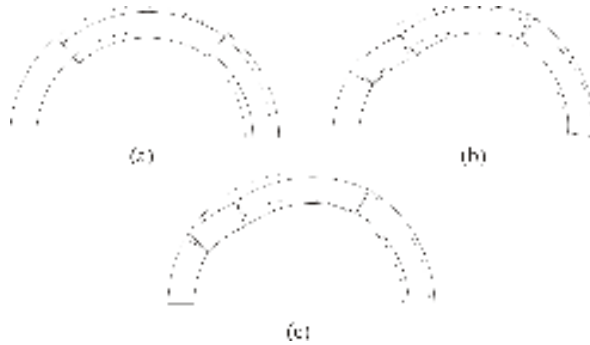


Fig 2. Collapse modes

The sign conventions for the horizontal and vertical deflections u and v are taken to be the same as for the x-axis and y-axis, the rotation φ is taken to be positive for clockwise rotations. Both axes are defined in Fig. 1. If the deflections at both supports are specified, these equations can be used to calculate the internal forces $N(\theta)$, $V(\theta)$ and $M(\theta)$.

To calculate the stability of an arch, a failure criterion is added to the model. According to [2], an arch bridge can collapse as a result of three possible collapse mechanisms: a shear mechanism (Fig. 2(a)), a hinge-mechanism (Fig. 2(b)) and a combined shear-hinge mechanism (Fig. 2(c)).

Experiments by Hendry et al. [15] and by others show that, due to most arches being well-buttressed, the hinge-mechanism (Fig. 3) can be considered as by far the most likely collapse mechanism for arches. Therefore this paper focuses on this mechanism.



Fig 3. Collapse of the Bridgemill arch bridge. (Source: http://archive.niees.ac.uk/talks/dem/nenad_bicanic.ppt)

The formation of hinges and the complementary propagation of cracks needs we assume that the first crack will appear at the position for which a defined tensile strength σ_t is reached. When the compressive strength σ_d is reached, the material behaves perfectly plastic finally resulting in a hinge. How these material properties are included into the analytical model is thoroughly explained in [12]. Many homogenisation technics exist to assess the material behavior of masonry as shown by [17], [18] and [19]. The solution of the differential equations can be found in [13].

2.3 Quick scan analysis

Current research predicts arch bridge behavior under a given load. However, such studies do not offer information about the bearing capacity of an arch bridge. The analytical model from Section 2 allows for an efficient sensitivity analysis on loading positions to determine the corresponding collapse load. The minimum of these collapse loads determines the ultimate allowable load for a given bridge geometry. In this section we will examine whether the critical load position differs for bridges with different geometry.

Bridges have primarily been built to resist vertical loads. Thus, in the case of a vertical concentrated load, the safety assessment of the bridge

should be performed for all possible load positions α . The position giving rise to the smallest value for the collapse load determines the bearing capacity of the arch. Figure 4 visualizes this bearing capacity procedure. Hence, we perform a search of the load position and compute the collapse load. The advantage of this approach is that it quickly gives the ultimate overview of the arch bridge's behavior.

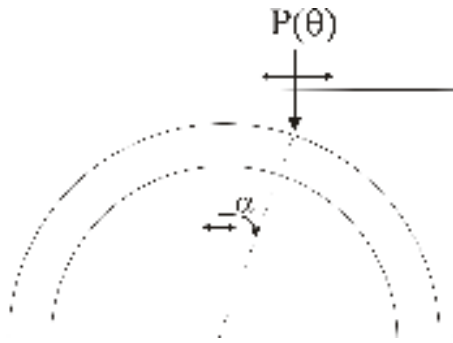


Fig 4. Algorithm to determine the bearing capacity of the arch in function of the load position.

To illustrate the necessity and workings of the quick scan methodology, consider the following set of semi circular masonry arch bridge features. All bridges are assumed to have the same material properties, infill, and radius of the outer curve of the arch (extrados r_e) (see Fig. 5).

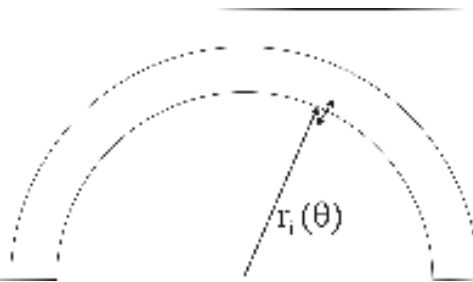


Fig 5. Algorithm to determine the variation in critical load position in function of the arch geometry

The bridges with unit width only differ in their radius of the inner curve of the arch, the so-called intrados r_i . The height of the infill h is 2 meters and the specific weight of the infill γ_2 is 21600 N/m^3 corresponding to a traditional sandbased infill. The masonry barrel is assumed to have a

specific weight of 21000 N/m^3 , an elasticity coefficient of 5 GPa, compressive strength of -8 MPa and tensile strength of zero reflecting the masonry's inability to resist tensile. The radius of the intrados r_i varies between 1.4 and 1.7 meters implying a variation in barrel thickness in the range of $[0.5-0.2]$ given the fixed radius of the extrados $r_e = 1.9\text{m}$. The position of the point load P varies from $-1 = -57.30^\circ$ to $1 = 57.30^\circ$.

In Table 1 the numerical results for the load position giving rise to the smallest collapse load end the corresponding collapse load are listed for some arch geometries.

$r_i [\text{m}]$	$P_{\text{collapse}} [\text{kN}]$	$\alpha [\text{rad}]$	$\alpha [^\circ]$
1.4	56.51	0.25	14.32
1.5	30.80	0.2	11.46
1.6	14.94	0.15	8.59
1.7	5.22	0.1	5.73

Table 1. Collapse loads and most dangerous load positions for different arch geometries.

In Fig.6 the collapse load is shown in function of all position of the load, α , and in function of the radius of the intrados, r_i . The results support common wisdom that thicker arch barrels can resist higher collapse loads.

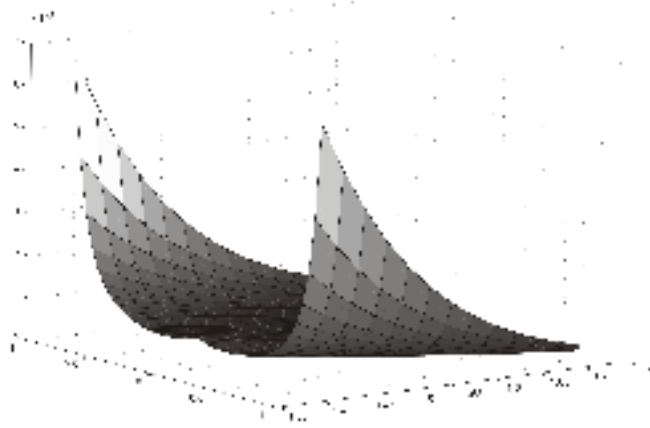


Fig 6. Collapse load in function of α and r_i

Also the higher the barrel of the arch, the more the weakest point of the arch (i.e., the location α corresponding with the smallest collapse load) differs from the crown of the arch.

For the biggest and the smallest arch the collapse load in function of the load position are shown respectively in Fig. 8 and in Fig. 7.

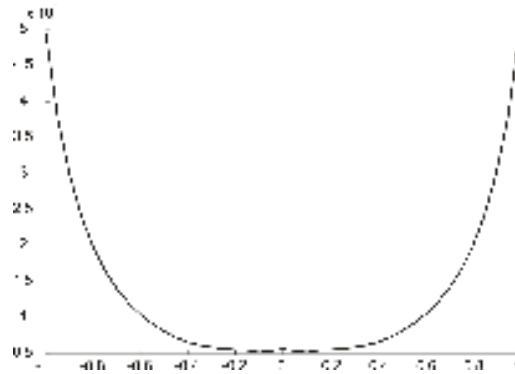


Fig 7. Collapse load in function of α for the smallest arch, $r_i = 1.7$

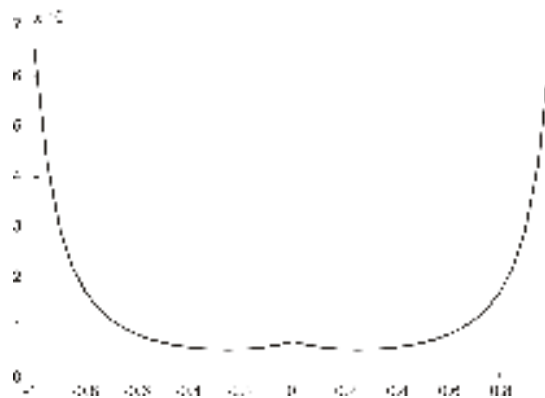


Fig 8. Collapse load in function of α for the thickest arch, $r_i = 1.4$

In Table 2 the overall collapse load is compared to the collapse load corresponding to the crown of the arch ($\alpha = 0$).

$P(r_i; \alpha)$ [kN]		α [rad]					%error
		0	0.1	0.15	0.2	0.25	
r_i [m]	1.4	73.44				56.51	29.96
	1.5	35.38			30.80		14.87
	1.6	16.08		14.94			7.63
	1.7	5.67	5.22				8.62

Table 2. Influence of α on the collapse load for different arch geometries.

For each arch geometry only these two values are listed in the table. Also the corresponding %-error is calculated. Especially for a thick arch

geometry ($r_i = 1.4$) the overall collapse load for a load positioned at $\alpha = 0.25$ differs significantly, approximately 30%, from the collapse load obtained for $\alpha = 0$. For smaller arch barrels the error-percentage diminishes, but remains larger than 8%. These results indicate that for every bridge one should conduct a sensitivity analysis based on an investigation of the weakest position of the arch.

The latter findings are in accordance with the findings of Brencich et al. [16] demonstrating that $\alpha = 0$ is not the weakest point of the bridge. As Brencich used a FEM-program for this calculations he was only able to evaluate a limited number of load positions. The computation time requirements of his FEM approach therefore excludes an extensive sensitivity analysis required for assessing the safety of arch bridges. Our method is capable of doing a fast and very accurate parameter analysis in function of the load position.

2.4 Conclusions

Determining a “safe load carrying capacity” of arch bridges continues to challenge civil servants and researchers. The proposed method of strength assessment described in this paper advances the state-of-the-art in two-dimensional modeling of masonry arch bridges. Our two-dimensional assessment algorithm serves as a conservative, computationally efficient approach to tackle safety of such masonry arch bridges.

Previous models from the literature assess arches under a given specific load condition. Because of the complexity of finite element modeling and the computation time requirements of 3D finite elements models for a large number of load positions, it is regularly assumed that collapse positions are situated at the crown of the arch.

Our methodology investigates the safety of arches under different conditions using sensitivity analysis. Computational testing shows that the weakest point of an arch is different for every geometry, which means that no assumptions about this matter can be made beforehand. The critical position differs from the crown of the arch and should therefore be determined for every single arch geometry. A small variation in geometry or material properties is shown to have a significant influence on the bearing capacity of the arch.

The approach developed in this paper allows for a quick scan of any type of masonry bridge, thus providing the necessary information on the critical range of load positions to be analyzed in more detail by a 3D modeling approach.

References

1. Boothby T (2001) Analysis of masonry arches and vaults, *Progress in Structural Engineering and materials*, 3: 246-256
2. Boothby T (1995) Collapse modes of masonry arch bridges, *Journal of the british masonry society*, 9(2): 62-69
3. Gilbert M and Melbourne C (1994) Rigid-block analysis to masonry arches, *Structural Engineering*, 72: 356-361
4. Hughes TG and Blackler MJ (1995) A review of the UK masonry assessment methods, *Proceedings of the Institution of Civil Engineers*, 110: 373-382
5. Wunderlich W (2002) *Mechanics of Structures: Variational and Computational Methods*, CRC, ISBN0849307007
6. Choo BS and Coutie MG and Gong NG (1991) Finite-element analysis of masonry arch bridges using tapered elements, *Proceedings of the Institution of Civil Engineers*, 91: 755-770
7. Mollins C and Roca P (1998) Capacity of masonry arches and spatial frames, *Journal of structural engineering*, 124: 653-663
8. Boothby TE, Domalik DE and Dalal VA (1998), Service load response of masonry arch bridges, *Journal of structural engineering*, 124: 17-23
9. Ng KH, Fairfield CA and Sibbad A (1999) Finite-element analysis of masonry arch bridges, *Proceedings of the Institution of Civil Engineers: Structures and Buildings*, 134: 119-127
10. Fanning PJ and Boothby TE and Roberts BJ (2001) Longitudinal and transverse effects in masonry arch assessment, *Construction and Building Materials*, 15: 51-60
11. Fanning PJ, Sobczak L, Boothby TE and Salomoni V (2005) Load Testing and Model Simulations for a Stone Arch Bridge, *Bridge Structures*, 1(4), ISSN 1573-2487
12. Audenaert A, Peremans H and De Wilde WP (2004) Static determination of the internal forces and displacements in arch bridges, *The Masonry Society Journal*, 22(1): 97-109
13. Audenaert A, Peremans H and Reniers G (2007) An analytical model to determine the ultimate load on masonry bridges, *Journal of Engineering Mathematics*, 59:3: 323-336
14. Audenaert A, Fanning P, Sobczak L and Peremans H (2008) 2D analysis of masonry arch bridges using an elasto-plastic material model, *Engineering structures: the journal of earthquake, wind and ocean engineering*, 30(3): 845-855

15. Hendry AW, Davies SR and Royles R (1985) Test on a Stone Masonry Arch at Bridgemill-Girvan. Transport and Road Research Lab, Contractors Report 7 United Kingdom
16. Brencich A, De Francesco U and Gambarotta L (2001) Elastic no tensile resistant-plastic analysis of masonry arch bridges as an extension of Castigliano's method, 9th Canadian masonry symposium, Fredericton (Canada)
17. Cecchi A (2006) CFRP reinforced masonry walls: analytical and numerical homogenized modes, WSEAS Transactions on Applied and Theoretical Mechanics, 1(1): 7-15
18. Cecchi A and Barbierri A (2008) Homogenisation procedure to evaluate the effectiveness of masonry strengthening by CFRP repointing technique, WSEAS Transactions on Applied and Theoretical Mechanics, 3(1):12-27
19. Milani G (2007) A simple equilibrated homogenization model for the limit analysis of masonry structures, WSEAS Transactions on Applied and Theoretical Mechanics, 2 (5):119-125

Chapter 3

Moving Surface Reconstruction by Self-Organization

M. V. Bessmeltsev

Mechanics and Mathematic Department, Novosibirsk State University,
Pirogova st.,2, Novosibirsk, 630090, Russia, bmpix@mail.ru

Abstract. In this paper we introduce a novel self-organizing method of moving surface reconstruction from the data obtained from measuring real-world objects. The core of the approach is using Kohonen's Self Organizing Maps model. We propose a modification of the model enforcing input data approximation and time-space smoothing. Due to stochastic nature of the proposed method, in many cases it is not necessary to employ additional denoising or data filtering. Underlying self-organizing principles make the technique human-free, efficient and easy to parallelize. The method was implemented using AITricks GeomBox® package.

Keywords. Animation, surface reconstruction, self-organization.

3.1 Introduction

Modeling of realistic three dimensional objects is one of the fundamental problems in geometry processing. Its applications include 3D computer graphics and movie making, industrial computer-aided design, numerical modeling of physical processes on real objects, acquiring CAD model of a part with lost CAD data, etc. [8]

Many approaches of 3D shape acquisition by measuring were developed; the most common of those are laser scanning [1], structured light techniques [2], and passive multi-view stereo [7]. The result of scanning is usually a point cloud or a height map. In this paper, we focus on one of the problems of reverse engineering of CAD models: moving surface reconstruction from point clouds.

Currently, numbers of methods exist that allow to reconstruct surfaces of static objects from unorganized set of points, particularly, methods of surface reconstruction of unknown topology, e.g. [8], methods of surface reconstruction via deformation [3], methods of topology estimation [5], etc.

Recently, several approaches have been proposed to extend measuring techniques to capture animated scenes in real-time. That evidently opens up a large variety of interesting new applications, such as creating special effects for movies or creating content for interactive applications and games. Besides, new approaches for moving surface reconstruction appear, such as [17], [16], [11]. However, despite high research interest in the subject, currently available technologies of automatic moving surface reconstruction impose some significant restrictions. First of all, point clouds often contain remarkable level of noise, the input data suffer from oversampling or undersampling. Second, it is not evident how to evaluate the topology of an object being scanned and how it is changing in time. Third, in attempt to regenerate a moving surface using a static surface reconstruction method we may obtain not a moving mesh, but a sequence of meshes of varying connectivity. And finally, the methods of surface reconstruction are often difficult to parallelize.

Those are the problems this paper addresses. The main contribution of the paper is a new method EDSOM (Expanding Dynamic SOM) which allows us to automatically reconstruct a moving surface with non-uniform mesh nodes density, fixed topology and fixed number of nodes from unorganized set of points. Due to underlying self-organizing principles and stochastic nature, the new method is resistant to noise, can efficiently process oversampled or undersampled data and is easy to parallelize.

The core of the proposed method is Kohonen's Self-Organizing Maps (SOM), a well-known neural network model [9]. In order to properly adjust the model to be applicable to the problem of moving surface reconstruction, the parameters of the model were fine-tuned. Also, some essential modifications are proposed which allow us to refine the surface approximation and make the quality of the final animation better. Particularly, we propose a stochastic technique which refines the mesh and decreases the gaps between the surface and the point cloud, moving the nodes in surface normal direction towards the point cloud and then additionally refining it with geodesic SOM [10]. To smooth trajectories, we use a stochastic K-means-like technique proposed in [4], which allows us to deform the mesh only in the areas of real surface deformations in time.

3.2 Related Work

We can divide the related work into two parts: moving surface reconstruction and surface reconstruction using self-organizing maps.

Moving surface reconstruction. Though many approaches were proposed to reconstruct surfaces from unorganized set of clouds, starting with [8], only a few techniques were developed to extend surface reconstruction to animated point clouds. Nevertheless, a full survey on the area is beyond the scope of the paper, and we will mention only some noticeable techniques.

A close strategy to ours is to generate moving meshes by fitting template meshes to the data [11], [16], [7]. The limitation of those methods is that those methods are able to work only with low-noise or noise-free data and sufficient sampling density. Those limitations are overcome by Wand et al. [17] at the cost of significant performance issues, though producing impressive results. Moreover, most of those methods cannot be efficiently parallelized.

Surface reconstruction using Self-Organizing Maps. One of the pioneer papers in the area is [18]. In this paper, static surface reconstruction using Kohonen's SOM is considered. The algorithm proposed in that paper yields reasonable though non-optimized meshes: the number of triangles in the mesh tends to the number of points, what is unacceptable in most applications. Moreover, even with proposed technique called edge swap, the method possibly can not be applied to large model with small features. In this paper we solve this problem by employing a geodesic metric.

The idea of surface deformation by SOM is not new itself [6], [15]. Though the key difference of our approach that the initial mesh for deformation is irregular, unstructured and has non-uniform nodes density when most of the papers concerning SOM deal with regular uniform grids.

Finally, in [4], moving meshes are generated using Self-Organizing Maps. As stated in the paper, the technique is limited to surface without holes in it, the input data is a height map while in our paper we deal with unstructured sets of points. Though, our paper may be considered as a significant extension of that method.

3.3 The EDSOM reconstruction pipeline

3.3.1 Scheme

The proposed method of moving surface reconstruction belongs to the class of methods of moving mesh generation via deformation. The main goal of the deformation at each time step is to fit the template mesh to current point cloud, preserving structure and topology of the template mesh.

The initial mesh may be generated by any suitable method of static surface reconstruction, e.g. mentioned above. It has to be noted that the only condition we require is that the initial mesh has the same topology as the real object, because we use the initial mesh as a template one, generating moving mesh by its deformation.

Our moving surface reconstruction pipeline at each time step consists of three major components: estimation of changes in the point cloud in time (Section 4), deformation of the current mesh in the areas of point cloud changes with Self-Organizing Maps algorithm (Section 5), and the so-called Expanding - the final refinement of the mesh (Section 6).

3.4 Motion Tracking

The main idea of the technique is to investigate changes of sample points near each node by its Voronoi cell centroid variation in time. Following this technique, we calculate the distance between each node and the centroid of its Voronoi cell computed using the point cloud at the next time step.

This distance actually shows how much the object under consideration has changed in the node's neighborhood. The last step is to normalize this distance and perform Gaussian smoothing on it. The obtained factor $v(t, x)$ is called *the function of deformation area*:

$$v_c(t, q_i) = \begin{cases} w(t, q_i) / \max_{i \in w(t, q_i)} w(t, q_i), & \text{if } n_i(t) > 0 \\ 1, & \text{if } n_i(t) = 0. \end{cases} \quad (1.1)$$

where $w(t, q_i)$ is Euclidian distance between the centroids of Voronoi cells of i -th node at two adjacent time steps and n_i is the number of points in corresponding Voronoi cell. The obtained factor will be used as a mask for the deformation.

3.5 Deformation via SOM

The main goal of SOM deformation stage is to roughly fit the current mesh to a new point cloud, making the global structure of the mesh correspond the point cloud.

Let $U(t)$ be the surface of an animated object in a 3D Euclidean space with spatial coordinates $x \in \mathbb{R}_U^3$, where t is a discrete time step. A moving triangle mesh $M(t) = \{m_1(t), \dots, m_N(t)\}$ is to be generated on $U(t)$, where $m_i(t)$, $i = 1, \dots, N$ are moving mesh nodes. Let $Q = M(0)$ be a fixed template mesh. The moving mesh will be acquired via deformation of that template in time.

For simplicity, here we assume that the global topology of $U(t)$ is not changing in time (i.e. no large holes appear or disappear in time in the object under consideration). This is actually a weak requirement since most of contemporary applications imply 3D modeling of a single moving object, e.g. actor in movie making.

The sets of sample points $X(t)$ are assumed to lie on the unknown surface $U(t)$ or near it. We do not make assumptions about the level of noise in the input data since SOM deformation algorithm is stable even on data with high noise level. Since the process is stochastic, the influence of noise may be lower than in conventional algorithms of surface reconstruction.

Taking into account that we deform the mesh only in the areas of object variation (Stage 1), the SOM deformation algorithm is as follows:

Algorithm 1. Procedure Deform(S_0, S). For each $s = S_0, \dots, S$:

0. *Point selection.* Take a random point $y \in X(t)$ from the point cloud.
1. *Winner determination.* Calculate all the Euclidean distances between y and all the nodes $m_i^s(t)$ and choose the node $m_{\psi(t)}^s(t)$ which is the closest to y , i.e.

$$\|y - m_{\psi(t)}^s(t)\|_U \leq \|y - m_i^s(t)\|_U, \quad (1.2)$$

for all $i = 1, \dots, N$. The node $m_{q_\psi}^s(t)$ is called a *winner*.

2. *Node coordinates correction*. Adjust locations of the mesh nodes according to the following rule:

$$m_{i(t)}^{s+1} = m_{i(t)}^s + \theta_{q_\psi}^S(s, q_i)(y - m_i^s(t)), \quad (1.3)$$

$i = 1, \dots, N$, where $\theta_{q_\psi}^S(s, q_i) \in [0, 1]$ is a *learning rate*.

The learning rate controls the overall quality of the mesh and affects the preservation of mesh structure throughout the iterative process. According to [12] and our experiments, the learning rate has been thoroughly selected to provide acceptable mesh quality with reasonable computational speed, and looks like $\theta_{q_\psi}^S(s, q_j) = \delta(s)\eta_{q_\psi}^{S(s, q_j)}$, where $\delta(s) = s^{-0.2}\chi(s)$ is a

learning step, $\eta_{q_\psi}^{S(s, q_i)} = \zeta^{\|q_\psi - q_i\|_Q / r(s)^2}$ (ζ is close to zero, e.g. $\zeta = 0.001$) is a neighborhood function, $r(s) = r(S) + \chi(s)(r(1)0.05^{s/S} - r(S))s^{-0.25}$ is a learning radius, $\chi(s) = 1 - e^{5(s-S)/S}$.

The maximum number of iterations S is fixed beforehand proportional to N , e.g. $S = 10N$; $r(1)$ and $r(S)$ are values of a learning radius that are selected depending on the distances between nodes of the template mesh Q , $r(1) > r(S)$.

The learning step indicates the displacement magnitude the winner node receives, and the learning radius controls the radius of the neighborhood of the winner node that moves along with the winner. Obviously, there is always a compromise between low and high learning radius. When the radius is high, a large neighborhood of the winner node moves, therefore producing smooth, but poorly approximating the point cloud surface. On the other hand, low learning radius may yield fine approximation, though possibly making mesh non-smooth and even disordering its structure.

In our experiments we have chosen the starting value of radius greater than maximum Euclidian distance between nodes of the mesh:

$$r(1) = 3 \max_{i, j \in [1; N]} \|m_i - m_j\|_U, \text{ and final radius value } r(S) \text{ equal to an average}$$

edge length in the initial triangle mesh $M(0) = Q$.

Following Kohonen [9], the learning process of the SOM model can be divided into 2 parts: ordering stage and refinement stage. During the ordering stage, mesh nodes get roughly distributed on the surface of the input geometry, since nodes get large displacements.

During the refinement stage, mesh nodes' displacements are relatively low, and that makes the mesh correspond to the probability distribution and approximate surface more precisely. This stage affects the final overall quality of the mesh (smoothness, topology correctness, etc.).

To avoid regeneration of the mesh from scratch at each time step, only deforming it to fit a new point cloud, we skip the ordering stage, performing only last iterations of the SOM deformation process. In our experiments, we chose $S_0 = S/4$, as the beginning iteration of the refinement stage, therefore using only 75% of whole iterative process.

3.6 Expanding

In this section, we introduce Expanding technique which is aimed to refine the approximation of the mesh, obtained as the result of Stage 1. The goal of the technique is to eliminate a distance between the mesh and the point cloud or at least decrease it.

Formally, the problem we have left is expanding the mesh to fit the point cloud which is an optimization problem. Even though after stage 1 we have a good guess of the reconstructed mesh, in the paper we prefer to avoid explicit minimization of non-linear functions since it would lead to notorious efficiency and parallelization problems, especially in the case of oversampling.

It has to be noted that the mesh after Stage 1 of reconstruction approximates the geometry of the surface and the distance between the mesh and the point cloud is greater in the areas where the surface is complex, i.e. has larger surface curvature. Taking that into account, the proposed additional refinement technique goes as follows:

Step 1. Move all the nodes of the mesh in the normal direction. The displacement value is proportional to surface curvature in node's position:

$$m'_i(t) = m_i(t) + n_i(t)\xi(i,t), \quad (1.4)$$

where $n_i(t)$ is normal of the surface in the i -th node. Since normal is a differential entity, it is not defined at the vertices of a mesh, instead of it we use pseudo-normal [19] which is angle-weighted sum of all the triangles adjacent to the node. $\xi(i,t)$ is a factor which is proportional to surface curvature. In our experiments we have chosen $\xi(i,t)$ proportional to maximum angle between neighborhood triangles' normals.

Step 2. Perform last iterations of SOM deformation process with a special learning radius: instead of calculating Euclidian distance, we can obtain better refinement using geodesic distance over the triangle. It is vital to use the geodesic distance only in the end of iterative process since then the learning radius is low and we can easily optimize the mesh deformation without need to employ complex algorithms of real geodesic distance calculations. It is also important to underline that Step 1 is

essential: last iterations of SOM with geodesic distance would be unable to refine the mesh. Also, due to computation reasons, it is inefficient to use whole SOM with geodesic distance.

In our experiments, instead of exact geodesic distance, we used a graph-based Dijkstra algorithm for shortest paths.

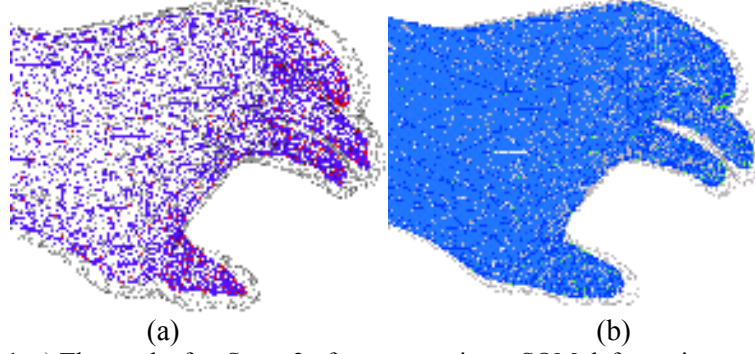


Fig.1. a) The mesh after Stage 2 of reconstruction – SOM deformation.
b) Stage 3 refines the mesh even in the areas of complex geometry.

3.7 Experiments and Results

The proposed method was implemented in C++ using AITricks GeomBox® package [20] and tested on synthetic data - random points generated on the surface of a moving gorilla model. The initial mesh is the triangulation of the gorilla at the first frame. As shown in (Fig. 2), the mesh successfully adapts to the animated point cloud, the mesh quality is satisfactory. Moreover, the surface has holes in it (mouth, eyes) and the method successfully tracks their motion. The parameters of the proposed method were: $S = 20 \cdot N$, $S_0 = S/4$ for SOM deformation, $S = 20 \cdot N$, $r(1) = \max_{i,j \in [1;N]} \|m_i - m_j\|_U / 8$, $r(S) = 3.0$, $S_0 = S/2$ for geodesic SOM. For

Stage 3 we used $\xi(i, t) = 1.5\sqrt{\rho(i, t)}$, where $\rho(i, t)$ is the maximum angle between the i -th node's normal and normals of its neighbors.

Computation time on a Pentium 4 3.4GHz/2GB RAM for mesh with 23152 triangles and animated point cloud consisting of 20 frames, 155000 points per frame is 420 seconds per frame. The proposed method has been also parallelized using MPI library and the efficiency of parallelization is above 95%.



Fig. 2. The reconstructed moving surface at $t = 2$ and $t = 13$. The mesh consists of 23152 triangles.

3.8 Conclusions and future work

We have presented a novel self-organizing method of moving surface reconstruction from an animated point cloud based on self-organization. The method is human free, resistant to noise or oversampling, topologically correct and efficiently parallelizable. It is able to reconstruct complex geometry with holes from noisy animated point cloud. Instead of regenerating the surface at each time step, the method only deforms the existing mesh in order to fit a new point cloud. The proposed method was successfully parallelized.

The distribution we use for point selection step in Algorithm 1 influences the final nodes density. To preserve the initial mesh nodes density, we choose uniformly among all the sampled points. In the other case, to make the animated mesh track the model features appearing in time, we generate a random point from the point cloud the more often the more complex the point cloud in that area is. One of the possible future directions is to carry out experiments with non-uniform density.

One of the limitations of method is that the gap between the reconstructed surface and the point cloud is still noticeable; this is the problem we are aiming at in future. Also we are going to improve smoothness of the obtained meshes.

References

1. Levoy, M., et al. , The digital Michelangelo project: 3D scanning of large statues. In Proc. of 27th Conference on CG&IT, NY, 131-144.

2. Rusinkiewicz, S., Hall-Holt, O., and Levoy, M., Real-time 3D model acquisition. *ACM Trans. Graph.* 21, 3 (Jul. 2002), 438-446.
3. Alexa, M., Cohen-Or, D., Levin, D., As-rigid-as-possible shape interpolation. In *Proc. SIGGRAPH 2000*.
4. Bessmeltsev M., Moving adaptive mesh generation on surface in 3d space based on self-organization. In *Proc. AISC - 2008*.
5. George Kamberov, Gerda Kamberova, Topology and Geometry of Unorganized Point Clouds, 3DPVT'04
6. Igwe, P.C., Knopf, G.K., Self-Organizing Feature Map (SOFM) based Deformable CAD Models In Neural Networks. *IJCNN '06* (2006).
7. Carranza J., Theobalt C., Magnor M., Seidel H., Free-viewpoint video of human actors. *ACM TOG* 22, 3 (2003), 569-577.
8. Hoppe H., DeRose T., Duchamp T., McDonald J., Stuetzle W., Surface reconstruction from unorganized points. In *SIGGRAPH '92*, ACM.
9. Kohonen T., *Self-organizing Maps*. Springer, Berlin, Heidelberg, 2001.
10. Wu, Y. and Takatsuka, M., The Geodesic Self-Organizing Map and its error analysis. In *Proceedings of 28th ACCS - vol. 38*. 343-351.
11. Li Z., Noah S., Brian C., Steven S. M., Spacetime faces: High-resolution capture for modeling and animation. In *ACM CG'04*.
12. Nechaeva O., Composite algorithm for adaptive mesh construction based on self-organizing maps. *ICANN 2006*, LNCS.
13. Park S. I., Hodgins J. K., Capturing and animating skin deformation in human motion. *ACM Trans. Graph.* 25, 3 (July 2006), 881-889.
14. Sand P., McMillan L., Popovic J., Continuous capture of skin deformation. *ACM Trans. Graph.* 22, 3 (2003), 578-586.
15. Morooka, K., Matsui, S., and Nagahashi, H., Self-organizing Deformable Model for Mapping 3D Object Model onto Arbitrary Target Surface. In *Proc. of the 6th Conf. on 3D Dig. Im. and Mod.*, IEEE CS, 193-200.
16. Stephen R. Marschner Brian Guenter S. R., Modeling and rendering for realistic facial animation. *Proc. of 11th EG Workshop on Rendering*.
17. Wand M., Jenke P., Huang Q., Bokeloh M., Guibas L., Schilling A., Reconstruction of deforming geometry from time-varying point clouds. In *Proc. SGP '07* pp. 49-58.
18. Yu Y., Surface reconstruction from unorganized points using self-organizing neural networks. In *Proc. IEEE Vis.* 99, pp. 61-64.
19. H. Aan's and J. A. B.rentzen Pseudo-Normals for Signed Distance Computation In *Vision, Modeling, and visualization 2003*, Munich, Germany, 2003.
20. M.Bessmeltsev, AITricks GeomBox package. <http://www.aitricks.com/>
21. J.Vieira et al. Neuro-Fuzzy Systems: A Survey, *WSEAS Tr.on Syst.* '04, Is.2
22. A. De Medeiros Brito Jr., A Self-Organized Neural Network for 3D Surface Reconstruction, *WSEAS Tr.on Syst.* '04, Is.2
23. R.Kohut, B.Steinbach. Boolean Neural Networks. *WSEAS Tr.on Syst.* '04, Is.2
24. S.Maeda, S.Ishii. Optimization of Product Code, *WSEAS Tr.on Syst.* '04, Is.2
25. E.A. Sarhan, et al. Evolution Strategy Programming (ESP), *WSEAS Tr.on Syst.* '04, Vol.3, Is.2, pp.921-928.

Chapter 4

Highest Label Preflow Algorithm for the Parametric Minimum Flow Problem - A Linguistic Rule-Based Network Partitioning Approach

L. Sangeorzan¹, M. N. Parpalea², M. M. Parpalea³

¹Department of Computer Science, *Transilvania University of Brasov*,
25, Eroilor blvd, 500030, Brasov, sangeorzan@unitbv.ro

²National College *Andrei Şaguna*,
1, Şaguna street, 590000, Brasov, parpalea@gmail.com

³Department of Foreign Languages, *Transilvania University of Brasov*, 25,
Eroilor blvd, 500030, Brasov, parpalea@yahoo.com

Abstract. The article presents a preflow algorithm for the parametric minimum flow problem working in a parametric residual network with linear lower bound functions of a single parameter. On each of the iterations, the highest label preflow partitioning-pull (HLPPP) algorithm pulls flow from an active node with the highest distance label over a conditionally admissible arc. After each pull of flow, either the parametric residual capacity of the arc or the parametric deficit of the node becomes zero for at least a subinterval of the parameter values. If the two situations take place for different subintervals, the algorithm is continued in two different parametric residual networks generated by this partitioning pull. The algorithm runs as the template-like structure of a dialogue act which reveals a design where information about the items (part-of-speech) is a multiple section vector with one segment for each of the used part of speech categories.

Keywords. Parametric minimum flow, preflow algorithm, fractal partitioning, generative linguistics.

4.1 Introduction

The proposed algorithm for the parametric minimum flow problem uses a fractal-like approach [14]. A non-cancelling pull of flow from any node of the network might leave the node unbalanced for only a subinterval of the parameter. We call this situation a *partitioning pull*. Like in all fractal approaches a partitioning pull is followed by separating the problem into disjoint subintervals allowing the algorithm to continue after the same rules independently on each of the partitioned subintervals. The idea of the approach derives both from the rules and recursion of generative linguistics on one hand and the collective behaviour of neurons on the other hand.

4.2 Linguistic rule-based approach for network partitioning

Constraint-based approaches to language bring a fresh perspective to this important problem because they share a clear analogue with pattern of activation in neural networks. Some of these theories (i.e. Optimality Theory (OT)), take their inspiration from the constraint optimisation that occurs in networks of neurons [2]. The principle of constraint satisfaction provides a clear mechanism by which we can construct neural models of high-level, linguistic processes-models which are independent of the complexity of biological neurons, and rely only upon the presence of constraint-satisfaction behaviour. The template-like structure of a dialogue act which reveals a design where information about the items (part-of-speech) is a multiple section vector with one segment for each of the used part of speech categories. These categories are divided into groups, according to their importance regarding the task, enabling each segment to use its own representations for the words within it.

The concept which has proved most useful in the description of German word order has become known under the name of Functional Sentence Perspective (FSP), working with the concepts of the “theme” of a sentence (that which is spoken about in the sentence) and of the “rheme” (that which is said about the theme in the sentence). Its principal idea is that information is not transmitted in random order, but that the speaker seeks to give his information to his interlocutor in portions, normally starting from what he assumes is common to both (the THEME, topic) and proceeding to what he regards as important new information (the RHEME, comment) [12]. From the standpoint of the need to establish the “informational content” of the text an analysis of this kind is more important than a traditional examination of a subject-predicate relations etc.

4.3 The parametric minimum flow problem

Given a capacitated network $G = (N, A, l, u, s, t)$, let n being the number of nodes, $n = |N|$ and m the number of arcs, $m = |A|$. The *upper bound* function and the *lower bound* function are nonnegative functions $u(i, j)$ and $\ell(i, j)$ associated with each arc $(i, j) \in A$. The minimum flow problem is to determine a flow \hat{f} for which v is minimized. A natural generalization of the minimum flow problem is obtained by making the lower bounds $\ell(i, j; \lambda)$ linear functions of a single, nonnegative, real parameter λ : $\ell(i, j; \lambda) = \ell_0(i, j) - \lambda \cdot \xi(i, j)$, where $\xi(i, j)$ is a real valued function associated with each arc $(i, j) \in A$, referred to as the *parametric part of the lower bound* of the arc (i, j) . The parameter λ takes values in the interval $[0, \Lambda]$ where Λ is chosen so that: $0 \leq \ell(i, j; \lambda) \leq u(i, j)$, $\forall (i, j) \in A$, $\forall \lambda \in [0, \Lambda]$. The parametric minimum flow problem is to compute all minimum flows for every possible value of $\lambda \in [0, \Lambda]$:

$$\text{minimize } v(\lambda) \text{ for all } \lambda \in [0, \Lambda] \quad (1.1)$$

$$\text{with } \sum_{j|(i,j) \in A} f(i, j; \lambda) - \sum_{j|(j,i) \in A} f(j, i; \lambda) = \begin{cases} v(\lambda), & i = s \\ 0, & i \neq s, t \\ -v(\lambda), & i = t \end{cases} \quad (1.2)$$

$$\ell(i, j; \lambda) \leq f(i, j; \lambda) \leq u(i, j) \quad \forall (i, j) \in A. \quad (1.3)$$

The variables $f(i, j; \lambda)$ of this problem are piecewise linear functions instead of real numbers. On the set of all piecewise linear functions $f(\lambda)$ an ordering cannot be defined for the whole interval $[0, \Lambda]$ since two piecewise linear functions are not necessarily comparable. Therefore a partitioning J_k of the interval of the parameter $[0, \Lambda]$ into disjoint subintervals $J_1 \cup \dots \cup J_K = [0, \Lambda]$ with $J_p \cap J_q = \emptyset$, $\forall p \neq q$ must be defined so that on each of the subintervals J_k an ordering to be defined as: $f_1 \leq f_2 \Leftrightarrow f_1(\lambda) \leq f_2(\lambda)$, $\forall \lambda \in J_k$.

A *parametric cut partitioning* $[S_k; J_k]$ is a finite set of cuts $[S_k, T_k]$, $k = 1, \dots, K$ together with a partitioning J_k of the interval of the parameter $[0, \Lambda]$ into disjoint subintervals so that $J_1 \cup \dots \cup J_K = [0, \Lambda]$. The capacity of a

parametric $s-t$ cut partitioning for the minimum flow problem is a piecewise linear function $\hat{c}[S_k; J_k]$ defined for all λ of every subinterval, $\lambda \in J_k$, $k = 1, \dots, K$: $\hat{c}[S_k; J_k] = \sum_{(i,j) \in (S_k, T_k)} \ell(i, j; \lambda) - \sum_{(i,j) \in (T_k, S_k)} u(i, j)$.

A parametric $s-t$ cut for which the subintervals of the parameter values \hat{J}_k assure that every $s-t$ cut is a maximum cut $[\hat{S}_k; \hat{T}_k]$ for all $\lambda \in \hat{J}_k$ is referred to as a *parametric maximum $s-t$ cut*, $[\hat{S}_k; \hat{T}_k]$ for the whole interval of the parameter values, $[0, \Lambda]$: $\hat{c}[\hat{S}_k; \hat{J}_k] = \hat{c}[\hat{S}_k; \hat{T}_k]$ for all $\lambda \in \hat{J}_k$, $k = 1, \dots, K$.

Theorem 1 (*Parametric Min-Flow Max-Cut Theorem*) [9]: *If there is a feasible flow in the parametric network, the value function of the parametric minimum flow from a source s to a sink t in a capacitated network with parametric lower bounds equals the capacity of the parametric maximum $s-t$ cut.*

Let $f(\lambda) = (\dots f(i, j; \lambda), \dots)_{(i,j) \in A}$ be a vector of flow functions defined on the interval $[0, \Lambda]$. The *parametric residual capacity* $\hat{r}(i, j; \lambda)$ of any arc $(i, j) \in A$, with respect to a given flow $f(\lambda)$, is given by: $\hat{r}(i, j; \lambda) = u(j, i) - f(j, i; \lambda) + f(i, j; \lambda) - \ell(i, j; \lambda)$. For a network $G(\lambda) = (N, A, \ell(\lambda), u, s, t)$ and a feasible solution $f(\lambda)$, the network denoted by $\hat{G}(\lambda, f) = (\hat{N}, \hat{A})$, with $\hat{N} = N$ and \hat{A} being the set of arcs consisting only of arcs with $\hat{r}(i, j; \lambda) > 0$ for at least a subinterval of $[0, \Lambda]$, is referred to as the *parametric residual network* with respect to the given flow $f(\lambda)$ for the parametric minimum flow problem.

The sets: $\hat{I}(i, j) = \{\lambda \mid \hat{r}(i, j; \lambda) > 0\}$ for $(i, j) \in \hat{A}$ describe subintervals of $[0, \Lambda]$, $\hat{I}(i, j) \subseteq [0, \Lambda]$ where a decreasing of flow along an arc (i, j) in $\hat{G}(\lambda, f)$ is possible, based on $f(\lambda)$. If an arc (i, j) doesn't belong to $\hat{G}(\lambda, f)$ then $\hat{I}(i, j) = \emptyset$ is set.

The *parametric deficit* of a node $i \in N$ is defined as: $\hat{e}(i; \lambda) = \sum_{j \mid (i,j) \in A} f(i, j; \lambda) - \sum_{j \mid (j,i) \in A} f(j, i; \lambda)$.

The sets: $\hat{I}(j) = \{\lambda \mid \hat{e}(j; \lambda) < 0\}$ for $j \in N - \{s, t\}$ describe subintervals of $[0, \Lambda]$, $\hat{I}_j \subseteq [0, \Lambda]$ where the deficit of node j is negative. An arc $(i, j) \in \hat{A}$ in the parametric residual network $\hat{G}(\lambda, f)$ is referred to as *conditionally*

admissible arc if $\hat{d}(j) = \hat{d}(i) + 1$ and $\hat{I}(i, j) \cap \hat{I}(j) \neq \emptyset$; otherwise it is *conditionally inadmissible*.

4.4 Highest-label preflow partitioning-pull (HLPPP) algorithm for the parametric minimum flow problem

The first phase of finding a parametric minimum flow consists in establishing a feasible flow in a nonparametric network $G'=(N, A, \ell', u, s, t)$ obtained from the initial network $G(\lambda)=(N, A, \ell(\lambda), u, s, t)$ by only modifying the parametric lower bounds as follows: $\ell'(i, j) = \max\{\ell(i, j; \lambda) \mid \lambda \in [0, \Lambda]\}$. In the second phase, the algorithm maintains a set, L of active nodes as a priority queue. In the *initialisation* step of the algorithm, the flow is set to the lower bound value for every arc (i, t) , $f(i, t; \lambda) := \ell(i, t; \lambda)$ and all the active nodes are added to the priority queue L with priorities given by the distance labels.

```

procedure partitioning pull ( $L, \mathcal{J}_p$ ) :
  if ( $(L \neq \emptyset)$  and ( $\mathcal{J}_p \neq \emptyset$ )) then
    begin
      remove node  $j$  with the highest priority from  $L$ ;
      if (does not exists any admissible arc  $(i, j)$ )
        then
           $\hat{d}(j) := \min\{\hat{d}(i) \mid (i, j) \in \hat{A}\} + 1$ ;
          add  $j$  with priority  $\hat{d}(j)$  to  $L$ ;
          partitioning pull ( $L, \mathcal{J}_p$ ) ;
        else
          select an admissible arc  $(i, j)$ ;
          pull  $\hat{g}(i, j; \lambda) = \min\{-\hat{e}(j; \lambda), \hat{f}(i, j; \lambda)\}$  from node  $j$  to node  $i$ ;
          if ( $(i \notin L)$  and ( $i \neq s$ ) and ( $i \neq t$ )) then add  $i$  with priority  $\hat{d}(i)$  to  $L$ ;
           $\mathcal{J}_{p1} = \{\lambda \mid -\hat{e}(j; \lambda) \leq \hat{f}(i, j; \lambda)\}$ ;
          do in parallel
            partitioning pull ( $L, \mathcal{J}_{p1}$ ) ;
            partitioning pull ( $L \cup \{j\}, \mathcal{J} - \mathcal{J}_{p1}$ ) ;
          end;
    end;

```

Fig. 1.1. The partitioning pull (L, \mathcal{J}) procedure of the HLPPP algorithm

The active nodes are then removed from the priority queue in the descending order of their priorities. For an active node $j \in \hat{G}(\lambda, f)$, if there exists a conditionally admissible arc (i, j) , the flow will be pulled on this arc and if

$i \neq s$, $i \neq t$ and $i \notin L$, node i will be added to the priority queue L with the priority $\hat{d}(i)$; otherwise the node j will be relabelled so that at least one conditionally admissible arc to be created and node j is added to L with its new priority $\hat{d}(j)$. The algorithm terminates when the queue of active nodes is empty. A pull of flow from node j to node i is referred to as a *cancelling pull* if it deletes the arc (i, j) from the residual network; otherwise it is a *non-cancelling pull*. For any node in $\hat{G}(\lambda, f)$, the expressions: *active node* and *balanced node* holds only for subintervals of the parameter values. While both the parametric residual capacity $\hat{r}(i, j; \lambda)$ of any arc, $(i, j) \in A$ and the parametric deficit $\hat{e}(j; \lambda)$ of any node $j \in N - \{s, t\}$ are piecewise linear functions, cancelling or non-cancelling pulls are defined only for certain subintervals of the parameter values. A non-cancelling pull of flow from a node $j \in N - \{s, t\}$ along an arc $(i, j) \in A$ in a subinterval $J_p = (\lambda_{p-1}, \lambda_p] \subseteq [0, \Lambda]$ which leaves the node j unbalanced is referred to as a *partitioning pull*. Whenever the algorithm performs a *partitioning pull* in $\hat{G}_p(\lambda, f)$, a new partitioning of J_p in at most two subintervals J_{p1} and J_{p2} , with $J_{p1} \cup J_{p2} = J_p$ and $J_{p1} \cap J_{p2} = \emptyset$ may take place. Let J_{p1} be the subinterval inside which the partitioning pull balances the node j , i.e. $J_{p1} = \{\lambda \mid -\hat{e}(j; \lambda) \leq \hat{r}(i, j; \lambda)\}$. If $J_{p2} \neq \emptyset$ then, as on every subinterval J_p both $\hat{r}(i, j; \lambda)$ and $\hat{e}(j; \lambda)$ are linear functions of λ , the partitioning pull generates two parametric residual networks, $\hat{G}_{p1}(\lambda, f)$ for $\lambda \in J_{p1}$ and $\hat{G}_{p2}(\lambda, f)$ for $\lambda \in J_{p2}$, so that node j is balanced in $\hat{G}_{p1}(\lambda, f)$ and active in $\hat{G}_{p2}(\lambda, f)$ while arc (i, j) does not belong to $\hat{G}_{p2}(\lambda, f)$ since after the pull $\hat{r}(i, j; \lambda) = 0$. The algorithm then continues separately in each of the parametric residual networks and for each of the two subintervals. Under these observations, the pull/relabel procedure of the non-parametric HL-PP algorithm is replaced with a recursive call of a *partitioning pull* (L, J) procedure.

```

HLPPP Algorithm;
  Begin
    initialization;
    partitioning pull (  $L, J$  );
  End.

```

Fig. 1.2. The Highest-label preflow partitioning-pull (HLPPP) algorithm

Theorem 2: *If there is a feasible flow in the network $G(\lambda) = (N, A, \ell(\lambda), u, s, t)$, the Highest-label preflow partitioning-pull algorithm computes correctly a minimum flow.*

Proof: The proof of the proposed theorem follows from the correctness of the general HL preflow algorithm for each of the subintervals of the parameter values. ■

A *breakpoint* is a place where the slope of the piecewise linear minimum flow value function is changing. In the worst case the number of breakpoints may be exponential in the size of the problem [13]. The *Highest-label preflow partitioning-pull algorithm* overcomes this inconvenient by using the multi-thread parallel implementation of a non parametric algorithm [16]. The main idea of this implementation is to assign a processor to each newly generated subinterval J_p which will carry out the problem forward from the current configuration of the problem. For each of the new generated subintervals a copy of the current distance labels values is generated so that they can be independently modified in the further parallel evolution of the algorithm.

Theorem 3: *The Highest-label preflow partitioning-pull algorithm solves the parametric minimum flow problem in $O(n^2 m^{1/2} + Kn)$ time.*

Proof: The complexity of the non-parametric HL preflow-pull algorithm is $O(n^2 m^{1/2})$ [1]. The *HL preflow partitioning-pull algorithm* generates new copies of distance label values every time a breakpoint occurs, i.e. copying distance labels takes $O(Kn)$ time where K is the number of breakpoints.

Thus, the total complexity of the algorithm is $O(n^2 m^{1/2} + Kn)$. ■

4.5 Example

The algorithm is illustrated for the parameter λ taking values in the interval $[0,1]$, i.e. $\Lambda=1$ on the parametric network presented in Fig.1.3 where node 1 is the source node s and node 4 is the sink node t .

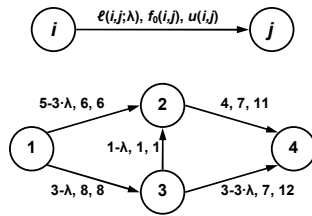


Fig. 1.3. The feasible flow f_0 in network $G(\lambda) = (N, A, u, \ell(\lambda), s, t)$

After the *initialization* procedure, the parametric residual network $\hat{G}(\lambda, f_0)$ is presented in Fig.1.4.a. The priority queue L contains two active nodes: node 2 and node 3 with the priorities $\hat{d}(2)=\hat{d}(3)=1$: $L=\{2,3\}$. The call of *partitioning pull* ($\{2,3\}, [0,1]$) removes node 2 from the list L and selects the *conditionally admissible* arc $(1,2)$.

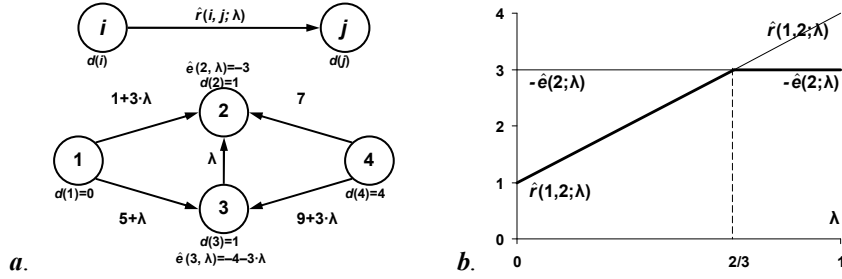


Fig. 1.4. *a.* The parametric residual network $\hat{G}(\lambda, f_0)$ after the initialization procedure; *b.* Generating the partitioning of the parameter values by the partitioning pull procedure on arc $(1,2)$ in $\hat{G}(\lambda, f_0)$

A pull of an amount of flow, $\hat{g}(1,2;\lambda) = \min\{-\hat{e}(2,\lambda), \hat{r}(1,2;\lambda)\}$ from node 2 to node 1, computed as in Fig.1.4.b, generates a partition of $J = [0,1]$ in the subintervals $J_1 = [2/3, 1]$ and $J_2 = [0, 2/3]$. The parametric residual networks, $\hat{G}_1(\lambda, f)$ and $\hat{G}_2(\lambda, f)$, for the two corresponding subintervals, are presented in Fig.1.5. Then the algorithm makes a parallel call of *partitioning pull*(L_1, J_1) and *partitioning pull*(L_2, J_2) procedures with $L_1 = \{3\}$ and $L_2 = \{2,3\}$.

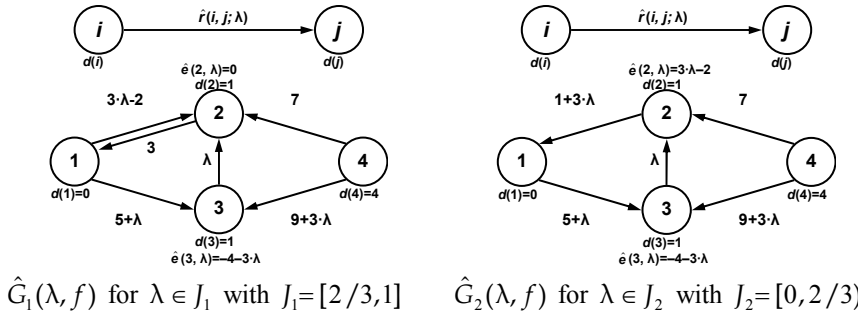


Fig. 1.5. The parametric residual networks for the two subintervals of the parameter values, generated by the partitioning pull over the arc $(1,2)$

After further iterations the algorithm will stop when the priority queues for all the subintervals of the parameter values are empty. Actually, the algorithm stops with the subintervals and with the corresponding parametric residual networks from which the parametric maximum $s-t$ cut partitioning can be derived. The way the *partitioning pull* procedure of the HLPPP algorithm generate the subintervals between consecutive breakpoints of the piecewise linear minimum flow value function is illustrated in Fig.1.6. If no partitioning of the parameter values occurs, for each of the generated subintervals, the algorithm goes on independently until the corresponding priority list becomes empty.

$[0, 1]$ <i>partitioning pull</i> over the arc (1,2)				$[2/3, 1]$
$[0, 2/3]$ <i>partitioning pull</i> over the arc (3,2)				
$[0, 1/2]$ <i>partitioning pull</i> over the arc (1,3)		$[1/2, 2/3]$ <i>partitioning pull</i> over the arc (2,3)		
$[0, 1/3]$	$(1/3, 1/2)$	$\{1/2\}$	$(1/2, 2/3)$	

Fig. 1.6. Illustrating of the partitioning pulls of the HLPPP algorithm for the parametric minimum flow problem in the parametric network in Fig.1.3.

For every subinterval of $[0, \Lambda]$, the parametric minimum flow value function equals the capacity of the maximum $s-t$ cut. The parametric minimum flow value function, $\hat{v}(\lambda)$ for the entire interval of the parameter values is given by the expression presented in Fig.1.7.

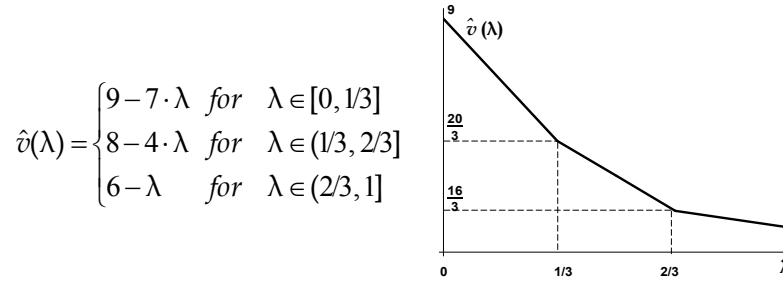


Fig. 1.7. The parametric minimum flow value function for the network in Fig.1.3.

4.6 Conclusion

The minimum flow problem on parametric networks turns out to be an important scenario in practice since the complexity of its solving algorithm

can be reduced to that of the equivalent non-parametric algorithm considering the approach of fractal partitioning of the network on one hand and the rules and recursion of generative linguistics and the collective behaviour of neurons in a network on the other.

References:

1. Ahuja R, Magnanti T, Orlin J (1993) Network Flows. Theory, algorithms and applications. Prentice Hall, Inc., Englewood Cliffs, New Jersey
2. Brill, E (1992) A simple rule-based part of speech tagger. Proceedings of the Third Conference on Applied Natural Language Processing, pp 152–155
3. Chang-Rui Yu and Yan Luo (2008) An Improved Nested Partitions Algorithm Based on Simulated Annealing in Complex Decision Problem Optimization. WSEAS Transactions on Computers 3/7: 75-82
4. Che-Chern Lin (2008) Implementation Feasibility of Convex Recursive Deletion Regions Using Multi-Layer Perceptrons. WSEAS Transactions on Computers 1/7: 24-31
5. Ciurea E, Ciupală L (2008) About preflow algorithms for the minimum flow problem. WSEAS Transactions on Computer Research 1/3: 35-42
6. Ciurea E, Ciupală L (2008) A parallel algorithm for the minimum flow problem in bipartite networks. 12th WSEAS International Conference on COMPUTERS, Heraklion, Greece, pp 203-207
7. Ciurea E, Ciupală L (2007) A Highest-Label Preflow Algorithm for the Minimum Flow Problem. Proceedings of the 11th WSEAS International Conference on Computers, Agios Nikolaos, Crete Island, Greece, pp 168-171
8. Deng Z, Zhang Y (2007) Collective Behavior of a Small-World Recurrent Neural System with Scale-Free Distribution. IEEE Transactions on Neural Networks 18(5), pp 1364-1375
9. Hamacher HW, Foulds LR (1989) Algorithms for Flows with Parametric Capacities. ZOR - Methods and Models of Operations Research, (33), pp 21-37
10. Negishi M (1998) Everything that Linguists have Always Wanted to Know about Connectionism. Department of Cognitive and Neural Systems, Boston University, 1998
11. Nitschke GS (2009) Neuro-evolution approaches to collective behavior. Proceedings of the 11th conference: Congress on Evolutionary Computation. Trondheim, pp 1554-1561
12. Parpalea MM (2009) German Word Order. Bulletin of the Transilvania University of Braşov, Series IV, Vol. 2 (51), pp 175-182
13. Ruhe G (1988) Complexity Results for Multicriterial and Parametric Network Flows Using a Pathological Graph of Zadeh, Zeitschrift für Operations Research, Vol. 32, pp 9- 27

14. Sangeorzan L, Kiss-Iakab K, Parpalea M (2007) Runtimes for Parallel Fractals, Proceedings On the 31st Annual Congress of the American Romanian Academy of Arts and Sciences, No.5, pp 549-552
15. Tarjan R, Ward J, Zhang B, Zhou Y, Mao J (2006) Balancing Applied to Maximum Network Flow Problems. Algorithms–ESA, Springer 4168 pp 612-623
16. Yong Chen, Chang-le Lu (2006) Using Multi-Thread Technology Realize Most Short-Path Parallel Algorithm. World Academy of Science, Engineering and Technology, No.15, pp 11-13

Chapter 5

Information Process in the Biosensor of the Angular Motion of the Head

V.V. Alexandrov^{1,2}, T.B. Alexandrova^{1,2}, G. Castillo-Quiroz¹, A. Ortega¹, R. Vega¹, N. Shulenina², G Sidorenko², E. Soto¹

¹ Universidad Autónoma de Puebla, Apartado postal 406, Puebla, Pue. C. P. 72000, México, e-mail: esoto@siu.buap.mx

² Moscow State University, Leninskie Gory, 119991, Moscow, Russian Federation, e-mail: vladimiralexandrov366@hotmail.com

Abstract. A compartmental model of the sensory activation process in the biosensor of angular motion of the head is presented. The functional and numerical parameters of the model have been identified by physiological and morphological experiments. Model includes four compartments and comparison with experimental results of Fernandez and Goldberg [1] that describes the changes in the firing frequency of the primary afferent neurons of the left lateral semicircular canal in response to a mechanical stimulus as trapezoidal-shape changes of angular velocity when the head turns around the vertical axis. Our model shows that the biosensor of angular motion of the head must include at least a pair of semicircular canals in each plane considered in the system.

Keywords. Mechanical stimulus, hair cell, primary neuron, vestibular mechanoreceptor, afferent impulses.

5.1 Introduction

Since 2001, mathematicians and physiologists from Lomonosov Moscow State University and the Autonomous University of Puebla, México, began collaborating for the development of a mathematical model of informative processes in the biosensors of the vestibular system [2, 3, 4, 5, 6, 10, 11, 12, 13].

Based on this work we present a mathematical model of the biosensor of angular acceleration when the head turns around a vertical axis z .

Figure 1 shows a functional scheme of the information output in the biosensor of angular acceleration in response to a stimulus that arises when a movement of the head occurs in the horizontal plane; therefore, the input is the angular acceleration $\dot{\omega}_z(t)$. The scheme has two input blocks describing the dynamics of the cupula-endolymphatic system (CES) of the horizontal semicircular canals. Assuming that the dynamics of displacement of the cilia bundle top x_L and x_R in the left and the right semicircular canals coincides with the dynamics of CES, it constitutes the input for two vestibular mechanoreceptors forming a string of hair cells and primary afferent neurons. The vestibular mechanoreceptor transforms the mechanical stimulus into a change in the membrane potential V_{IL} and V_{IR} of the hair cell in both left and right semicircular canals, and this is the primary output from this model. Subsequently, in the following blocks of the model scheme, this output activates the synaptic transmission I_{synL} and I_{synR} and the action potential discharge in the primary afferent neurons V_{2L} and V_{2R} . The scheme has two signal outputs that form the primary input to the vestibular nuclei and to the oculomotor system that controls the eye movements in the horizontal plane.

5.2 Mathematical model formulation

For simplification we consider only the top row of the functional scheme (Fig.1.1 A) which corresponds to the left horizontal semicircular canal.

The mathematical model of the CES is in the form of a modified equation of Steinhausen [8] of order two (1.1) where τ_1, τ_2 are time constants ($\tau_2 \ll \tau_1$), x is the displacement (μm) of the cilia bundle top.

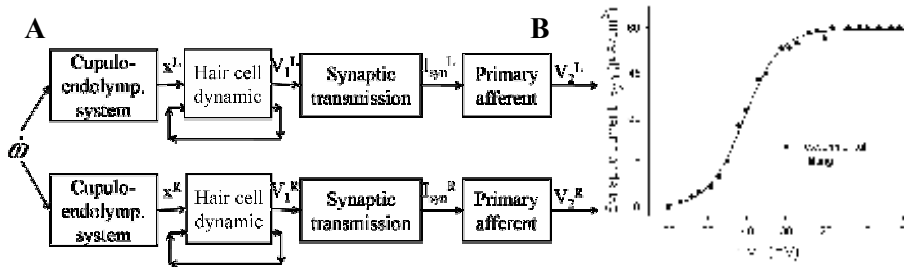


Fig 1.1. Model scheme. (A) The Scheme of the vestibular mechanoreceptor compartments considered in the angular acceleration sensor model. (B) The

relationship between hair cell voltage and synaptic current in the afferent neuron displays a sigmoid dependence (from Keen and Hudspeth, 2006).

This displacement produces the transduction current I_{Tr} . The equations (1.2), (1.3), (1.4), (1.5) describe the dynamics of the membrane potential V_I and the total ionic current I_T produced by the mechanical displacement x . The variables m , h_1 , h_2 are physiological parameters that correspond to the parameters of activation and inactivation of ionic currents [3,4].

The block corresponding to the synaptic transmission is represented by the algebraic model shown in (1.6), obtained from experimental data reported by Keen and Hudspeth, 2006 [7]. The graph corresponding to

(1.6) is shown in Fig. 1.1 B) where $I_{Syn}^{max} = 60 \frac{\mu A}{sm^2}$.

The equations (1.7), (1.8), (1.9) describe the output (from the afferent neurons) in the form of a change in the frequency of auto-oscillations produced by sodium (I_{Na}), potassium (I_K) and synaptic (I_{Syn}) currents. The variables of the output block are: V_2 is the membrane potential at the first node of Ranvier; n , h_K are the parameters of activation and inactivation for the potassium current [2].

$$\frac{d^2x}{dt^2} + \frac{1}{\tau_2} \frac{dx}{dt} + \frac{1}{\tau_1 \tau_2} x = k_0 R \dot{\omega} \quad (1.1)$$

$$C_{m1} \frac{dV_1}{dt} = -I_{Tr} - I_T - I_{L1}, \quad (1.2)$$

$$\frac{dm}{dt} = \left(\frac{m_{ST}(V_1) - m}{\tau_m(V_1)} \right) Q_{10}, \quad (1.3)$$

$$\frac{dh_1}{dt} = \left(\frac{q_1 h_{ST}(V_1) - h_1}{\tau_{h1}(V_1)} \right) Q_{10}, \quad (1.4)$$

$$\frac{dh_2}{dt} = \left(\frac{q_2 h_{ST}(V_1) - h_2}{\tau_{h2}(V_1)} \right) Q_{10}, \quad (1.5)$$

$$I_{Syn}(V_1) = 59.6962 \left(1 + \exp \left(\frac{-(V_1 + 40.6031)}{4.5979} \right) \right)^{-1}, \quad (1.6)$$

$$C_{m2} \frac{dV_2}{dt} = I_{Syn}(V_1) - I_{Na} - I_K - I_{L2}, \quad (1.7)$$

$$\frac{dn}{dt} = \left(\frac{n_{\infty}(V_2) - n}{\tau_n(V_2)} \right) Q_{10}, \quad (1.8)$$

$$\frac{dh_K}{dt} = \left(\frac{h_{K\infty}(V_2) - h_K}{\tau_{h_K}(V_2)} \right) Q_{10}. \quad (1.9)$$

$$\begin{aligned} \text{where } I_{Tr} &= g_{Tr}(x)(V_1 - E_{Tr}), \quad g_{Tr} = \bar{g}_{Tr} p(x), \\ I_T &= g_T m^3 (h_1 + h_2)(V_1 - E_T), \quad I_{L1} = g_{L1} V_1, \quad p(x) = \left(1 + \exp\left(-\frac{x - x_0}{s_1}\right)\right)^{-1} \\ I_{Na} &= g_{Na} (m_\infty(V_2))^3 (C(V_2) - n)(V_2 - V_{Na}), \\ I_K &= g_K n^4 h_K (V_2 - V_K), \quad I_{L2} = g_{L2} (V_2 - V_L) \end{aligned}$$

Here $x = R\xi$ where ξ is the angular deviation of the endolymph [8] relative to the center of the semicircular canal, R - the outer radius of the canal, the coefficient k_0 has physiological sense and is expressed by geometrical parameters of the semicircular canal [6].

The steady state variables, when the mechanic stimulus is absent ($\dot{\omega} \equiv 0$) are: $I_T = 145$ nA, $V_I = -56$ mV, $I_{Syn} = 1.82$ pA/cm². Noteworthy, the stationary hair cell potential is $V_1^0 = -56,5$ mV and the basal frequency of discharge of the neuron (V_2) in resting condition is $\nu_0 = 56$ Hz.

In the presence of angular accelerations ($\dot{\omega}(t) \neq 0$), the change of action potential discharge frequency $\Delta\nu(t) = \nu(t) - \nu_0$ provides the output of the biosensor. The functional and numerical parameters are as in [13].

The model (1.1-1.9) was tested for long stimuli analog to those used by Fernandez and Goldberg [1], and short duration stimuli with the constant accelerations (0.1 - 1 s). For long stimulus we present the experimental result of Goldberg and Fernandez, 1971 [1] (Fig. 1.2) and the function of our model. The change of the output discharge of the model based on eq. 1.1-1.9 was calculated for each second and is presented in Fig. 1.3. If we compare the results in Fig. 1.2 with those in Fig. 1.3 A, it is possible to observe a qualitative coincidence between the experimental result and the result obtained by our model (1.1-1.9).

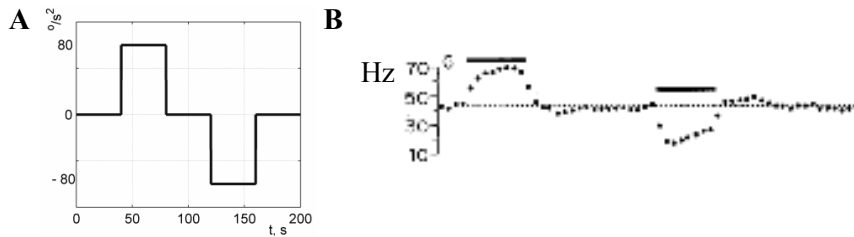


Fig. 1.2. (A) The velocity trapezoid stimulus with acceleration magnitude of 80 deg/s². (B) Experimental response of the afferents of the left lateral semicircular canal [1].

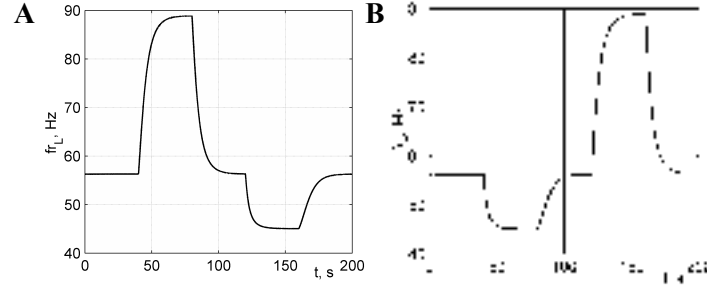


Fig.1.3. (A) Response of the model (1.1–1.9) to the long stimulus (corresponding to the left semicircular canal). (B) Response of the model (1.1–1.9) to the long stimulus (corresponding to the right semicircular canal).

5.3 Model response to short stimuli

With the assumptions adopted in the development of some blocks of the model (1.1-1.9) and described in [2,3,4,5,6,13], our model is intended for reproduction of the informative process when $\dot{\omega}(t) \neq 0$ in intervals of seconds, using stimuli that are analogous to the head turn under natural conditions. For this reason we studied the model response (1.1-1.9) to trapezoidal stimulus of 1 s duration (Fig.1.4), corresponding to head turning around vertical axis at about 12° .

The dynamics of the response of (1.1-1.9) to this stimulus is presented in Fig. 1. 5. The change on the firing frequency in the left semicircular canal $\Delta v(t) = 20$ Hz in response to a steady acceleration $150^\circ/s^2$ can lead to the contraction of the oculomotor muscles producing a movement of the eyes of 12° to the opposite side of the head movement (Fig 1.5 C).

With the use of short stimuli (Figs. 1.4 A,B), the output of the biosensor supplies information about the angular velocity of the head. In the case of a long stimulus (Fig. 1.3), the head angular acceleration is the output information.

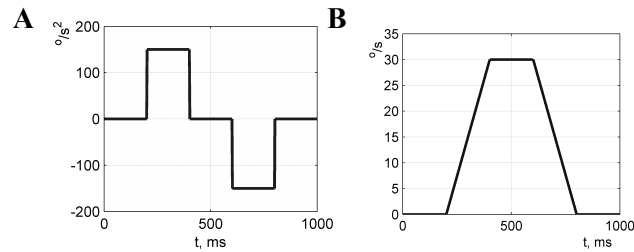


Fig 1.4. Short stimulus. A) Head angle acceleration; B) Head angle velocity.

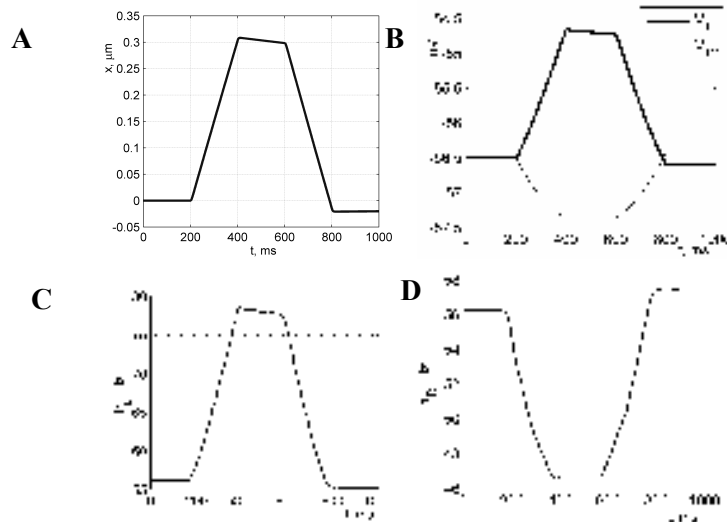


Fig. 1.5. Response of the model (1.1–1.9) to the short stimulus: A) displacement of the cilia bundle top; B) change of hair cell potential in the left and right canals; C) change of the afferent neuron discharge frequency corresponding to the left semicircular canal; D) change of discharge frequency corresponding to the right semicircular canal.

In the literature this fact is known (for example [9]). However the fundamental difference between the left and right semicircular canals outputs represented in Fig. 1.3 and Fig. 1.5 C),D) wasn't known.

Based on our results it is concluded that the biosensor of the angular motion of the head must include two semicircular channels since the information provided by each channel is needed to have a bidirectional movement detection.

Acknowledgements. This material is based on work supported by a grant from the University of California Institute for México and the United States (UC-MEXUS) and the Consejo Nacional de Ciencia y Tecnología de México (CONACyT), grant of RFFI 07-01-00216 (Russia) and State Grant 02.740.11.0300 (Russia).

References

1. Fernández C, Goldberg JM (1971) Physiology of peripheral neurons innervating semicircular canals of the squirrel monkey. I. Resting discharge and response to constant angular accelerations. *J Neurophysiol* 34: 635–660
2. Alexandrov VV, Mikhaleva EYu, Soto E, García Tamayo R (2006) Modification of Hodgkin-Huxley mathematical model for the primary neuron of vestibular apparatus. *Moscow University Mechanics Bulletin* 5:65–68
3. Alexandrov VV, Alexandrova TB, Vega R, Castillo Quiroz G, Reyes Romero M, Soto E (2007) Information process in vestibular system. *WSEAS Transactions on Biology and Medicine* 4:193-203
4. Sadovnichii VA, Alexandrov VV, Soto E, Alexandrova TB, Astakhova TG, Vega R et al (2007) A mathematical model of the response of the semicircular canal and otolith to vestibular system rotation under gravity. *J Mathematical Sciences*, 146, 3: 5938-5947
5. Sadovnichii VA, Alexandrov VV, Alexandrova TB, Vega R, Castillo Quiroz, Reyes Romero M, Soto E. Shulenina NE (2008) A mathematical model for the generation of output information in a gravito-inertial mechanoreceptor when moving in a sagittal plane. *Moscow University Mechanics Bulletin* 63,6: 53-60
6. Vega R, Alexandrov VV, Alexandrova TB, Soto E (2008) Mathematical model of the cupula-endolymph system with morphological parameters for the axolotl (*Ambystoma tigrinum*) semicircular canals. *The Open Medical Informatics Journal*, 2: 138-148
7. Keen EC, Hudspeth AJ (2006) Transfer characteristic of the hair cell's afferent synapse. *Proc. Natl. Acad. Sci. USA*, 103: 5537-5542
8. Steinhausen W (1933) Über die beobachtung der cupula in den bogengang-sampullen des labirinth des lebenden Heechts. *Pflug. Arch. ges. Physiol* 232: 500-512
9. Orlov IV (1998) The vestibular function. Nauka, Saint Petersburg, p 211
10. Grebennikov A, Alexandrov V, Leyva J F (2006) Computer modeling of the neuronal activity described with the simplified Hodgkin-Huxley system of differential equations. *Proceedings of the 10th WSEAS International Conference on Computers, Vouliagmeni Athens, Greece*, pp. 998- 1001
11. Alexandrov VV, Alexandrova TB, Vega R, Castillo Quiroz G, Reyes Romero M, Soto E (2008) Mathematical model of information process in vestibular mechanoreceptor. *WSEAS Mathematical Biology and Ecology (MABE'08)*, 4: 86-91
12. Grebennikov A, Alexandrov V, Soto E, Castillo G (2009) Local method in computer modeling of mechanoreceptor from vestibular system. *WSEAS Biomedical Electronics and Biomedical Informatics*, pp 136-142
13. Alexandrov V, Alexandrova TB, Castillo G, Sidorenko G, Ortega A, Vega R, Soto E (2009). Mathematical modeling of the informative process in the biosensor of angular acceleration. *WSEAS Biomedical Electronics and Biomedical Informatics*, pp 105-110

Chapter 6

Numerical Analysis of Electromagnetic (EM) Absorption Reduction by using Material Attachment

M. R. I. Faruque^{1,2}, M. T. Islam², N. Misran^{1,2}

¹Institute of Space Science (ANGKASA),

²Dept. of Electrical, Electronic and Systems Engineering,

Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia.

rashedgen@yahoo.com, titareq@yahoo.com, bahiah@vlsi.eng.ukm.my

Abstract— The numerical analysis is used to reduce the electromagnetic absorption with materials attachment. The finite-difference time-domain method with lossy-Drude model is adopted in this study. The methodology of SAR reduction is addressed and the effects of attaching location, distance, and size of ferrite sheet material on the SAR reduction are investigated. Materials have achieved a 47.68% reduction of the initial SAR value for the case of 1 gm SAR. These results suggest a guideline to choose various types of materials with the maximum SAR reducing effect for a phone model.

Keywords — antenna, human head model, lossy-Drude model, materials, specific absorption rate (SAR), symmetry.

6.1 Introduction

The present day electromagnetic solution is possible due to the progress of door due to numerical analysis. RF/MW sources are part of daily life, but they also reason for concern regarding the possible biological special effects of microwaves. It is important that the biological effects of RF/MW fields are minimal, at least at the level of their clinical significance, so that

health risk can be assessed. Because the potential shock of RF/MW fields on human health has not yet been well characterized, the basic knowledge from laboratory studies based on cellular and animal test systems are invaluable. The interaction of handset antennas with human body is a great consideration in cellular communications. The user's body, especially head and hand, influence the antenna voltage standing wave ratio (VSWR), gain and radiation patterns. Furthermore, thermal effects, when tissues are exposed to unlimited electromagnetic energy, can be a serious health hazard. Therefore standards organizations have set exposure limits in terms of SAR [1-3].

The exposure limits are defined commonly in terms of the spatial peak SAR averaged either over any one gram or ten grams of tissue. Since 1997, the U.S. Federal Communication Commission (FCC) requires the routine SAR evaluation of phone model prior to device authorization or use. So there is a need to reduce the spatial peak SAR in the design stage of a phone model because the possibility of a spatial peak SAR exceeding the recommended exposure limit cannot be completely ruled out [2-4]. The interaction of the cellular handset with the human head has been investigated by many published papers considering; first, the effect of the human head on the handset antenna performance including the feed-point impedance, gain, and efficiency [4-7], and second, the impact of the antenna EM radiation on the user's head due to the absorbed power, which is measured by predicting the induced SAR in the head tissue [7-9].

The most used method to solve the electromagnetic problem in this area is the finite-difference time-domain (FDTD) technique [5-8]. Although, in principle, the solution for general geometries does not require any additional effort with respect to the standard method, the technique requires the definition of a discretized space by assigning to each cell its own electromagnetic properties, which is not an easy process [7-10]. Specifically, the problems to be solved in SAR reduction need a correct representation of the cellular phone; anatomical representation of the head; alignment of the phone and the head, and suitable design of materials.

Human exposure to electromagnetic (EM) radiation, as well as the pertinent health effects, constitutes a matter of raised public concern, and this issue has undergoing continuous scientific investigation. Various studies on this subject exist [9-10], most of which mainly investigate into the consequences of mobile-phone usage. Yet, devices and communication terminals operating in other frequency bands have also gained substantial interest in the last 15 years. In [5], a ferrite sheet was adopted as protection between the antenna and the human head. A reduction of over 13% for the spatial peak SAR over 1 gm averaging was achieved. A study on the effects of attaching a ferrite sheet for SAR reduction was presented in [10],

and it was concluded that the position of shielding plays an important role in the reduction effectiveness. This paper is structured as follows. Modeling and analyzing technique will be described in Section 1.2. Numerical method FDTD will be analyzed in section 1.3 and SAR reduction with numerical schemes will be discussed in section 1.4. Simulation and comparing results of materials will be summarized in Section 1.5 and finally in Section 1.6 concludes the paper.

6.2 Simulation Model and Numerical Techniques

6.2.1 Model Description

The simulation model which includes the handset with PIFA type of antenna and the SAM phantom head provided by CST Microwave Studio® (CST MWS) is shown in Fig.1.1 Complete handset model composed of the circuit board, LCD display, keypad, battery and housing was used for simulation. The relative permittivity and conductivity of individual components were set to comply with industrial standards. In addition, definitions in [7,10] were adopted for material parameters involved in the SAM phantom head. In order to accurately characterize the performance over broad frequency range, dispersive models for all the dielectrics were adopted during the simulation [7]. The electrical properties of materials used for simulation are listed in Table 1.1 PIFA type antenna constructed in a helical sense operating at 900 MHz for GSM application was used in the simulation model. In order to obtain high-quality geometry approximation for such helical structure, predictable meshing scheme used in FDTD method usually requires large number of hexahedrons which in turn makes it extremely challenging to get converged results within reasonable simulation time.



Fig.1.1. Complete model used for simulation including handset and SAM phantom head.

Table 1.1. Electrical properties of materials used for simulation

Phone Materials	ϵ_r	$\sigma(S/m)$
Circuit Board	4.4	0.05
Housing Plastic	2.5	0.005
LCD Display	3.0	0.02
Liquid @ 900MHz	40	1.42
Rubber	2.5	0.005
SAM Phantom Head		
Shell	3.7	0.0016

6.2.2 Numerical Technique

CST MWS, which adopted finite integral time-domain technique (FITD) proposed was used as the main simulation instrument. In permutation of the perfect boundary approximation (PBA) and thin sheet technique (TST), significant development in geometry approximation with computation speed is achieved squashy highly accurate results. Non-uniform meshing scheme was adopted so that major computation endeavor was dedicated to regions along the inhomogeneous boundaries for fast and perfect analysis. The minimum and maximum mesh sizes were 0.3 mm and 1.0 mm, respectively. A total of 2,097,152 mesh cells were generated for the complete model, and the simulation time was 1163 seconds (including mesh generation) for each run on an Intel Core™ 2 Duo E 8400 3.0 GHz CPU with 4 GB RAM system.

The analysis workflow started from the design of antenna with complete handset model in free space. The antenna was designed such that the S_{11} response was less than -10 dB over the frequency band of interest. SAM phantom head was then included for SAR calculation using the standard definition as

$$SAR = \frac{\sigma}{2\rho} E^2$$

where E is the induced electric field (V/m); ρ is the density of the tissue (kg/m³) and σ is the conductivity of the tissue (S/m). The resultant SAR values averaged over 1 gm and 10gm of tissue in the head were denoted as SAR 1 gm and SAR 10gm, respectively. These values were used as a benchmark to appraise the effectiveness in peak SAR reduction.

6.3 FDTD Method

While many electromagnetic simulation techniques are applied in the frequency domain, FDTD solves Maxwell's equations in the time domain. This means that the calculation of the electromagnetic field values progresses at discrete steps in time. One benefit of the time domain approach is that it gives broadband output from a single execution of the program. However the main reason for using the FDTD approach is the excellent scaling performance of the method as the problem size grows. As the number of unknowns increases, the FDTD approach quickly outpaces other methods in efficiency. FDTD has also been identified as the preferred method for performing electromagnetic simulations for biological effects from wireless devices [3-5]. The FDTD method has been shown to be the most efficient approach and provides accurate results of the field penetration into biological tissues.

As indicated by its name, FDTD method solves Maxwell's equations directly in time domain. Assuming a piecewise uniform, homogeneous, isotropic and lossy media, source-free Maxwell's curl equations: [5]

$$\mu_0 \frac{\delta \vec{H}}{\delta t} = -\nabla \times \vec{E}.$$

(1)

$$\varepsilon \frac{\delta \vec{E}}{\delta t} = \nabla \times \vec{H} - \sigma \vec{E}.$$

(2)

are discretized both in space and time with (i, j, k) and n representing the discrete space and time nodes,

$$E_x(x, y, z, t) = E_x^n(i, j, k); x = i \times \Delta x, y = j \times \Delta y, z = k \times \Delta z, t = n \times \Delta t.$$

(3)

respectively. All six components of electromagnetic fields are located at specific points of unit cells as suggested by Yee [5], where \vec{E} and \vec{H} are interleaved in both space and time. This FDTD arrangement is shown in Fig.1.2 Here, electric field components E_x, E_y, E_z are located on the three edges and magnetic field components H_x, H_y and H_z are located in the middle of the three surfaces and are assumed constant. In addition to one-half spatial-cell displacement between \vec{E} and \vec{H} , there is also one-half time-cell displacement. Applying the discretization and rearranging the equations suitable to iterative calculations yield,

$$H_x^n(i, j, k) = H_x^{n-1}(i, j, k) - \frac{\Delta t}{\mu_0 \Delta z} [E_y^n(i, j, k) - E_y^n(i, j, k-1)]$$

$$- \frac{\Delta t}{\mu_0 \Delta y} [E_z^n(i, j, k) - E_z^n(i, j-1, k)]$$

(4)

$$H_y^n(i, j, k) = H_y^{n-1}(i, j, k) - \frac{\Delta t}{\mu_0 \Delta x} [E_z^n(i, j, k) - E_z^n(i-1, j, k)]$$

$$- \frac{\Delta t}{\mu_0 \Delta z} [E_x^n(i, j, k) - E_x^n(i, j, k-1)]$$

(5)

$$H_z^n(i, j, k) = H_z^{n-1}(i, j, k) - \frac{\Delta t}{\mu_0 \Delta y} [E_x^n(i, j, k) - E_x^n(i, j-1, k)]$$

$$- \frac{\Delta t}{\mu_0 \Delta x} [E_y^n(i, j, k) - E_y^n(i-1, j, k)]$$

(6)

$$E_x^n(i, j, k) = \frac{2\varepsilon - \sigma\Delta t}{2\varepsilon + \sigma\Delta t} E_x^{n-1}(i, j, k) - \frac{2\Delta t}{(2\varepsilon - \sigma\Delta t)\Delta z} [H_y^n(i, j, k) - H_y^n(i, j, k-1)]$$

$$+ \frac{2\Delta t}{(2\varepsilon - \sigma\Delta t)\Delta y} [H_z^n(i, j, k) - H_z^n(i, j-1, k)]$$

(7)

$$E_y^n(i, j, k) = \frac{2\varepsilon - \sigma\Delta t}{2\varepsilon + \sigma\Delta t} E_y^{n-1}(i, j, k) - \frac{2\Delta t}{(2\varepsilon - \sigma\Delta t)\Delta x} [H_z^n(i, j, k) - H_z^n(i-1, j, k)]$$

$$+ \frac{2\Delta t}{(2\varepsilon - \sigma\Delta t)\Delta z} [H_x^n(i, j, k) - H_x^n(i, j, k-1)]$$

(8)

$$E_z^n(i, j, k) = \frac{2\varepsilon - \sigma\Delta t}{2\varepsilon + \sigma\Delta t} E_z^{n-1}(i, j, k) - \frac{2\Delta t}{(2\varepsilon - \sigma\Delta t)\Delta y} [H_x^n(i, j, k) - H_x^n(i, j-1, k)]$$

$$+ \frac{2\Delta t}{(2\varepsilon - \sigma\Delta t)\Delta x} [H_y^n(i, j, k) - H_y^n(i-1, j, k)]$$

(9)

where $\vec{n} = n+1/2$. In equations (4)-(9), it is assumed that no magnetic materials are present so that each cell is characterized by constant medium parameters μ_0 , ε , and σ given for the reference point of the cell.

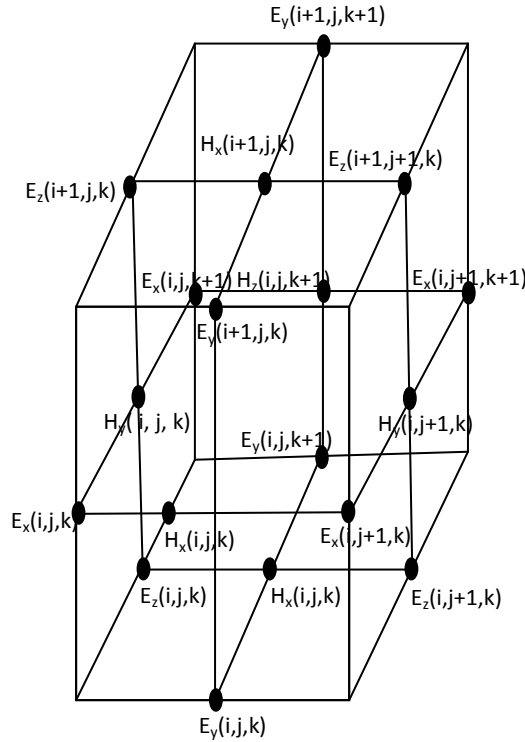


Fig.1.2.FDTD unit cell suggested by Yee and locations of EM field components

FDTD method is mainly based on calculation of equations (4)-(9), with suitable excitation representation and with satisfaction of the necessary boundary conditions at $t_n = n \times \Delta t$ time instants. The most important points in FDTD calculations are the stability and numerical dispersion [18]. Since, FDTD algorithm is based on iterative calculations of six field components, a certain relation between spatial and time intervals

$$\Delta t \leq \frac{1}{c \sqrt{\left(\frac{1}{\Delta x}\right)^2 + \left(\frac{1}{\Delta y}\right)^2 + \left(\frac{1}{\Delta z}\right)^2}}; \quad c = 3 \times 10^8 \text{ m/s free- space EM velocity.} \quad (10)$$

needs to be satisfied for stability (named as Courant criteria) [5]. Similarly, the spatial discretization must be done to allow tracing the highest frequency (minimum wavelength) component in the desired frequency band (numerical dispersion).

6.4 SAR reduction with Numerical Schemes

6.4.1 Lossy-Drude Model

The SAR reduction effectiveness and antenna performance with different positions, sizes and materials properties of materials will be analyzed. The head models used in this study were obtained from a MRI-based head model through the whole brain Atlas website. Six types of tissues, i.e., bone, brain, muscle, eye ball, fat, and skin were involved in this model [5-7]. This paper was considered a horizontal cross-section through the eyes of this head model. The electrical properties of tissues were taken from [10]. Numerical simulation of SAR value was performed by the FDTD method. The parameters for FDTD computation were as follows. In our lossy-Drude simulation model, the domain was $128 \times 128 \times 128$ cells in the FDTD method. The cell sizes were set as $\Delta x = \Delta y = \Delta z = 1.0$ mm. The computational domain was terminated with 8 cells perfect matched layer (PML). A PIFA antenna was modeled for this paper by the thin-wire approximation. Simulations of materials are performed by the FDTD method with the lossy-Drude Model [7-9]. The method is utilized to understand the wave propagation characteristics of materials.

6.4.2 Analysis Method

A portable telephone model at 900 MHz was considered in this paper. It was considered to be a quarter wavelength PIFA antenna mounted on a rectangular

conducting box. The conducting box was 10 cm long, 4 cm wide and 3 cm thick. The PIFA antenna was located at the top surface of the conducting box. A ferrite sheet of height of 90 mm, width of 40 mm and thickness of 3.5 mm was attached to the conducting box. The SAM head model was considered for this research where it consists about 2,097,152 cubical cells with a resolution of 1 mm.

6.5 Impact on SAR of ferrite sheet attachment

In this section, a ferrite sheet is placed between the antenna and a human head thus reducing the SAR value. In order to study SAR reduction of an antenna operated at the GSM 900 band, different positions, sizes, and ferrite sheet materials for SAR reduction effectiveness are also analyzed by using the FDTD method in conjunction with a detailed human head model.

This paper considered that the simulation model which includes the handset with monopole type PIFA antenna and the SAM phantom head provided by CST MWS. The dispersive models for all the dielectrics were adopted during the simulation in order to accurately characterize the ferrite sheet. The antenna was arranged in parallel to the head axis; the distance is varied from 5 mm to 20 mm; and finally 20 mm was chosen for comparison with the ferrite sheet. Besides that, the output power of the mobile phone model need to be set before SAR is simulated. In this paper, the output power of the cellular phone is 500 mW at the operating frequency of 0.9 GHz. In the real case, the output power of the mobile phone will not exceed 250 mW for normal use, while the maximum output power can reach till 1 W or 2 W when the base station is far away from the mobile station (cellular phone). The SAR simulation is compared with the results in [3, 10] for validation, as shown in Table 1.2. The calculated peak SAR 1 gm value is 2.002 W/Kg, and SAR 10 gm value is 1.293 W/Kg when the phone model is placed 20 mm away from the human head model without a ferrite sheet. This SAR value is better compared with the result reported in [1], which is 2.43 W/Kg for SAR 1 gm. The ferrite sheet material is utilized in between the phone and head models, and it is found that the simulated value of SAR 1 gm and SAR 10 gm are 1.043 W/Kg and 0.676 W/Kg respectively. The reduction about of 47.68% was observed in this study when a ferrite sheet is attached between the phone and human head models for SAR 1 gm. This SAR reduction is better than the result reported in [5], which is 13% for SAR 1 gm. This is achieved using different radiating powers and impedance factors. Figs. 1.3-1.6 show the SAR value compared with the distance between phone and head models, width of ferrite sheet between 20-40 mm, thickness of ferrite sheet between 2-3.5 mm and height between 40-90 mm respectively.

The reduction efficiency of the SAR depends on its width and height. In order to definitely confirm this, 1 gm and 10 gm average SAR versus distance, width, thickness and height are plotted in the Figs. 1.3-1.6. In Fig. 1.3, it is shown that if the distance between phone and human head models is varied then the SAR value

decreases. This is because the dielectric constant, conductivity, density and magnetic tangent losses are also varied. In Fig. 1.4, it can be observed that the SAR value reduces with the increase of the width of the ferrite sheet. As shown in Fig. 1.5, the SAR value decreases until a thickness of 3 mm, and then a different tendency i.e., it started to increase after 3 mm. The height is varied up to 90 mm in Fig. 1.6. From this figure it can be shown that if the height of the ferrite sheet increases then the SAR value also decreases up to a height of 80 mm, and it started to increase after 80 mm. The results implies that only suppressing the maximum current on the front side of the conducting box contributes significantly to the reduction of spatial peak SAR. This is because the decreased quantity of the power absorbed in the head is considerably larger than that dissipated in the ferrite sheet.

Table 1.2. Comparisons of peak SAR with ferrite sheet

Tissue	SAR value (W/kg)
SAR value for [3]	2.17
SAR value for [10]	2.28
SAR value with ferrite sheet for 1gm	1.043

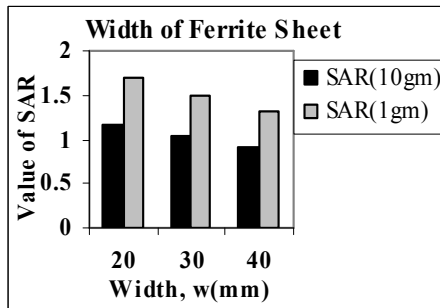


Fig.1.4.SAR value versus the width of the ferrite sheet.

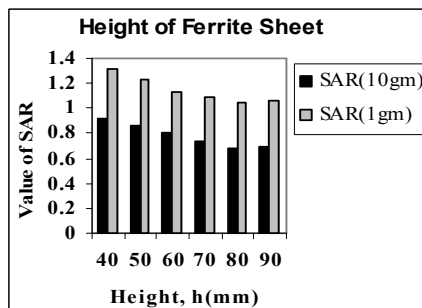


Fig.1.6.SAR value versus the height of the ferrite sheet.

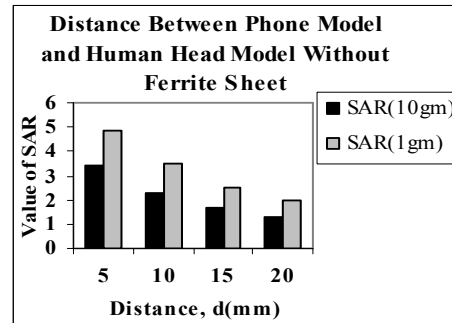


Fig.1.3.SAR value versus the distance between phone model and human head model without ferrite sheet

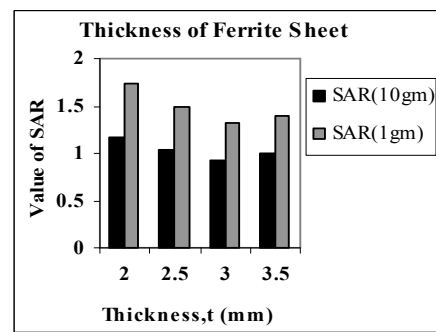


Fig.1.5.SAR value versus the thickness of the ferrite sheet.

6.6 Conclusions

The EM interaction between an antenna and the human head with materials has been discussed in this paper. Utilizing material in the phone model a SAR value is achieved about 0.676 W/Kg for SAR 10 gm and 1.043 W/Kg for SAR 1 gm. Based on the 3-D FDTD method with lossy-Drude model, it is found that the peak SAR 1 gm and SAR 10 gm of the head can be reduced by placing the materials between the antenna and the human head. Numerical results can provide useful information in designing communication equipment for safety compliance.

References

1. Hawang J. N., Fu-chiang chen (2006) Reduction of the peak SAR in the human head with metamaterials. IEEE Trans. on antenna and propagation 54, pp. 3763-3770
2. Huang S. Y., Chen S. Y., Sun J. S. (2005) Far-field effects with human head evaluation of EM emission. Proceedings of the 5th WSEAS Int. Conf. on Applied Electromagnetics, Wireless and Optical Communications. pp. 47-51
3. Wang J., Fujiwara O. (1999) FDTD computation of temperature rise in the human head for portable telephones. IEEE Trans. Microwave Theory Tech. 47 pp. 1528-1534
4. Mimoza I., Luan A., Enver H. (2010) Comparative SAR assessment in adults and children exposed to electromagnetic fields of radio frequency devices. WSEAS Transactions on Communications. 9 (2): pp 105-114
5. Wang J., Fujiwara O. (1997) Reduction of electromagnetic absorption in the human head for portable telephones by a ferrite sheet attachment. IEICE Trans. Commun., E80b pp. 1810-1815
6. Raghavan S., Jayanthi N. (2009) Design of Planar Inverted -F Antenna for wireless applications. WSEAS Transactions on Communications. 8(8):pp.863-872
7. Vaughan R. G., Scott N. L. (1999) Evaluation of antenna configurations for reduced power absorption in the head. IEEE Trans. On Vehicular Technology 48 :(5)
8. Ibrani-Pillana M., Ahma L., Hamiti E. (2008) Human exposure in vicinity of 900 MHz GSM base station antenna. WSEAS Transactions on Communications. 7(4):pp.229-234
9. Ahma L., Ibrani M., Hamiti E. (2009) Assessment of SAR in a human exposed to GSM electromagnetic fields. Proceedings of 13th WSEAS International Conference on Communications. pp. 47-51
10. Kuo C .M, Kuo C. W. (2003) SAR distribution and temperature increase in the human head for mobile communication in IEEE-APS Int. Symp. Dig., Columbus, OH, pp. 1025-1028

Chapter 7

Comparison of a Pressure Correction based Solver VoF Method in Moving Floor Container

R. Ghandriz,¹ M. Goudarzi²

¹MAPNA Boiler Company (MAPNA Group), #16, 16th St, Argentina Sq, Tehran, Iran, r.ghandriz@gmail.com

²Mechanical Department, Bu- Ali Sina University, Mahdie Blvd, Hamadan, Iran, mohsengood@gmail.com

Abstract. In this paper, free surface flow problems in a 2-D container with a moving floor are solved by VoF method. This numerical method (VoF) is used to employ impacted factors on dimensional analysis in laminar and turbulence cases. Comparing laminar and turbulence cases (in free surface flows), in laminar flows, steady state happened sooner and more symmetric circulation is formed. Then, in turbulence cases (with and without surface tension) have been considered and inconsiderable effect of surface tension will be presented. Also VoF method is utilized in laminar flows to study about some impacted factors on dimensional analysis such as different velocity for moving floor, different height of fluid in the container, different fluid properties (density and viscosity), etc. So, in free surface flows, Reynolds number has no magnitude in comparison with without free surface flows and fluid properties should be considered in advance.

Keywords. Cavity, Free Surface Flow, Laminar Flow, Turbulence Flow, VoF Method.

7.1 Introduction

Applying numerical models for flows with free surface has less research background and less attention is paid to numerical models. By considering oil industry and vast offshore facility that exist in some countries, it is necessary to believe numerical models in this state of flows.

The application of this problem is in area of survey and investigation the situation of fluid inside the controller (capacity) in transfer pipe line, also in cases, which the bottom wall of oil or fuel capacitor for some reason such as earthquake or unwanted movement are shocked or fluctuated. Research about these flows like other flows is possible in two methods. First method is using laboratory and experimental models; usage of this choice is so difficult and costly. Point of view, it is not reachable by all researchers. Thus, applying numerical models which could simulate the behavior of flow have to recommend such as the second method. Of course it should be mentioned that the numerical research in the field of flow with free surface has less background than numerical research in the field of flow without free surface. Purpose of flow with free surface is that kind of flows that have a common boundary between two different phases. Both of these phases could be liquid or one of them could be gas. Prevalent equations of flows for both phases are Navier- Stocks equations. As the common boundary was unknown and the complexity of forces that exist in common boundary of these two phases make it harder to model these flows in comparison with classic flows without free surface.

Numerical models, which apply for free surface flows are mainly categorized in two parts as follow:

1. Following the free surface inside a constant grid in order to free surface is always inside the grid. In this method the scalar quantity that introduces the location of free surface is used. Among the applied methods we can cite MaC [1], and VoF [2] methods. These methods are suitable for situation that difference height of free surface at different point is very high or intensive slopes in free surface exist to horizontal surface are suitable.
2. Following the free surface as a solving grid boundary. In these methods, solving grid is exclusive to liquid flow and points of grid do not exist outside of fluid and where free surface is as one of the boundary condition of field should be distinguished continuously. Thus, solving grid in each order of calculation would be regenerated. In this method free surface is nominated like a less thickness surface.

These methods are suitable where difference height of free surface at different points is not very high or sloppily.

Goudarzi and Azimian [3], [4], [5], introduced a numerical method in 2003 and tested this procedure for channel flow. This method is placed in second part and it is practical for free surface flow and also for flow without free surface because in this manner momentum equation has been solved by introducing none hydrostatic pressure term. Also, they have taken advantage of approximate boundary condition instead of exact boundary condition in free surface for some auxiliary channel flows [4], [5]. As their results, because none hydrostatic pressure gradient in direction of channel axis is dominant whereas there is the inconsiderable fluctuations in free surface we can ignore such as these negligible displacements.

In this paper we compared the results of one of laminar cases (that has been solved by VoF Method) with this method for free surface flow in moving floor cavity. In this problem unlike the channel flow, none hydrostatic pressure gradient is important in all directions.

7.2 Problem formulation

Differential equations for a laminar and incompressible fluid flow are Navier- Stocks equations. By choosing Cartesian as reference, dominant equations on 3-D incompressible flows could be introduced. These equations consist of continuous equation, three momentum equation, fraction volume equation and turbulence models equations for turbulence cases. In this Cartesian coordination x-y axes laid on horizontal directions and z axis laid on vertical direction supposed as opposite of gravity acceleration. By considering all above the dominate equation at steady state are written as follow:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0.0 \quad (1.1)$$

$$\begin{aligned} \rho(u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z}) = -\frac{\partial P}{\partial x} \\ + (\mu + \mu_1)(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}) \end{aligned} \quad (1.2)$$

$$\begin{aligned} \rho(u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z}) = -\frac{\partial P}{\partial y} \\ + (\mu + \mu_1)(\frac{\partial^2 v}{\partial x^2} + v \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}) \end{aligned} \quad (1.3)$$

$$\rho(u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z}) = - \frac{\partial P}{\partial z} - \rho g + (\mu + \mu_1) \left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \right) \quad (1.4)$$

Where, u, v, w presented velocity component (m/s) at x, y, z directions, ρ presented fluid density (kg/m^3), g presented gravity intensity (m/s^2) and P presented pressure value (Pa).

It should be noted; in 2-D cavity the third equation in these formulations is unconsidered.

Suppose, two uncombined and incompressible fluids are available, so divergence of the velocity is:

$$\nabla \cdot u = 0 \quad (1.5)$$

For using the fraction volume, the positions of two fluids are important, so for inside one of the fluid $C=1$ and for inside another fluid $C=0$. Cells which are between these two phases are verified as $0 < C < 1$. The first fluid fraction volume is expressed as follow:

$$\frac{\partial C}{\partial t} + \nabla \cdot (uC) = 0 \quad (1.6)$$

No slip condition is another boundary condition for studding the lid driven cavity flow that surrounded the fluid completely. But in free surface flow problem, one of the boundary condition is free surface. In this boundary, kinematic and dynamic conditions should be met. If we assume that free surface flow is a scheme (z) without any breakdown, it will be possible to consider the height of different points proper to a reference surface (figure 1) as following equation:

$$z = h(x, t) \quad (1.7)$$

Where, h denoted liquid height, x presented horizontal direction and t denoted time dependent.

Without any evaporation through free surface, no mass transfer from free surface would be occurred, that is presented as:

$$\frac{Dz}{Dt} = \frac{\partial h}{\partial t} + u_{fs} \frac{\partial h}{\partial x} \quad (1.8)$$

Left hand term illustrates free surface velocity (w_{fs}), so we can rewrite it such as:

$$w_{fs} = \frac{\partial h}{\partial t} + u_{fs} \frac{\partial h}{\partial x} \quad (1.9)$$

Therefore, kinematic conditions for free surface would be demonstrated by no mass transferring.

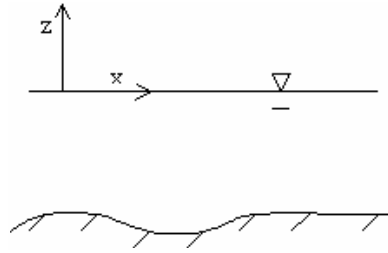


Fig.1.1 Cartesian coordination and reference surface

In order to dynamic condition the free surface forces should be equal which were consisted of the consequent of the tangent forces such as surface tension and shear stress, and the consequent of the vertical forces affected by atmosphere pressure and fluid pressure have to be zero. Frequently, because of negligible density and viscosity of surrounded permanent, shear stress has to be inconsiderable:

$$\partial u / \partial \zeta = 0.0 \quad (1.10)$$

Where, ζ denoted vertical vector through the free surface.

Also, surface tension is negligible for extensive free surface flow. Thus, in this case, dynamic condition is equality of atmosphere pressure (in this paper) and fluid pressure.

$$p_{fs} = p_{atm} \quad (1.11)$$

For atmospheric and free surface attribute, we use *atm*, *fs* subscripts.

By variable replacement we have:

$$p_{fs} = \rho g z_{fs} \quad (1.12)$$

7.3 Problem solution

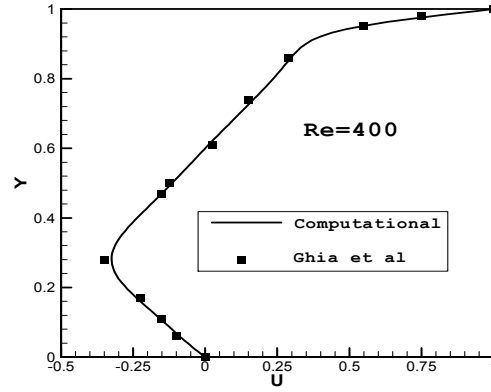
Numerical method is based on grid generation. In free surface flow, the boundary condition of free surface is not obvious; in fact one of the unknown parameter is the configuration and geometry of free surface. First, for solving this problem, we should suppose a model for free surface. Considering it; first grid has to be generated that doesn't confirm to the boundary condition of free surface. Then, by numerical solving and employment of kinematic and dynamic boundary conditions in free surface, the correct boundary geometry and solving grid should be exhibited. Some of the numerically models for solving VoF method were presented by scientist that based on Hert and Nickels' method. This solver estimated the curve of medial surface by vertical and horizontal lines and was structured by upwind and downwind methods. The advantage of upwind method convention is its stability, but this method is much spacious and might be extend in some middle surface cells. Although downwind method is not stable, this able to form the middle surface and it is very useful in that's function. Some of the Vof methods were incremented which are able to offset the stability of upwind method and the advantage for front surface modeling by downwind method [6], [7], [8].

7.3.1 Numerical results

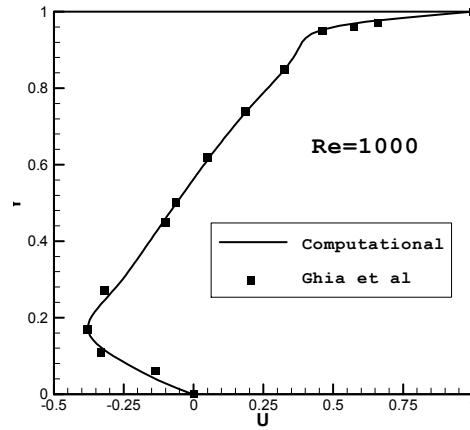
To confirm the accuracy of VoF Method, we solved numerically a 2-D lid driven cavity. The numerical results in comparison with Ghia's [9] results are shown in figure (1.2). The numerical accordance obtained from this method and which one explained by Ghia's [9], portended acceptable accuracy of employed this method, so this procedure would be applicable. By assuring this numerical method ability, we have investigated such as 2-D cavity with moving floor and free surface flow instead of upon rigid wall (figure 1.3). First, we suppose laminar case and Reynolds Number equal to 1000. Therefore, once grid mesh 100*100 and second grid mesh 200*200 has to be used and boundary conditions were defined by steady lateral walls, moving floor by velocity equal to 1 m/s and pressure outlet condition for upon wall. Then, consider the half of this cavity was full of a fluid. Where μ presented viscosity value (kg/ms^{-1}).

$$\frac{\rho_{gas}}{\mu_{gas}} = \frac{1.22}{1.78e-5} \approx 10^5 \text{ s/m}^2$$

$$\frac{\rho_{liq}}{\mu_{liq}} = \frac{10}{0.01} = 10^3 \text{ s/m}^2$$



(a)



(b)

Fig.1.2 Horizontal velocity component diagram in the middle of lid driven cavity. a) $Re=400$. b) $Re=1000$

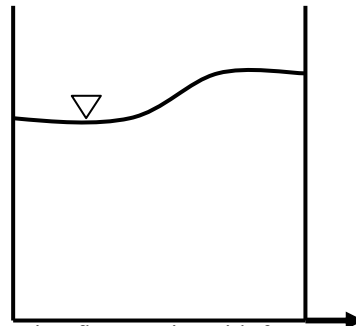


Fig.3 Geometry of the moving floor cavity with free surface

Consider laminar unsteady flow and time step = 0.001, by solving this problem with VoF Method after 36.28s (CPU time = 259200s) the steady state was achieved. Obviously, in figure 1.4 (stream function and vorticity contours) the right corner of the surface flow rises and the circulation is closer to the right side and a visible dip springs in the middling of flow. In turbulence case, we consider the physical theorem as same as laminar case except that the fluid in the cavity was liquid water.

$$\frac{\rho_{liq}}{\mu_{liq}} = \frac{998.2}{0.001} \approx 10^6 \text{ s} / \text{m}^2$$

$$\frac{\rho_{gas}}{\mu_{gas}} = \frac{1.22}{1.78e-5} \approx 10^5 \text{ s} / \text{m}^2$$

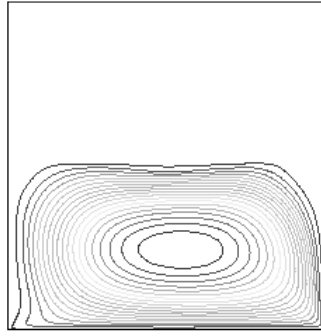


Fig.1.4 Laminar flow in the moving floor cavity with free surface.

Hence, the mesh regeneration has to be harmonized to that statue. Fig. (1.5) demonstrate unsteady solution via VoF Method by time step = 0.03 verged to steady state after 88.76s (CPU time = 432000s) and the circulation traversed the same auxiliary orbit and stood close to the right side.

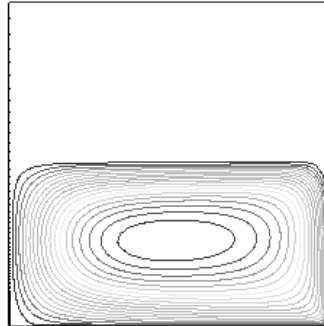


Fig.1.5 Turbulence flow in the moving floor cavity with free surface and without surface tension

Although, the ledge in the right corner have been emerged, the visible dip in the midline of flow was not obvious. Afterwards, this problem was solved by involving surface tension up to 0.07 and considering unsteady solution via VoF Method with time step = 0.03. In this situation, after 118.76s (CPU time = 518400s) that redounded steady state and the same consequences happened such as previous problem (Figure 1.6). These outcomes are derived from stream function and vorticity contours, pressure contours, free surface profile contour, horizontal velocity vector component in the middle of cavity (Figure 12,13a), velocity vectors diagrams (Figure 10,11a) in the steady state for laminar and turbulence cases.

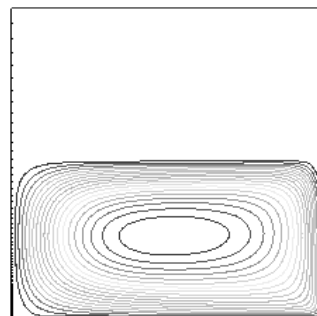


Fig.1.6 Turbulence flow in the moving floor cavity with free surface and surface tension

- Laminar flow achieved to steady state earlier than turbulence flow.

- Comparing the steady state results expose that formed circulation in laminar case is closer to the right side; also, the right lower corner in stream function and vorticity contours is sharp.

- Although in laminar case a small circulation was formed in right upon corner, which was not appeared in the turbulence mode.

- According to horizontal velocity vector component diagram, in laminar flow the circulation is more symmetric than turbulence flow. So, in laminar case these vectors mitigated from bottom to upon side.

- Because of larger amount of diffusion terms in comparison with inertia terms in laminar case with further density content, free surface velocity vector in the middle of the cavity in turbulence case is -0.2 m/s while in laminar mode it is -0.7 m/s. Also, in laminar case velocity vectors diverge to zero value without continuance; these facts actualized the fluctuated free surface laminar flow profile. (Figure 12, 13a)

- Velocity vector of turbulence flow converged to zero value in higher height.

- More continues velocity vectors in turbulence flow demonstrate steadier free surface profile.

- Considering the left part of the velocity diagram concluded that fluctuation near the floor in laminar case is more continuous and further than turbulence mode in the same distance, this reason was occasion of sharpness near the right bottom corner (Figure 1.4).

- Obviously, the vortex flow is the dominant flow in cavity. In this flow, pressure gradient is important in all of directions and there is not any horizontal or vertical conqueror pressure gradient in the fluid flow. According to pressure counters, the difference between minimum and maximum pressure value in laminar and turbulence cases have been 38.15 k Pa, 2774 k Pa. This fact illustrated a little free surface fluctuation is able to vary pressure dominate completely. Therefore, the pressure value reduces in turbulence case more than in laminar case.

- Velocity vectors for both cases in the cavity conceded the previous results. Also, the descent gradient in figure (1.11a), near the floor is less in laminar flow than turbulence flow, hence velocity vectors is more continuous.

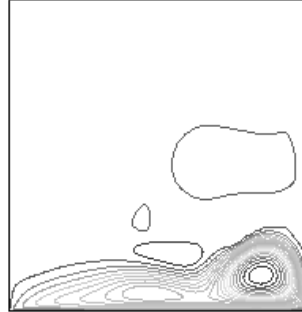


Fig.1.7 Laminar flow in the moving floor cavity with free surface and the height of liquid was

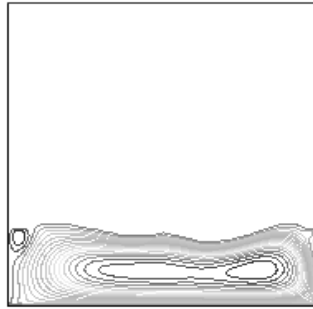


Fig.1.8 Laminar flow in the moving floor cavity with free surface, the height of liquid was $\frac{1}{4}$, by ignoring gas density versus from liquid density

Next, for investigating about free surface flow, which is closer to the moving floor, we suppose the same physical problem and the same property for liquid and gas, by considering $\frac{1}{4}$ height of liquid in the cavity and floor velocity is equal to 2 m/s ($Re = 2000$) and solving this problem via VoF Method, after 26.11 s (CPU time= 172800 s), the steady state was occurred. This moment is lower than previous case. According to figure (1.7), there are many circulations and the initial one is closer to the right side corner. Comparing this result and another one solved numerically by Goudarzi and Azimian [3], [4], [5] (Figure 1.9), verified this method for free surface flow. By examining the results from stream function and velocity contours, pressure contours, phase contours, horizontal velocity component vector in the midline of the cavity (Figures 1.13a, b) and the velocity vectors (Figures 1.11a, b), from outset to steady state, in the same time for both laminar cases, we mentioned such as:

- Laminar flow in the second case (which the height of liquid was $\frac{1}{4}$ and the floor velocity was 2 m/s) attained to the steady state rather than the first case, because of lower height of liquid.

- In the second case, more circulations have been formed, because the velocity of moving floor increased.

- Comparing the circulation promenade to achieve the steady state in both cases, we are able to represent the circulation in the second case has not grow and has not orbit completely opposite the first one.

- Real circulation in the first case of laminar flow was more symmetric than another one, which was derived by the horizontal velocity component diagram. (In the further section, you would follow by receding from floor, velocity vectors in the first case opposite of second one decreased gently up to achieved to free surface.)

Comparing the horizontal velocity component in the midline of the cavity (Figures 1.13 a, b), these results could be decelerated:

- In the first case, after circulation grew and promoted completely, according to C.C.W revolution, horizontal velocity components descended up to free surface, though the descend course in the second one changed in the middle of fluid and incepted the ascendant course earlier than another one, that explained the center of the circulation was closer to free surface in the first case.

- Minimum velocity in the first case occurred on free surface (-0.7 m/s), though in the second one because of forming the adverse circulation, minimum velocity happened in the middle of fluid height (-0.65 m/s).

- According to the phase contours, free surface fluctuated more in the first case that is able to describe by attending to the horizontal velocity component diagram.

- Obviously, the vortex flow is the dominant flow in cavity. In this flow, pressure gradient is important in all of directions and there is not a conqueror pressure gradient in the fluid flow. Due to quicker moving floor, pressure value decreased more in the second laminar case than the first laminar case which is able to consent by pressure difference (38.15 k Pa in the first case and 99.56 k Pa in the second one).

- According to the horizontal velocity component diagram, the descendent gradient of the curve in the first laminar case was much less than another one which proved the continuous variation in velocity vectors.

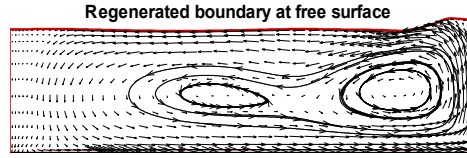


Fig.1.9 Stream function contours for laminar flow in the moving floor cavity with free surface and the height of liquid was 1/4. (Solving numerically by Goudarzi and Azimian [3], [4], [5].

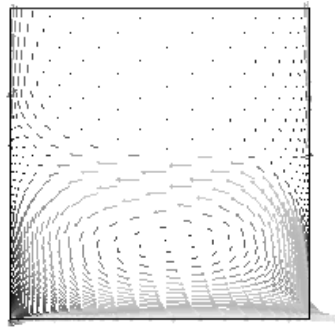


Fig.1.10 Velocity vectors for turbulence flow in the moving floor cavity with free surface

For laminar flow study in this cavity, when the difference between liquid density and gas density was much enough to ignore gas density, first we consider the same physical problem (the height of liquid is equal to $\frac{1}{4}$ and the velocity of moving floor is equal to 2 m/s), if properties of liquid and gas change such as below, it would be approached by these properties and density difference, the results must be unusual. That was because of ignorable gas density. After 235.39 s (CPU time= 777600 s), we get the steady state.

$$\frac{\rho_{liq}}{\mu_{liq}} = \frac{1}{0.001} = 10^3 \text{ s/m}^2$$

$$\frac{\rho_{gas}}{\mu_{gas}} = \frac{0.01}{1.42e-7} \approx 10^5 \text{ s/m}^2$$

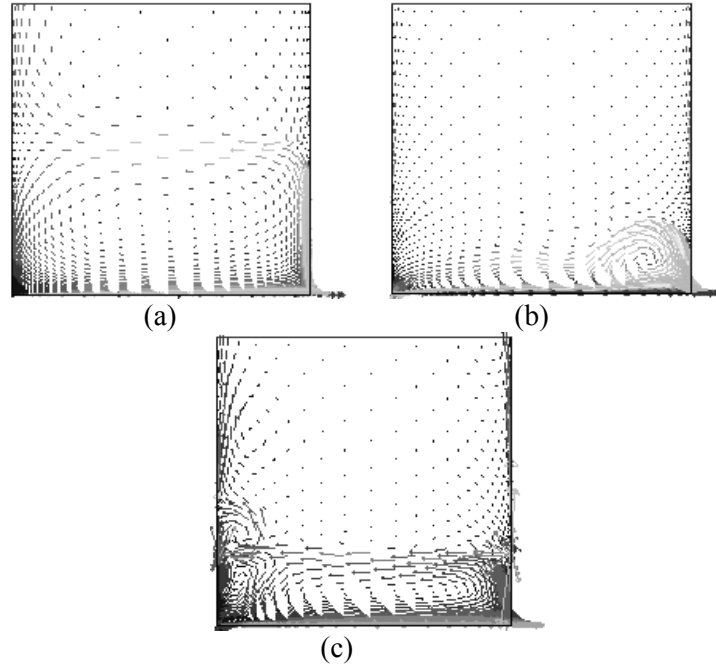


Fig.1.11 Velocity vectors for laminar flows in the moving floor cavity with free surface. a) First case. b) Second case. c) Third case

According to figure (1.8) the result is so different from other laminar cases and the circulation was more extended because of high distinctive between gas viscosity and liquid viscosity. Next, by examining the results of second laminar flow and third laminar flow, which consist of stream function and velocity contours (Figure 1.7, 1.8), phase contours, horizontal velocity component distribution diagram in the midline of cavity (Figure 1.13b, c), velocity vectors diagram (Figures 1.11b, c) for flow in the cavity.

- Second flow achieved to the steady state (26.11 s) rather than third flow.
- The scale and the location of circulation in the second flow are more adjacent than the scale and location of circulation in the third one and the circulation is closer to the right side.
- The circulation of the second flow is more symmetric than another one that is also arising from parabolic zone of horizontal velocity component (Figure 1.13).

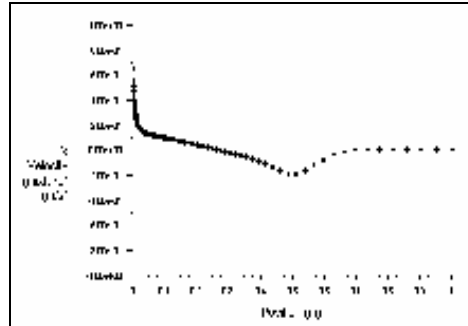


Fig.1.12 Horizontal velocity component distribution in the midline of the cavity for turbulence flow

By ignoring the density of air opposite of water we were not able to persuade the accurate results that should be attended in solving problems by VoF method.

Comparing horizontal velocity component distribution diagram in the midline of cavity with moving floor velocity equal to 2 m/s (Figures 1.13b, c), in the laminar flows following results were stated.

- Due to opposite revolving circulation from the first circulation in the second flow, the minimum velocity occurred almost in the middle of liquid height (-0.65 m/s), whereas the minimum velocity of third flow happened near free surface and it is because of homogenous revolving circulations and it is more than minimum velocity in the second flow (-1.25 m/s).

- Parabolic zone in the horizontal velocity component diagram in the second flow is more symmetric. So, free surface in this flow fluctuated less and the circulation is more symmetric than the third one.

- In the second flow, the value of horizontal velocity component is positive because the powerful circulation near free surface was formed

- According to figures (1.7, 1.8), it is obvious that free surface in the second one is steady and the formed dip is in the reduced region. That could be obtained by observing the horizontal velocity component distribution diagram.

- Because of ignorable gas and liquid density in the third case opposite of second one, the pressure decrease in the third case must be so little. The pressure value reduction in the second one is 99.56 k Pa whereas in the third case, it is 11.53 k Pa. The maximum pressure reduction occurred in free surface.

- Due to less downward slope of parabolic curve in the horizontal velocity component distribution diagram (Figure 1.13b, c), we expected continuous variation for velocity vectors near the moving floor in the third

case (Figure 1.11b, c). Certainly this anticipation realized by comparing these figures, near the lateral walls discontinues distribution of velocity vectors was formed admitted by different assumption in fluid properties.

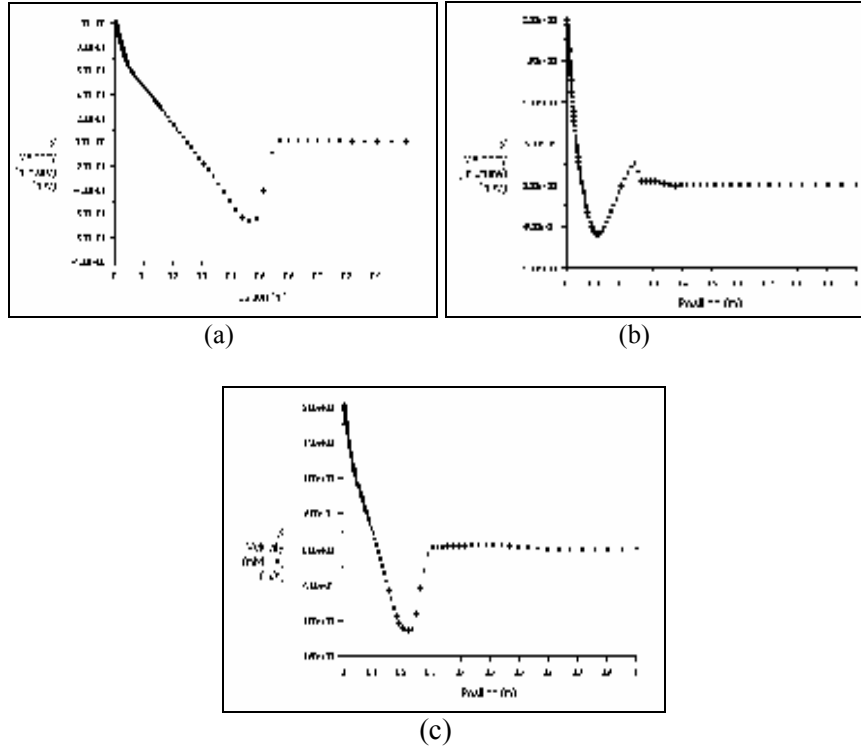


Fig.1.13 Horizontal velocity component distribution in midline of the cavity. a) First case. b) Second case. c) Third case

7.4 Conclusions

As we knew, laminar flows achieve to the steady state rather than turbulence flows. Although, there was the same pressure gradient, the free surface fluctuation gauge in laminar case was further than turbulence case that occurred because of less fluid density value in laminar flow. Also, because of the ignorable surface tension between air and liquid water there was no discrepancy in both turbulence cases.

By this investigation, we can observe this problem in Taylor fluids is similar to Benard problem in convection branch, which occurred because

of the temperature difference between upper and lower part that offered such Benarth case in fluids.

Also, in free surface flows opposite of other flows, Reynolds number has no effect on problem analyses and viscosity and density parameters are important in solving this problem exclusively, by ignoring gas density versus from liquid density, unusual results were achieved. Also, we have to examine considerably in these cases for Reynolds number, as, each steady turbulent in free surface flows could not result steady state in the problem.

References

1. F. H. Harlow, and J. E. Welch, "Numerical Calculation of Time-dependent Viscous Incompressible Flow of Fluid with Free Surface", *Physics of Fluid*, 8 (12): pp.2182-2189, 1965.
2. C. W. Hirt, and B. D. Nichols, "Volume of Fluid (VoF) Method for the Dynamics of Free Boundaries", *Journal of Computational Physics*, 39: pp. 201- 225, 1981.
3. M. Goudarzi, "*Investigation about Diffusion Phenomena in Incorporate Stead of Two Channels*". Doctoral Thesis, Isfahan Industrial University, 2003.
4. M. Goudarzi, A. R. Azimian, "*Numerical Flow Modeling in Open Channel with Free Surface Following Method*", Engineering Department of Tabriz University, 2004.
5. M. Goodarzi, and A. R. Azimian, " Mixing of Two Parallel Flows with Free Surfaces ", *Proceeding of 9ACFM*, 27-31 May, 2002.
6. J. Farmer, L. Martinelli, and A. Jameson, "Fast Multigrid Method for Solving Incompressible Hydrodynamic Problems with Free Surface", *AIAA J.*, 32(6): pp. 1175-1182, 1994.
7. S. E. Norris, "*A Parallel Navier-Stokes Solver for Natural Convection and Free Surface Flow*", Ph.D. Thesis, University of Sydney, pp. 195-215, 2000.
8. J. H. Ferziger, and M. Peric, "*Computational Methods for Fluid Dynamics*", Corrected 2nd printing, Springer, pp. 333-343, 1997.
9. U. Gia, K. Ghia and C. Shin, "High- resolutions for incompressible flow using the Navier-Stocks Equations and a multigrid method", *J. of Computational Physics*, Vol. 48, pp.387-411, 1982.

Chapter 8

Possibilities of Air Quality Modelling based on IF-Sets Theory

V. Olej, P. Hajek

Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentská 84, 532 10 Pardubice, Czech Republic, vladimir.olej@upce.cz, petr.hajek@upce.cz

Abstract. The paper presents a design of parameters for air quality classification of districts into classes according to their pollution. Therefore, the design of air quality classification is presented by IF-sets introduced by K.T. Atanassov. In the next part of the paper we present basic notions of IF-sets for the design of the tree hierarchical IF-inference systems of Mamdani type. Moreover, the paper describes air quality modelling, the design of membership functions and non-membership functions, if-then rules of individual subsystems, inference mechanism and analysis of the results.

Keywords. Air quality, IF-sets, hierarchical IF-inference systems of Mamdani type, classification.

8.1 Introduction

At this time there are several generalizations of fuzzy set theory for various objectives [1, 2]. Intuitionistic fuzzy sets (IF-sets) theory represents one of the generalizations, the notion introduced by K.T. Atanassov [3, 4]. The concept of IF-sets can be viewed as an alternative approach to define a fuzzy set in cases where available information is not sufficient for the definition of an imprecise concept by means of a conventional fuzzy set. In this article we will present IF-sets as a tool for

reasoning in the presence of imperfect fact and imprecise knowledge. The IF-sets are for example also suitable for the air quality modelling as they provide a good description of object attributes by means of membership functions μ and non-membership functions ν . They also present a strong possibility to express uncertainty.

The paper presents the basic notions of IF-sets and the design of the hierarchical IF-inference systems of Mamdani type. Hereby, it points out the reduction of if-then rules. Based on [5], the output of the IF-inference system is defined in general. In the next part of the paper, we design and formalize the tree hierarchical IF-inference system of Mamdani type for the classification of the i -th district $o_i^t \in O$, $O = \{o_1^t, o_2^t, \dots, o_i^t, \dots, o_n^t\}$ in time t to the j -th class $\omega_{i,j}^t \in \Omega$, $\Omega = \{\omega_{1,j}^t, \omega_{2,j}^t, \dots, \omega_{i,j}^t, \dots, \omega_{n,j}^t\}$. Moreover, the classification of the i -th district $o_i^t \in O$ in time t to the j -th class $\omega_{i,j}^t \in \Omega$ presented in the paper assists state administration to evaluate air quality. The knowledge of notable experts in the field of air quality measuring gives support to the results of the classification.

8.2 Problem formulation

Harmful substances in the air represent the parameters of air quality modelling. They are defined as the substances emitted into the external air or formed secondarily in the air which harmfully influent the environment directly, after their physical or chemical transformation or eventually in interaction with other substances. Except the harmful substances, other components also influence the overall air pollution. For example, solar radiation, the speed or the direction of wind, air humidity and air pressure represent these components. Both, the parameters concerning the harmful substances in the air and the meteorological parameters influence the air quality development. The interaction of both types of parameters can cause an increase of air pollution and influence human health this way. The design of parameters can be realized as presented in Table 1.1. Based on the presented facts, the following data matrix \mathbf{P} can be created

$$\mathbf{P} = \begin{matrix} & \begin{matrix} x_1^t & \dots & x_k^t & \dots & x_m^t \end{matrix} \\ \begin{matrix} o_1^t \\ \dots \\ o_i^t \\ \dots \\ o_n^t \end{matrix} & \left[\begin{array}{ccccc} x_{1,1}^t & \dots & x_{1,k}^t & \dots & x_{1,m}^t \\ \dots & \dots & \dots & \dots & \dots \\ x_{i,1}^t & \dots & x_{i,k}^t & \dots & x_{i,m}^t \\ \dots & \dots & \dots & \dots & \dots \\ x_{n,1}^t & \dots & x_{n,k}^t & \dots & x_{n,m}^t \end{array} \right] \end{matrix} \begin{matrix} \omega_{1,j}^t \\ \vdots \\ \omega_{i,j}^t \\ \vdots \\ \omega_{n,j}^t \end{matrix},$$

Table 1.1. Parameters design for air quality modelling

Parameters	
Harmful substances	$x_1^t = \text{SO}_2$, SO_2 is sulphur dioxide. $x_2^t = \text{O}_3$, O_3 is ozone. $x_3^t = \text{NO}$, NO_2 (NO_x) are nitrogen oxides. $x_4^t = \text{CO}$, CO is carbon monoxide. $x_5^t = \text{PM}_{10}$, PM_{10} is particulate matter (dust).
Meteorological	$x_6^t = \text{SW}$, SW is the speed of wind. $x_7^t = \text{DW}$, DW is the direction of wind. $x_8^t = T_3$, T_3 is the temperature 3 meters above the Earth's surface. $x_9^t = \text{RH}$, RH is relative air humidity. $x_{10}^t = \text{AP}$, AP is air pressure. $x_{11}^t = \text{SR}$ is solar radiation.

where $o_i^t \in O$ are objects (districts) in time t , x_k^t is the k -th parameter in time t , $x_{i,k}^t$ is the value of the parameter x_k^t for the i -th object $o_i^t \in O$, $\omega_{i,j}^t \in \Omega$ is the j -th class assigned to the i -th object $o_i^t \in O$, $\mathbf{p}_i^t = (x_{i,1}^t, x_{i,2}^t, \dots, x_{i,k}^t, \dots, x_{i,m}^t)$ is the i -th pattern, $\mathbf{x}^t = (x_1^t, x_2^t, \dots, x_k^t, \dots, x_m^t)$ is the parameters vector. The air quality evaluation is based on the results of weight concentrations' measures of substances in the air. The evaluation takes the possible effects on human health into account. New limits specified in the government order of the Czech Republic No: 350/2002 Coll. set the limits of pollutants, the conditions and the procedure of air quality monitoring, evaluation and management. The classes $\omega_{i,j}^t \in \Omega$ for air quality evaluation defined in this order are presented in Table 1.2.

Table 1.2. Air quality classes $\omega_{i,j}^t \in \Omega$

	Class description
$\omega_{i,1}^t$	Clean air, very healthy environment.
$\omega_{i,2}^t$	Satisfactory air, healthy environment.
$\omega_{i,3}^t$	Slightly polluted air, acceptable environment.
$\omega_{i,4}^t$	Polluted air, environment dangerous for sensitive population.
$\omega_{i,5}^t$	High polluted air, environment dangerous for the whole population.

Air quality modelling [6] represents a classification problem. By the defining the problem in this manner it is possible for it to be modelled by unsupervised methods (if classes $\omega_{i,j}^t \in \Omega$ are not known). The districts (bus stops: Cihelna, Dubina, Polabiny, Rosice, Rybitví, Srnojedy, crossroads: Palacha-Pichlova, Square Republic, Spa Bohdaneč, chemical factory of Paramo) in the city of Pardubice, the Czech Republic, have no class $\omega_{i,j}^t \in \Omega$ assigned.

8.3 Hierarchical IF-inference systems design

The concept of IF-sets is the generalization of the concept of fuzzy sets, the notion introduced by L.A. Zadeh [7]. The theory of IF-sets is well suited to deal with vagueness. Recently, the IF-sets have been used to IF-sets classification models which can accommodate imprecise information.

Let a set X be a non-empty fixed set. An IF-set A in X is an object having the form [3, 4]

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in X \}, \quad (1.1)$$

where the function $\mu_A: X \rightarrow [0,1]$ defines the degree of membership function $\mu_A(x)$ and the function $\nu_A: X \rightarrow [0,1]$ defines the degree of non-membership function $\nu_A(x)$, respectively, of the element $x \in X$ to the set A , which is a subset of X , and $A \subset X$, respectively; moreover for every $x \in X$, $0 \leq \mu_A(x) + \nu_A(x) \leq 1$, $\forall x \in X$ must hold.

The amount $\pi_A(x) = 1 - (\mu_A(x) + \nu_A(x))$ is called the hesitation part, which may cater to either membership value or non-membership value, or both. For each IF-set in X , we will call $\pi_A(x)$ as the intuitionistic index of the element x in set A . It is a hesitancy degree of x to A . It is obvious that $0 \leq \pi_A(x) \leq 1$ for each $x \in X$. The value denotes a measure of non-determinacy. The intuitionistic indices $\pi_A(x)$ are such that the larger $\pi_A(x)$ the higher a hesitation margin of the decision maker. Intuitionistic indices allow us to calculate the best final results (and the worst one) we can expect in a process leading to a final optimal decision.

Next we define an accuracy function H to evaluate the degree of accuracy of IF-set by the form $H(A) = \mu_A(x) + \nu_A(x)$, where $H(A) \in [0,1]$. From the definition H , it can be also expressed as follows $H(A) = \mu_A(x) + \nu_A(x) = 1 - \pi_A(x)$. The larger value of $H(A)$, the more the degree of accuracy of the IF-set A .

Let there exists a general IF-inference system defined in [5]. Then it is possible to define its output y_η as

$$y_\eta = (1 - \pi_A(x)) \times y_\mu + \pi_A(x) \times y_\nu, \quad (1.2)$$

where y_μ is the output of the fuzzy inference system (FIS) using the membership function $\mu_A(x)$, y_v is the output of the FIS using the non-membership function $v_A(x)$.

Let $x_1^1, x_2^1, \dots, x_1^l, \dots, x_m^l$ be input variables, and let $y_{\eta}^{1,1}, y_{\eta}^{1,2}, \dots, y_{\eta}^{q,1}$ be the outputs of subsystems $FIS_{\eta}^{1,1}, FIS_{\eta}^{1,2}, \dots, FIS_{\eta}^{q,1}$, where $\eta=\mu$ are membership functions ($\eta=v$ are non-membership functions). Then, if-then rules $R^{h_{1,1}}, R^{h_{1,2}}, \dots, R^{h_{q,1}}$ of the tree hierarchical IF-inference system, presented in Fig. 1.1,

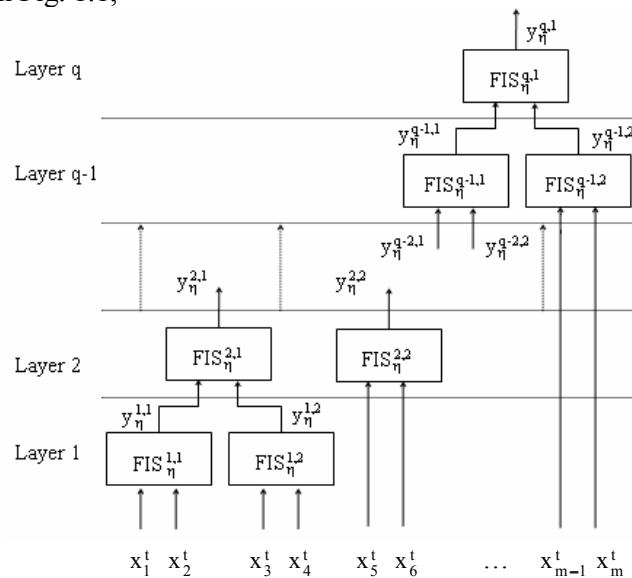


Fig. 1.1. A tree I-hierarchical fuzzy inference system

where q is the number of layers, can be defined as follows

$$\begin{array}{ll}
\text{Layer 1: FIS}_{\eta}^{1,1} & R^{h_{1,1}} : \text{if } x_1^t \text{ is } A_1^{h_{1,1}} \text{ AND } x_2^t \text{ is } A_2^{h_{1,1}} \\
& \quad \text{then } y_{\eta}^{1,1} \text{ is } B^{h_{1,1}}, \\
\text{FIS}_{\eta}^{1,2} & R^{h_{1,2}} : \text{if } x_3^t \text{ is } A_3^{h_{1,2}} \text{ AND } x_4^t \text{ is } A_4^{h_{1,2}} \\
& \quad \text{then } y_{\eta}^{1,2} \text{ is } B^{h_{1,2}}, \\
\text{Layer 2: FIS}_{\eta}^{2,1} & R^{h_{2,1}} : \text{if } y_{\eta}^{1,1} \text{ is } B^{h_{1,1}} \text{ AND } y_{\eta}^{1,2} \text{ is } B^{h_{1,2}} \\
& \quad \text{then } y_{\eta}^{2,1} \text{ is } B^{h_{2,1}}, \\
\text{FIS}_{\eta}^{2,2} & R^{h_{2,2}} : \text{if } x_5^t \text{ is } A_5^{h_{2,2}} \text{ AND } x_6^t \text{ is } A_6^{h_{2,2}} \\
& \quad \text{then } y_{\eta}^{2,2} \text{ is } B^{h_{2,2}}, \dots,
\end{array}$$

Layer q: $FIS_{\eta}^{q,1} \quad R^{h_{q,1}} : \text{if } y_{\eta}^{q-1,1} \text{ is } B^{h_{q-1,1}} \text{ AND } y_{\eta}^{q-1,2} \text{ is } B^{h_{q-1,2}}$
 then $y_{\eta}^{q,1}$ is $B^{h_{q,1}}$,

where: - $h_{1,1} = h_{1,2} = \dots = h_{q,u} = \{1, 2, \dots, k^m\}$, $u=1, 2$,
 - $A_1^{h_{1,1}}, A_2^{h_{1,1}}, \dots, A_n^{h_{q,1}}$ are linguistic variables corresponding to fuzzy sets represented as $\eta_1^{h_{1,1}}(x_1^t), \eta_2^{h_{1,1}}(x_1^t), \dots, \eta_m^{h_{q,1}}(x_1^t)$,
 - $B^{h_{1,1}}, B^{h_{1,2}}, \dots, B^{h_{q,1}}$ are linguistic variables corresponding to fuzzy sets represented as $\eta^{h_{1,1}}(y_{\eta}^{1,1}), \eta^{h_{1,2}}(y_{\eta}^{1,2}), \dots, \eta^{h_{q,1}}(y_{\eta}^{q,1})$,
 - $\eta_{B^{h_{1,1}}}(y_j^{1,1}), \eta_{B^{h_{1,2}}}(y_j^{1,2}), \dots, \eta_{B^{h_{q,1}}}(y_j^{q,1})$ are membership function $\eta=\mu$ (non-membership function $\eta=v$) values of aggregate fuzzy set for outputs $y_j^{1,1}, y_j^{1,2}, \dots, y_j^{q,1}$.

Similarly, it is possible to design and define a cascade hierarchical IF-inference system and various others, hybrid hierarchical IF-inference systems. The outputs $y_{\eta}^{1,1}, y_{\eta}^{1,2}, \dots, y_{\eta}^{q,1}$ of particular subsystems $FIS_{\eta}^{1,1}, FIS_{\eta}^{1,2}, \dots, FIS_{\eta}^{q,1}$ of the tree hierarchical IF-inference system can be expressed by using defuzzification method Center of Gravity (COG) [8, 9] as

$$y_{\eta}^{1,1}(B^{h_{1,1}}) = \frac{\sum_{j=1}^q y_j^{1,1} \times \eta_{B^{h_{1,1}}}(y_j^{1,1})}{\sum_{j=1}^q \eta_{B^{h_{1,1}}}(y_j^{1,1})}, \dots, y_{\eta}^{q,1}(B^{h_{q,1}}) = \frac{\sum_{j=1}^q y_j^{q,1} \times \eta_{B^{h_{q,1}}}(y_j^{q,1})}{\sum_{j=1}^q \eta_{B^{h_{q,1}}}(y_j^{q,1})},$$

and the outputs of particular subsystems $FIS_{\eta}^{1,1}, FIS_{\eta}^{1,2}, \dots, FIS_{\eta}^{q,1}$ in each layer of the hierarchical IF-inference system are calculated as follows

$$y_{\eta}^{r,s}(B^{h_{r,s}}) = (1 - \pi_{\mu}^{r,s}) \times y_{\mu}^{r,s}(B^{h_{r,s}}) + \pi_{\nu}^{r,s} \times y_{\nu}^{r,s}(B^{h_{r,s}}), \quad (1.3)$$

for $r = 1, 2, \dots, q$, $s = 1, 2$.

8.4 Modelling and analysis of the results

Air quality modelling is realized by the tree hierarchical IF-inference system with inputs parameters $\mathbf{x}^t = (x_1^t, x_2^t, \dots, x_k^t, \dots, x_m^t)$, $m=11$, outputs $y_{\eta}^{1,1}, y_{\eta}^{1,2}, \dots, y_{\eta}^{q,1}$ of individual subsystems $FIS_{\eta}^{1,1}, FIS_{\eta}^{1,2}, \dots, FIS_{\eta}^{q,1}$, $q=6$. The design of specific tree hierarchical IF-inference system results from the recommendation of experts in given field. Thus, it simulates their decision-making process. The design of input (output) membership functions μ and non-membership functions ν is based on the limits specified by the government and on the recommendation of experts.

As an example, the input (output) membership functions μ for input parameters x_1^t and x_2^t ($y_\mu^{1,1}$) of the particular subsystem $FIS_\mu^{1,1}$ are presented in Fig. 1.2 (Fig. 1.3). In a similar manner, the input (output) non-membership functions ν for input parameters x_1^t and x_2^t ($y_\nu^{1,1}$) of the subsystem $FIS_\nu^{1,1}$ are presented in Fig. 1.4 (Fig. 1.5). These functions are designed for an example of intuitionistic index $\pi=0.05$.

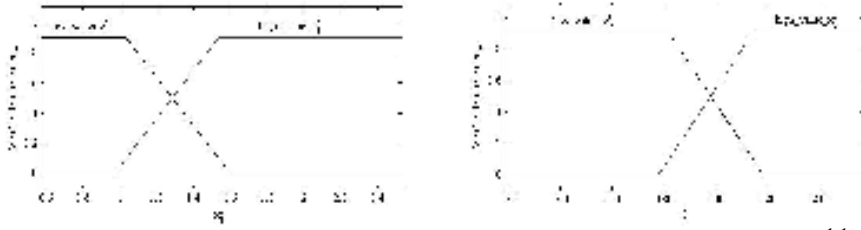


Fig. 1.2. Input membership functions μ for x_1^t and x_2^t of subsystem $FIS_\mu^{1,1}$

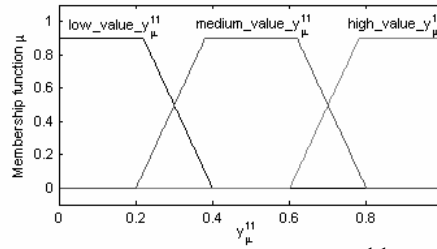


Fig. 1.3. Output membership functions μ for $y_\mu^{1,1}$ of subsystem $FIS_\mu^{1,1}$



Fig. 1.4. Input non-membership functions ν for x_1^t and x_2^t of subsystem $FIS_\nu^{1,1}$

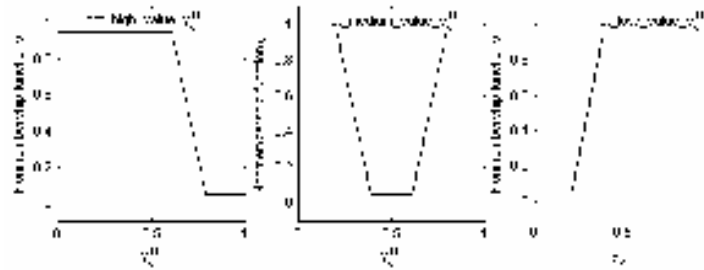


Fig. 1.5. Output non-membership functions ν for $y_\nu^{1,1}$ of subsystem $FIS_\nu^{1,1}$

Base of if-then rules for example for subsystem $FIS_{\eta}^{1,1}$ is defined as follows:

for $\eta=\mu$

$FIS_{\mu}^{1,1}$ R^1 : if x_1^t is low_value_ x_1^t AND x_2^t is low_value_ x_2^t
then $y_{\mu}^{1,1}$ is low_value_ $y_{\mu}^{1,1}$,

$FIS_{\mu}^{1,1}$ R^2 : if x_1^t is low_value_ x_1^t AND x_2^t is high_value_ x_2^t
then $y_{\mu}^{1,1}$ is medium_value_ $y_{\mu}^{1,1}$,

$FIS_{\mu}^{1,1}$ R^3 : if x_1^t is high_value_ x_1^t AND x_2^t is low_value_ x_2^t
then $y_{\mu}^{1,1}$ is medium_value_ $y_{\mu}^{1,1}$,

$FIS_{\mu}^{1,1}$ R^4 : if x_1^t is high_value_ x_1^t AND x_2^t is high_value_ x_2^t
then $y_{\mu}^{1,1}$ is high_value_ $y_{\mu}^{1,1}$.

for $\eta=v$

$FIS_v^{1,1}$ R^1 : if x_1^t is not_low_value_ x_1^t AND x_2^t is not_low_value_ x_2^t
then $y_v^{1,1}$ is not_low_value_ $y_v^{1,1}$,

$FIS_v^{1,1}$ R^2 : if x_1^t is not_low_value_ x_1^t AND x_2^t is not_high_value_ x_2^t
then $y_v^{1,1}$ is not_medium_value_ $y_v^{1,1}$,

$FIS_v^{1,1}$ R^3 : if x_1^t is not_high_value_ x_1^t AND x_2^t is not_low_value_ x_2^t
then $y_v^{1,1}$ is not_medium_value_ $y_v^{1,1}$,

$FIS_v^{1,1}$ R^4 : if x_1^t is not_high_value_ x_1^t AND x_2^t is not_high_value_ x_2^t
then $y_v^{1,1}$ is not_high_value_ $y_v^{1,1}$.

The inference mechanism of particular subsystems $FIS_{\eta}^{1,1}, FIS_{\eta}^{1,2}, \dots, FIS_{\eta}^{q,1}$ involves also the process of implication (MIN method) and aggregation (MAX method) within if-then rules, and the process of defuzzification by COG method of obtained outputs to the crisp values. Data measured monthly during six years ($n=720$) was used for air quality modelling. Yearly means of these measurements was applied for the classification, i.e. $n=60$. Further, the classifiers assigned the best $\omega_{i,1}^t$ and the worst $\omega_{i,5}^t$ class (from the original five classes) to a minimum number of districts. Moreover, the classification into three classes ($\omega_{i,2}^t, \omega_{i,3}^t, \omega_{i,4}^t$) considers also the recommendation of experts. The output $y_{\eta}^{6,1}$ (the classification of the i -th district $\omega_i^t \in O$ in time t to the j -th class $\omega_{i,j}^t \in \Omega$, $j=2,3,4$) of the designed tree hierarchical IF-inference system (the frequencies f of the classes $\omega_{i,j}^t \in \Omega$) is presented in Fig. 1.6. The classification problem works with the set of input patterns \mathbf{p}_i^t assigned to one of the classes $\omega_{i,j}^t \in \Omega$. From Fig. 1.6 it also results that in the neighbourhood of class $\omega_{i,3}^t$ the highest occurrence of classified districts is located in time t . Similarly, classes $\omega_{i,2}^t$ and $\omega_{i,4}^t$ can be described with the difference that more districts are located in the neighbourhood of the class $\omega_{i,4}^t$. Thus, it is obvious that the tree hierarchical IF-inference system designed this way better models imperfect fact and imprecise knowledge

better models imperfect fact and imprecise knowledge than standard hierarchical fuzzy inference system.

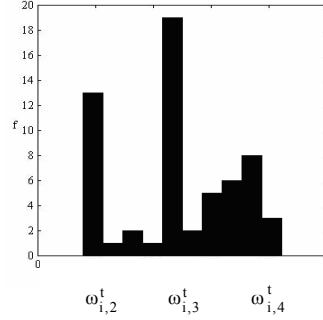


Fig. 1.6. The output $y_{\eta}^{6,1}$ of the tree hierarchical IF-inference system (the frequencies f of the classes)

8.5 Conclusions

The air quality modelling has been focused for example on the air quality parameters prediction [10] and modelling by multi-agents systems [11] so far while classification of the district [12, 13, 14]. Air quality modelling in generally is currently realized by methods combining mathematical-statistical methods and expert opinion. Some of the methods do not make it possible to work with uncertainty and the expert knowledge; the others feature imprecise classification of assessed objects. Therefore, the model based on IF-sets is designed in the paper as they allow processing uncertainty and the expert knowledge. IF-sets can be viewed in the context as a proper tool for representing hesitancy concerning both membership and non-membership of an element to a set. Based on [5], the paper presents the design of the tree hierarchical IF-inference system of Mamdani type. Similarly, it is possible to design and define a cascade and various others, hybrid hierarchical IF-inference systems. The hierarchical IF-inference system defined this way works more effective than the standard hierarchical fuzzy inference systems as it provides stronger possibility to accommodate imprecise information and better model imperfect fact and imprecise knowledge. For example, we can express the fact that parameters $\mathbf{x}^t = (x_1^t, x_2^t, \dots, x_k^t, \dots, x_m^t)$ of the i -th object $o_i^t \in O$ changes and are not quite clear. The measurements of the air quality parameters were realized by the mobile monitoring system.

The model design was carried out in Matlab in MS Windows XP operation system.

Acknowledgement. This work was supported by the scientific research project of Ministry of Environment, the Czech Republic under Grant No: SP/4i2/60/07.

References

- 1 Dubois D, Gottwald S, Hajek P, Kacprzyk J, Prade H (2005) Terminological difficulties in fuzzy set theory - The case of Intuitionistic Fuzzy Sets. *Fuzzy Sets and Systems* 156: 485-491
- 2 Atanassov KT (2005) Answer to D. Dubois, S. Gottwald, P. Hajek, J. Kacprzyk, H. Prade's Terminological difficulties in fuzzy set theory - The case of Intuitionistic Fuzzy Sets. *Fuzzy Sets and Systems* 156: 496-499
- 3 Atanassov KT (1986) Intuitionistic fuzzy sets. *Fuzzy Sets and Systems* 20: 87-96
- 4 Atanassov KT (1999) Intuitionistic fuzzy sets. Springer, Berlin Heidelberg Germany
- 5 Montiel O, Castillo O, Melin P, Sepúlveda R (2008) Mediative fuzzy logic: A new approach for contradictory knowledge management. *Soft Computing* 20: 251-256
- 6 Olej V, Hájek P (2008) Air quality modelling by Kohonen's self-organizing feature maps and intuitionistic fuzzy sets. In: Pobil AP (ed) *Artificial Intelligence and Soft Computing*. ACTA Press, Calgary, 22-27
- 7 Zadeh LA (1965) Fuzzy sets. *Information and Control* 8: 338-353
- 8 Olej V, Hájek P (2007) Hierarchical structure of fuzzy inference systems design for municipal creditworthiness modelling. *WSEAS Transactions on Systems and Control* 2: 162-169
- 9 Pedrycz W (1993) Fuzzy control and fuzzy systems. Research Studies Press Ltd., England
- 10 Nicolescu CL et al (2008) Mapping air quality: An assessment of the pollutants dispersion in inhabited areas to predict and manage environmental risks. *WSEAS Transactions on Environment and Development* 4: 1078-1088
- 11 Iliadis LS, Papaleonidas A (2009) Intelligent agents networks employing hybrid reasoning: Application in air quality monitoring and improvement. *Communications in Computer and Information Science* 43: 1-16
- 12 Vongmahadlek CH, Satayopas B (2007) Applicability of RAMS for a simulation to provide inputs to an air quality model: Modelling evaluation and sensitivity test. *WSEAS Transactions on Environment and Development* 8: 129-138
- 13 San Jose R, Perez JL, Gonzalez RM (2006) Air quality CFD and mesoscale modelling simulations: Madrid case study. *WSEAS Transactions on Environment and Development* 10: 1291-1296

- 14 Hájek P, Olej V (2008) Air quality modelling by Kohonen's self-organizing feature maps and LVQ neural networks. WSEAS Transactions on Environment and Development 4: 45-55

Chapter 9

Reduction Method with Finite-Difference Approximation for the Model of Small Transverse Vibrations in Thin Elastic Plates and some Applications

A. A. Kuleshov

Keldysh Institute of Applied Mathematics
of Russian Academy of Sciences,
University Miuskaya sq., 4-A, Moscow, 125047 Russia
E-mail: andrew_kuleshov@mail.ru

Abstract. The problem of small transverse vibrations in a thin elastic plate of variable thickness with a bending moment and a shearing force on the plate contour is considered. The elaborated numerical method for the problem solution is described. It is based on a reduction of the initial partial differential equation to a system of equations with the first-order time derivatives. The results of numerical simulation for the problem of floating ice vibrations caused by moving loads are discussed.

Keywords. Thin elastic plate, transverse vibrations, numerical method and applications.

9.1 Introduction

The mathematical model of transverse vibrations in thin elastic plates has applications in various fields of science and engineering [1-3]. In particular, in geocology it can be applied to the problem of floating ice vibrations caused by various moving loads. It can also be applied to the

problem of a plate material strength and to the problem of transverse vibrations in oceanic lithosphere plates [4] because thin oceanic lithosphere (as compared with thickness on the continental lithosphere) consists of a relatively uniform basalt layer and can be considered as the thin elastic plate. In addition to the above mentioned applications there have been other applications (e.g., in structural mechanics). And yet despite the fact that there is a wide range of applications, the numerical methods to solve the above problem have not been developed sufficiently. In present paper, a new efficient numerical method to solve the above problem is described and some results of numerical simulation for the problem of vibrations in floating ice caused by moving loads are discussed.

9.2 Problem formulation

The equation of small transverse vibrations in a thin isotropic elastic plate of variable thickness $h(x,y)$ lying on the Winkler elastic base has the form [5]:

$$\rho h W_{tt} + \Delta(D\Delta W) - (1-\sigma)(D_{yy}W_{xx} - 2D_{xy}W_{xy} + D_{xx}W_{yy}) + aW = F, \quad (1)$$

where $(x,y) \in \Omega$, $W(x,y,t)$ is the plate deflection measured along the z axis, ρ is the density of the plate material, $D = Eh^3/[12(1-\sigma^2)]$ is the cylindrical stiffness of the plate, E is the modulus of elasticity, σ is Poisson's ratio of the plate material, aW is the reaction of elastic base (reactive pressure) proportional to the plate deflection according to the Winkler model, $a = \text{const} > 0$, F is the external force given on a surface of a plate.

On the curvilinear contour of a plate $\partial\Omega$ general (Kirchhoff-generalized) conditions: the bending moment $M_{\partial\Omega}$ and the vertical shearing force $N_{\partial\Omega}$ are given:

$$B_1 W = M_{\partial\Omega}, \quad (x,y) \in \partial\Omega,$$

$$B_2 W = N_{\partial\Omega}, \quad (x,y) \in \partial\Omega,$$

where

$$B_1 W = -D\Delta W - D(1-\sigma)[\sin 2\theta W_{xy} - \sin^2\theta W_{xx} - \cos^2\theta W_{yy}],$$

$$B_2 W = -\frac{\partial(D\Delta W)}{\partial \mathbf{n}} -$$

$$-(1-\sigma)\frac{\partial}{\partial \mathbf{l}}\left[D(\sin\theta\cos\theta(W_{yy}-W_{xx})+\cos 2\theta W_{xy})\right]+$$

$$+(1-\sigma)\left[\sin\theta D_y W_{xx}+\cos\theta D_x W_{yy}-(\sin\theta D_x+\cos\theta D_y)W_{xy}\right],$$

\mathbf{n} is the outward normal to the contour $\partial\Omega$, \mathbf{l} is the tangent to the contour, θ is the angle between normal to the contour and the positive OX axis.

The initial conditions:

$$W|_{t=0} = \varphi(x, y), \quad W_t|_{t=0} = \psi(x, y).$$

A finite-difference approximation of the equation (1) based on a three-level finite-difference scheme will have a multipoint stencil and its numerical implementation will be complicated. Our new approach [6] to construction of a numerical method for the problem solution based on reduction method with finite-difference approximation is described in the next section.

9.3 Numerical method for the problem solution

We use the well-known approach to reducing a high-order partial differential equation to a system of equations of smaller order. We reduce the initial equation, which is of the second order on time, by a system of equations with first-order on time and a two-level implicit finite-difference scheme is developed for solving this system. For this purpose at first we will write down the equation (1) in an equivalent form [5]:

$$\rho h \frac{\partial^2 W}{\partial t^2} = \frac{\partial^2 M_x}{\partial x^2} - 2 \frac{\partial^2 M_{xy}}{\partial x \partial y} + \frac{\partial^2 M_y}{\partial y^2} - aW + F, \quad (2)$$

where $M_x = -D(W_{xx} + \sigma W_{yy})$, $M_y = -D(W_{yy} + \sigma W_{xx})$ are bending moments, $M_{xy} = (1-\sigma)DW_{xy}$ is a torsion moment. Then in the equation of plate vibrations (2), we make the substitution $S = W_t$. Further we differentiate on t the formulas for the bending moments M_x and M_y and

make the substitution $S = W_t$, too. As a result, we receive the system of the equations of the first order on time:

$$\rho h \frac{\partial S}{\partial t} = \frac{\partial^2 M_x}{\partial x^2} - 2 \frac{\partial^2 M_{xy}}{\partial x \partial y} + \frac{\partial^2 M_y}{\partial y^2} - aW + F, \quad (3)$$

$$\frac{\partial M_x}{\partial t} = -D \frac{\partial^2 S}{\partial x^2} - \sigma D \frac{\partial^2 S}{\partial y^2} \quad (4)$$

$$\frac{\partial M_y}{\partial t} = -\sigma D \frac{\partial^2 S}{\partial x^2} - D \frac{\partial^2 S}{\partial y^2} \quad (5)$$

$$\frac{\partial W}{\partial t} = S \quad (6)$$

with boundary conditions:

$$M_x \cos^2 \theta + M_y \sin^2 \theta - M_{xy} \sin 2\theta = M_{\alpha\Omega}, \quad (x, y) \in \partial\Omega,$$

$$Q_x \cos \theta + Q_y \sin \theta - \frac{\partial M_{nl}}{\partial l} = N_{\alpha\Omega}, \quad (x, y) \in \partial\Omega,$$

where

$$Q_x = \frac{\partial M_x}{\partial x} - \frac{\partial M_{xy}}{\partial y}, \quad Q_y = \frac{\partial M_y}{\partial y} - \frac{\partial M_{xy}}{\partial x}$$

are vertical shearing forces,

$$M_{nl} = M_{xy}(\cos^2 \theta - \sin^2 \theta) + (M_x - M_y) \sin \theta \cos \theta.$$

The initial conditions:

$$W|_{t=0} = \varphi, \quad S|_{t=0} = \psi,$$

$$M_x|_{t=0} = -D(\varphi_{xx} + \sigma \varphi_{yy}), \quad M_y|_{t=0} = -D(\varphi_{yy} + \sigma \varphi_{xx}).$$

For numerical solution of the system (3)-(6) the two-layer implicit finite-difference approximation [6] on a rectangular mesh has been created. For interior cell of the mesh, the system of the finite-difference equations has the form:

$$\rho h_{ij} s_{t,ij}^n = u_{\bar{x}x,ij}^{n+1/2} - 2r_{\circ \circ \circ \circ x y,ij}^{n+1/2} + v_{\bar{y}y,ij}^{n+1/2} - a w_{ij}^{n+1/2} + f_{ij}^{n+1/2},$$

$$u_{t,ij}^n = -D_{ij} s_{\bar{x}x,ij}^{n+1/2} - \sigma D_{ij} s_{\bar{y}y,ij}^{n+1/2},$$

$$v_{t,ij}^n = -\sigma D_{ij} s_{\bar{x}x,ij}^{n+1/2} - D_{ij} s_{\bar{y}y,ij}^{n+1/2},$$

$$w_{t,ij}^n = s_{ij}^{n+1/2},$$

where $[w_{ij}^n], [s_{ij}^n], [u_{ij}^n], [v_{ij}^n], [r_{ij}^n], [h_{ij}], [D_{ij}], [f_{ij}^n]$ are the finite-difference analogues of the corresponding functions $W, S, M_x, M_y, M_{xy}, h, D, F$,

$$\phi_{ij}^{n+1/2} = \frac{\phi_{ij}^{n+1} + \phi_{ij}^n}{2}, \quad \phi_{ij}^n = (w_{ij}^n, s_{ij}^n, u_{ij}^n, v_{ij}^n, r_{ij}^n)^T,$$

$$\phi_{\bar{x}x,ij} = \frac{\phi_{i+1,j} - 2\phi_{ij} + \phi_{i-1,j}}{(\Delta x)^2}, \quad \phi_{\bar{y}y,ij} = \frac{\phi_{i,j+1} - 2\phi_{ij} + \phi_{i,j-1}}{(\Delta y)^2},$$

$$\phi_{\circ \circ \circ \circ xy, i+1/2, j+1/2} = \frac{\phi_{i+1,j+1} - \phi_{i,j+1} - \phi_{i+1,j} + \phi_{ij}}{\Delta x \Delta y}, \quad \phi_{t,ij}^n = \frac{\phi_{ij}^{n+1} - \phi_{ij}^n}{\Delta t}.$$

This scheme is more convenient for the numerical solution of the problem, than the three-layer scheme for the equation (1). The system of finite-difference equations is suitable to solve by splitting into two subsystems. In the first subsystem are implicitly approximated only the finite-difference derivatives on x :

$$\rho h_{ij} \frac{s_{ij}^{n+1/2} - s_{ij}^n}{0.5\Delta t} = u_{\bar{x}x,ij}^{n+1/2} + P_{ij}^n,$$

$$\frac{u_{ij}^{n+1/2} - u_{ij}^n}{0.5\Delta t} = -D_{ij} s_{\bar{x}x,ij}^{n+1/2} + Q_{ij}^n,$$

$$\frac{v_{ij}^{n+1/2} - v_{ij}^n}{0.5\Delta t} = -\sigma D_{ij} s_{\bar{x}x,ij}^{n+1/2} + R_{ij}^n,$$

$$\frac{w_{ij}^{n+1/2} - w_{ij}^n}{0.5\Delta t} = s_{ij}^{n+1/2},$$

and in the second – the finite-difference derivatives on y :

$$\rho h_{ij} \frac{s_{ij}^{n+1} - s_{ij}^{n+1/2}}{0.5\Delta t} = v_{yy,ij}^{n+1} + P_{ij}^{n+1/2},$$

$$\frac{u_{ij}^{n+1} - u_{ij}^{n+1/2}}{0.5\Delta t} = -\sigma D_{ij} s_{yy,ij}^{n+1} + Q_{ij}^{n+1/2},$$

$$\frac{v_{ij}^{n+1} - v_{ij}^{n+1/2}}{0.5\Delta t} = -D_{ij} s_{yy,ij}^{n+1} + R_{ij}^{n+1/2},$$

$$\frac{w_{ij}^{n+1} - w_{ij}^n}{0.5\Delta t} = s_{ij}^{n+1}.$$

To solve each of these subsystems efficiently, we renumber the unknowns in a certain order and reduce the system of linear algebraic equations to a system with a nine-diagonal matrix, which is solved by Gaussian elimination. In this case, the coefficients are calculated in four stages. Test calculations have shown the high efficiency of the method proposed for solving the problem.

Stability of the finite-difference scheme from the input data has been proved [6,7]. The function F describing the forces acting on the plate surface can be discontinuous. Therefore, solutions of the initial problem are regarded as generalized solutions in corresponding function spaces [8]. The strong convergence of the solution of the finite-difference problem to a generalized solution of the initial differential problem has been proved and the rate of convergence has been estimated [7].

9.4 Numerical criterion of a plate destruction

In the process of transverse vibrations in a plate its material is stretched and compressed. For a thin elastic plate the stress tensor components determined by Hooke's law and have the forms: $\sigma_{xx} = 12z M_x / h^3$, $\sigma_{yy} = 12z M_y / h^3$, $\sigma_{xy} = -12z M_{xy} / h^3$. The finite-difference analogues of their corresponding moments M_x , M_y , M_{xy} are computed at each time

step of the computer program. Since maxima in absolute magnitude values of tensile and compression stress occur on the plate's surfaces, the maxima in absolute magnitude values of stress for the mesh at $z = \pm h/2$ are computed. When one of these values reaches its critical magnitude, we conclude that at this time step an irreversible deformation and destruction of the plate will take place.

9.5 Some applications of the method

As an example of the developed numerical method application, let us consider the problem of vibrations of a floating ice in response to a moving load. The mathematical model of the floating ice vibration problem is described by the equation for transverse vibrations of a thin elastic plate (1) lying on a liquid base, where $a = \rho_l g$, ρ_l is the water density, $\rho_l g W$ is the buoyancy Archimedean force on the lower surface of the ice, F is the moving load on the upper surface of the ice. Notice that the floating ice vibration problem so formulated has been considered by many authors (See, e.g., [9-12]). In [9-12] the analytical methods have been used to solving the problem: the plane pressure front advance on the ice was specified in the right-hand side of the equation by a δ -function of the form $P\delta(x-vt)$, where P is the force acting upon the ice surface, the δ -function was represented in the form of a Fourier integral, and the solution of the problem for the plate's deflection and for liquid flow potential under the ice was also represented in the form of integrals. The focus was mainly on the study of the resulting dispersion relations, dependence of the wave amplitude on the load's velocity, and existence of critical velocities. Consequently it may be concluded that for the above applied problem no numerical methods with direct approximations of the equation for transverse vibrations in thin elastic plates were used.

The elaborated method and computer code were used to carry out numerical simulation of the problem of propagation of floating ice vibrations caused by moving cars. The parameters used in the computations were as follows: linear dimensions of an ice plate with constant thickness were 100m×40m, the size of the mesh was 500×200 mesh points, the time step was chosen corresponding with the stability condition of the method used for solving the system of difference equations [5]; the values of physical parameters of the ice were as follows: modulus of elasticity $E = 5.1 \cdot 10^8$ n/m², Poisson's ratio $\sigma = 0.35$, ice tensile strength was 0.5–1.0 MPa, ice compression strength was 2–3 MPa.

Minimal values of ice tensile and compression strength were selected as the destruction criteria. The boundary conditions are given on the free border: $M_{\partial\Omega} = 0$, $N_{\partial\Omega} = 0$.

The motion of one car with a mass of 2.2 tons moving at a speed of 15m/sec on a floating 0.26m thick ice plate was numerically simulated (See Fig.1). Tensile and compression strength was computed at each time step of the computer program. No ice destruction happened in this case.

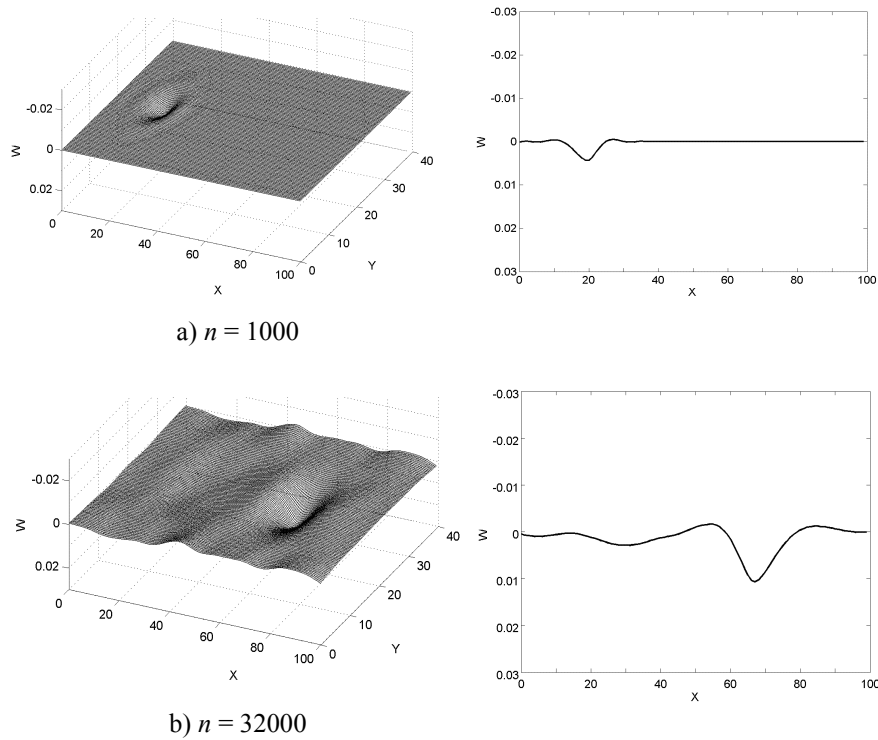


Fig. 1. The computed results for the case when the only car is moved across the ice cover. Wave dynamics of the floating ice deflection W [m]. On the right the cross-section in a direction of movement of the car's centre of mass is shown.

Of particular interest is the case of the combined motion of two vehicles or more on a floating ice. Depending on the distance between these vehicles, the waves propagating from the vehicles can be quenched or added increasing their amplitudes. Numerical simulation was carried out of the combined motion of two identical cars each having a mass of 2.2 tons moving at the same speed of 15m/sec on a floating 0.29m thick ice, the

distance between the cars being 20m (See Fig. 2). When the wave due to the motion of the first car reached the other car, there was an increase in amplitude of vibration (See Fig. 2b) and the computed maximum stress values exceeded the minimum tensile strength value of 0.5 MPa, and it means that the destruction of the ice will take place.

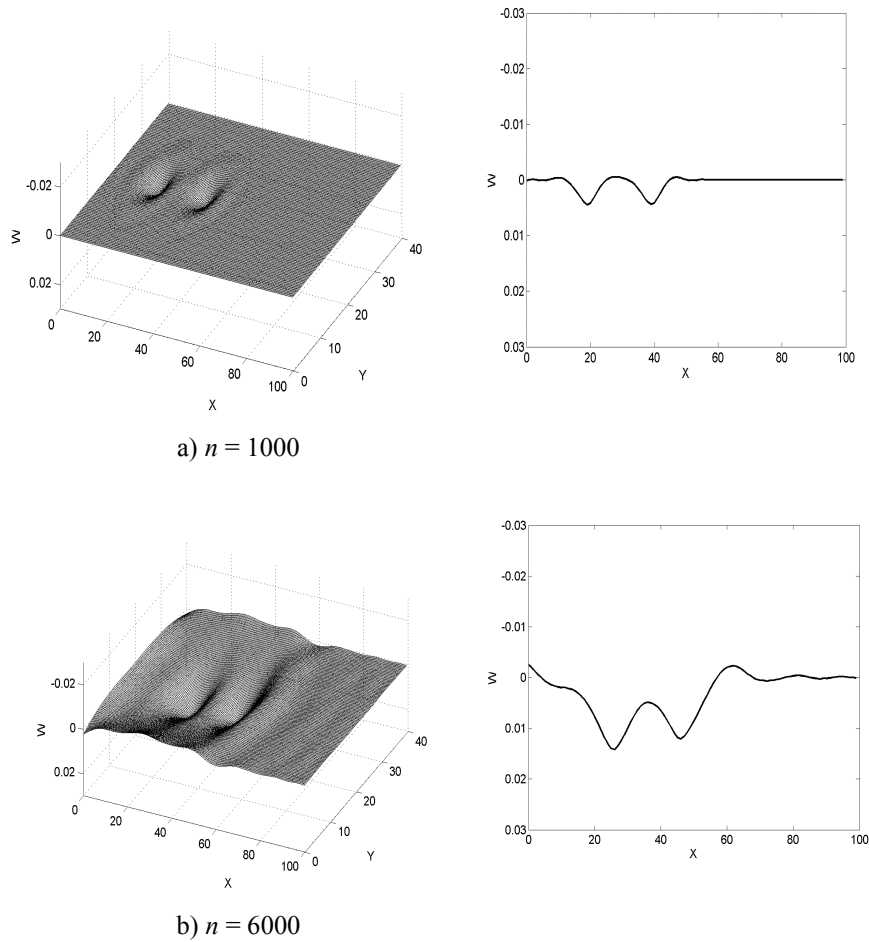


Fig. 2. The computed results for the case when two cars are moved across the ice cover. Wave dynamics of the floating ice deflection W [m].

9.6 Conclusions

The elaborated numerical method and computer software to solve the problem of small transverse vibrations in thin elastic plates gives an opportunity to study wave dynamics of transverse vibration in such plates under the action of various moving loads and also enables us to make inferences about possible destruction of the plate material.

References

1. Rosca A, Rosca D, Nastasescu V (2008) Consideration on stainless steel plate plastic deformed by fluid impact substances and pressured nitrogen blasting. WSEAS Transactions on Applied and Theoretical Mechanics, Issue 1, V 3
2. Sapountzakis EJ, Mokos VG (2006) Influence of the interface forces to the analysis of beam stiffened plates. Proceedings of the IASME/WSEAS International Conference on Continuum Mechanics, Chalkida, Greece
3. Bugaru M, Vasile O (2006) Non-linear dynamic behavior of thin rectangular plates parametrically excited using the asymptotic method, Part 1: Computation of the amplitude. Proceedings of the 10th WSEAS International Conference on Applied Mathematics, Dallas, Texas, USA
4. Jin S, Park P-H (2008) Tectonic activities and deformation in South Korea constrained by GPS observations. Proceedings of the 2nd IASME / WSEAS International Conference on Geology and Seismology, Cambridge, UK
5. Timoshenko SP, Woinovsky-Krieger S (1959) Theory of plates and shells. McGraw Hill, New York Toronto London
6. Kuleshov AA (2005) Difference approximation of the problem of bending vibrations of a thin elastic plate. Comput. Math. and Math. Phys. 4: 694–715
7. Kuleshov AA, Mymrin VV, Razgulin AV (2009) Strong convergence of difference approximation in the problem of transverse vibrations of thin elastic plates. Comput. Math. and Math. Phys. 1: 146–171
8. Lions JL, Magenes E (1972) Nonhomogeneous boundary value problems and applications V 2. Springer, Berlin
9. Schulkes RMSM, Sneyd AD (1988) Time-dependent response of floating ice to a steadily moving load. J. Fluid Mech. 186: 25–46
10. Miles J, Sneyd AD (2003) The response of a floating ice sheet to an accelerating line load. J. Fluid Mech. 497: 435–439
11. Milinazzo F, Shinbrot M, Evans NW (1995) A mathematical analysis of the steady response of floating ice to the uniform motion of a rectangular load. J. Fluid Mech. 287: 173–197
12. Părau E, Dias F (2002) Nonlinear effects in the response of a floating ice plate to a moving load. J. Fluid Mech. 460: 281–305

Chapter 10

Principal Component Analysis for Pose Estimation using Range Data

Galina Okouneva¹, Donald J. McTavish², Aradhana Choudhuri³

Department of Aerospace Engineering, Ryerson University, 350 Victoria Street, Toronto, Ontario, CANADA

¹gokounev@ryerson.ca ²mctavish@ryerson.ca ³a2choudh@ryerson.ca

Abstract. This paper presents an application of Principal Component Analysis (PCA) to the problem of pose estimation in computer vision. Continuum Shape Constraint Analysis (CSCA), a theoretical development based on PCA, generates a variety of numerical measures that can be used to assess the shape and predict the accuracy of pose estimation. The proposed approach was developed for LIDAR-based scanning that samples non-specific points from the object across the area observed from a single view. Based on CSCA measures, the paper answers the question: what views of an object can be expected to lead to the lowest pose estimation error computed via the Iterative Closest-Point Algorithm, or conversely, what level of error can be expected for a particular scan view. A novel measure, the Expectivity Index, presented in this paper, is used to assess and predict the pose estimation accuracy. The approach is demonstrated in both numerical simulation and experimental studies using the Stanford Bunny, cuboid and asymmetrical cuboctahedron shapes. The continuum nature of the CSCA formulation produces measures that are pure shape properties of an object.

Keywords. Computer Vision, LIDAR, ICP, Pose Estimation, Principal Component Analysis

10.1 Introduction

Computer vision has become an integral part of numerous aerospace industry applications. The need for rapid development of on-orbit autonomous robotic operations entails further improvement of computer vision systems and algorithms. One of these innovations is the use of active sensors that scan an object with a laser beam and deliver 3D range data in form of a point cloud. The practical application of the present work supports the use of LIDAR (light detection and ranging) based computer vision for

spacecraft rendezvous operations [1, 2, 3 and 4]. Pose estimation is performed using the Iterative Closest-Point (ICP) algorithm [5] which generally seeks to minimize a cost function that quantifies the registration error between a model and the data points. A continuous surface model of the object, generally assumed to be a triangulated surface mesh model, is registered against the point cloud to produce a pose estimate for the object. Our emphasis and contribution is focused on the potential terminal accuracy of the algorithm in the context of imperfect data. To this end, our invocation of geometric constraint as a shape property will focus its attention on the sensitivity of the pose solution in the vicinity of the true pose to the inevitable presence of noise-like error.

Pose estimate can also be computed using non-iterative algorithms which perform matching of shape descriptors such as spin images, geometric histogram, surface signatures, etc. For details, see [6 and 7].

Principal Component Analysis (PCA) is a powerful tool which is used in many applications in computer vision such as surface inspection [8] and face recognition [9]. Geometric constraint analysis is an application of PCA that directly assesses the sensitivity of shape registration error to variation in pose, providing a powerful way of assessing the expected accuracy of iterative registration algorithms. Both ICP and constraint analysis share the same attractive feature of being based on bulk calculation of data, avoiding feature (primitives and local shape invariants) detection tasks, and is thus generally applicable to any object shape.

Simon [10] is credited with introducing the application of constraint analysis to the ICP process of pose estimation in computer vision. In his work, constraint analysis was used to optimize the selection of target points on human bones for scanning during radiation therapy. The key metric used in by Simon was the Noise Amplification Index (NAI), developed by Nahvi & Hollerbach [11]. A high value of NAI indicates “good” geometric constraint balancing the extremities of pose vector error in pose space and has been used as a relative measure comparing scenes under comparable conditions. While Simon uses a sparse set of key points for the bone problem, Shahid & Okouneva [12] later used the same form of discrete-point constraint analysis applied to point-clouds collected from the uniform projection of points onto “windowed” areas of spacecraft objects in order to identify optimal local scan areas for pose estimation. Another related paper by Gelfand et al. [13] considers sampling strategies over an area to produce an effective set of points for use by ICP.

A more recent paper by McTavish & Okouneva [14] generalizes the concept of discrete-point self-registration to a surface-integral based approach referring to its use for pose estimation assessment as Continuum-Shape Constraint Analysis (CSCA). The application was single-view LIDAR-based scanning that generates a return of uniformly distributed, but non-specifically located sample (data) points, from the target on a given measurement set. The authors account for the directional nature of a single scan by incorporating a view factor into the CSCA cost matrix calculation. The result is a cost matrix that is a well-defined property of the shape geometry and also dependent on view direction. The authors followed Simon in the use of NAI as a relative indicator of expected pose estimate accuracy. Rather than for designing a target point pattern on an object, CSCA was used to assess or compare different shapes and views of shapes for good pose estimation.

In our present work we have found that while very high NAI values are generally associated with “good views” (or “good shapes”), the index does not provide an adequate general indication of pose estimate accuracy. The main contribution of this paper is the development of a new constraint

analysis measure, the Expectivity Index, which provides more informative predictions for registration accuracy.

Section 2 of this paper reviews (briefly) the discrete-point basis for the surface-integral forms of CSCA, defines self-registration and provides the definition of NAI. Section 3 describes the CSCA formulas that represent the discrete-point computed cost function minimized by ICP. Section 4 derives different expectivity indices that seek to provide general predictions of pose error (in the neighborhood of true pose). Section 5 provides numerical and experimental demonstrations of the CSCA/expectivity approach including a comparison with NAI.

10.2 Point Based Registration

The fundamental problem considered in this paper is the registration of a continuous surface model M to a set of discrete data points $\{z_i\}$ measured in a reference frame \mathcal{F}_0 (shown in Figure 1).

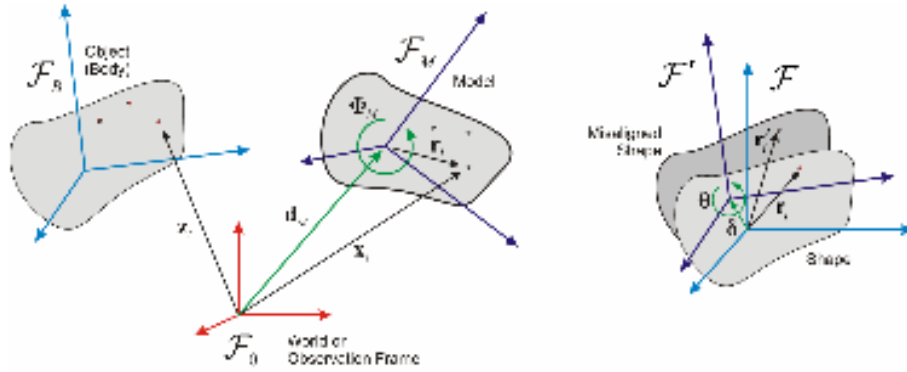


Fig. 1. Point-based shape registration: general (left) and self-registration (right)

The pose of model defined by the translation \mathbf{d}_M and rotation Φ_M vectors of its describing reference frame F_M , is determined to minimize the closest-point distances as defined by the cost function

$$E = \frac{1}{2} \sum_i \Delta \mathbf{r}_i^T \Delta \mathbf{r}_i \quad (1)$$

where $\Delta \mathbf{r}_i = \mathbf{x}_i - \mathbf{z}_i$ and the $\{\mathbf{x}_i\}$ are the transformed closest points $\{\mathbf{r}_i\}$ on the model's surface to the data points $\{\mathbf{z}_i\}$. The basic ICP algorithm 10 selects a new set of closest points $\{\mathbf{r}_i\}$ for each iteration based on the previous iteration's pose to refine the pose estimate toward its optimal value based on the transformation $\mathbf{x}_i = \mathbf{d}_M + \mathbf{C}_M \mathbf{r}_i$. The final position pose of F_M then represents the pose of the target body frame F_B . Regarding orientation, note that we use the rotation vector Φ_M as a concise reference noting its alternate invocations as either a rotation matrix \mathbf{C}_M used for writing transformation or as an Euler-parameter quaternion \mathbf{q}_M used for the calculation thereof (as per Horn's method [15]), for example).

Though not named therein, Simon [10] introduces the concept of *self-registration* – the perturbation of pose from its truth value. The concept of self-registration is shown in Figure 1 wherein the difference between the frames \mathcal{F} and \mathcal{F}' represents the small pose difference. For Simon's

problem, the self-registration cost function (for noiseless data) takes a form equivalent¹ to

$$\hat{E}_{\perp} = \frac{1}{2} \cdot \frac{1}{N} \sum_{i=1}^N \Delta \mathbf{r}_{\perp i}^T \Delta \mathbf{r}_{\perp i} = \frac{1}{2} \mathbf{p}^T \hat{\mathbf{E}}_{\perp} \mathbf{p} \quad \text{with} \quad \Delta \mathbf{r}_{\perp i} = \mathbf{\Psi}_{\perp i} \mathbf{p} \quad (2)$$

In (2), $\mathbf{p} = (\boldsymbol{\delta} \ \boldsymbol{\theta})^T$ is the 6×1 small pose vector, consisting of a small translation and rotation vectors $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$, and $\hat{\mathbf{E}}_{\perp} = \frac{1}{N} \sum \mathbf{\Psi}_{\perp i}^T \mathbf{\Psi}_{\perp i}$ is the 6×6 cost matrix built from the projection matrices $\mathbf{\Psi}_{\perp i} = \mathbf{n}_i \mathbf{n}_i^T (\mathbf{1} - \mathbf{r}_i^{\times})$ corresponding to the truth position $\{\mathbf{r}_i\}$ of each point in the model frame with the normal \mathbf{n}_i . The matrix cross construct is defined as

$$\mathbf{a}^{\times} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} = \begin{pmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{pmatrix} \quad (3)$$

With the form of the $\{\mathbf{\Psi}_{\perp i}\}$ above, this self-registration emulates the closest-point difference used in ICP, becoming a infinitesimal form of the point-to-plane distance consistent with that as developed by Chen&Menioni [16]. We note our deliberate choice of normalizing the cost matrix by the number of points N , and the use of the over-caret notation to indicate this.

10.3 Point Measurement Noise

Application of constraint analysis to pose estimation involves the addition of the vector of noise $\{\boldsymbol{\epsilon}_i\}$ to the closest-point difference vector from (2) written here as

$$\Delta \mathbf{r}_{\perp i} = \mathbf{\Psi}_{\perp i} \mathbf{p} - \boldsymbol{\epsilon}_i \quad (4)$$

The self-registration cost function (1) then becomes

$$\hat{E}_{\perp} = \frac{1}{2} \hat{E}_0 - \mathbf{p}^T \hat{\mathbf{e}} + \frac{1}{2} \mathbf{p}^T \hat{\mathbf{E}}_{\perp} \mathbf{p} \quad \text{with} \quad \hat{E}_0 = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_i \quad \text{and} \quad \hat{\mathbf{e}} = \frac{1}{N} \sum_{i=1}^N \mathbf{\Psi}_{\perp i}^T \boldsymbol{\epsilon}_i \quad (5)$$

Before proceeding further, note the new definition $\hat{\mathbf{\Psi}}_{\perp} = \frac{1}{\sqrt{N}} \text{col}\{\mathbf{\Psi}_{\perp i}\}$

where $\text{col}\{\mathbf{\Psi}_{\perp i}\} = \begin{pmatrix} \mathbf{\Psi}_{\perp 1} \\ \vdots \\ \mathbf{\Psi}_{\perp N} \end{pmatrix}$. The relationships (5) are reorganized as

$$\hat{\mathbf{E}}_{\perp} = \sum_1^N \hat{\mathbf{\Psi}}_{\perp}^T \hat{\mathbf{\Psi}}_{\perp} \quad \text{and} \quad \hat{\mathbf{e}} = \frac{1}{\sqrt{N}} \hat{\mathbf{\Psi}}_{\perp}^T \text{col}\{\boldsymbol{\epsilon}_i\} \quad (6)$$

¹ Our notation and use of the perturbation form is developed from 14 whose work is extended in this paper.

The noise, now embedded in $\hat{\mathbf{e}}$, may be considered to represent some combination of measurement or model error that leads to an error in pose estimate. So long, as the cost matrix is non-singular, the small pose solution for the self-registration context is $\mathbf{p}_\varepsilon = \hat{\mathbf{E}}_\perp^{-1} \hat{\mathbf{e}}$.

10.4 Rotation Scaling

It is generally required to dimensionally scale the rotation part of \mathbf{p} to balance its numerical influence on cost gradient with translation. This is done in a simple manner by using a scaling distance D as and defining a modified pose vector $\tilde{\mathbf{p}}$ as follows:

$$\tilde{\mathbf{p}} = \begin{pmatrix} \delta \\ D\theta \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & D\mathbf{1} \end{pmatrix} \begin{pmatrix} \delta \\ \theta \end{pmatrix} = \mathbf{D}\mathbf{p} \quad (7)$$

leading also to the modified definitions $\hat{\tilde{\mathbf{E}}}_\perp \mathbf{p} = \mathbf{D}^{-1} \hat{\mathbf{E}}_\perp \mathbf{D}$ and $\hat{\tilde{\Psi}} = \hat{\Psi} \mathbf{D}^{-1}$. Generally, we assume the use of rotation scaling and drop the explicit over-tilde notation.

10.5 Continuum-Shape Constraint Analysis

The basic idea behind CSCA is to “re-tool” self-registration with continuous surface integration rather than a discrete-point summation. While the latter is explicitly experiment specific (for LIDAR scanning), and hence unknown a priori, the CSCA approach leads to well-defined shape properties.

10.6 Directional Closest-Point Registration

Corresponding to Equation **Error! Reference source not found.**, we have

$$E_\perp = \frac{1}{2} \int_s \Delta \mathbf{r}_\perp^T \Delta \mathbf{r}_\perp dS = \frac{1}{2} \mathbf{p}^T \mathbf{E}_\perp \mathbf{p}, \quad \mathbf{E}_\perp = \int_s \Psi_\perp^T \Psi_\perp dS \quad (8)$$

where $\Delta \mathbf{r}_\perp = \Psi_\perp \mathbf{p}$ and $\Psi_\perp = \mathbf{n} \mathbf{n}^T (\mathbf{1} - \mathbf{r}^\times)$.

When the discrete-points arise from a uniform directional scan, Equations (8) are modified with a view factor in the integrand as follows:

$$E_{\perp v} = \frac{1}{2} \int_s \Delta \mathbf{r}_\perp^T \Delta \mathbf{r}_\perp dS = \frac{1}{2} \mathbf{p}^T \mathbf{E}_{\perp v} \mathbf{p}, \quad \mathbf{E}_{\perp v} = \int_s \Psi_\perp^T \Psi_\perp v dS \quad (9)$$

The view factor is defined as

$$v = \begin{cases} \mathbf{v}^T \mathbf{n} & \text{when } dS \text{ is not obstructed, } \mathbf{v}^T \mathbf{n} > 0 \\ 0 & \text{when } dS \text{ is obstructed, } \mathbf{v}^T \mathbf{n} \leq 0 \end{cases}$$

where \mathbf{v} is the direction vector (unit-length) of the viewpoint from the surface location \mathbf{r} . As a practical approximation for viewpoints at a significant distance, \mathbf{v} can be taken to be constant (orthographic view). Similar to the point-number normalization of the discrete-point cost and cost matrix, the continuum-shape analog is normalized by area:

$$\bar{E}_{\perp v} = \frac{1}{A_p} E_{\perp v} \text{ and } \bar{\mathbf{E}}_{\perp v} = \frac{1}{A_p} \mathbf{E}_{\perp v}$$

Where A_p is the un-obscured projected area as seen from the viewpoint and can be computed from the upper-left translation-translation partition $\mathbf{E}_{\perp v, \delta \delta}$ of the raw cost matrix as $A_p = \text{trace}\{\mathbf{E}_{\perp v, \delta \delta}\}$. The eigenvalues of the raw and area-normalized cost matrices are $\{\lambda_k\} = \text{eig}\{\mathbf{E}_{\perp v}\}$ and

$$\{\bar{\lambda}_k\} = \text{eig}\{\bar{\mathbf{E}}_{\perp v}\} = \left\{ \frac{1}{A_p} \lambda_k \right\}.$$

Note, that there is a direct correspondence between the point-number normalized discrete-point cost matrix of (2) and the area-normalized continuum-shape cost matrix $\bar{\mathbf{E}}_{\perp v} : \hat{\mathbf{E}}_{\perp} \leftrightarrow \bar{\mathbf{E}}_{\perp v}$ and $\{\hat{\lambda}_k\} \leftrightarrow \{\bar{\lambda}_k\}$. The key implication is that the continuum-based quantities, being well-defined shape properties and known a-priori, can be used in the place of the discrete-point summation based quantities that depend, in principle, upon unknown future data to be collected by a measurement scan. In the case of NAI, we use

$$\hat{K}_{NAI} \approx \bar{K}_{NAI} = \frac{\bar{\lambda}_{\min}}{\sqrt{\bar{\lambda}_{\max}}}. \text{ If the pose rotation is dimensionally-scaled}$$

according to **Error! Reference source not found.**, the source cost matrix used for the $\{\bar{\lambda}_k\}$ must be similarly scaled.

CSCA measures are based on surface integrals and as such are definitions. In practice, and for the examples provided in this paper, our choice is to work from triangulated mesh models of a shape's surface. The process generally involves projecting the full 3D mesh onto a view and re-meshing as required to remove all back-facing or occluded surface regions. Analytic integration is then available, summing over the triangular faces of the re-meshed model.

10.7 The Expectivity Index

The following development is based upon the characterization of noise present in the scan points by a common standard deviation from zero-mean error. If the actual point locations are $\{\mathbf{y}_i\}$, then the measured locations are $\{\mathbf{z}_i\} = \{\mathbf{y}_i + \boldsymbol{\varepsilon}_i\}$. Assuming further that the components of the $\{\boldsymbol{\varepsilon}_i\}$ are uncorrelated with equal standard deviations each of $\{\sigma_{\varepsilon}\}$, we can write for the covariance of $\{\boldsymbol{\varepsilon}_i\}$: $\mathbf{E} = \{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T\} = \sigma_{\varepsilon}^2 \cdot \mathbf{1}$.

The pose solution and its expected value in the noise added case are

$$\mathbf{p}_{\varepsilon} = \hat{\mathbf{E}}_{\perp}^{-1} = \frac{1}{\sqrt{N}} \hat{\mathbf{E}}_{\perp}^{-1} \hat{\boldsymbol{\Psi}}_{\perp}^T \text{col}\{\boldsymbol{\varepsilon}_i\}, \quad \mathbf{E}\{\|\mathbf{p}_{\varepsilon}\|^2\} = \mathbf{E}\{\mathbf{p}_{\varepsilon}^T \mathbf{p}_{\varepsilon}\} = \mathbf{E}\{\text{trace}\{\mathbf{p}_{\varepsilon} \mathbf{p}_{\varepsilon}^T\}\} \quad (10)$$

Applying (10), the above covariance for $\{\boldsymbol{\varepsilon}_i\}$ and using $\hat{\mathbf{E}}_{\perp} = \hat{\boldsymbol{\Psi}}_{\perp}^T \hat{\boldsymbol{\Psi}}_{\perp}$, leads to

$$E\{\|\mathbf{p}_\varepsilon\|^2\} = \frac{\sigma_\varepsilon^2}{N} \left\{ \text{trace}(\hat{\mathbf{E}}_\perp^{-1}) \right\} \quad (11)$$

With the discrete-point cost matrix isolated, we can use its continuum-shape counterpart in its place

$$E\{\|\mathbf{p}_\varepsilon\|^2\} = \frac{\sigma_\varepsilon^2}{N} \left\{ \text{trace}(\hat{\mathbf{E}}_{\perp v}^{-1}) \right\} = \frac{\sigma_\varepsilon^2}{N} \sum_k \frac{1}{\bar{\lambda}_k} \quad (12)$$

We now define the “standard deviation” of the pose-error magnitude as

$$\sigma_{p_e} = \sqrt{E\{\|\mathbf{p}_\varepsilon\|^2\}} = \frac{\sigma_\varepsilon}{\sqrt{N}} \sqrt{\sum_k \frac{1}{\bar{\lambda}_k}} \quad (13)$$

and define the area-normalized and non-normalized Expectivity Indices for pose accuracy as

$$\bar{K}_{EI} = \left(\sqrt{\sum_k \frac{1}{\bar{\lambda}_k}} \right)^{-1} \quad \text{and} \quad K_{EI} = \left(\sqrt{\sum_k \frac{1}{\lambda_k}} \right)^{-1} = A_p \bar{K}_{EI} \quad (14)$$

Using the definition of the Expectivity Indices above, we recognize two useful forms of (13):

$$\sigma_{p_e} = \frac{\sigma_\varepsilon}{\sqrt{N}} \frac{1}{\bar{K}_{EI}} \quad \text{and} \quad \sigma_{p_e} = \frac{\sigma_\varepsilon}{\sqrt{\rho_p}} \frac{1}{K_{EI}} \quad (15)$$

The first applicable when the number of scan points is specified, and the second when the point density of scan is given ($N = \rho_p A_p$). Again, if the pose orientation is dimensionally scaled, the source cost matrix used for the $\{\lambda_k\}$ or $\{\bar{\lambda}_k\}$ must be similarly scaled.

10.8 Edge Effect in ICP

Point-to-plane registration upon which self-registration constraint analysis is based, is one of the contributors to the computed cost function of the ICP algorithm. *ICP also detects edges*. A simple example of this is the exterior boundary on the object as projected in the view of the scanner. For a particular pose, data points that are “past the edge” will be matched to the closest-points on the edge, deviating from the small pose assumption of point-to-plane matching used in the self-registration cost functions. The same applies to internal edges as well. The spacing of points in the scan effectively provides an additional constraint that holds onto the edges. A simple example is the view of a plane rectangular surface, geometrically unconstrained in its own plane but absolutely held within the data points by the matching of data points to the outer boundary as a trial pose shifts those points beyond the boundary.

10.9 Validation of CSCA Approach

We present results from three numerical studies: an illustrative cuboid shape, the Stanford Bunny and the asymmetrical cuboctahedron. The numerical studies employ simulated LIDAR scans. The cuboid and cuboctahedron are further used in the experimental study described below.

10.10 Numerical Simulation

CSCA indices can be graphically represented as using the sphere. Viewing directions are mapped onto the sphere of a desired resolution to create a sphere-mesh as shown in Figure 2 (left).

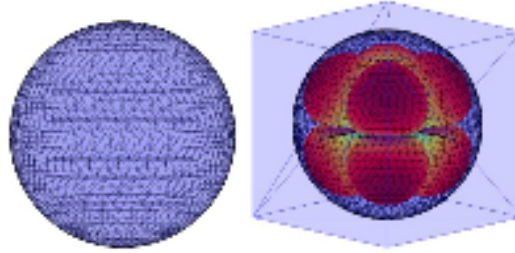


Fig. 2. Graphical Representation of CSCA Indices

The index value for a given shape for a particular view is found by projecting the view direction out from the origin through the sphere surface, and the radial distance of the intersection is then taken as the index value. Figure 2 (right) shows the spherical map of the Expectivity Index, along with the cube polyhedron and the view-sphere. In order to read the function value from any given view vector, one can simply read the radial distance to the surface of the function-map along the direction of the view vector. For additional clarity, the plots are color-coded with blue as minimum and red as maximum. For clarity, the sphere can be omitted. In Figure 2, one can see that the highest values of the Expectivity Index (red) are found at the vertices of the cube and the lowest at the faces.

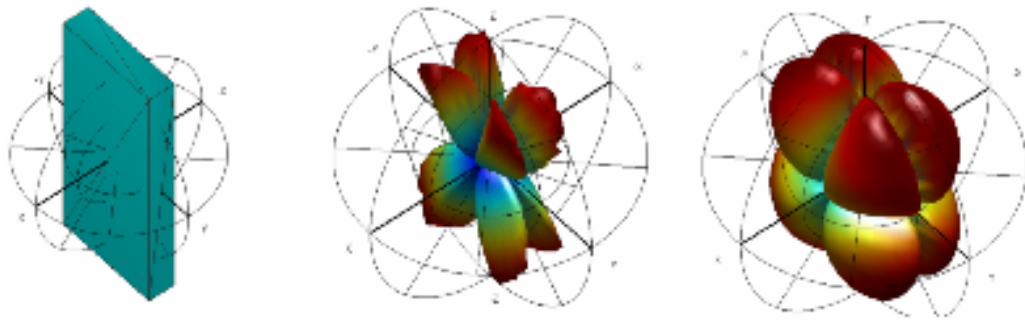


Fig 3. (a) Cuboid shape with (b) NAI function and (c) Expectivity Index function, over all possible views.

The cuboid, shown in 3, is chosen for study due to its extreme range of geometric constraint based on view direction. The relative dimensions of the shape sides are $\{1, 4, 8\}$, with a characteristic distance of 2 used for rotation scaling. The figure also shows the index functions for the area-normalized NAI and the Expectivity Index. The index value for a particular view direction \mathbf{v} is found by projecting that vector out from the origin to intersect the function surface. The two shown indices both range

from values of zero (unconstrained when less than three faces of the cuboid are visible) to very high values (well-constrained when all three orthogonal faces are highly visible). Qualitatively, the \bar{K}_{NAI} function is much more selective about view quality than the \bar{K}_{EI} .

We restrict ourselves to the question... what views produce the most accurate pose estimate? The plots of Figure 4 record the pose estimates returned by ICP for up to 10 different views for each index. The views were selected to cover the range of index values presented in the corresponding function plots of

Fig, but otherwise randomly selected from a large pool of views.

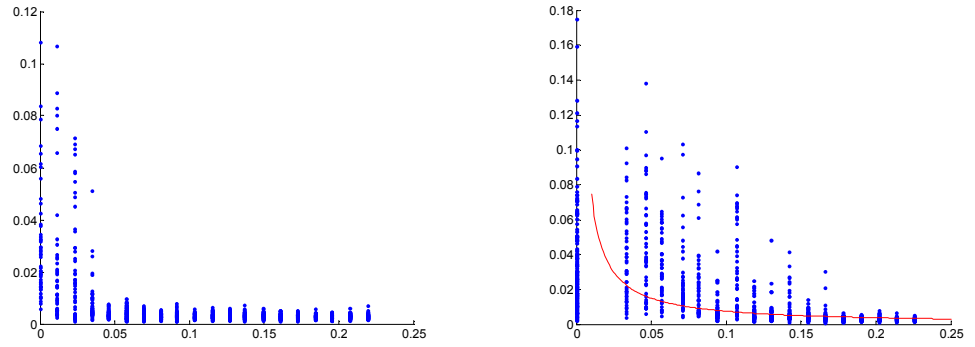


Fig. 4. Pose error norm vs. CSCA index value (a) NAI and (b) Expectivity index

At each view, 50 runs with different added noise ($\sigma_\varepsilon = 0.01$) and initial conditions randomly specified. A nominal 60 simulated scan hits were generated independently for each pose estimate run. The set of runs on any view generates a distribution of pose norm error $\|\mathbf{p}_\varepsilon\|$ according to the distribution of weak and strong constraint directions associated with that view. The RMS average value of this distribution is the goal of the Expectivity Index \bar{K}_{EI} to predict, and shown as the red curve in 4(b)) with good agreement. Regarding the ability of NAI to predict pose estimate accuracy, we observe in 4(a) that the view with the largest value of NAI does produce a pose estimate with low error, however it is also noted that there is no perceptible variation of pose estimate accuracy with decreasing NAI until the index value becomes very low near a completely unconstrained view (where $\bar{\lambda}_{\min}$ becomes zero).

Although used by the cited authors to detect well constrained situations, the NAI index does only what it is able to do. Only in situations where $\lambda_{\min} \approx \lambda_{\max}$ can it be considered a general indicator of pose accuracy. A

quick examination of the form $K_{NAI} = \frac{\lambda_{\min}}{\sqrt{\lambda_{\max}}}$

reveals its limitations immediately when $\lambda_{\min} \approx \lambda_{\max}$ is not true. For one, an increase in the value of any eigenvalue with all others held constant represents a reduction in the

volume of the pose-error space and will result in small pose errors. NAI follows this rule when λ_{\min} increases, but breaks it when λ_{\max} increases. Second, NAI provides limited information regarding the other four eigenvalues that lie between λ_{\min}

and λ_{\max} : the greater the spacing, the greater the possible range of the unaccounted-for pose space directions, each contributing in the same relative measure to the overall volume of the pose-error space.

The Stanford Bunny represents a more general shape to which CSCA expectivity can be applied. Fig.2 shows a model of the bunny superimposed with the \bar{K}_{EI} expectivity function. Unlike the cuboid, the bunny is fairly well constrained on any view. As indicated in the figure six views were selected spanning from the minimum to maximum value.

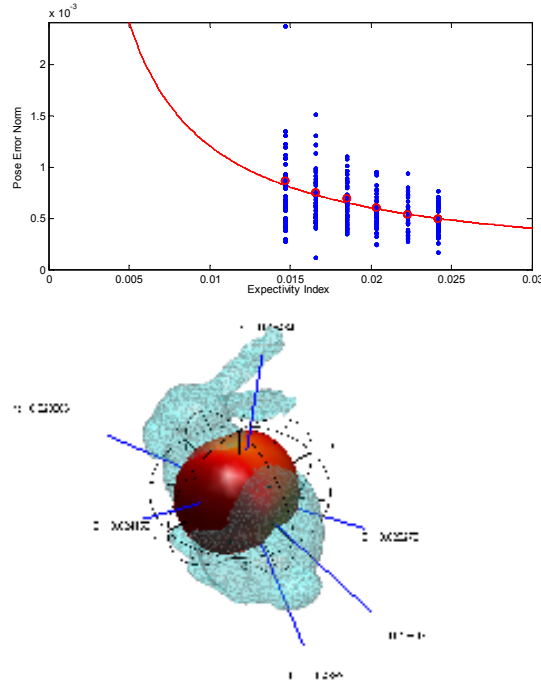


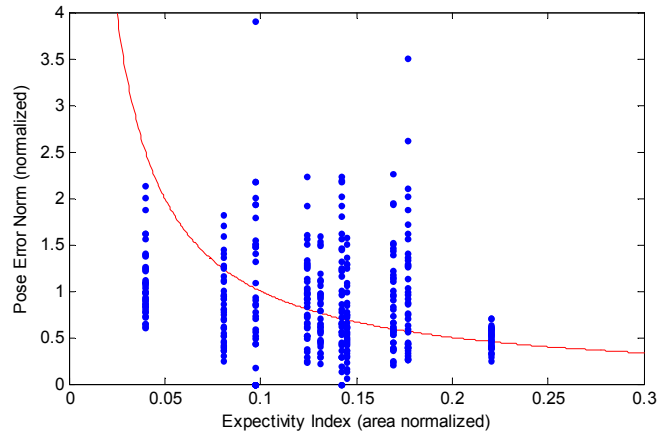
Fig.2. Stanford Bunny with CSCA Expectivity Index function and results from simulated scans.

ICP was applied with simulated noisy scans and fully randomized initial conditions to produce the plot in Fig.2. The blue points are the pose error norm values obtained from 50 trials at each view. The RMS values are indicated by red circles and match well to the red curve generated from the second relation in.

10.11 Experimental Study

Large data sets were collected for a cuboid test article, approximately 50x200x400mm in size (same relative dimensions as numerical study) using a TriDAR scanning system developed by Neptec Design Group. For the views considered, a truth pose of the object and effective measurement noise level were estimated by using a large number of scan points (5000) out of a larger set. Typically, the data exhibits a noise level of 2-3mm standard deviation. For each view, 40 independent sets of 100 scan points were extracted

from the larger set and ICP used for pose estimation. Fig 3 presents the results compared to the Expectivity Index function prediction. As the estimated noise levels at each view differ the data is normalized to the noise



level, i.e., plotting $|\mathbf{p}|/\sigma_\epsilon$.

Fig 3. Experimental results for a cuboid shape

Similar experimental validation of the CSCA approach using Neptec's TriDAR scanning system was performed for the asymmetrical cuboctahedron. Figure 7 shows the shape and the graph of the Expectivity Index vs. the norm of the pose error. The cuboctahedron was selected because it is a well-constrained shape which delivers a consistent pose error from all around views.

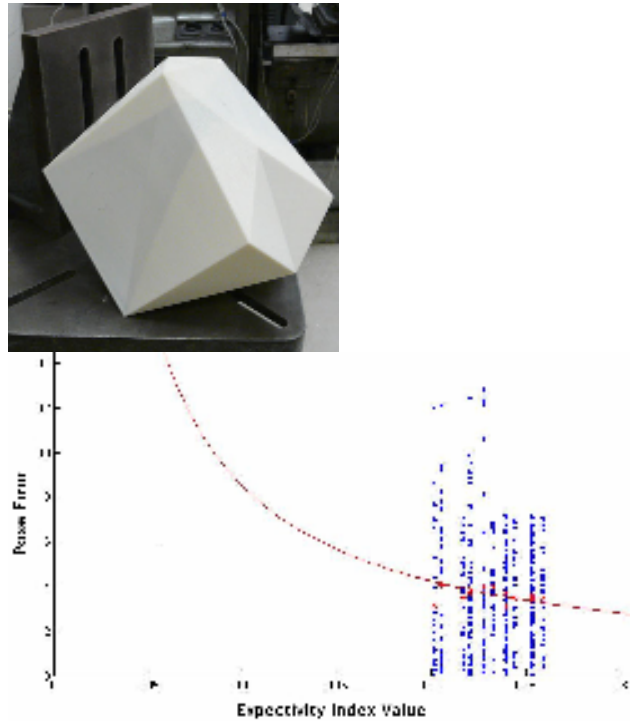


Fig 7. Experimental Results for asymmetrical Cuboctahedron

It can be seen that the mean pose error from the experiment (red dots) matches the theoretical curve (15).

10.12 Concluding Remarks

The paper applies CSCA to directly assess the sensitivity of shape registration error to variation in pose, providing a powerful shape-property based approach to estimating the expected accuracy of iterative registration algorithms. CSCA can be applied to generally complex shapes represented as triangular meshes. The newly developed CSCA measure, the Expectivity Index, provides estimation of the expected norm of the pose error. The paper presents the CSCA analysis of two shapes, the Stanford Bunny and a cuboid, which were used to validate the concept of the Expectivity Index. The Bunny is a shape which is well-constrained from all views. On the contrary, the cuboid is well-constrained from only a limited range of views. To obtain the expected pose error norms via the ICP algorithm, simulated and experimentally collected scans were used. The norm of the expected pose error computed from scans showed a consistent agreement with the norm obtained via formulas (15) and with visual assessment of the shapes.

References

1. C. Samson, C. English, A. Deslauriers, I. Christie, F. Blais and F. Ferrie. Neptec 3D Laser Camera System: From Space Mission STS-105 to Terrestrial Applications (NRC Report 46565), *Canadian Aeronautics and Space Journal*, Vol.50, No.2, 2004.
2. A. C. M. Allen, C. Langley, R. Mukherji, A.B. Taylor, M. Umasuthan and T. D. Barfoot. Rendezvous Lidar Sensor System for Terminal Rendezvous, Capture, and Berthing to the International Space Station, *Sensors and Systems for Space Applications II*, SPIE Proceedings Vol.6598, 2008.
3. P. Jasiobedzki, M. Abraham, P. Newhook and J. Talbot. Model Based Pose Estimation for Autonomous Operations in Space, *Proceedings of the IEEE International Conference on Intelligence, Information and Systems*, 1999.
4. C. Rossi, M. Abderrahim, J. C. Diaz. Evolutionary Model-based Pose Estimation for Variable Configuration Objects. *Proc. of the 5-th WSEAS International Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED 2006)*, Spain, February, 2006.
5. P. J. Besl and N. D. McKay. A Method for Registration of 3D Shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No.2, 1992.
6. C. Torre-Ferrero, S. Robla, E. G. Sarabia, J. R. Llata. A Coarse-to-Fine Algorithm for 3D Registration based on Wavelet Decomposition. *WSEAS Transactions on Systems*, issue 7, vol. 7, pp. 763-770, July 2008.
7. F. Samadzadegan, S. Saeedi, T. Shenk. A Feature Based Generic Model for Georeferencing of High Resolution Satellite Imagery. *Proc. of the 2-nd WSEAS International Conf. on Remote Sensing*, December 16-18, 2006.
8. H. D. Lin, C. Y. Chung, W. T. Lin. Principal Component Analysis Based on Wavelet Characteristics Applied to Automated Surface Defect Inspection. *WSEAS Transactions on Computer Research*, issue 4, vol. 3, pp. 195-202, 2008.
9. O. Khayat, H.R. Shahdoosti, A. J. Motlagh. An overview on Model-based Approaches in Face recognition. *Proc. of the 7-th WSEAS International Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED 2008)*, UK, February, 2008.
10. D. A. Simon. Fast and Accurate Shape-Based Registration, Ph.D. Thesis, Report CMU- RI-TR-96-45, Carnegie Mellon University, 1996.
11. A. Nahvi and J. M. Hollerbach. The Noise Amplification Index for Optimal Pose Selection in Robot Calibration, *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, Minneapolis MN, April 1996.
12. K. Shahid and G. Okouneva. Intelligent LIDAR Scanning Region Selection for Satellite Pose Estimation, *Computer Vision and Image Understanding*, Vol.107, 2007.

13. N. Gelfand, L. Ikemoto, S. Rusinkiewicz and M. Levoy, Geometrically Stable Sampling for the ICP Algorithm, *Proceedings of the 4th International Conference on 3-D Digital Imaging and Modeling (3DIM 2003)*, Stanford CA, October 2003.
14. D. J. McTavish and G. Okouneva. A New Approach to Geometrical Feature Assessment for ICP-Based Pose Measurement: Continuum Shape Constraint Analysis, *Proceedings of the IEEE International Machine Vision and Image Processing Conference (IMVIP 2007)*, Maynooth, Ire., September 2007.
15. B. K. P. Horn. Closed-Form Solution of Absolute Orientation Using Unit Quaternions, *Journal of the Optical Society of America A*, Vol.4, April 1987.
16. Chen and G. Mendioni. Object Modelling by Registration of Multiple Range Images, *Image and Vision Computing*, Vol.10, No.3, April 1992.

Chapter 11

A Parametric Approach to Matrix Root Clustering

V. G. Melnikov

Department of Theoretical and Applied Mechanics, The National
Research University of Information Technologies, Mechanics and Optics,
Sablinskaya 14, 197101, St.- Petersburg, Russia, melnikov@mail.ifmo.ru

Abstract. This paper presents a new criteria for matrix root-clustering in sophisticated, multiply-connected and non-convex regions of the complex plane. A concept of the *closed forbidden region* is defined as a complement of the root-clustering region to \mathbb{C} and the concepts of the *forbidden subregions* and the expanded *root-clustering regions* are introduced to formulate the main results of the paper. The set of the modified three-parametrical Cassini regions are offered to use as the subregions covering the forbidden region. In these terms the generalizations of the Gutman theorem and the Jury-Ahn theorem onto the intersection of the systems of nonlinear algebraic inequalities are obtained. An application to the problem of root-clustering in outside of a system of the forbidden frequency bands is shown for illustration.

Keywords. matrix root clustering, Lyapunov inequality, forbidden subregions, expanded root-clustering regions, modified Cassini regions.

11.1 Notation and introduction

$s(A)$ the spectrum of the matrix $A \in R^{n \times n}$

D the open root-clustering region

$\bar{S} = \mathbb{C} \setminus D$ the closed forbidden region

\bar{S}_α the forbidden subregion number α
 $D_\alpha = \mathbb{C} \setminus \bar{S}_\alpha$ the expanded root-clustering region number α
 $\bar{S}(\mu)$ the forbidden subregion depending on parameter μ
 $\psi_\alpha(s), \phi_\alpha(s)$ the polynomials of a complex variable
 X_α the solution of a generalized Lyapunov matrix inequality number α
 $X(\mu)$ the parameter-dependent solution of the generalized Lyapunov matrix inequality
 $\Gamma(a, c, \mu)$ the three-parametric modified Cassini oval or a pair of the modified Cassini ovals
 $\bar{H}(a, c, \mu)$ the modified oval regions with a border $\Gamma(a, c, \mu)$
 $G(a, c, \mu) = \mathbb{C} \setminus \bar{H}(a, c, \mu)$, external open Cassini region
 $z(s), w(s)$ the functions of conformal mapping of external Cassini region respectively into the central unit disk and into the left half-plane
 $Q > 0$

The properties of linear and quasi-linear control systems strongly depend on the location of eigenvalues of a system matrix on the complex plane. In this connection the problem of root-clustering acquires interest and importance for linear and robust systems analysis and has received much attention in recent years [1-9]. There exist several approaches to this problem. The first approach is based on the Kronecker product of matrices. The second approach uses a conformal mapping of the root-clustering region into the left half-plane or into the unit central disk and leads to the common Lyapunov inequality or Hurwitz criteria for functional transformed system matrices. The third and the most-used approach is based on the generalized Lyapunov matrix inequality [1-5], [8]. The necessary and sufficient conditions for matrix root-clustering in Γ -transformable and Ω -transformable algebraic regions obtained in the work of Gutman and colleagues [1] made the fundamental contributions in this direction. The problem of root clustering for robust systems was investigated in the work [2], and was further developed in a number of issues; some of them are: [3-4]

This paper is devoted to a problem of matrix root-clustering in the sophisticated regions given as an intersection of several algebraic regions of the fourth and smaller order. Such regions can have one or several boundaries. The boundaries may be smooth or piecewise smooth, convex or nonconvex; the boundaries may be formed by segments of algebraic curves or by curves enveloping a continual set of curves. The region may be multiply- or simply-connected, nonconvex or convex.

Thus we consider more general statement of the problem than in the literature. In this connection the approach to the problem is changed. The main attention is focused on a complement of the root-clustering region to \mathbb{C} these regions are called the *closed forbidden region*. The forbidden region we cover or approximate by a set of algebraic subregions; we offer to use a set of simply-connected and multiply-connected three-parametric regions of the fourth order called the *modified Cassini oval regions* as these algebraic subregions. The simply-connected and multiply-connected modified Cassini regions are not harder in use than the regions bounded by simply-connected algebraic curves of the second order. Functions of a simple form conformally map the modified Cassini regions onto the unit central disk and onto the left half-plane [6]. The modified Cassini regions are eminently suitable for covering of the multiply-connected regions symmetrically located about the real axis.

In the introduced terms the root-clustering region is an intersection of the complements of a set of the forbidden subregions. Necessary and sufficient conditions of matrix spectrum localization in the specified region are obtained as a union of a set of conditions of spectrum location in outside of each forbidden subregion.

11.2 Root-clustering in outside of the closed forbidden region \bar{S} and generalization of root-clustering theorems

In the literature a problem of matrix root - clustering is usually considered as a problem of a membership of all roots to a specified open region D of the complex plane \mathbb{C} . Let us consider another treatment of this problem. Let the closed region \bar{S} be the complement to \mathbb{C} of D , i.e., $\bar{S} = \mathbb{C} \setminus D$; then the region \bar{S} is called *the forbidden region*. Let us consider 2 cases.

Case 1 Suppose the region \bar{S} is defined as below:

$$\bar{S} = \bigcup_{\alpha} \bar{S}_{\alpha} \quad \forall \quad \alpha = 1, 2, \dots, \nu, \quad (1.1)$$

where $\{\bar{S}_{\alpha}, \quad \forall \alpha = 1, \nu\}$ is a *final* set of mutually intersecting or nonintersecting subregions; the envelope of this set forms the border of \bar{S} . Then by definition put

$$\{D_{\alpha} : \quad D_{\alpha} = \mathbb{C} \setminus \bar{S}_{\alpha} \supseteq D\} \quad \forall \quad \alpha = 1, 2, \dots, \nu, \quad (1.2)$$

where the regions D_α are called *the expanded root-clustering regions*. Note that $D_\alpha \supseteq D$; this follows from $\bar{S}_\alpha \subseteq \bar{S} \Rightarrow D_\alpha = \mathbb{C} \setminus \bar{S}_\alpha \supseteq \mathbb{C} \setminus \bar{S} = D$.

Case 2 Let $\{\bar{\Omega}(\mu), \mu \in [0, \tau]\}$ be a *continual* set of the subregions covering the region \bar{S}

$$\bar{S} = \cup \bar{\Omega}(\mu) \quad \forall \quad \mu \in [0, \tau]. \quad (1.3)$$

Then by definition, put

$$D(\mu): \quad D(\mu) = \mathbb{C} \setminus \bar{\Omega}(\mu) \quad \forall \quad \mu \in [0, \tau], \quad (1.4)$$

where $D(\mu) \supseteq D \quad \forall \mu \in [0, \tau]$ is the continual parametric set of the expanded root-clustering regions.

Theorem 1 *Let the forbidden region \bar{S} for the matrix spectrum $s(A)$ be a union of the regions (1.1); then the root-clustering region $D = \mathbb{C} \setminus \bar{S}$ satisfies the equality*

$$D = D^*, \quad D^* = \cap D_\alpha = \cap (\mathbb{C} \setminus \bar{S}_\alpha), \quad \alpha = 1, 2, \nu. \quad (1.5)$$

i.e., D is the intersection of the set of the expanded root-clustering regions (1.2).

Proof The proof is trivial. It follows from the point set property $A \setminus (B \cup \mathbb{C}) = (A \setminus B) \cap (A \setminus \mathbb{C})$. \square

Theorem 2 *Let the forbidden region S be a union of the parametric continual set (1.3); then the root-clustering region D satisfies the equality*

$$D = D^*, \quad D^* = \cap D(\mu), \quad D(\mu) = \mathbb{C} \setminus \bar{\Omega}(\mu) \quad \forall \mu \in [0, \tau]. \quad (1.6)$$

Proof The proof is in 2 steps.

Step1: First let us prove that $D^* \subseteq D$. Assume the converse. Then there exists a point $M \in D^*$ such that $M \notin D$. From here follows that $M \in \bar{S}$ and taking into account (1.3), we obtain that there exists $\mu = \mu'$ such that $M \in \bar{\Omega}(\mu')$. This means that $M \notin D(\mu')$ and $M \notin D$. The contradiction proves that $D^* \subseteq D$.

Step2: On the other hand, by (1.3) and $D^* = \cap D(\mu)$, it follows that $D^* \supseteq D$. Finally, taking into account the result of Step 1, we obtain $D^* = D$. This completes the proof of Theorem 2. \square

Extensive works during last decades on the subject of the matrix root clustering with respect to various algebraic regions led to the standard methods developed by Gutman, Jury and colleagues [1] enlarging Lyapunov theory on the regions defined by algebraic inequalities. This paper offers the further generalization of these results on a case of a final set of algebraic inequalities.

By A' denote the transpose of a matrix $A \in R^{n \times n}$. Suppose the forbidden subregion \bar{S}_α in the union (1.1) satisfy the inequalities

$$\bar{S}_\alpha = \{s \in \mathbb{C} : f_\alpha(s) = |\psi_\alpha(s)|^2 - |\phi_\alpha(s)|^2 \leq 0\}, \quad \forall \alpha = 1, 2, \dots, \nu, \quad (1.7)$$

where $f_\alpha(s) \neq \text{const}$, and $\psi_\alpha(s), \phi_\alpha(s)$ are some polynomials of s . Then the expanded regions $D_\alpha = \mathbb{C} \setminus \bar{S}_\alpha$ satisfy the inequalities

$$D_\alpha = \{s \in \mathbb{C} : f_\alpha(s) = |\psi_\alpha(s)|^2 - |\phi_\alpha(s)|^2 < 0\}, \quad \forall \alpha = 1, \dots, \nu. \quad (1.8)$$

Theorem 3 (Generalization of results of Jury): *Let the root-clustering region D be defined as (1.5), (1.8). Then a necessary and sufficient condition for the matrix spectrum $s(A)$ to be clustered in D is that there exist ν symmetric positively determined matrices X_α satisfying the matrix inequalities*

$$\bar{\psi}_\alpha(A')X_\alpha\psi_\alpha(A) - \bar{\phi}_\alpha(A')X_\alpha\phi_\alpha(A) = -Q, \quad Q > 0, \quad \alpha = 1, \dots, \nu. \quad (1.9)$$

Proof Actually, any region D_α from set (1.8) satisfies the Jury-Ahn theorem; therefore conditions (1.9) are necessary and sufficient conditions for root-clustering in any D_α from the set (1.8). In light of Theorem 1 the region D is the intersection of the final set of the expanded regions (1.8). From here follows, that a necessary and sufficient condition for root-clustering in D is the intersection of conditions (1.9) for all $\alpha = 1, 2, \dots, \nu$. This completes the proof of Theorem 3.

□

Further, let's consider the forbidden region \bar{S} covered by a continual set of intersecting and non-intersecting parametric subregions $\bar{S}(\mu)$. In other words, let the subregion $\bar{S}(\mu)$ completely sweeps the region \bar{S} as parameter μ ranges over $[0, \nu]$, i.e.,

$$\bar{S}(\mu) = \{s \in \mathbb{C} : \text{at } f_\alpha(s) = |\psi(s, \mu)|^2 - |\phi(s, \mu)|^2 \leq 0\}, \quad \mu \in [0, \nu]. \quad (1.10)$$

Theorem 4 Assume that the open region D is a complement to \mathbb{C} of the forbidden region $\bar{S} = \cup \bar{S}(\mu) \forall \mu \in [0, \tau]$, where the continual set of the forbidden subregions $\bar{S}(\mu)$ is determined by inequality (1.10). The spectrum $s(A)$ lies in the region $D = \mathbb{C} \setminus \bar{S}$ if and only if there exists a symmetric positively determined depending on parameter μ matrix $X(\mu) > 0$ such that at all values of the parameter μ it satisfies to a matrix inequality

$$\bar{\psi}(A', \mu)X(\mu)\psi(A, \mu) - \bar{\phi}(A', \mu)X(\mu)\phi(A, \mu) = -Q \quad \forall \mu \in [0, \nu]. \quad (1.11)$$

These obtained theorems can be easily generalized on a case of complex functions ψ, ϕ and on a case of a complex matrix A .

Theorem 5 Let the forbidden region is given as an association of subregions (1.1)

$$\bar{S}_\alpha = \{s \in \mathbb{C} : \sum_{0 \leq k, l \leq m} c_{kl}^\alpha s^k \bar{s}^l \leq 0\}, \quad c_{kl}^\alpha = \bar{c}_{lk}^\alpha \in \mathbb{C}, \quad \alpha = 1, \dots, \nu. \quad (1.12)$$

It corresponds to the root-clustering region

$$D = \cap D_\alpha \quad \forall \alpha = 1, \dots, \nu, \quad D_\alpha = \{s \in \mathbb{C} : \sum_{k, l} c_{kl}^\alpha s^k \bar{s}^l < 0\}, \quad (1.13)$$

The spectrum $s(A)$ is clustered in D , if and only if there exist a set of symmetric matrices $\{X_\alpha, \forall \alpha = 1, 2, \dots, \nu\}$ satisfying

$$\sum_{k, l} c_{kl}^\alpha A^k X_\alpha (A')^l < 0, \quad X_\alpha > 0, \quad \forall \alpha \in 1, \dots, \nu. \quad (1.14)$$

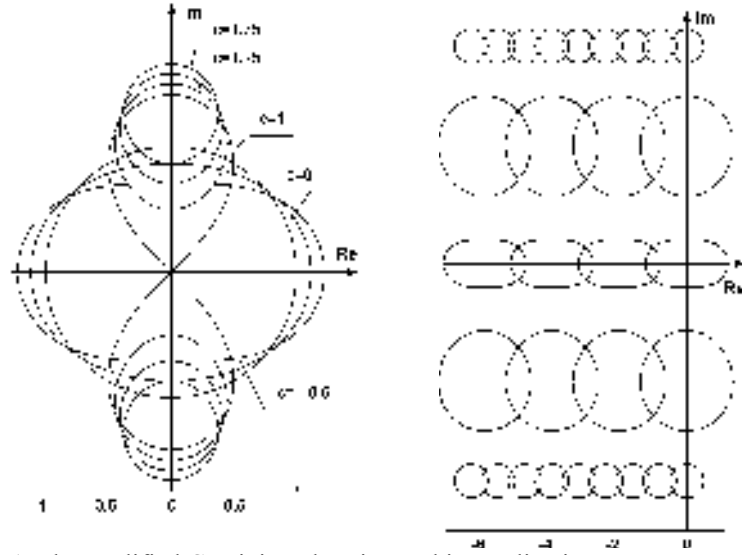


Fig. 1.1. The modified Cassini oval region and its applications

11.3 Root clustering theorems for Cassini regions

Suppose the closed region H satisfies the inequality

$$\bar{H}(a, c, \mu) = \{s \in \mathbb{C} : |(s + \mu)^2 + c|^2 - a^2 \leq 0\}, \quad (1.15)$$

where $a > 0, \mu \geq 0, c \in [-a/2, 0] \cup [a, \infty)$. Then the region $\bar{H}(a, c, \mu)$ is called the *modified three-parametric Cassini oval region*. At $c < 0$ we obtain the oval with semi-axis $(\sqrt{a-c})$ and $(\sqrt{a+c})$, extended along the real axis. At $c = -a/2$ the curvature of the oval in the top and in the bottom points are equal to zero. At $c = a$ we have the lemniscates. Finally, at any values $c > a$ we have the pair of ovals symmetrically located at both sides of the real axis having the height of $(\sqrt{c+a} - \sqrt{c-a})$ and having the maximal width of (a/\sqrt{c}) (Fig 1.1.).

Suppose the forbidden region \bar{S} is covered by the final set of regions (1.15), i.e.,

$$\bar{S} = \cup \bar{H}(a_\alpha, c_\alpha, \mu_\alpha), \quad \forall \alpha = 1, 2, \dots, \nu, \quad (1.16)$$

where

$$\bar{H}(a_\alpha, c_\alpha, \mu_\alpha) = \{s \in \mathbb{C} : |(s + \mu_\alpha)^2 + c_\alpha|^2 - a_\alpha^2 \leq 0\}. \quad (1.17)$$

Then the set of external oval regions is defined as follows

$$G(a_\alpha, c_\alpha, \mu_\alpha) = \{s : f_\alpha(s) = a_\alpha^2 - |(s + \mu_\alpha)^2 + c_\alpha|^2 < 0\}, \quad \alpha = 1, \dots, \nu. \quad (1.18)$$

Let $\psi_\alpha(s)$ and $\phi_\alpha(s)$ in inequality (1.7) be defined as $\psi_\alpha(s) = a_\alpha$, $\phi_\alpha(s) = (s + \mu_\alpha)^2 + c_\alpha$.

Then by Theorem 3, we get the following statement.

Theorem 6 *All roots of a real matrix A lie in the region $D = \mathbb{C} \setminus \bar{S}$, where \bar{S} is the forbidden region (1.16), (1.17), if and only if there exist a set of positively definite symmetric matrices $\{X_\alpha, \forall \alpha = 1, 2, \dots, \nu\}$ satisfying*

$$a_\alpha^2 X_\alpha - ((A' + \mu_\alpha E)^2 + c_\alpha E) X_\alpha ((A + \mu_\alpha E)^2 + c_\alpha E) = -Q \quad (1.19)$$

Let the parameters a and c be functions of the third parameter μ , and the oval region completely sweeps the forbidden region as μ ranges over $[0, \tau]$, i.e.,

$$\bar{S} = \bigcup \bar{H}(\mu), \bar{H}(\mu) = \{s : a^2(\mu) - |(s + \mu)^2 + c(\mu)|^2 \geq 0\}, \mu \in [0, \tau]. \quad (1.20)$$

Then using Theorem 4, we get the following statement.

Theorem 7 *All roots of a real matrix A are clustered in $D = \mathbb{C} \setminus \bar{S}$, where \bar{S} is the forbidden region (1.20), if and only if for all $\mu \in [0, \tau]$ there exists a positively definite symmetric parameter-dependent matrix $X(\mu) > 0$ satisfying the following parametric matrix inequality*

$$a^2 X(\mu) - ((A' + \mu E)^2 + cE) X(\mu) ((A + \mu E)^2 + cE) = -Q \quad (1.21)$$

11.4 Example

This example is adopted from a band-filtration problem and illustrates the developed method. Let a band filter has three suppression frequency bands: low, high and intermediate all having different width (Fig. 1.1). In terms of the developed method these bands form the *forbidden region*. Now we will obtain the algebraic conditions for matrix root-clustering in outside of this forbidden region. First we approximate the forbidden bands

by a set of modified three-parametric Cassini oval regions (1.16), (1.17). The low frequency band we approximate by four simply-connected oval regions with the values of the parameters $a=1, c=-0.5, \mu=[0,2,4,6]$. The intermediate and high bands we approximate by multiply-connected oval regions with the values of the parameters $\{a=10, c=14.5, \mu=[0,2,4,6]\}$ and $\{a=65, c=42.5, \mu=[0,0.7,.,6.3]\}$. As a whole we approximate the forbidden region by seventeen simply-connected and conjugate oval regions. Secondly in the light of Theorem 7 all roots of matrix A are clustered in outside of this forbidden region, if and only if there exist seventeen positively-definite symmetric matrices $\{X_\alpha, \forall \alpha=\overline{1,17}\}$ satisfying (1.19) for all given above values of the parameters $a_\alpha, c_\alpha, \mu_\alpha$.

11.5 Conclusions

In this paper the new theorems for matrix root-clustering in sophisticated and multiply-connected regions with piecewise-smooth boundary have been established. We introduce a new approach to the problem of root-clustering based on the application of a new class of regions of the forth order. The example illustrates an application of the developed method to the problem of band filtration.

Acknowledgement. The research was supported by the grant of the Russian Foundation for Basic Research (RFBR).

References

1. Gutman S, Jury EI (1981) A general theory for root-clustering in subregions of the complex plane, IEEE Trans. Automat. Contr. 26, , pp.~853-863.
2. Sezer ME, Siljak DD (1989) A note on robust stability bounds, IEEE Trans. on Automat. Contr., 34, , pp.1212-1215.
3. Rejichi O, Bachelier O, Chaabane M, Mehdi D (2008) Robust root-clustering analysis in a union of subregions for descriptor systems, IET Control Theory Appl., 2(7), pp. 615-624.
4. Bachelier O, Mehdi D (2003) Robust matrix DU-stability analysis, Int. J. of Robust and Nonlinear Control, 13, pp.533-558.
5. Lee CH, Lee ST (2001) On estimation of solutions of the generalized Lyapunov equations and the robust root clustering for the linear perturbed systems, Int. J. Control, 74, pp.996-1008.

6. Melnikov VG (2010) About root-clustering in sophisticated regions, 14th WSEAS International Conference Latest Trends on Systems, 1, pp. 297-300.
7. Melnikov VG (2010) Chebyshev economization in transformations of nonlinear systems with polynomial structure, 14th WSEAS International Conference Latest Trends on Systems, 1, pp.301-303.
7. Prepelita V (2008) Structural Properties of Linear Generalized Systems, WSEAS Transactions on Systems and Control, 1(3), pp. 701-711
8. Spitsa V, Alexandrovitz A, Zeheb E (2005) Root clustering method for a small-signal stability analysis of power systems, 7th WSEAS Int. Conf. on Math. Meth. and Comp. Tech. in El. Eng., pp.275-280
9. Koşar K, Durmaz S, Jafarov EM (2007) Longitudinal displacement autopilot design for Boeing 747-400 by root-locus Proc. of the 9th WSEAS International Conf. on Automatic Control, Modeling & Simulation, Istanbul, Turkey, May 27-29, pp.88-93

Chapter 12

Towards a Deterministic Model for Course Timetabling

P. Pozos-Parra, O. Chavez-Bosquez, J. L. Gomez Ramos

Department of Informatics and Systems, University of Tabasco, Carretera Cunduacan-Jalpa Km. 1, Tabasco, Mexico,
[pilar.pozos,oscar.chavez,jose.gomez}@dais.ujat.mx](mailto:{pilar.pozos,oscar.chavez,jose.gomez}@dais.ujat.mx)

Abstract. The Course Timetabling Problem consists in the weekly scheduling for all the lectures (events) of a set of university courses, subject to certain constraints. Unfortunately, course timetabling problems vary from university to university, and as far as we know, no standard formulation has been proposed from the community. Nevertheless, the International Timetabling Competitions, ITC-2002 and ITC-2007, have been organized with the aim of creating the common formulation for comparison. The formulation has become quite standard, and many solutions have been proposed in the literature. Most of the provided solutions utilize non-deterministic techniques. In this paper, we introduce a new deterministic algorithm for the solution of timetabling problems. The instances upon which the algorithm was tested are the official ones of the ITC-2002 web page. Almost all solutions run in less than 10% of the ITC-2002 benchmark time. The analysis is still ongoing, and it includes suitable extensions for tackling problems of the ITC-2007.

Keywords. Course Timetabling Problem, International Timetabling Competition, STF Algorithm, Scheduling Problem, Deterministic Search, Implementation.

12.1 Introduction

The Course Timetabling Problem consists of fixing a sequence of events (lectures) of a set of university courses within a number of rooms and time periods, usually weekly, satisfying some constraints. During the recent years a large number of works has been directed to automated timetabling. Unfortunately, course timetabling problems vary from university to university, and as far as we know, no standard formulation has been proposed from the community. Nevertheless, throughout the years it has been possible to characterize common underlying formulations that could be used for comparing algorithms.

In particular the International Timetabling Competitions, ITC-2002 and ITC-2007, have been organized with the aim of creating the common formulation for comparison. The formulations have become quite standard, and many solutions have been proposed in the literature [14, 6]. Even when the competition is open to stochastic and deterministic approaches, all the proposed solutions appearing in the competition web pages are stochastic [4], they utilize non-deterministic techniques, such as tabu search, genetic algorithms, simulated annealing, ant colony optimization, etc., thus the competitions missed of deterministic approaches; as far as we know there is no record of some deterministic approach to find feasible solutions and effectively participate in the contest.

The motivation of this work was to develop a deterministic algorithm that solves the hard constraints of the 20 instances of the ITC-2002 in a timely manner, so we introduce a new deterministic algorithm, called the Sort Then Fix (STF) algorithm, for solving the timetabling problems. The instances upon which the algorithm was tested are the official ones of the ITC-2002. Even when the proposal solves only instances from ITC-2002, the analysis is still ongoing, and we expect suitable extension of the algorithm for ITC-2007. One of the major advantages of the proposal is that the best and worst cases are the same; this means that the running time does not need to be averaged over some number of runs.

Although there is evidence of the implementation of an initial algorithm that preprocess the information problem in order to find a feasible solution [6, 7], authors do not provide the overall penalty nor the time spent in this initial phase. This deterministic algorithm may create standard feasible inputs for other metaheuristics in order to measure their real performance.

This paper is organized as follows, the next section introduces the target problem; the third section presents the deterministic algorithm; the fourth section provides the results in detail, and finally we conclude and discuss some possible research directions.

12.2 Target problem

We consider the problem of weekly scheduling a set of single events (or lectures). The problem has been discussed in [1] and it was the topic of ITC-2002 [2], where twenty artificial instances were proposed. The instances are available from the ITC-2002 web page. In addition, some instances have been proposed and made available via web in [5]. The data format used is an ad hoc fixed-structure text-only one. The input data comes in a single file containing the scalar values (number of events, rooms, room features, and students), followed by the elements of the input arrays, one per line. The output format must be done in a single file containing two scalar values (room and timeslot) per line, which indicate the schedule for events.

Formally, the problem consists of finding an optimal timetable within the following framework: there is a set of events $E = \{E_1, E_2, \dots, E_{nE}\}$ to be scheduled a set of rooms $R = \{R_1, R_2, \dots, R_{nR}\}$, where each room has 45 available timeslots, nine for each day in a five day week. There is a set of students $S = \{S_1, S_2, \dots, S_{nS}\}$ who attend the events, and a set of features $F = \{F_1, F_2, \dots, F_{nF}\}$ satisfied by rooms and required by events. Each event is attended by a number of students, and each room has a given size, which is the maximum number of students the room can accommodate. A feasible timetable is one in which all events have been assigned a timeslot and a room so that the following hard constraints are satisfied:

1. no student attends more than one event at the same time;
2. the room is big enough for all the attending students and satisfies all the features required by the event; and
3. only one event is scheduled in each room at any timeslot.

In contest instance files there were typically 10-11 rooms, hence there are 450-495 available places. There were typically 350-400 events, 5-10 features and 200-300 students.

The problem proposes to penalize a timetable for each occurrence of some soft constraint violations, which are the followings:

1. a student has to attend an event in the last timeslot on a day;
2. a student has more than two classes in a row; and
3. a student has to attend solely an event in a day.

The problem may be precisely formulated as:

- let $F = \{F_1, F_2, \dots, F_{nF}\}$ be a set of symbols representing the features;
- $R_i = \{F'_1, F'_2, \dots, F'_{n_{Ri}}\}$ where $F'_j \in F$ for $j = 1, \dots, n_{Ri}$ and n_{Ri} is the number of features satisfied by room R_i ,

- $N = \{N_1, \dots, N_{nR}\}$ be a set of integer numbers indicating the maximum of students each room can accommodate;
- $E_i = \{F''_1, F''_2, \dots, F''_{n_{Ei}}\}$ where $F''_j \in F$ for $j = 1, \dots, n_{Ei}$ and n_{Ei} is the number of features required by event E_i ;
- $S_i = \{E'_1, E'_2, \dots, E'_{n_{Si}}\}$ where $E'_j \in E$ for $j = 1, \dots, n_{Si}$ and n_{Si} is the number of events student S_i attends; and
- $T = \{T_1, \dots, T_{45}\}$ be a set of timeslots.

Find a feasible solution, i.e. a set of pairs $\{(T'_1, R'_1), \dots, (T'_{nE}, R'_{nE})\}$ such that:

- $T'_i \in T$ and $R'_i \in R$
- $\neg(T'_i = T'_j)$ if $E_i, E_j \in S_k$ and $\neg(i = j)$
- $E_i \subseteq R'_i$ and $|\{S_j \mid j = 1, \dots, n_S \text{ and } E_i \in S_j\}| \leq N_k$ where $R'_i = R_k$,
- $\neg \exists_i \exists_j (\neg(i = j) \wedge T'_i = T'_j \wedge R'_i = R'_j)$.

The competition adds a constraint over execution time, i.e. given the information about rooms, events, features, and students; find the best possible feasible solution within a given time limit. The time limit is given by a benchmark tool provided by the organizers.

12.3 Algorithm STF

We propose the Algorithm 1 to solve timetabling problems; it is based on events and rooms sorting.

The first step in the algorithm, apart from the obvious actions such as reading the problem file, is to find a binary matrix that represents the available rooms for every event. The following actions are performed:

- the number of students for each event is calculated and stored, $n_i = |\{S_j \mid j = 1, \dots, n_S \text{ and } E_i \in S_j\}|$
- a list of available rooms is created for each event, $e_i = \{R_j \mid E_i \subseteq R_j \text{ and } n_i \leq N_j\}$.

This first step allows us to reduce the problem by eliminating the information concerning features and room capacity, defining a new event set $\{e_1, \dots, e_{nE}\}$, called in the algorithm Event-Room, that includes the eliminated information. The new event set will be used for defining the most constrained event.

Our approach differs from direct heuristics [13], which usually fill up the complete timetable with one event at a time as far as no-conflicts arise,

Data:*Data-file***Result:***Solution-file***begin***Timeslots* $\leftarrow 45$ Read from *Data-file*: [*Events*: Number of events*Rooms*: Number of rooms*Features*: Number of features*Students*: Number of students*RoomCapacity* Vector of number of available places of rooms*EventStudent* Matrix of attendance to events by students*Feature-Room* Matrix of satisfied features by rooms*Feature-Event* Matrix of required features by events]*EventCapacity* \leftarrow find number of places required by every event*Event-Room* \leftarrow find binary matrix of suitable rooms for events*EventsARoom* \leftarrow find for every room the number of potential events to be allocated in*RoomSlot* \leftarrow create a matrix of ones_{*Rooms* * *timeslost*}*EventSlot* \leftarrow create a matrix of ones_{*Rooms* * *timeslost*}*R_S* \leftarrow create a matrix of zeros_{2 * *Events*}*IX-Rest* \leftarrow indexes of events**for** *i* = 1 ... *Events* **do***IX_i* \leftarrow find the index of the most constrained event in *IX-Rest**IX-Room* \leftarrow find the indexes of suitable rooms for event *IX_i*sort *IX-Room* from the least to the most constrained room *Events4Room* \leftarrow *Events4Room* - *Event-Room_{IX_i}***for** *r* \in *IX_Room* **do****for** *s* = 0 ... *Timeslots* - 1 **do***OK* $\leftarrow 1$, *k* $\leftarrow 1$ *IX_Student* \leftarrow find indexes of students associated to event *IX_i* **while** *k* < *i***and** *OK* **do****if** *R_S_{2,IX_i}* = *r* **and** *R_S_{1,IX_i}* = *s* **then***OK* $\leftarrow 0$, break**for** *j* \in *IX_Student* **do****if** *Event_Student_{IX_i,j}* **and** *R_S_{1,IX_i}* = *s* **then***OK* $\leftarrow 0$, break*k* $\leftarrow k + 1$ **if** *OK* **then***R_S_{1,IX_i}* = *s**RoomSlot_{r,s}* $\leftarrow 0$ **for** *j* = 1 ... *Events* **do****if** any(*Event_Student_{j,IX_Student}*) **then***EventSlot_{j,s}* $\leftarrow 0$ **break****if** *OK* **then***RS_{2,IX_i}* $\leftarrow r$, break*IX_Rest* \leftarrow *IX_Rest* - *IX_i*write to *Solution-file* : [*R_S*]**end**

```

Data:
  IX_Rest
  Event_Room
  Room_Slot
  Event_Slot
  EventCapacity
  Events4Room
Result:
  IX_i
begin
  sort IX_Rest in ascending order for  $Event\_Room_{IX\_Rest} * Room\_Slot. * Event\_Slot_{IX\_Rest}$ 
  for in case of tie do
    sort indexes with equal values in descending order for
       $EventCapacity_{IX\_Rest}$ 
  for in case of tie do
    sort indexes with equal values in ascending order for  $\sum_{Rooms} Event\_Room_{IX\_Rest}$ 
  for in case of tie do
    sort indexes with equal values in ascending order for  $Event\_Room_{IX\_Rest} * Events4Room$ 
     $IX_i \leftarrow$  first index of the sorted IX_Rest
end

```

Algorithm 2: find the index of the most constrained event

and then at that point they start making some swapping so as to accommodate other events.

Our approach does not consider swapping; if an event cannot be scheduled then the next most constrained event is scheduled letting the event in question without room and timeslot. Thus we focus in the definition of the “most constrained event” as precise as possible in order to avoid letting an event without schedule.

The core strategy of the algorithm is to assign the most constrained event to the least constrained timeslot found for that event. In order to avoid strategies of movement that provide a limited form of backtracking to recover from mistakes, our aim is not to make mistakes of selection of placement and event.

Algorithm 2 finds the most constrained event, the intuitive idea is as follows: sort *IX-Rest* (indexes of the unassigned events) in ascending order for the number of free slots, then for any index with equal number of free slots, sort them in descending order for the number of students attending the event, then for any index with equal number of students attending the associated event, sort them in ascending order for the number of available rooms for the event, then for any index with equal number of suitable rooms associated to the event, sort them in ascending order for the number of potential accommodation of events in its available rooms, finally the

most constrained event is the one appearing in the first place of the sorted *IX-Rest*.

We have developed an extension of STF algorithm for considering two soft constraints. The extension uses the same criteria of Algorithm 1.

12.4 Results

The STF algorithm was developed in GNU Octave 3.0 on a Toshiba Satellite A105-S433(R)laptop computer, with a 1.60 GHz Centrino Duo processor T5200, Mobile Intel(R)945GM Express Chipset, 2 GB in RAM, 160 GB in HDD and the Ubuntu Linux 9.10 operating system.

All tests were done in this box, according to the time provided by the benchmark tool provided by the organizers of the ITC-2002. The benchmark program took 558 seconds to end, so the STF algorithm has this time to find a feasible solution.

The Table 1.1 shows the results obtained by the STF algorithm in the 20 problem instances of the ITC-2002. The table display the hard constraints broken, the soft constraints, the time in seconds taken by the algorithm running on GNU Octave 3.0, an the time taken running on a trial version of Matlab® R2009b.

As the Table 1.1 shows, the STF algorithm finds a feasible solution in a timely manner, according to the benchmark program. It is worth noticing that the algorithm has a better performance on Matlab®, almost all solutions run in less than 10% of the benchmark time.

The Table 1.2 shows the results obtained by the extension STF algorithm in the 20 problem instances of the ITC-2002. The table display the hard constraints broken, the soft constraints broken, the time in seconds taken by the algorithm running on a trial version of Matlab® R2009b. It is worth noticing that the only soft constraint broken concerning students having single event in a day. Some events might not be placed in the timetable in order to ensure that no hard constraints are being violated. Such events are then be considered unplaced. It is worth noticing that in ITC-2007 it is allowed to generate timetables with unplaced events.

In order to verify the results, the participants of the ITC-2002 provide the executable file of their implementations. Nevertheless, the details are hidden in this file format. Our results can be reproduced using the source code of the STF algorithm.¹

¹ The STF implementation is available under request via email.

Table 1.1. Results of the STF algorithm in the 20 problem instances. Time in seconds.

Instance	Hard constraints	Soft constraints	GNU Octave time	Matlab time
01	0	940	139.491	35.775
02	0	870	129.923	33.507
03	0	926	132.475	34.354
04	0	1191	144.784	40.401
05	0	1395	130.249	33.199
06	0	1387	128.457	32.857
07	0	1736	104.068	27.671
08	0	1135	135.687	35.811
09	0	937	154.627	40.409
10	0	952	145.488	36.641
11	0	958	138.342	35.398
12	0	829	126.545	32.471
13	0	1097	141.385	36.905
14	0	1605	125.638	30.660
15	0	1448	131.486	31.743
16	0	1032	163.606	42.263
17	0	1491	134.854	35.476
18	0	926	138.835	35.992
19	0	1325	156.347	40.449
20	0	1450	141.738	34.448

12.5 Conclusions

STF algorithm has been proposed and appears to solve timetabling problems in a natural way. The idea is intended to sort the events before scheduling. Unlike other approaches STF is defined in a deterministic manner.

The STF algorithm solves effectively the 20 instances of the ITC-2002 within the time established by the benchmark program provided by the organizers of the contest. The approach utilized in the algorithm focuses only in the hard constraints, so the algorithm finds feasible solutions to the problem regardless the soft constraints.

As a main contribution, the STF algorithm may be used to find an initial solution to another optimization algorithm, so it can be combined with a non-deterministic approach, such as a metaheuristic, to build a hybrid

algorithm that can find a better solution faster than a metaheuristic by itself.

The metaheuristics used by the other participants of the contest do a preprocessing of the problem data before starting the search process, so all of the algorithms start with a different initial solution. To tackle this issue, the timetables generated by the STF algorithm can be used as an initial solution for all of these metaheuristics, in order to measure the real effectiveness of each metaheuristic with respect to the other ones.

Table 1.2. Results of the extension of STF algorithm in the 20 problem instances. Time in seconds. HC, Hard constraints broken. SC, Soft constraints broken.

Instance	HC	SC	Unplaced	Matlab time
01	0	80	66	31.07
02	0	91	53	29.22
03	0	88	59	91.43
04	0	148	76	72.51
05	0	137	69	49.74
06	0	149	69	104.82
07	0	143	42	61.47
08	0	82	45	44.62
09	0	103	57	62.45
10	0	79	63	60.93
11	0	109	71	69.57
12	0	84	48	49.40
13	0	125	68	78.12
14	0	139	46	58.20
15	0	133	54	58.45
16	0	88	66	139.45
17	0	146	58	40.12
18	0	72	53	47.95
19	0	104	62	82.83
20	0	139	61	73.11

The STF algorithm finds a feasible timetable in much less than the time available. A preliminary approach for finding perfect solutions has been proposed. This extension solves two soft constraints and reduces considerably the number of soft constraints broken.

Also, the STF algorithm will be extended for solving instances of the ITC-2007, in order to manage the new constraints proposed in the competition and thus close the gap between the real problems and theoretical solutions.

References

1. Rossi-Doria O, Sampels M, Birattari M, Chiarandini M, Dorigo M, Gambardella LM, Knowles J, Manfrin M, Mastrolilli M, Paechter B, Paquette L, Stutzle T (2003) A comparison of the performance of different metaheuristic-tics on the timetabling problem. PATAT 2002, LNCS, 2740 pp 329-351
2. Paechter B, Gambardella LM, Rossi-Doria O (2003) International Timetabling Competition Webpage (viewed January, 2010) (updated July 10, 2003), <http://www.idsia.ch/Files/ttcomp2002/>
3. McCollum B (2007) International Timetabling Competition Webpage (2007) (viewed January, 2010) (updated October 1, 2008), <http://www.idsia.ch/Files/ttcomp2002/>
4. di Gaspero L, Schaerf A (2003) Timetabling Competition TTComp 2002: Solver Description, (viewed January, 2010) (updated July 10, 2003), <http://www.idsia.ch/Files/ttcomp2002/schaerf.pdf>
5. Socha K, Knowles J, Sampels M (2002) A MAX-MIN Ant System for the University Timetabling Problem. Algorithmic Number Theory. LNCS, 2369
6. Kostuch P (2005) The University course timetabling problem with a three-phase approach. PATAT 2004, LNCS, 3616
7. Frausto-Solis JF, Alonso-Pecina F, Larre M, Gonzalez-Segura C, Gomez-Ramos JL (2006) Solving the timetabling problem with three heuristics. WSEAS Transactions on Computers. Issue 11, Volume 5, November 2006 Pireaus, Greece. ISSN 1109-2750
8. de Werra D (1985) An introduction to timetabling, European Journal of Operational Research
9. Gotlieb H (1963) The construction of class-teacher timetables. IFIP Congress
10. Kostuch P, Socha K (2004) Hardness prediction for the University Course Timetabling problem. Proceedings of the Evolutionary Computation in Combinatorial Optimization, EvoCOP
11. Rossi-Doria O, Blum C, Knowles J, Sampels M, Socha K, Paechter B (2002) A local search for the timetabling problem, PATAT 2002
12. Socha K (2003) Metaheuristics for the Timetabling Problem. DEA Thesis, Univesite Libre de Bruxelles
13. Schaerf A (1995) A survey of automated timetabling, Technical Report. Centrum voor Wiskunde en Informatica
14. Chiarandini M, Birattari M, Socha K, Rossi-Doria O (2006) An effective hybrid approach for the university course timetabling problem. Journal of scheduling, 9 (2006) pp 403-432

15. Chavez Bosquez O, Pozos Parra P, Lengyel F (2010) Towards a Deterministic Algorithm for the International Timetabling Competition. Proceedings of the 17th International Workshop on Experimental Evaluation of Algorithms for Solving Problems with Combinatorial Explosion, RCRA
16. Stanarevic N, Tuba M, Bacanin N (2010) Enhanced Artificial Bee Colony Algorithm Performance. In 14th WSEAS International Conference on Computer (Part of the 14th WSEAS CSCC Multiconference), Corfu, Greece
17. Stoica F, Cacovean L (2010) An Evolutionary Approach for Optimizing Stochastic Learning Automata. In 14th WSEAS International Conference on Computer (Part of the 14th WSEAS CSCC Multiconference), Corfu, Greece
18. Leros A, Andreatos A, Zagorianos A (2010) Matlab-Octave Science and Engineering Benchmarking and Comparison. In 14th WSEAS International Conference on Computer (Part of the 14th WSEAS CSCC Multiconference) Corfu, Greece
19. Rezaeian J, Javadian N (2010) Designing an Incremental Cellular Manufacturing System Based on Heuristic Methods. In 14th WSEAS International Conference on Computer (Part of the 14th WSEAS CSCC Multiconference), Corfu, Greece.

Chapter 13

Periodic Flows Associated with Solutions of Hill Equation

G. Bogнар¹, S. H. Sohrab²

¹Department of Analysis, University of Miskolc, 3515
Miskolc-Egyetemváros, Hungary, matvbg@uni-miskolc.hu

²Department of Mechanical Engineering, Northwestern University,
Evanston, Illinois 60208, USA, s-sohrab@northwestern.edu

Abstract. Periodic solutions of Hill's differential equation and its associated ordered eigenvalue pairs with 2π and π periodicity are investigated. Also, the occurrence of Hill equation in a number of problems of laminar boundary layer flows is demonstrated and the relevant solutions are described.

Keywords. Hill's Differential Equation, Boundary Layer, Inverse Problems of Fluid Mechanics

13.1 Introduction

Hill's differential equation is a homogeneous, linear, second order differential equation

$$f'' + q(t)f = 0, \quad (1.1)$$

with real periodic coefficients q . It is named after G.W. Hill, who investigated it in his study of lunar motion [6]. The motion of particles or massive bodies in periodic external fields has been studied extensively by astronomers in the past century especially in connection with the three-body problem. It is also well known in the quantum theory of metals and semi-conductors [1], or in optics when ultra short optical pulses are

examined [2]. The value of the period of the solution plays an important role in the discussion of periodic solutions. A specific question is the case of solutions of period π and 2π [3-5]. This equation involves Mathieu differential equation and Lamé's differential equation as particular cases and by suitable transformations, Legendre's differential equation and the confluent hypergeometric differential equation as well.

There exist hundreds of applications of Hill's equation to problems in engineering, physics, astronomy, the theories of electric circuits and electric conductivity. Our aim here is to show some applications of the Hill's equation to the boundary layer theory of laminar fluid flows.

13.2 Hill's Differential Equation

We will discuss in this part the equation called Hill's equation, its solutions, and properties. If $q(t) = \lambda^2$, where λ is a positive constant, then (1.1) is the simple wave equation with solutions $C \sin[\lambda(t - t_0)]$ which are clearly oscillatory in \mathbf{R} . We have to point out two properties. First, that the zeros of linearly independent solutions are interlaced. Between two consecutive zeros of one of the solutions there is only one zero of the other linearly independent solution. The second feature is that the distance between consecutive zeros is π/λ and becomes smaller the larger λ is. It is possible to generalize these properties to the case of solutions of (1.1). Let us consider the differential equation (1.1), where $q(t)$ is an integrable, real function of period π . We apply the Prüfer's method, which provides a peculiar type of polar coordinates in order to handle the zeros of linear second order equations. For solution f of (1.1) we introduce A and φ defined by

$$\begin{aligned} f(t) &= A(t) \sin \varphi(t) \\ f'(t) &= A(t) \cos \varphi(t) \end{aligned} \quad (1.2)$$

The differential equations for A and φ are

$$\begin{aligned} \varphi'(t) &= 1 - [\sin \varphi(t)]^2 + q(t) [\sin \varphi(t)]^2 \\ A'(t) &= A(t)(1 - q(t)) [\cos \varphi(t)]^2 \end{aligned} \quad (1.3)$$

From the initial conditions $f(0) = 0$, $f'(0) = a \neq 0$ we get $A(0) = A_0$ and $\varphi_0 = \varphi(0) \in [0, \pi]$. We find φ from the first equation of (1.3), since this equation is independent of A . Once φ is determined, A , is found by

integration. A has to be exponential and therefore it cannot vanish. Moreover, $A(t)/A(0)$ is a positive function. We obtain from (1.3) that $|A'(t)/A(t)| \leq |1 - q(t)|$ from where $\log A(t)$ is of bounded variation and $\lim_{t \rightarrow \infty} A(t) = A_0$. $A(t)$ has to be exponential and therefore it cannot vanish.

M. Hukuhara and M. Nagumo [7] showed that all solutions of (1.1) are bounded if q satisfies the condition: there exists constant $M > 0$ such that

$\int_0^\infty |q(t) - M| dt < \infty$. In this case the solution of (1.1) with initial conditions

$f(0) = 0, \quad f'(t_0) = a \neq 0$ can be given by $f(t) = (A_0 + o(t)) \sin(t + \varphi_0 + o(t))$, where $\lim_{t \rightarrow \infty} o(t) = 0$.

In the theory of Hill's equation (1.1) the solutions of period π and 2π play an exceptional role. Let us consider the case $q(t) = \lambda + Q(t)$. The main problem is the determination of those values of λ for which the equation $f'' + [\lambda + Q(t)]f = 0$ has a solution of period π or 2π (see Floquet's theorem [9]). If $f(t, \lambda)$ is periodic with period π , then it satisfies conditions $f(\pi, \lambda) = f(0, \lambda)$, $f'(\pi, \lambda) = f'(0, \lambda)$ and from (1.2) we get that $A(\pi, \lambda) = A(0, \lambda)$, $\varphi(\pi, \lambda) - \varphi(0, \lambda) = 2k\pi$, where k is a positive integer. If $f(t, \lambda)$ is periodic with period 2π , then it satisfies the so-called semi-periodic conditions: $f(\pi, \lambda) = -f(0, \lambda)$, $f'(\pi, \lambda) = -f'(0, \lambda)$, hence $A(\pi, \tilde{\lambda}) = A(0, \tilde{\lambda})$, $\varphi(\pi, \tilde{\lambda}) - \varphi(0, \tilde{\lambda}) = (2k - 1)\pi$, where k is a positive integer. To every differential equation (1.1), there belong two monotonically increasing infinite sequences of real numbers $\lambda_0, \lambda_1, \lambda_2, \dots$ and $\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3, \dots$ such that (1.1) has a solution of period π if and only if $\lambda = \lambda_n$, $n = 0, 1, 2, \dots$, and a solution of period 2π if and only if $\lambda = \tilde{\lambda}_n$, $n = 1, 2, 3, \dots$. The λ_n and $\tilde{\lambda}_n$ satisfy the inequalities

$$\lambda_0 < \tilde{\lambda}_1 \leq \tilde{\lambda}_2 < \lambda_1 \leq \lambda_2 < \tilde{\lambda}_3 \leq \tilde{\lambda}_4 < \lambda_3, \dots, \quad (1.4)$$

and the relations $\lim_{n \rightarrow \infty} (\lambda_n)^{-1} = 0$, $\lim_{n \rightarrow \infty} (\tilde{\lambda}_n)^{-1} = 0$. The real numbers λ_n shall be called characteristic values of first kind of (1.1) and the $\tilde{\lambda}_n$ shall be called characteristic values of second kind. The solutions are stable in the intervals $(\tilde{\lambda}_2, \lambda_1)$, $(\lambda_2, \tilde{\lambda}_3)$, $(\tilde{\lambda}_4, \lambda_3)$ (see [3]).

The asymptotic behavior of the characteristic values was examined in [8-10]. In that case when $Q(t)$ is periodic with period π such that

$\int_0^\pi |Q(t)|/\pi dt = B$ exists and $\int_0^\pi Q(t)dt = 0$, then for any integer $n > B/2$ for

the characteristic values: $|\sqrt{\lambda_{2n-1}} - 2n| \leq B/(4n)$, $|\sqrt{\lambda_{2n}} - 2n| \leq B/(4n)$,

$|\sqrt{\tilde{\lambda}_{2n-1}} - 2n + 1| \leq B/(4n - 2)$, $|\sqrt{\tilde{\lambda}_{2n}} - 2n + 1| \leq B/(4n - 2)$ hold (see G.

Borg [10]).

We note that many of the properties valid for (1.1) can be extended to the nonlinear version of the Hill's equation of the form $f''|f|^{p-1} + q(t)f|f|^{p-1} = 0$, where $p > 0$ ([11, 12]).

13.3 Hill Equation and the Modified Form of the Equation of Motion

In this section, some examples of applications of Hill's equation to the problems of laminar boundary layer in fluid mechanics will be discussed. The scale invariant modified forms of the conservation equations were recently introduced as [13]

$$\frac{\partial \rho_\beta}{\partial t} + \mathbf{w}_\beta \cdot \nabla \rho_\beta - D_\beta \nabla^2 \rho_\beta = \Omega_\beta \quad (1.5)$$

$$\frac{\partial T_\beta}{\partial t} + \mathbf{w}_\beta \cdot \nabla T_\beta - \alpha_\beta \nabla^2 T_\beta = -\tilde{h}_\beta \frac{\Omega_\beta}{\rho_\beta c_{p\beta}} \quad (1.6)$$

$$\frac{\partial \mathbf{v}_\beta}{\partial t} + \mathbf{w}_\beta \cdot \nabla \mathbf{v}_\beta - \nu_\beta \nabla^2 \mathbf{v}_\beta = -\frac{\nabla p_\beta}{\rho_\beta} + \frac{1}{3} \nu_\beta \nabla (\nabla \cdot \mathbf{v}_\beta) - \mathbf{v}_\beta \frac{\Omega_\beta}{\rho_\beta} \quad (1.7)$$

where Ω_β is the reaction rate and the last equation is the modified form of the equation of motion that is linear as compared to the classical non-linear Navier-Stokes equation of motion

$$\frac{\partial \mathbf{v}_\beta}{\partial t} + \mathbf{v}_\beta \cdot \nabla \mathbf{v}_\beta - \nu_\beta \nabla^2 \mathbf{v}_\beta = -\frac{\nabla p_\beta}{\rho_\beta} + \frac{1}{3} \nu_\beta \nabla (\nabla \cdot \mathbf{v}_\beta), \quad (1.8)$$

The main difference between the modified (1.7) and the classical (1.8) forms of the equation of motion is the distinction between the convective

versus the local velocity in the former. In the following some examples of fluid flow models that lead into Hill's equation will be examined.

An important feature of (1.7) is that one can first determine the convective velocity \mathbf{w}_β from the solution of the inviscid potential flow $\mathbf{v}_{\beta+1}$ at the outer scale $(\beta+1)$. Next, with the known convective velocity one can obtain the associated local velocity \mathbf{v}_β from the solution of (1.7) in the presence of viscous term. Let us consider the classical problem of *Blasius*, i.e. laminar flow over a flat plate [14, 15]. For an incompressible steady laminar flow in the absence of reactions the last two terms on the right-hand side of (1.7) vanish and it simplifies to

$$\mathbf{w}_\beta \cdot \nabla \mathbf{v}_\beta - \nabla^2 \mathbf{v}_\beta = -\nabla p_\beta, \quad (1.9)$$

involving the dimensionless velocities $(v_x, v_y, w_x, w_y) = (v'_x, v'_y, w'_x, w'_y) / w'_o$, $p = p' / \rho w_o'^2$ and coordinates $x = x'_e / l_H$, $y = y' / l_H$, $l_H = \nu / w'_o$ where l_H is the hydrodynamic diffusion length and ν is the kinematic viscosity. First, in the outer potential flow one sets the viscosity to zero thus neglecting the diffusion term and hence the diffusion velocity [13] $v_x - w_x = V_x = -\nu \nabla \ln(\rho \nu) = 0$ such that $v_x = w_x$ and with $w_x = 0$, (1.9) reduces to

$$(1/2) \nabla \mathbf{v}_{\beta+1}^2 = -\nabla p_{\beta+1} \quad (1.10)$$

For uniform outer velocity $v_{x\beta+1} = \text{constant}$, the solution of (1.10) gives the constant convective velocity and pressure as

$$v_{x\beta+1} = w_{x\beta} = 1, \quad p = \text{constant} \quad (1.11)$$

Next, for the inner flow within the boundary layer the usual assumption $\partial^2 / \partial x'^2 \ll \partial^2 / \partial y'^2$ and constant pressure from (1.11) simplify (1.9) to

$$w_{x\beta} \frac{\partial v_{x\beta}}{\partial x_\beta} = \frac{\partial^2 v_{x\beta}}{\partial y_\beta^2} \quad (1.12)$$

Further substitution from (1.11) for the convective velocity leads to

$$\frac{d^2 v_x}{d\xi^2} + 2\xi \frac{\partial v_x}{\partial \xi} = 0, \quad (1.13)$$

where $\xi = y/\sqrt{4x} = \eta/2$ that with the boundary conditions $v_x(0) = 0$, $v_x(\infty) = 1$ leads to the solution

$$v_x = \text{erf}(\xi) = \text{erf}(\eta/2) \quad (1.14)$$

Defining the edge of the boundary layer as the location where $v_x(\xi^*) = 0.99938$ one obtains from (1.14) $\xi^* \approx 2.5$, $\eta^* \approx 5$ in exact agreement with the classical result of Blasius [14]. It is emphasized that the modified theory leads to analytical solution (1.14) while the classical nonlinear problem requires numerical solution.

As a second example let us consider the classical problem of stagnation-point flow with the outer convective velocity field [14, 16] $w'_x = \Gamma x'$, $w'_y = -\Gamma y'$, where Γ is the stagnation flow velocity gradient. One introduces the dimensionless quantities $(v_x, v_y, \mathbf{w}) = (v'_x, v'_y, \mathbf{w}')/\sqrt{\nu\Gamma}$, $p = p'/\rho w_o'^2$ and coordinates $x = x'_e/\delta$, $y = y'/\delta$, $\delta = \sqrt{\nu/\Gamma}$. For the potential flow at the outer scale $\beta+1$ the vanishing of viscosity leads to zero diffusion velocity hence $v_x = w_x$ and (1.9) reduces to the Euler equation

$$v_x \frac{\partial v_x}{\partial x} + v_y \frac{\partial v_x}{\partial y} = -\frac{\partial p}{\partial x}, \quad v_x \frac{\partial v_y}{\partial x} + v_y \frac{\partial v_x}{\partial y} = -\frac{\partial p}{\partial y} \quad (1.15)$$

with the local solutions close to stagnation-point

$$v_{x\beta+1} = x, \quad v_{y\beta+1} = -y, \quad p = 1 - (x^2 + y^2)/2 \quad (1.16)$$

The stream function for the outer potential flow (1.16) is $\Psi_o = xy$. Next, the flow within the boundary layer at the lower scale β is considered and the local velocity of the outer potential flow (1.16) provides for the convective velocity of the inner viscous flow as

$$v_{x\beta+1} = w_{x\beta} = x, \quad v_{y\beta+1} = w_{y\beta} = -y \quad (1.17)$$

Assuming that the pressure gradient within the boundary layer is negligible and $\partial^2 v_x / \partial x'^2 \ll \partial^2 v_x / \partial y'^2$, $\partial^2 / \partial x'^2 \ll \partial^2 / \partial y'^2$ one obtains from (1.9)

$$w_y \frac{\partial v_x}{\partial y} = \frac{\partial^2 v_x}{\partial y^2}. \quad \text{That with the boundary conditions } v_x(0) = 0, \quad v_x(\infty) = x$$

and upon substitution from (1.17) leads to the solution [16]: $v_x = x \text{erf}(y/\sqrt{2})$. Also, from the continuity equation (1.5) for an

incompressible flow one obtains $v_y = -\int_0^y \operatorname{erf}(y/\sqrt{2}) dy$ leading to the stream function

$$\Psi = -x \int \operatorname{erf}(y/\sqrt{2}) dy \quad (1.18)$$

The above procedures show that for any convective velocity field associated with an outer potential flow one can determine the corresponding solution of the equation of motion for the inner viscous boundary layer. Therefore, one next searches for possible solutions of (1.12) for flow fields when the convective and local velocities have the forms

$$w_x = a(x)q(y), \quad v_x = b(x)f(y) \quad (1.20)$$

and satisfy the condition

$$a(x) = -b(x)/b'(x) \quad (1.21)$$

Substituting from (1.20)-(1.21) into (1.13) results in

$$f'' + q(y)f = 0 \quad (1.22)$$

that becomes *Hill's* equation (1.1) if $q(y)$ is a periodic function of y .

The assumption of incompressibility $\nabla \cdot \mathbf{w}_\beta = 0$, $\nabla \cdot \mathbf{v}_\beta = 0$ and hence the condition (1.21) will be satisfied for certain pairs the functions $a(x)$ and $b(x)$ in (1.20) with a few examples given in Table 1.1.

Table 1.1. Pairs of conjugate functions $a(x)$ and $b(x)$

$a(x)$	x	$\tan x$	$(\tan x)/2$	$(\tan x)/n$	$(\sin^2 x + K)/(2 \sin x \cos x),$	$\sin x \cos x$	$x \ln x$
					$K = \text{const.}$		
$b(x)$	x	$\sin x$	$\sin^2 x$	$\sin^n x$	$\sin^2 x + K$	$\tan x$	$\ln x$

For the case of $a(x) = b(x) = x$ from Table 1.1, let us consider an example of *Hill's* equation when the outer potential flow has periodicity in the y -coordinate and given as

$$w_x = xq(y) = -2x(1 - \tan^2 y) \quad (1.23)$$

such that by (1.5) we have

$$w_y = 2(y - \tan y) \quad (1.24)$$

The stream function for outer potential flow becomes $\Psi_o = -2x(2y - \tan y)$. The flow fields (1.23)-(1.24) are somewhat similar to (1.17) except that due to certain obstacles periodically positioned in the transverse direction they have acquired periodicity in the y -coordinate. However, the exact nature of the outer potential flow requires further examination and is as yet not fully understood. With the outer flow (1.23) and the inner boundary layer velocity of the form (1.20) equation (1.12) reduces to (1.22) and becomes the Hill's equation with the solution

$$v_x = -2x \cos^2 y \quad (1.25)$$

The result (1.25) and (1.5) give $v_y = y + (1/2)\sin 2y$ that lead to the stream function $\Psi_i = -x[y + (1/2)\sin 2y]$. Near the wall $y = 0$, the outer (1.23) and the inner (1.25) velocities become identical.

Even though the flow fields could indeed represent a model for a realistic hydrodynamic problem involving an accelerating basic flow $w_x = -2x$ with periodic transverse modulations due to certain obstacles.

A conjugate flow field of the form (1.20) that also satisfies (1.22) is the outer potential flow

$$w_x = -2x(1 - \cot^2 y), \quad w_y = 2(2y + \cot y), \quad \Psi_o = -2x(2y + \cot y) \quad (1.26)$$

and the associated inner viscous flow

$$v_x = 2x \sin^2 y, \quad v_y = 2(2y - \sin 2y), \quad \Psi_i = -x[y - (1/2)\sin 2y] \quad (1.27)$$

In view of the results (1.20)-(1.27), one can identify a general procedure whereby a known outer solution with $a(x)$ and $q(y)$ are substituted in (1.21) and (1.22) for the determination of the corresponding inner solution $b(x)$ and $f(y)$. On the other hand, an *inverse problem* is encountered when a known inner flow field given by $b(x)$ and $f(y)$ are applied to obtain the corresponding outer potential flow $a(x)$ and $q(y)$ from (1.21) and (1.30) by direct substitutions. The latter problem arises in design of hydrodynamic systems when one wants to know what outer potential flow field should be imposed in order to induce a desired local boundary layer flow field. If one considers an outer potential flow that is periodic in two spatial directions such as

$$w_x = 2 \tan x(1 - \cot^2 y), \quad w_y = -\frac{2}{\cos^2 x}[2y + \cot y] \quad (1.28)$$

by (1.21) and Table 1.1, the corresponding inner velocity field will have the form

$$v_x = 2 \sin x f(y) \quad (1.29)$$

By substitutions from (1.28)-(1.29) into (1.12) one obtains the solution

$$v_x = 2 \sin x \sin^2 y, \quad v_y = -\cos x [y - (1/2) \sin 2y] \quad (1.30)$$

The stream function for the outer flow fields (1.28) will be $\Psi_o = 2 \tan x (2y + \cot y)$. The exact nature of the potential flow (1.28) requires further future investigation. The stream function for the inner flow (1.30) will be $\Psi_i = \sin x [y - (1/2) \sin 2y]$. Although we do not address unsteady flows in the present study it is clear that if instead of (1.12) one considers

$$\frac{\partial v_{x\beta}}{\partial t} + w_{x\beta} \frac{\partial v_{x\beta}}{\partial x_\beta} = \frac{\partial^2 v_{x\beta}}{\partial y_\beta^2} \quad (1.31)$$

and neglects the time dependence of the outer solution then unsteady solutions of the form $v_x = b(x)f(y)e^{-\lambda t}$ substituted in (1.31) will lead to equations of type $f'' + [\lambda + Q(t)]f = 0$ discussed in the previous section. Therefore, it is reasonable to expect that the ordered pairs of eigenvalues identified in (1.4) for solutions with 2π and π periodicity correspond to the zeros of the outer and the inner periodic solutions such as (1.23) and (1.25) discussed above. In other words, between each pair of zeros of the outer flow there is a pair of zeros of the inner flow.

13.4 Concluding Remarks

Periodic solutions of the Hill equations with periodicity π and 2π were described and their ordered pairs of eigenvalues were identified. Also, some examples of the Hill equation encountered in a number of problems of laminar boundary layer flows are demonstrated and the relevant solutions are described.

References

1. Caspersen LW (1984) Solvable Hill equation. *Phys. Rev. A.* 30:2749-2751
2. Krylov V, Rebane A, Kalintsev AG, Schwoerer H, Wild UP (1995) Secondharmonic generation of amplified femtosecond Ti: sapphire laser pulses. *Optic Lett.* 20:198-200
3. W. Magnus W, Winkler S (1979) *Hill's Equation*. Dover Publ. Inc.
4. Coddington EA, Levinson N (1955) *Theory of Ordinary Differential Equations*. McGraw Hill, New York
5. Eastham MSP (1973) *The Spectral Theory of Periodic Differential Equations*. Scottish Academic Press, Edinburgh London
6. Hill GW (1886) On the Part of the Motion of Lunar Perigee Which is a Function of the Mean Motions of the Sun and Moon. *Acta Math.* 8:1-36
7. Hukuhara M, Nagumo M (1930) On a condition of stability for a differential equation. *Proc. Imp. Akad. Tokyo.* 6:131-132
8. Hochstadt H (1961) Asymptotic estimates for the Sturm-Liouville spectrum. *Commun. Pure Appl. Math.* 15:749-764
9. Floquet G (1883) Sur les équations différentielles linéaires à coefficients périodiques. *Ann. Sci. École Norm. Sup.* 12:47-82 (IV 6)
10. Borg G (1949) On a Liapunoff criterion of stability. *Amer. J. Math.* 71:67-70
11. Bognár G (2006) Lower bound for the eigenvalues of quasilinear Hill's equation. *Proc. Conf. on Differential and Difference Equations and Applications*, Edited by Ravi P. Agarwal and Kanishka Pereira, pp. 211-219
12. Bognár G (2008) Periodic and antiperiodic eigenvalues for quasilinear differential equations. *Recent Advances on Applied Mathematics*, WSEAS Press. pp. 307-313
13. Sohrab SH (2008) Derivation of invariant forms of conservation equations from the invariant Boltzmann equation. in *New Aspects of Fluid Mechanics and Aerodynamics*, S. H. Sohrab, H. J. Catrakis, N. Kobasko, N. Necasova, and N. Markatos (Eds.), WSEAS Press
14. Schlichting H (1968) *Boundary-Layer Theory*. McGraw Hill, New York
15. Sohrab SH (2005) Modified theory of laminar boundary layer flow over a flat plate. S. H. Sohrab, *IASME Transactions*. *IASME Transactions.* 2:1389
16. Sohrab SH (2005) Modified theories of axi-symmetric stagnation-point laminar boundary layer flow and counterflow jets. *IASME Transactions.* 2:1097

Chapter 14

Simulation Model of the JPEG Sub-Image Compression based on an Upgraded Petri Net

P. Strbac,¹ M. Tuba²

¹Faculty of Computer Science, Megatrend University Belgrade, Bulevar umetnosti 29, 11070, Belgrade, strbac@beotel.rs

²Faculty of Computer Science, Megatrend University Belgrade, Bulevar umetnosti 29, 11070, Belgrade, tubamilan@ptt.rs

Abstract. The objective of this paper is to develop a model of JPEG 8×8 sub-image compression and decompression for simulation and analysis by using an Upgraded Petri Net (UPN). The Upgrade Petri net is an extension of an ordinary Petri Net and a formal modeling tool appropriate for simulation and analysis of processes, particularly at the register transfer level. This paper describes an UPN model which can generate the compressed image for given input image and initial marking. From this model a new model can be derived which includes different parameters for simulation and analysis. As an example, a model that generates data for analysis of average absolute error is included. Original software for modeling and simulations of UPN, PeM (Petri Net Manager), is developed and used for all models described in this paper. This paper includes a preface on the UPN theory, the UPN model formulation, simulation and analysis.

Keywords. Upgraded Petri Net, Image Compression, JPEG, Modeling.

14.1 Introduction

Upgraded Petri nets (UPN) are a formal mathematical apparatus which enables modeling, simulation and process analysis [1]. They enable interactive monitoring of process operations. The hierarchical structure of an UPN gives wide possibilities for abstraction and facilitates the model implementation consisting, at the same time, of elaborate pieces essential for the analysis at a certain level, and also of some general pieces whose details are irrelevant for the analysis at the given level of abstraction.

This paper presents the usage of UPN for modeling, simulation and analysis of JPEG 8×8 sub-image compression [2, 3, 4]. The Upgraded Petri net model includes the visualization of data flow and data generation through execution of the UPN model.

An Upgraded Petri net is used to develop the low level model which refers to detailed representation of JPEG method for image compression and decompression on an 8×8 8-bit sub-image. The Upgraded Petri net model can be changed with respect to the object of analysis. This object can be: quality of compression, compression ratio, average absolute error, etc. [5, 6] or some other aspects of the algorithm [7, 8, 9]. As an example, this UPN model generates data for analysis of average absolute error.

For given input parameters the model generates data into a *.mem file. Input parameters are determined by: initial marking, arcs multiplicities, transition functions, transition firing levels, place attributes, and UPN conflict solving rules.

14.2 Upgraded Petri Net Formal Theory

Upgraded Petri net formal theory is based on functions. An Upgraded Petri net is a 9-tuple:

$$C = (P, T, F, B, \mu, \theta, TF, TFL, PAF) \text{ where :}$$

- $P = \{p_1, p_2, p_3, \dots, p_n\}$, $n > 0$ - a finite nonempty set of places p_i
- $T = \{t_1, t_2, t_3, \dots, t_m\}$, $m > 0$ - a finite nonempty set of transitions t_j
- $F: T \times P \rightarrow N_0$ - Input Function;
- $B: T \times P \rightarrow N_0$ - Output Function;
- $\mu: P \rightarrow N_0$ - Marking Function;
- $\theta: T \times \Delta \rightarrow \lambda$ - Timing Function;
- $TF: T \rightarrow A$ - Transition Function;
- $TFL: T \rightarrow N_0$ - Transition Firing Level;
- $PAF: P \rightarrow (x, y)$ - Place Attributes Function;

The *input function* assigns a non-negative integer number to an ordered pair (t_i, p_j) . This number defines how many times the place p_i is input into the transition t_i . N_0 represents the set of non-negative integer numbers. The set of places which are input into the transition t_j is labeled as ${}^*t_j = \{p_i \in P, F(t_j, p_i) > 0\}$. The following notation is introduced: ${}^*t_j^S$ for the place $p_i \in {}^*t_j$ which has the standard input into the t_j , $F^S(t_j, p_i)$ for the corresponding input function, ${}^*t_j^I$ for the places $p_i \in {}^*t_j$ with inhibitor arc into the t_j transition, and $F^I(t_j, p_i)$ for the corresponding input function.

The *output function* assigns a non-negative integer number to the ordered pair (t_i, p_j) . This number determines how many times the place p_i is input into to the transition t_i . The set of places which are input into the transition t_j is labeled as $t_j^* = \{p_i \in P, F(t_j, p_i) > 0\}$.

The *marking function* can be defined as an n -component vector: $\mu = (\mu_1, \mu_2, \dots, \mu_n)$, where $n = |P|$ and $\mu_{(p_i)}$ can be used instead of μ_i . The marking function assigns a non-negative integer number to the place p_i .

The *timing function* assigns the probability λ_j to the ordered pair (t_i, Δ_{ij}^q) . The transition t_i can be fired at the interval Δ_{ij}^q with the probability λ_j . The notation Δ_{ij}^q represents the q -th interval of the transition t_j .

The *transition function* $TF(t_j)$ assigns one operation $\alpha_j \in A$ to the transition t_j , where A is the set of operations which can be assigned to the transition. This operation can be: arithmetical, logical or file operation.

The *transition firing level function* assigns a non-negative integer number to the transition t_j . If that number is zero then all places $p_i \in {}^*t_j$ affect the transition t_j firing, otherwise it shows the number of places $p_i \in {}^*t_j$ that participate in the transition firing.

The *place attributes function* assigns an ordered pair (x, y) to the place $p_i \in {}^*t_j$ where $x \in R$, $y \in N_0$. Attributes x and y represent value and position, respectively, of an operand for the operation α_j assigned to the transition t_j .

An Upgraded Petri net execution performs system state changes triggered by firing of appropriate transitions. By Upgraded Petri net execution the marking, the contents of the *.mem file and attributes which belong to the places $p_i \in {}^*t_j$ of the fired transition t_j can change.

A transition $t_j \in T$ of the UPN $C = (P, T, F, B, \mu, \theta, TF, TFL, PAF)$ can be enabled if the next 3 conditions are true:

1. if timing function $\theta(t_j, \Delta_{ij}^q) > 0$ over an interval Δ_{ij}^q where transition t_j can become enabled.
2. if $TFL(t_j) > 0$, then $(\#p_i \in {}^*t_j) + (\#p_k \in {}^*t_j) = TFL(t_j)$
if $TFL(t_j) = 0$, then $(\#p_i \in {}^*t_j) + (\#p_k \in {}^*t_j) = |{}^*t_j|$
where:

- $\#p_i$ is the number of places $p_i \in {}^*t_j^S$ which satisfies: $\mu(p_i) \geq F^S(t_j, p_i)$
 $\#p_k$ represents number of places $p_k \in {}^*t_j^I$ which satisfy: $\mu(p_k) = 0$
 3. if a logical operation $\alpha_j \in A$ assigned to the transition t_j
 then the result of the operation α_j must be true.

A marking μ will be changed to the new marking μ' by firing of transition t_j where:

- $\mu'(p_i) = \mu(p_i) - F(t_j, p_i) + B(t_j, p_i)$, $p_i \in {}^*t_j^S$
- $\mu'(p_k) = \mu(p_k) + B(t_j, p_k)$, $p_k \in {}^*t_j^I$

By firing of the transition t_j a corresponding arithmetic or file operation is executed. A logical operation which is assigned to the transition t_j will be executed if conditions 1 and 2 related to t_j are true.

14.3 An Upgraded Petri Net Model of the JPEG Image Compression and Decompression

The JPEG algorithm is a well established standard for digital image compression which features an excellent compression ratio. It is unavoidable since even relatively low resolution images contain megabytes of data. The JPEG algorithm is applicable to wide range of continuous-tone digital images without any restrictions regarding image resolution or color representing scheme and with tractable computational complexity to make possible efficient software, as well as cheap hardware implementations.

The JPEG algorithm is based on discrete cosine transform (DCT) [10, 11, 12]. Basic steps of JPEG sequential encoding are forward DCT, quantization and entropy coding (where a variation of *run-length* encoding appears). Most digital images reveal strong coherence in horizontal and vertical direction that means intensity is usually slowly changing within large pixel blocks. Gradual changes of image digital signal means that it has low frequency components dominating, while most of high frequency components are equal 0. The JPEG algorithm converts the image to frequency domain where only small number of components is significant and high compression ratio can be attained.

The JPEG algorithm is not a single fixed standard, but rather a set of rules and recommendations with many adjustable parameters. Different applications require emphasis on different features of the JPEG algorithm. JPEG standard specifies that trade-off between compression ratio and image fidelity can be further fine-tuned by setting appropriate algorithm

parameters. An Upgraded Petri net model can be used to experiment with these parameters.

The Upgraded Petri net model of the JPEG image compression and decompression on a 8×8 8-bit sub-image is depicted in Fig. 1.

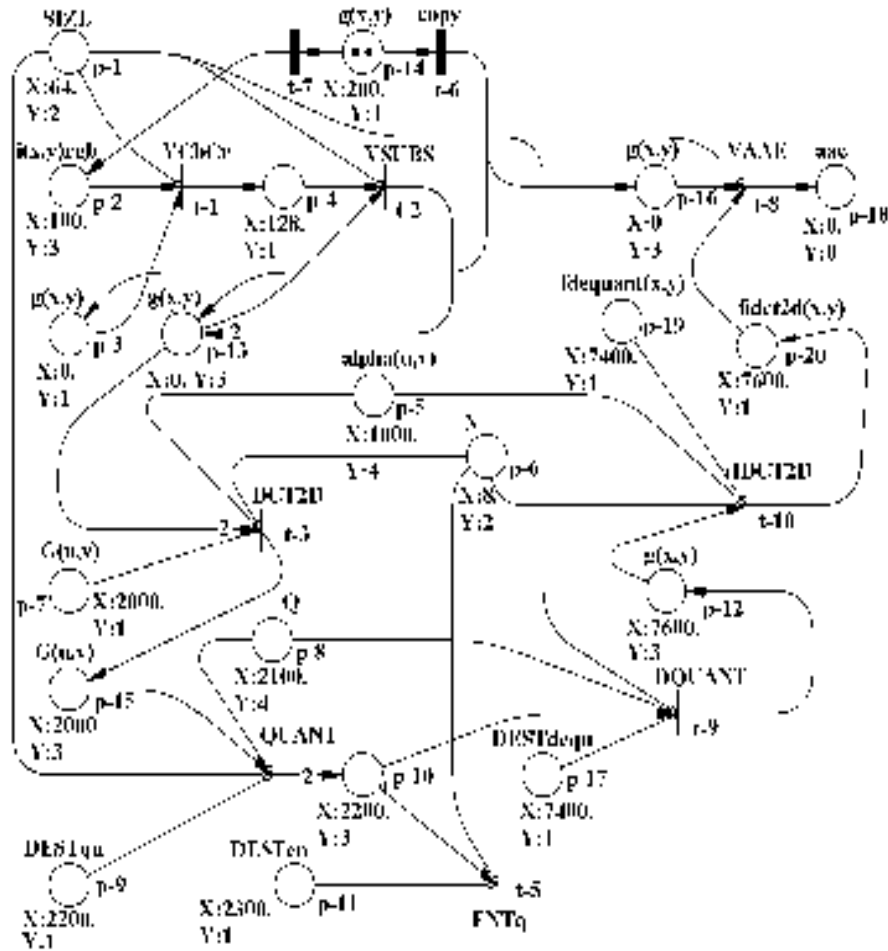


Fig. 1. UPN model of sub-image of JPEG image compression and decompression which generates an absolute average error

In the PeM software a file which is the target of file operation function has a *.mem extension. This file is a text file and it is used for simulation of computer system memory. One line inside the *.mem file represents the content of one memory location. Input data matrix $i(x,y)$ is stored in the *.mem file and it contains RGB (red, green, blue) values of the sub-image.

Places represent state of the system as follows: $p-2$: the input data matrix $i(x,y)$ starts from address 100 in the *.mem file, $p-1$: constant value 64, $p-6$: constant value 8, $p-4$: constant value 128, $p-5$: start address of matrix $\alpha(u,v)$ which includes normalizing function values (Eq. 3.1), $p-8$: start address of matrix Q which is a typical quantization matrix as specified in the JPEG standard.

Initial marking of the UPN model shown in Fig. 1. is $\mu_{(p14)}=2$. This marking enables transitions $t-7$ and $t-6$. After firing of this set of transitions the markings $\mu_{(p2)}$, $\mu_{(p3)}$, $\mu_{(p13)}$ and $\mu_{(p16)}$ all become equal to 1 and the x -attributes of the places $p-2$, $p-3$, $p-13$ and $p-16$ all become equal to 200, because of $TF(t_6)=copy$ function. This x -attribute value refers to the start address of $g(x,y)$ matrix in the *.mem file. This matrix is generated by firing of transitions $t-1$ and $t-2$ respectively. By firing of transition $t-1$ a destination matrix $g(x,y)$ is calculated as conversion of the input matrix $i(x,y)$ which contains RGB values into the YCbCr (luma, blue-difference and red-difference chroma components) values. By firing of transition $t-2$ all elements of the destination matrix $g(x,y)$ are decremented by 128, since $TF(t_2)=VSUBS$ (vector subtract by scalar) and x -attribute of the place $p-4$ is equal to 128. After this firing new markings are $\mu_{(p16)}=1$ and $\mu_{(p12)}=2$, so transition $t-3$ becomes enabled.

Function $DCT2D$ (discrete cosine transform, 2 dimensions) is assigned to this transition. After firing of transition $t-3$ the DCT function is applied to the matrix $G(u,v)$:

$$G(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^7 \sum_{y=0}^7 g(x,y) F(x,u) F(y,v) \quad (3.1)$$

where

- $F(x,u) = \cos \left[\frac{\pi}{8} \left(x + \frac{1}{2} \right) u \right]$
- $F(y,v) = \cos \left[\frac{\pi}{8} \left(y + \frac{1}{2} \right) v \right]$
- u is the horizontal spatial frequency, for the integers $0 \leq u < 8$,
- v is the vertical spatial frequency, for the integers $0 \leq v < 8$,
- $\alpha(n) = \begin{cases} \sqrt{\frac{1}{8}}, n = 0 \\ \frac{1}{2}, otherwise \end{cases}$ is a normalizing function
- $g(x,y)$ is the pixel value at coordinates (x,y)
- $G(u,v)$ is the DCT coefficient at coordinates (u,v)

The state of the UPN model after this firing is $\mu_{(p15)}=\mu_{(p16)}=I$, and transition $t-4$ is enabled. The corresponding transition function $TF(t_4)=QUANT$ (quantization). This function calculates a new matrix which starts at address 2200 in the *.mem file, as determined by the x -attribute of place $p-7$. Elements of this matrix are rounded integer values of $B(j,k)=G(j,k)/Q(j,k)$. After firing of transition $t-4$, set of transitions $\{t-5, t-9\}$ become enabled and the state of the UPN model is determined by $\mu_{(p10)}=\mu_{(p16)}=I$.

Transition function $TF(t_5)$ is $ENTq$ (entropy) function. By firing of this transition, final matrix is generated, starting at address 2300 in the *.mem file, as determined by x -attribute of place $p-11$. After that Run Length Encoding (RLE) can be done over this final matrix.

By firing of transition $t-9$ a $DQUANT$ (dequantization) function will be performed on the matrix which starts at address 2200, as determined by x -attribute of place $p-10$ and the resulting matrix is stored starting at address 7400 because of x -attribute of the place $p-17$. Typical quantization matrix Q , which starts at address 2100 is used again. This means that the place $p-8$ is included as the fourth input of the transition $t-10$ because of the y -attribute of this place. The place $p-8$ has the inhibitor arc to transition $t-9$ and its marking is equal to 0, thus this transition becomes enabled.

After parallel firing of transitions $t-5$ and $t-9$, the state of the UPN model is determined by $\mu_{(p12)}=\mu_{(p16)}=I$ and transition $t-10$ becomes enabled.

By firing of this transition an $IDCT2D$ (inverse-DCT2D) function will be applied to the matrix which starts at address 7400 and matrix which starts at address 7600 will be populated by resulting values. These addresses are determined by x -attributes of the places $p-19$ and $p-12$ respectively. Now, the state of the UPN model is determined by $\mu_{(p16)}=\mu_{(p20)}=I$ and transition $t-8$ becomes enabled. A transition function assigned to this transition is $VAAE$ i.e. vector absolute average error:

$$VAAE = \frac{1}{N} \sum_{i=1}^N |V_1(i) - V_2(i)| \quad (3.2)$$

where: N is vector length, and V_1, V_2 are vectors.

By firing of transition $t-8$ $VAAE$ function will calculate average absolute error. Start addresses of the vectors are 7600 and 200 as determined by x -attributes of places $p-16$ and $p-20$ respectively. Size of vectors is determined by x -attribute of the place $p-1$. And finally, x -attribute of the place $p-18$ is equal to the value of the absolute average error. After firing of transition $t-8$ a dead node is reached i.e. there is no enabled transition in the UPN model.

14.4 The Upgraded Petri Net Model Execution

Initial marking of the UPN model is shown in Fig. 1. By executing the UPN model for given initial marking, according to the UPN rules of transition firing, the next sequence of sets of firing transitions will happen:

$$\{t6, t7\} \rightarrow \{t1\} \rightarrow \{t2\} \rightarrow \{t3\} \rightarrow \{t4\} \rightarrow \{t5, t9\} \rightarrow \{t10\} \rightarrow \{t8\}$$

This sequence changes marking μ from the initial marking until dead node is reached. By execution of this UPN model it passes through many states (markings) as follows:

- $\mu(0,0,0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0) = (\text{initial marking}),$
- $\mu(0,1,1,0,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0,0),$
- $\mu(0,0,0,1,0,0,0,0,0,0,0,0,2,0,0,1,0,0,0,0),$
- $\mu(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,0,0),$
- $\mu(0,0,0,0,0,0,0,0,0,0,2,0,0,0,0,0,1,0,0,0,0),$
- $\mu(0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0),$
- $\mu(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,1),$
- $\mu(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0) = (\text{dead node}).$

The dead node refers to the state of the UPN model where x -attribute of place $p-18$ is equal to average absolute error between original and decompressed image (see Eq. 3.1 and Eq. 3.2).

14.5 The Upgraded Petri Net Model Analysis

The described UPN model was tested on two sub-images, one from a text image and one from a photorealistic image. For both test sub-images execution of the UPN model generated matrix which represents the uncompressed sub-image, absolute difference per pixel and average absolute error between original and uncompressed sub-image. This model consists of parts we want to analyze and of the parts which are not interesting for analysis, according to the chosen parameters. In example model which calculates average absolute error, entropy is not interesting so it is represented at the high level. The functions, however, which calculate the matrices for the JPEG sub-image compression and decompression are represented at the detailed level. At the same time the model consists of elaborated parts essential for the analysis at a detailed level and also of some general parts whose details are irrelevant for the analysis at the given level of representation. The results of the simulation give a good insight

into the JPEG sub-image compression and decompression algorithm. These results are shown in Fig. 2.

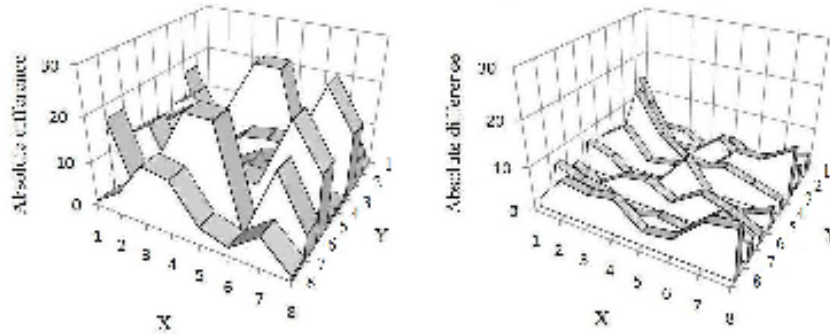


Fig. 2. Absolute difference between original and decompressed image and absolute average error for text image (left, $VAAE=9.0313$ [values/pixels]) and for photorealistic image (right, $VAAE=4.9063$ [values/pixels])

14.6 Conclusions

Upgraded Petri nets are formal modeling tool appropriate for simulation and analysis of processes, particularly at the register transfer level. An Upgraded Petri net low level model is developed which refers to the detailed representation of JPEG method for image compression and decompression on an 8×8 8-bit sub-image. This Upgraded Petri net model is represented at the register level so it can be used for analyzing hardware implementation of the JPEG sub-image compression.

The model can be adjusted for many different purposes, but as an example, it generates the average absolute error between original and decompressed image. It can be easily modified to generate data which can be used for analysis of some other parameters. Further, from this model a complex model can be developed at the register level, which will calculate JPEG compression and decompression over the large image.

Suitability of developed UPN model was checked by execution of the model and the results were generated dynamically during this execution. Original software for modeling and simulations of Upgraded Petri Nets, PeM (Petri Net Manager), is developed and used for modeling described in this paper. The PeM has the following features: supports the UPN theory, graphical modeling interface, interactive simulation by single or parallel transition firing and reachability tree generation. This allows that the whole model or only some parts are examined. Also, the features enable

interactive monitoring of the execution of the model and gradual improvement from the initial phase all the way to the final version.

Acknowledgement. This research was supported by the Ministry for Science and Technical Development of the Republic of Serbia, Project 144007.

References

1. Strbac P, Tuba T, Simian D (2009) Hierarchical model of a systolic array for solving differential equations implemented as an Upgraded Petri net, *WSEAS Transactions on Systems*, Volume 8, Issue 1, pp 12-21.
2. Salomon D (2002) *The Springer Guide to Data Compression Methods*, Springer-Verlag New York, LLC
3. Skodras A, Christopoulos C, Ebrahimi T (2001) The JPEG 2000 still image compression standard, *Signal Processing Magazine, IEEE* Volume 18, Issue 5, pp 36-58
4. Thao NT, Vetterli M (1998) Set theoretic compression with an application to image coding, *IEEE Transactions on Image Processing*, Volume 7, number 7, pp 1051-1056
5. Tuba M, Samardzic A, Tosic D (2002) JPEG Algorithm Parameter Adjustment, *Bulletins for Applied Mathematics (BAM)*, CI Budapest, pp 303-312
6. Tuba M, Lazic J (2004) JPEG and JPEG2000 Compression Algorithms Applied to Low-Bandwidth Channels, *Bulletins for Applied Mathematics (BAM)*, CVI, Budapest, pp 1-10
7. Aboalsamh H, Mathkour H, Mursi M, Assassa G (2008) Steganalysis of JPEG Images: An Improved Approach for Breaking the F5 Algorithm, *Proceedings of the 12th WSEAS International Conference on Computers*, Heraklion, Greece, July 23-25, 2008, pp. 1011-1018.
8. Carpentieri B (2009) Image Compression via Textual Substitution, *WSEAS Transactions on Information Science and Applications*, Issue 5, Volume 6, pp 768-777
9. Popa C, Gordan M, Vlaicu A, Orza B, Oltean G (2008) Computationally efficient algorithm for fuzzy rule-based enhancement on JPEG compressed color images, *WSEAS Transactions on Signal Processing*, Issue 5, Volume 4, pp. 310-319.
10. Strang G (1999) The Discrete Cosine Transform, *SIAM Review*, Volume 41, Number 1, pp. 135-147
11. Jahne B (2005) *Digital Image Processing*, 6th edition, Springer-Verlag, Berlin
12. Kuo-Ming H (2008) A Novel Robust Watermarking Technique Using IntDCT Based AC Prediction, *WSEAS Transactions on Computers*, Issue 1, Volume 7.

Chapter 15

BIOSARP – Bio-Inspired Self-Optimized Routing Algorithm using Ant Colony Optimization for Wireless Sensor Network – Experimental Performance Evaluation

Kashif Saleem¹, Norsheila Fisal,¹ M. Ariff Baharudin,¹ Adel Ali Ahmed,¹ Sharifah Hafizah,¹ Sharifah Kamilah¹

¹Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia, kashnet@hotmail.com

Abstract. Nowadays, wireless sensor networks (WSNs) are becoming increasingly beneficial, worthwhile and a challenging research area. The advancements in WSN enable a wide range of environmental monitoring and object tracking applications. Moreover, multihop (node by node) routing in WSN is affected by new devices constantly entering or leaving the network. Therefore, nature inspired self-maintained protocols are required to tackle the problems arising in WSN. We proposed ant colony stimulated routing, which shows an outstanding performance for WSNs. In this manuscript, a self-optimized (ACO) routing protocol for WSN and the results are presented. Link quality, energy level and velocity parameters are used to discover an optimal route. The signal strength, remaining power and timestamp metrics are trade in from physical layer to network layer. The emitted decision through the WSN discovery will establish the optimal route from source to destination. The adopted architecture helps ACO in improving the overall data throughput; especially in the case of real time traffic.

Keywords. Ant Colony Optimization, Energy, Multihop, Packet Reception Rate, Routing, Velocity, Wireless sensor Network.

15.1 Introduction

Wireless communication plays a significant role in the telecommunications sector and has huge importance for future research. Wireless communications enables many new applications for sensing and monitoring systems. Some infrastructure free networks like WSN serves an imperative task in monitoring. With the passage of time new gadgets and software advancements are becoming available to end-users on a frequent basis. The stated fast growth and the huge number of devices in the network make WSN more and more complex. The deployment area for WSNs is mostly out of the human reach. The above mentioned challenges, such as growing complexity and unreachable maintenance need new solutions.

The new self-maintained mechanism can maintain the features of WSNs such as multihop routing and dynamically environmental changes in a completely autonomous mode. In order to address autonomous capability for multihop WSNs, it has been visualized that self-maintained network applications can understand the operational objectives of the network. Additionally, probabilistic methods that provide scalability can be found in nature and adapted to technology.

Towards this vision, it is observed that various biological principles are capable to overcoming the above adaptability problems. The area of bio-inspired network engineering has the most well known approaches which are swarm intelligence (ANT Colony, Particle swarm), AIS and intercellular information exchange (Molecular biology)[1-4]. WSN routing algorithms based on ANT Colony Optimization (ACO) have been presented in the last few years, such as [5], Sensor-driven Cost-aware Ant Routing (SC), the Flooded Forward Ant Routing (FF) algorithm, and the Flooded Piggybacked Ant Routing (FP) algorithm [6], Adaptive ant-based Dynamic Routing (ADR) [7], Adaptive Routing (AR) and Improved Adaptive Routing (IAR) algorithm [8], and E&D ANTS [9].

The problem of the previous approaches is that the selected shortest path might not be a minimum energy cost route. Some other works concentrate on decreasing the energy consumption by replacing the hop-count routing with minimum energy routing. They compute a minimum-energy path for packet delivery in a multi-hop wireless network. However, the nodes on this path will get depleted soon. Radhika D.Joshi [10] given an idea about combination of least hops and minimum remaining energy.

This manuscript present a novel architecture by implementing the most well known and successful approaches. ACO method is utilized for the optimum route discovery in multihop WSN. Standard ACO is very com-

plex and heavy for WSN. Consequently, we come up with an ACO that can perform better optimization for WSN in terms of less load, energy consumption and high delivery rate. Generally, the ACO algorithm is agent based [11] as forward ant (FA), backward ant (BA), search ant (SA) and data ant (DA) agents. Under BIOSARP, the ACO is based on only two types of ant agents, which are, search ant (SA) and data ant (DA) agents as shown in Figure 1.1. The agents will work in a decentralized way to collect data on individual nodes and carry data to the required destination through multihop communication.

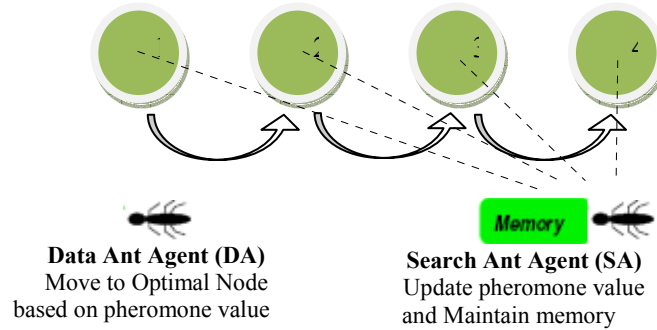


Fig. 1.1. BIOSARP mechanism based on Data Ant (DA) and Search Ant (SA) agents.

The signal strength, remaining power and timestamp parameters have been taken from the physical layer to the network. By assigning the above mentioned metrics to the ACO process running on the network layer allows an ultra effective optimal route for WSNs.

15.2 Related Research

15.2.1 Outline of Ant Colony Optimization

Dorigo et al [5] proposed the first ant colony algorithms as a multi-agent approach to difficult combinatorial optimization problems like the traveling salesman problem (TSP) and the quadratic assignment problem (QAP), Minimum Weight Vertex Covering Problem [12, 13], and later introduced the ACO meta-heuristic by Dorigo et al [5].

There are two types of ants applied in the algorithms, forward ants and backward ants. Forward ants, whose main actions are exploring the path

and collecting the information from the source nodes to destination node, have the same number as the source nodes. The paths that forward ants travel will construct a tree when they merge into each other or reach the destination and data is transmitted along the tree paths. There are two key factors that conduct the movement of the forward ants: one is pheromone trails that are deposited along the edges, and the other is the nodes potential. Whereas the backward ants, travelling back from destination node to source nodes contrary to the forward ants, perform their uppermost function of updating the information of their pass-by nodes.

15.2.2 Overview of ACO Based Routing Algorithms in WSN

Aghaei et al [8] proposes two adaptive routing algorithms based on ant colony algorithm, the Adaptive Routing (AR) algorithm and the Improved Adaptive Routing (IAR) algorithm. To check the suitability of the ADR algorithm in the case of sensor networks, they modified the ADR algorithm (removing the queue parameters) and used their reinforcement learning concept and named it the AR algorithm. The AR algorithm did not result in optimum solutions. In IAR algorithm by adding a coefficient, the cost between the neighbor node and the destination node, they further improve the AR algorithm.

Wen et al [9] proposed a dynamic adaptive ant algorithm (E&D ANTS) is based on Energy and Delay metrics for routing operations. Their main goal is to maximize the network lifetime while minimizing propagation delay by using a novel variation of reinforcement learning (RL). E&D ANTS results was evaluated with AntNet and AntChain schemes.

In [14] a novel routing approach using an Ant Colony Optimization algorithm is proposed for Wireless Sensor Networks consisting of stable nodes. The probabilistic decision rule depends on pheromone value and the value of the heuristic related to energy. They have also implemented their approach to a small sized hardware component named MICAz mote as a router chip.

15.2.3 Comparison between Ant Based Routing Mechanisms

In our proposed algorithm, the best values of velocity, PRR and remaining power mechanism [15] are used to select forwarding node because velocity alone does not provide the information about link quality. The best link quality usually provides low packet loss and energy efficient

[16]. Another novel feature is the remaining power parameter to select the forwarding candidate node.

15.3 Methodology

System design deals mainly with the development of state machine and flow chart diagram of the sections as power and neighbor management. Routing management will be dependent mostly on forwarding metric calculations. If any error occurred in this state, the generated error will be handled by the routing problem handler.

Further onwards the most important state in this routing mechanism is neighbor management. Selection of better neighbors will be handled by this state. Common functions in the neighbor management state are neighbor table maintenance, neighbor discovery, insert new neighbor, neighbor replacement, etc.

Our proposed self-maintained system is mainly based on route section. The optimal route discovery is tackled by ACO. Routing decision will be achieved by the probabilistic decision rule described in [17]. Two parameters delay and battery remaining are used in [17] while acquiring optimal decision.

BIOSARP is based on three metrics as, velocity, PRR and remaining power mechanism as given in Table 1.1. The link quality of the wireless medium determines the performance of WSN. In designing BIOSARP, the link quality is considered in order to improve the delivery ratio and energy efficiency.

We have added second heuristic value ω_{ij} in probabilistic rule to determine the link quality of neighboring nodes while making decision. Jovanovic Raka discuss about heuristic previously in [18]. The probabilistic rule is expressed mathematically as Equation (see Eq. 1.1).

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta \cdot [\omega_{ij}(t)]^\vartheta}{\sum_{h \in j_i^k} [\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta \cdot [\omega_{ij}(t)]^\vartheta} \quad (1.1)$$

$p_{ij}^k(t)$ overall desirability for ant k located in city i to choose to move to city j .

τ_{ij} depends on the delay parameter.

η_{ij} is an heuristic evaluation of edge (i,j) .

ω_{ij} is the 2nd heuristic evaluation of edge (i,j) .

α , β and \square are three parameters that control the relative weight of pheromone trail and heuristic values.

Table 1.1. Routing Metrics

	Velocity (End2End Delay)	Energy	Link Quality (PRR)
Node 1	τ^1	η^1	ω^1
Node 2	τ^2	η^2	ω^2
.	.	.	.
.	.	.	.
.	.	.	.
Node n	τ^n	η^n	ω^n

While data forwarding, the node first calls the DA. DA will select the optimal node based on the pheromone value stored in neighbour table. DA will move hop by hop on the base of pheromone values for neighbouring nodes until the destination as shown in Figure 1.1. At the time of selection, if DA could not find the entry, it will invoke SA as given in Figure 1.1. The SA will search for new nodes and calculates their pheromone value through the probabilistic rule as Equation (see Eq. 1.1).

Every ant agent is generated with a sequence ID and also is set with maximum time to live (TTL). By the help of assigning sequence ID, sensor node will not accept reappearance of the same ant agent to avoid routing cycles. Energy management is evolved to maintain the energy consumption of every sensor node in WSN.

15.4 Real Time Experiment

BIOSARP is verified by deploying a real time test bed of TelosB [19] sensor nodes. 10 wireless sensor nodes were deployed onto 15m x 10 m grid as shown in Figure 1.2. For the bio-inspired routing algorithm implementation under tinyos [20], the program is written in NesC programming language [21]. During experimentation, by animation we can examine the output of network. The traffic is produced from all the nodes to the sink. Node 0 is connected to the laptop to collect the data as shown in Figure 1.3. All nodes have neighboring tables which contains the information about the neighboring nodes. Depending on information, pheromone value is calculated via probabilistic rule. The parameter weights α , β and \square are adjusted to 0.6. The optimal node is selected based on the pheromone value. The experimental parameters used to configure the sys-

tem according to WSN are listed in Table 1.2. Table 1.3 shows the code size of BIOSARP and RTLD [15] in TELOSB. The code size of BIOSARP routing protocol in the TELOSB is 31498 bytes of flash memory and 1247 bytes of RAM.

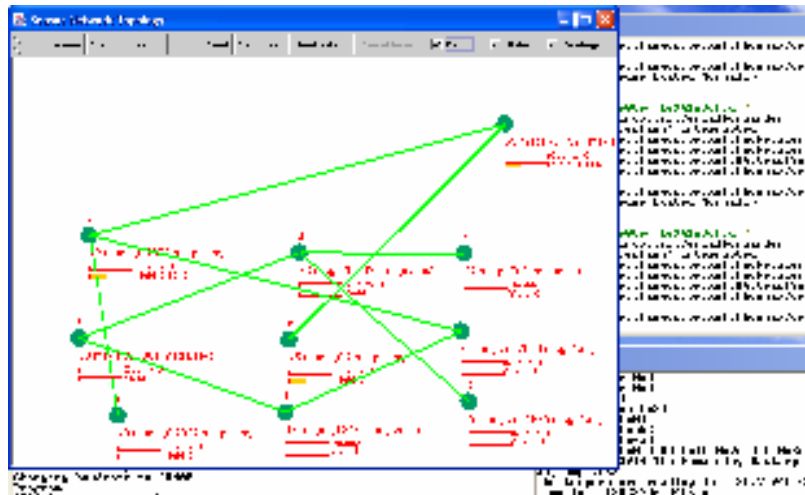


Fig. 1.2. Network Testbed.



Fig. 1.3. Sink Node.

Table 1.2. System Properties

Parameters	Values
Power Level	Selected 15 (lowest 3 – maximum 23)
Traffic	0.2, 0.25, 0.33 & 0.5 Packet rate per second

Table 1.3. Coding size of BIOSARP and RTLD

Comparison	TELOSB	
	Flash memory	RAM
BIOSARP	31496	1247
RTLD	32212	1207

1.4.1 Performance and Analysis of Experiment

BIOSARP is compared with RTLD routing protocol [22] because RTLD also makes the next hop decision based on same parameters. Also RTLD got the best performance results till yet over WSN. Packet delivery ratio is the metric used to analyze the performance of BIOSARP and the baseline RTLD. In experiment the impact of varying network load is accumulated by varying the packet rate. While the end-to-end deadline is fixed at 250ms. The traffic load is varied from 0.2 to 0.5 packet per sec to emulate low data rate in IEEE 802.15.4. The results in Figure 1.4 show higher delivery ratio about 20% to 40% compared to the RTLD.

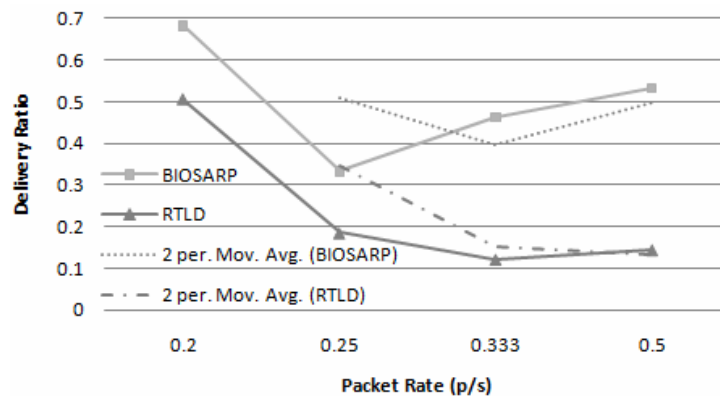


Fig. 1.4. Delivery ratio of BIOSARP and RTLD.

15.5 Conclusions

In this manuscript, we have proposed an enhanced ant colony inspired self-optimized routing protocol for WSN. Our specified mechanism is based on link quality, energy and velocity parameters. The adopted

architecture helps WSN in improving the overall data throughput; especially in the case of real time traffic. The design also assists WSN in better delivery ratio. The algorithm is also capable of avoiding permanent loops which promotes dead lock in the running networks. Experimental results demonstrate the protocol efficiency. Finally, this autonomic routing mechanism will come up with better delivery ratio over WSN. Our immediate future work evolved to enhance our routing mechanism with self-protection system.

Acknowledgement. The authors wish to express their sincere appreciation, sincerest gratitude to Ministry of Higher Education Malaysia for their full support and Research Management Center (RMC), Universiti Teknologi Malaysia (UTM) for their contribution. Special thanks to researchers in Telematic Research Group (TRG), UTM.

References

1. Balasubramaniam, S., Botvich, D., Donnelly, W., Foghluh, M., Strassner, J.: Biologically Inspired Self-Governance and Self-Organisation for Autonomic Networks. Proceedings of the 1st international conference on Bio inspired models of network, information and computing systems, Vol. 275. ACM, Cavalese, Italy (2006) 30
2. Balasubramaniam, S., Donnelly, W., Botvich, D., Agoulmine, N., Strassner, J.: Towards integrating principles of Molecular Biology for Autonomic Network Management. Hewlett Packard university Association (HPOVUA) conference, Nice, France. (2006)
3. Boonma, P., Suzuki, J.: MONSOON: A Coevolutionary Multiobjective Adaptation Framework for Dynamic Wireless Sensor Networks. In Proc. of the 41st Hawaii International Conference on System Sciences (HICSS), Big Island, HI (2008)
4. Mazhar, N., Farooq, M.: BeeAIS: Artificial Immune System Security for Nature Inspired, MANET Routing Protocol, BeeAdHoc. Springer-Verlag Berlin Heidelberg **LNCS 4628** (2007) 370–381
5. Chen, G., Guo, T.-D., Yang, W.-G., Zhao, T.: An improved ant-based routing protocol in Wireless Sensor Networks. Collaborative Computing: International Conference on Networking, Applications and Worksharing, 2006. CollaborateCom 2006. IEEE, New York, NY (2006) 1-7
6. Zhang, Y., Kuhn, L.D., Fromherz, M.P.J.: Improvements on Ant Routing for Sensor Networks. M. Dorigo et al. (Eds.): ANTS 2004, Springer-Verlag Berlin Heidelberg 2004 **LNCS 3172** (2004) 154-165

7. Lu, Y., Zhao, G., Su, F.: Adaptive Ant-based Dynamic Routing Algorithm. In Proceedings of the 5th World Congress on Intelligent Control and Automation. IEEE, Hangzhou, China (2004) 2694-2697
8. Aghaei, R.G., Rahman, M.A., Gueaieb, W., Saddik, A.E.: Ant Colony-Based Reinforcement Learning Algorithm for Routing in Wireless Sensor Networks. Instrumentation and Measurement Technology Conference - IMTC IEEE, Warsaw, Poland (2007)
9. WEN, Y.-f., CHEN, Y.-q., PAN, M.: Adaptive ant-based routing in wireless sensor networks using Energy*Delay metrics. Journal of Zhejiang University SCIENCE A **9** (2008) 531-538
10. D.JOSHI, R., P.REGE, P.: Energy Aware Routing in Ad Hoc Networks. In: (WSEAS), T.W.S.a.E.A.a.S. (ed.): 6th WSEAS International Conference on CIRCUITS, SYSTEMS, ELECTRONICS, CONTROL & SIGNAL PROCESSING, Cairo, Egypt (2007) 469-475
11. Kakkasageri, M.S., Manvi, S.S., Soragavi, G.D.: Mobile agent based event discovery in wireless sensor networks. In: (WSEAS), T.W.S.a.E.A.a.S. (ed.): Proceedings of the 5th WSEAS International Conference on Applied Computer Science, Hangzhou, China (2006) 731-735
12. Milan, T., Raka, J.: An analysis of different variations of ant colony optimization to the minimum weight vertex cover problem. WSEAS Trans. Info. Sci. and App. **6** (2009) 936-945
13. Jovanovic, R., Tuba, M., Simian, D.: Comparison of Different Topologies for Island-Based Multi-Colony Ant Algorithms for the Minimum Weight Vertex Cover Problem. WSEAS TRANSACTIONS on COMPUTERS **9** (2010)
14. Okdem, S., Karaboga, D.: Routing in Wireless Sensor Networks Using an Ant Colony Optimization (ACO) Router Chip. Sensors **9** (2009) 909-921
15. Ali, A., Latiff, L.A., Sarijari, M.A., Fisal, N.: Real-time Routing in Wireless Sensor Networks. The 28th International Conference on Distributed Computing Systems Workshops. IEEE, Beijing, China (2008)
16. Zhao, J., Govindan, R.: Understanding Packet Delivery Performance in Dense Wireless Sensor Networks. Proceedings of the 1st international conference on Embedded networked sensor systems, Los Angeles, USA (2003)
17. Okdem, S., Karaboga, D.: Routing in Wireless Sensor Networks Using Ant Colony Optimization. In: Okdem, I. (ed.): Proceedings of the First NASA/ESA Conference on Adaptive Hardware and Systems (AHS'06), Vol. 0-7695-2614-4/06, Istanbul (2006)
18. Raka, J., Milan, T., Dana, S.: Ant colony optimization applied to minimum weight dominating set problem. Proceedings of the 12th WSEAS international conference on Automatic control, modelling & simulation. World Scientific and Engineering Academy and Society (WSEAS), Catania, Italy (2010) 322-326
19. Chipcon: CC2420 low power radio transceiver, <http://www.chipcon.com>. (2010)
20. UC-Berkeley: TinyOS Tutorial, http://docs.tinyos.net/index.php/TinyOS_Tutorials. Vol. 16/08/2010 (2009)

21. Gay, D., Levis, P., Behren, R.v., Welsh, M., Brewer, E., Culler, D.: The nesc language: A holistic approach to networked embedded systems. In: 2003, A.S. (ed.): Conference on Programming Language design and implementation (2003) 1-1 1
22. Ali, A., Latiff, L.A., Fisal, N.: Simulation-based real-time routing protocol with load distribution in wireless sensor networks. Wireless Communications and Mobile Computing **9999** (2009)

Chapter 16

Characterization of all Optimal Flows in Networks

L. Ciupala,¹ E. Ciurea²

¹Department of Computer Science, University Transilvania of Braşov,
Iuliu Maniu Street 50, Braşov, laura_ciupala@yahoo.com

²Department of Computer Science, University Transilvania of Braşov,
Iuliu Maniu Street 50, Braşov, e.ciurea@unitbv.ro

Abstract. In this paper, we describe methods for determining all optimal flows in a given network: all minimum flows and all maximum flows. The minimum flow algorithms developed until now determine only a minimum flow, but, usually, there is more than one minimum flow in a network. There are several applications that can be modeled as minimum flow problems in some networks and solved by determining all the minimum flows in those networks. This is the main reason for which we focus on the problem of determining all minimum flows in networks.

The same problem arises when solving maximum flow problems. Starting with the well known labeling algorithm by Ford and Fulkerson, many maximum flow algorithms were developed in the past six decades, but they determine only a maximum flow, although, most of the times, there are several maximum flows.

Keywords. Network flow, network algorithms, maximum flow, minimum flow, minimum cut, maximum cut.

16.1 Introduction

Network flow problems are a group of network optimization problems with widespread and diverse applications. The literature on network flow problems is extensive ([1, 2, 6, 9, 10]). Over the past 60 years researchers

have made continuous improvements to algorithms for solving several classes of problems. From the late 1940s through the 1950s, researchers designed many of the fundamental algorithms for network flow, including methods for maximum flow and minimum cost flow problems. In the next decades, there are many research contributions concerning improving the computational complexity of network flow algorithms by using enhanced data structures, techniques of scaling the problem data etc. In the beginning of this century, the minimum flow problem starts to be studied.

One of the reasons for which the network flow problems were studied so intensively is the fact that they arise in a wide variety of situations and in several forms. But, the network flow algorithms developed until now determine only an optimal solution of the problem, although, most of the time, there are several optimal solutions.

We focus on determining all the optimal solutions of a network flow problem because there are several practical problems that can be modelled and solved by establishing all the maximum flows or all the minimum flows in a network. The problem of determining all the optimal flows in a network also arises as a subproblem when solving more complex network flows problems, for instance when solving inverse flow problems.

16.2 Maximum flow problem

The maximum flow problem is one of the fundamental problems in network flow theory and it was studied extensively. The importance of the maximum flow problem is due to the fact that it arises in a wide variety of situations and in several forms. Sometimes the maximum flow problem occurs as a subproblem in the solution of more difficult network problems, such as the minimum cost flow problem or the generalized flow problem. The maximum flow problem also arises in a number of combinatorial applications that on the surface might not appear to be maximum flow problems at all. The problem also arises directly in problems as far reaching as machine scheduling, the assignment of program modules to computer processors, the rounding of census data in order to retain the confidentiality of individual households, tanker scheduling and several others.

The maximum flow problem was first formulated and solved using the well known augmenting path algorithm by Ford and Fulkerson in 1956. Since then, two types of maximum flow algorithms have been developed: augmenting path algorithms and preflow algorithms:

1. The augmenting path algorithms maintain mass balance constraints at every node of the network other than the source node and the sink node. These algorithms incrementally augment flow along paths from the source node to the sink node. By determining the augmenting paths with respect to different selection rules, different algorithms were developed.
2. The preflow algorithms flood the network so that some nodes have excesses. These algorithms incrementally relieve flow from nodes with excesses by sending flow from the node forward toward the sink node or backward toward the source node. By imposing different rules for selecting nodes with excesses, different preflow algorithms were obtained. These algorithms are more versatile and more efficient than the augmenting path algorithms.

All these algorithms determine only one solution of the maximum flow problem, although, generally, the problem has several optimal solutions.

In [11] G. Ruhe characterized all the maximum flows in a given network. For many applications it is important to determine not only a maximum flow, but all the maximum flows in a given network. Also it is useful to know all the optimal solution of a network flow problem, if some of the original values of the network are changed or if additional constraints are considered. Besides, the problem of determining all the maximum flows in a given network arises as a subproblem when solving inverse network flow problems.

The method described by Ruhe is a decomposition method for determining of all optimal solutions of a maximum flow problem. Starting from a nontrivial subset of the set of all minimum cuts, this algorithm determines a decomposition of the given graph $G = (N, A)$ into several subgraphs $G_k = (N_k, A_k)$. If all these subgraphs are trees, then the maximum flow problem has a unique solution, otherwise there are several maximum flows. All these maximum flows are described in [11] by means of feasible flows in the subgraphs $G_k = (N_k, A_k)$ determined by the decomposition algorithm.

16.3 Minimum flow problem

We consider a capacitated network $G = (N, A, l, c, s, t)$ with a nonnegative capacity $c(i, j)$ and with a nonnegative lower bound $l(i, j)$ associated with each arc $(i, j) \in A$. We distinguish two special nodes in the network G : a source node s and a sink node t .

Let $n=|N|$, $m=|A|$ and $C = \max \{ c(i, j) \mid (i, j) \in A \}$.

A *flow* is a function $f: A \rightarrow \mathbf{R}_+$ satisfying the next conditions:

$$f(i, N) - f(N, i) = \begin{cases} v, & \text{if } i = s \\ 0, & \text{if } i \neq s, t \\ -v, & \text{if } i = t \end{cases} \quad (1.1)$$

$$l(i, j) \leq f(i, j) \leq c(i, j), (i, j) \in A \quad (1.2)$$

for some $v \geq 0$, where

$$f(i, N) = \sum_j f(i, j), i \in N$$

and

$$f(N, i) = \sum_j f(j, i), i \in N.$$

We refer to v as the *value* of the flow f . The minimum flow problem is to determine a flow f for which v is minimized.

Although it has its own applications, the minimum flow problem was not dealt so often as the maximum flow and the minimum cost flow problem. There are many problems that occur in economy that can be reduced to minimum flow problems.

The minimum flow problem in a network can be solved in two phases:

1. establishing a feasible flow, if there is one
2. from a given feasible flow, establish the minimum flow.

The problem of determining a feasible flow can be reduced to a maximum flow problem (for details see [1]). For the second phase of the minimum flow problem there are three approaches:

1. using decreasing path algorithms (see [6])
2. using preflow algorithms (see [3, 6])
3. finding a maximum flow from the sink node to the source node in the residual network (see [2, 4]).

The decreasing path algorithms maintain during their executions the mass balances constraints (1.1) at every node of the network other than the source or the sink node. These algorithms identify decreasing path and decrease flows on these paths until the network contains no such path, which means that the flow is a minimum flow. By establishing different rules for determining the decreasing paths, one obtains different decreasing path algorithms for minimum flow (see [6]).

The preflow algorithms pull flow along individual arcs. These algorithms do not satisfy the mass balances constraints (1.1) at intermediate stages. In fact, these algorithms permit the flow leaving a node to be fewer

than the flow entering the node. Any preflow algorithm for the minimum flow problem proceeds by pulling flow from the neighbor nodes of the sink node, creating deficits in these nodes. The basic step in any preflow algorithm is to select a node with deficit and to try to eliminate its deficit by pulling flow to its neighbors which are closer to the source node. Any preflow algorithm ends when all the intermediate nodes have no deficits, which means that a minimum flow was obtained. By establishing different rules for selecting the nodes with deficit, one obtains different preflow algorithms for minimum flow (see [3, 6]).

The minimax algorithm determines a minimum flow from the source node to the sink node by establishing a maximum flow from the sink node to the source node in the residual network (see [2, 6]).

All these algorithms determine only one solution of the minimum flow problem, although, generally, the problem has several optimal solutions.

16.3.1 Notation and definitions

Let f be a flow from the source node s to the sink node t in the network $G = (N, A, l, c, s, t)$.

For the minimum flow problem, the *residual capacity* $r(i, j)$ of any arc $(i, j) \in A$, with respect to a given flow f , is given by

$$r(i, j) = c(j, i) - f(j, i) + f(i, j) - l(i, j).$$

By convention, if $(j, i) \notin A$ then we add arc (j, i) to the set of arcs A and we set $l(j, i) = 0$ and $c(j, i) = 0$. The residual capacity $r(i, j)$ of the arc (i, j) represents the maximum amount of flow from the node i to node j that can be canceled. The network $G_f = (N, A_f)$ consisting only of the arcs with positive residual capacity is referred to as the *residual network* (with respect to flow f).

A *cut* is a partition of the node set N into two subsets S and $\bar{S} = N - S$; we represent this cut using the notation $[S, \bar{S}]$. We refer to a cut $[S, \bar{S}]$ as a *s-t cut* if $s \in S$ and $t \in \bar{S}$. We refer to an arc (i, j) with $i \in S$ and $j \in \bar{S}$ as a *forward arc* of the cut, and to an arc (i, j) with $i \in \bar{S}$ and $j \in S$ as a *backward arc* of the cut. Let (S, \bar{S}) denote the set of forward arcs and let (\bar{S}, S) denote the set of backward arcs of the s-t cut $[S, \bar{S}]$.

For the minimum flow problem, we define the *capacity* $c[S, \bar{S}]$ of an s-t cut $[S, \bar{S}]$ as the difference between the sum of the lower bounds of the forward arcs and the sum of the capacities of the backward arcs. That is,

$$c[S, \bar{S}] = l(S, \bar{S}) - c(\bar{S}, S).$$

We refer to an s - t cut whose capacity is maximum among all s - t cuts as a *maximum cut*.

The *residual capacity* of an s - t cut $[S, \bar{S}]$ is defined as

$$r[S, \bar{S}] = f(S, \bar{S}) - l(S, \bar{S}) + c(\bar{S}, S) - f(\bar{S}, S).$$

Theorem 1. (Min-Flow Max-Cut Theorem) [6]. If there is a feasible flow in the capacitated network with nonnegative lower bounds $G = (N, A, l, c, s, t)$, the value of the minimum flow from the source node s to the sink node t is equal to the capacity of the maximum s - t cut.

A forward consequence of the Min-Flow Max-Cut Theorem is the following:

Theorem 2. Let $[S, \bar{S}]$ be a maximum cut in the network $G = (N, A, l, c, s, t)$. Then for any minimum flow f hold the following relations:

$$f(i, j) = \begin{cases} l(i, j), & \text{if } (i, j) \in (S, \bar{S}) \\ c(i, j), & \text{if } (i, j) \in (\bar{S}, S) \end{cases} \quad (1.3)$$

Theorem 3. If f is a flow of value v in the network $G = (N, A, l, c, s, t)$, if $[S, \bar{S}]$ is a s - t cut and if f' is a flow of value v' , with $v' \leq v$ then $v - v' \leq r[S, \bar{S}]$.

Proof. From Theorem 1 it follows that

$$v' \geq c[S, \bar{S}] = l(S, \bar{S}) - c(\bar{S}, S).$$

But $v = f(S, \bar{S}) - f(\bar{S}, S)$. Consequently,

$$v - v' \leq f(S, \bar{S}) - f(\bar{S}, S) - l(S, \bar{S}) + c(\bar{S}, S) = r[S, \bar{S}].$$

16.3.2 Determining all minimum flows

Generally, the minimum flow problem in a network has more than one optimal solution. For many applications, it is important to determine all the minimum flows in a given network. There are also several optimization problems that can be solved if all the minimum flows are determined. But, the minimum flow algorithms developed until now determine only a minimum flow. We will describe a decomposition method for determining of all optimal solutions of a minimum flow problem.

Let K the set of all maximum cuts in the network $G = (N, A, l, c, s, t)$.

The method of determining all the minimum flows in a network is a decomposition method based on Theorem 2. We can apply this method only if a nontrivial subset K_1 of K is known. This restriction implies no loss of generality because we can identify two maximum cuts in the following manner:

1. We can determine a maximum cut and a minimum flow by applying a modified version of Ford-Fulkerson algorithm for the minimum flow problem. A description of this algorithm can be found in [6].
2. For determining a second cut, the minmax algorithm, described in [4], can be used. The main idea of this approach is to reverse the arcs in G and to compute a maximum flow from the sink node to the source node.

If these two cuts are the same, then there is only a minimum flow in the network; otherwise we can determine all the minimum flows through the use of the following algorithm:

Decomposition algorithm;

begin

let $N_1 = \bigcap_{[S, \bar{S}] \in K_1} S$;

$K_2 = \{N_1\}$;

$p = 2$;

while there is $S \in K_1 \setminus K_2$ such that $\exists S' \in K_2$ and $S' \subset S$ **do**

begin

let S be the minimal set such that

$S \in K_1 \setminus K_2$ such that $\exists S' \in K_2$ and $S' \subset S$;

let S' be the maximal set such that $S' \in K_2$ and $S' \subset S$;

$N_p = S \setminus S'$;

$K_2 = K_2 \cup \{N_p\}$;

$p = p + 1$;

end;

$N_p = N \setminus \bigcup_{N_i \in K_2} N_i$;

$K_2 = K_2 \cup \{N_p\}$;

end.

Let $K_2 = \{N_1, \dots, N_p\}$ be the set of node sets determined by the decomposition algorithm.

Let $G_k = (N_k, A_k)$ be the subgraph of G induced by the node set N_k , for each $N_k \in K_2$.

Property 4 The graphs $G_k = (N_k, A_k)$, $N_k \in K_2$ satisfy the following properties:

$$N_k \cap N_l = \emptyset, \text{ for all } k, l \in \{1, \dots, p\}, k \neq l$$

$$\bigcup_{N_k \in K_2} N_k = N$$

$$A_k \cap A_l = \emptyset, \text{ for all } k, l \in \{1, \dots, p\}, k \neq l$$

$$\text{If } A' = A \setminus \bigcup_{N_k \in K_2} A_k \text{ then } A' \subset A \setminus \bigcup_{[S, \bar{S}] \in K} (S, \bar{S}).$$

Let f be a flow in the network $G = (N, A, l, c, s, t)$. We associate to f a set of flows $F(f)$ that contains all the flows g_k satisfying the flowing constraints:

$$g_k(i, N_k) - g_k(N_k, i) = 0, i \in N_k \quad (1.4)$$

$$l(i, j) - f(i, j) \leq g_k(i, j) \leq c(i, j) - f(i, j), (i, j) \in A_k \quad (1.5)$$

for all $N_k \in K_2$.

Theorem 5. Let f^* be a minimum flow in the network $G = (N, A, l, c, s, t)$. Then all the minimum flows in G have the form

$$f(i, j) = \begin{cases} f^*(i, j) + g_k(i, j), & \text{if } (i, j) \in A_k, g_k \in F(f) \\ f^*(i, j), & \text{if } (i, j) \in A' \end{cases}$$

Proof. This theorem can be proved in the same manner as the Theorem 2 in [11].

Remark If all the subgraphs $G_k = (N_k, A_k)$ induced by the node sets N_k , $N_k \in K_2$ contain no cycles then set of flows $F(f)$ is an empty set and the minimum flow problem has an unique optimal solution.

16.4 Conclusions

In this paper we focused on characterizing all optimal flows in a network because this problem has several practical applications and because it

arises as subproblem when dealing with more complex network flow problems, such inverse network flow problems. Although minimum flow problem has generally several optimal solutions, the algorithms developed until now determines only one of them. We described a decomposition method for determining all the optimal solutions of a minimum flow problem. Starting from a nontrivial subset of the set of all maximum cuts, a partition of the node set N is determined. In this way, the given graph $G = (N, A)$ is decomposed into several subgraphs $G_k = (N_k, A_k)$. In theorem 5, we characterized all the minimum flows by using the flows g_k in those subgraphs.

References

1. Ahuja R, Magnanti T, Orlin J (1993) Network flows. Theory, algorithms and applications. Prentice Hall, Inc., Englewood Cliffs, New Jersey
2. Bang-Jensen J, Gutin G (2001) Digraphs: Theory, Algorithms and Applications, Springer-Verlag, London
3. Ciupală L (2006), A deficit scaling algorithm for the minimum flow problem. *Sadhana* 31 (3):1169-1174
4. Ciupală L, Ciurea E (2001) An approach of the minimum flow problem, The Fifth International Symposium of Economic Informatics, pp. 786-790.
5. Ciurea E, Deaconu A (2009) Minimum Flows in Bipartite Networks with Unit Capacities. In N. Mastorakis et al.(eds) Recent Advances in Computers, Proceedings of the 13th WSEAS International Conference on COMPUTERS (part of the 13th WSEAS CSCC Multiconference) Rodos, Greece, July 23-25, 2009, pp. 313-315
6. Ciurea E, Ciupală L (2004) Sequential and parallel algorithms for minimum flows. *Journal of Applied Mathematics and Computing* 15 (1-2): 53-78
7. Deshpande A, Patkar S, Narayanan H (2005) Submodular Theory Based Approaches For Hypergraph Partitioning. *WSEAS Transactions on Circuit and Systems* 6 (4): 647-655.
8. Guo M, Qu H (2009) Constrained Optimization Evolutionary Algorithm. In N. Mastorakis et al.(eds) Applied Computer & Applied Computational Science, Proceedings of the 8th WSEAS International Conference on Applied Computer and Applied Computational Science (ACACOS '09), Hangzhou, China, May 20-22, 2009, pp. 446 -450
9. Mazlack L (2009) Graph Weaknesses in Commonsense Causal Representation. In N. Mastorakis et al.(eds) Recent Advances in Computers, Proceedings of the 13th WSEAS International Conference on COMPUTERS (part of the 13th WSEAS CSCC Multiconference) Rodos, Greece, July 23-25, 2009, pp. 435-440

10. Patkar S, Sharma H, Narayanan H (2004) Efficient Network Flow based Ratio-cut Netlist Hypergraph Partitioning. WSEAS Transactions on Circuits and Systems 3 (1): 47-53
11. Ruhe G (1985) Characterization of all optimal solutions and parametric maximum flows in network. Optimization 16 (1): 51-61
12. Zerovnik J (2010) Edge Fault-Diameter of Graph Product. In N. Mastorakis et al. (eds) Latest Trends on Computers, 14th WSEAS International Conference on Computers, Corfu Island, Greece, July 23-25, 2010, pp.630-635
13. Zerovnik J (2010) Fault Diameter of Graph Products and Bundles. In N. Mastorakis et al. (eds) Latest Trends on Computers, 14th WSEAS International Conference on Computers, Corfu Islands, Greece, July 23-25, 2010, pp. 636-641.

Chapter 17

Parameterization to avoid the Gibbs Phenomenon

E. Chicurel-Uziel

Instituto de Ingeniería, Universidad Nacional Autónoma de México, Av.
Universidad 3000, Coyoacan, 04510 México, D. F., México,
ecu@pumas.ii.unam.mx

Abstract. Series expansions of functions with discontinuities are plagued by spurious oscillations, this is the well known Gibbs phenomenon. A simple parameterization scheme is proposed to circumvent this phenomenon. The original discontinuous function is represented by parametric equations, with a special parameter such that this representation is continuous, exact and closed. Since the cause of the spurious oscillations are the discontinuities and they are removed, the Gibbs phenomenon simply does not ever arise. There is no significant change in the nature of the original function. Furthermore, it is possible to reconvert the pair of parametric equations into a single non-parametric equation in terms of the original independent variable.

Keywords. discontinuous functions, piecewise continuous, Gibbs phenomenon, spurious oscillations, Fourier series.

17.1 Introduction

In many applications it is convenient, and often necessary, to represent a discontinuous function by a series expansion. It is frequently required that, in addition, the representing equation be expressed in terms of transcendental functions, i.e., a Fourier series. The discontinuities, however, give rise to spurious oscillations, the magnitude of which remains essentially constant no matter how much the number of terms of the series is increased. This is the well known Gibbs phenomenon.

Many schemes have been proposed to diminish the effect of the Gibbs phenomenon, for instance: the classical Fejer and Lanczos averaging methods [12]; and, in recent times: the Gegenbauer polynomials direct

[1,5,6] and inverse methods [1,14], variations of the Lanczos σ factors [11,15], Padé methods [2,9], method associated with the sharp transition FIR filter [13], Jacobi, Laguerre and Hermite polynomial methods[6], wavelet methods [7,8], to mention only a few. Apparently, in all of these methods the original function is *first expanded in series*, and *after this expansion the series is reconstructed* to diminish the effect of the Gibbs phenomenon.

17.2 Concept

An entirely different concept is proposed in this paper. A very simple scheme is established to *represent the original discontinuous function by continuous, closed and exact parametric equations* and this parametric representation is then expanded into a Fourier series. *But neither the original function nor the series expansion are ever reconstructed*. Since the discontinuities are the cause of the Gibbs phenomenon, and they are removed, *the Gibbs phenomenon simply does not arise*.

But there are additional advantages: there are no reconstruction distortions and the number of terms in the series that are required to obtain an acceptable convergence is significantly reduced.

Incidentally, this approach made possible the establishment of exact, closed parametric equations to represent the Dirac delta [4] as well as the equations to represent rotational mechanical elements [3].

17.3 Procedure

This paper deals with piecewise continuous functions, PCF's, i.e., composite functions made up of continuous component functions with jump discontinuities between them, and possibly, at the beginning of the first and the end of the last component functions.

The value of the PCF at a jump discontinuity is usually undetermined and therefore, in a plot, it should be associated with a vertical gap. Such a representation will be assumed in this paper. In many books and in several software, however, the jump discontinuity is improperly represented graphically by an analytically non-existing vertical line segment.

The vertical gap associated with a jump discontinuity may be removed geometrically by simply drawing a vertical line segment to link the adjacent ends of two consecutive component functions. The analytical

existence of this line segment may be established by a suitable parametric representation.

It is worth pointing out that the modified, i.e., linked PCF is multi-valued at the junction point and, therefore, it is not continuous there with respect to the independent variable and thus the gap has been removed but the discontinuity, or rather, the lack of continuity remains. *The parameter, however, may be chosen so that both the dependent variable and the independent variable are continuous functions of the parameter and thus the parametric equations of these functions constitute an exact, closed, continuous representation of the discontinuous unlinked original PCF.* It is convenient that the parameter be as simple as possible, of course.

The addition of the vertical links is a step in the process of obtaining a continuous representation but it does not alter the original function in any other way.

17.3.1 Example

The parametric Fourier series expansion will be determined for the PCF defined by the following two component functions:

$$\begin{aligned} y(x) &= 2 + \frac{(x-3)^2}{2} & 3 \leq x < 7 \\ y(x) &= 4 & 7 \leq x < 11 \end{aligned} \quad (1)$$

The Heaviside unit step function :

$$\begin{aligned} h(x, a) &= 0 & x < a \\ h(x, a) &= 1 & x \geq a \end{aligned} \quad (2)$$

is used in this work in order to “switch in” or “switch out” functions which are valid only on a finite interval. In accordance with this, all six relations (2) may be expressed as a single equation, without the need of inequalities, as follows:

$$y(x) = [h(x, 3) - h(x, 7)] \left[2 + \frac{(x-3)^2}{2} \right] + [h(x, 7) - h(x, 11)] \quad (3)$$

See Fig. 1a.

The first step in the procedure is to eliminate the vertical gaps at the jump discontinuities, i.e. at $x = 3$, $x = 7$ and $x = 11$ by drawing vertical line

segments to link the adjacent component functions, Fig 1b. Parameterization will be used to represent these line segments analytically.

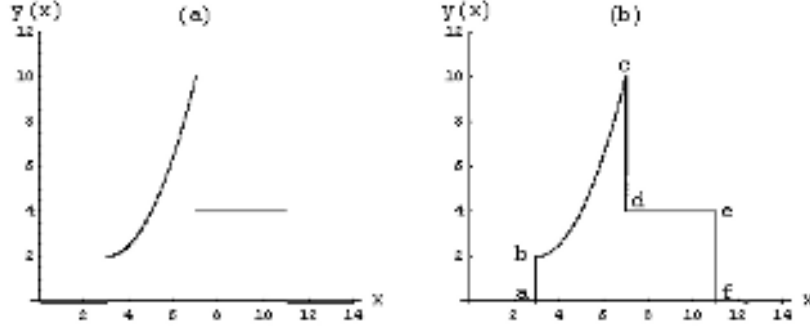


Fig. 1 (a) Piecewise continuous function, PCF, Eq. (3), (b) Linked PCF, the result of adding vertical line segments to eliminate the vertical gaps of the jump discontinuities. The letters refer to the junction points.

17.3.2 The chosen parameter

The special parameter chosen for the proposed representation is u :

$$u = \Delta_{ab} + \Delta_{bc} + \Delta_{cd} + \Delta_{de} + \dots \quad (4)$$

where:

$$\Delta_{jk} = \begin{cases} |x_k - x_j| & \text{if } jk \text{ is a component} \\ \text{function.} \\ |y_k - y_j| & \text{if } jk \text{ is a vertical link} \end{cases} \quad (5)$$

Thus, for instance for $x = 5$, Fig. 2a:

$$u = |y_b - y_a| + |x - x_b| \quad (6)$$

$$u = |2 - 0| + |5 - 3| = 4$$

for $x = 7$, $y = 6$, Fig. 2b:

$$u = |y_b - y_a| + |x_c - x_b| + |y - y_c|$$

$$u = |2 - 0| + |7 - 3| + |6 - 10| = 10$$
(7)

In this case the start point chosen was point *a*, but any convenient point could have been chosen instead, for instance, the origin.

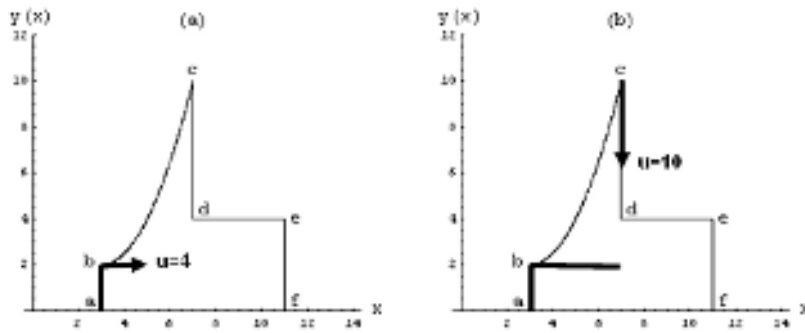


Fig. 2. Values of u for: (a) $x=5$, Eq. (6), (b) $x=7$, $y=6$, Eq.(7).

-By reference to Fig. 1b, the parametric relations may now be easily determined graphically, Figs 3a and 3b.

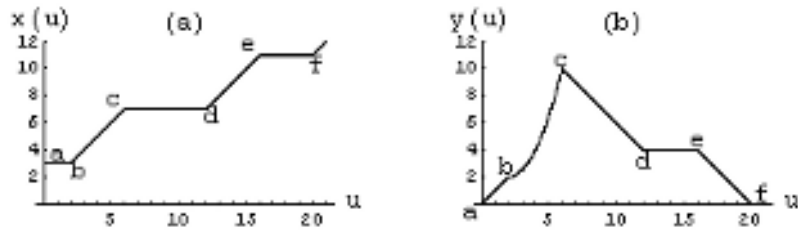


Fig. 3. These plots refer to the *continuous* parametric representation of the given *discontinuous* function. **(a)** The independent variable vs. the parameter u , plot of equation (8a), **(b)** The dependent variable vs. the parameter u , plot of equation (8b).

From these Figs. (3a) and (3b) the following parametric equations are established:

$$\begin{aligned}
 x(u) = & [h(u,0) - h(u,2)](3) + [h(u,2) - h(u,6)](u+1) \\
 & + [h(u,6) - h(u,12)](7) + [h(u,12) - h(u,16)](u-5) \\
 & + [h(u,16) - h(u,20)](11) + [h(u,20)](u-9)
 \end{aligned} \tag{8a}$$

$$\begin{aligned}
 y(u) = & [h(u,0) - h(u,2)](u) + [h(u,2) - h(u,6)] \left[2 + \frac{(u-2)^2}{2} \right] \\
 & + [h(u,6) - h(u,12)](-u+16) + [h(u,12) - h(u,16)](4) \\
 & + [h(u,16) - h(u,20)](-u+20)
 \end{aligned} \tag{8b}$$

Equations (8) are the ***continuous, closed and exact parametric equations of the given discontinuous function***. As a matter of fact, Fig. 1b is actually a parametric plot of Eq.(8a) as abscissa and Eq.(8b) as ordinate.

17.3.3 Fourier series expansion

We now proceed to expand the parametric function $y(u)$ as a Fourier sine series. The coefficients are:

$$b_n = \frac{2}{L} \int_0^L y(u) \sin \frac{n\pi u}{L} du \tag{9}$$

or equivalently:

$$\begin{aligned}
 b_n = \frac{2}{L} \bigg\{ & \int_0^2 u \sin \frac{n\pi u}{L} du + \int_2^6 \left[2 + \frac{(u-2)^2}{2} \right] \sin \frac{n\pi u}{L} du \\
 & + \int_6^{12} (-u+16) \sin \frac{n\pi u}{L} du + \int_{12}^{16} 4 \sin \frac{n\pi u}{L} du \\
 & + \int_{16}^{20} (-u+20) \sin \frac{n\pi u}{L} du \bigg\}
 \end{aligned} \tag{10}$$

$L = 20$, see Fig. 3b

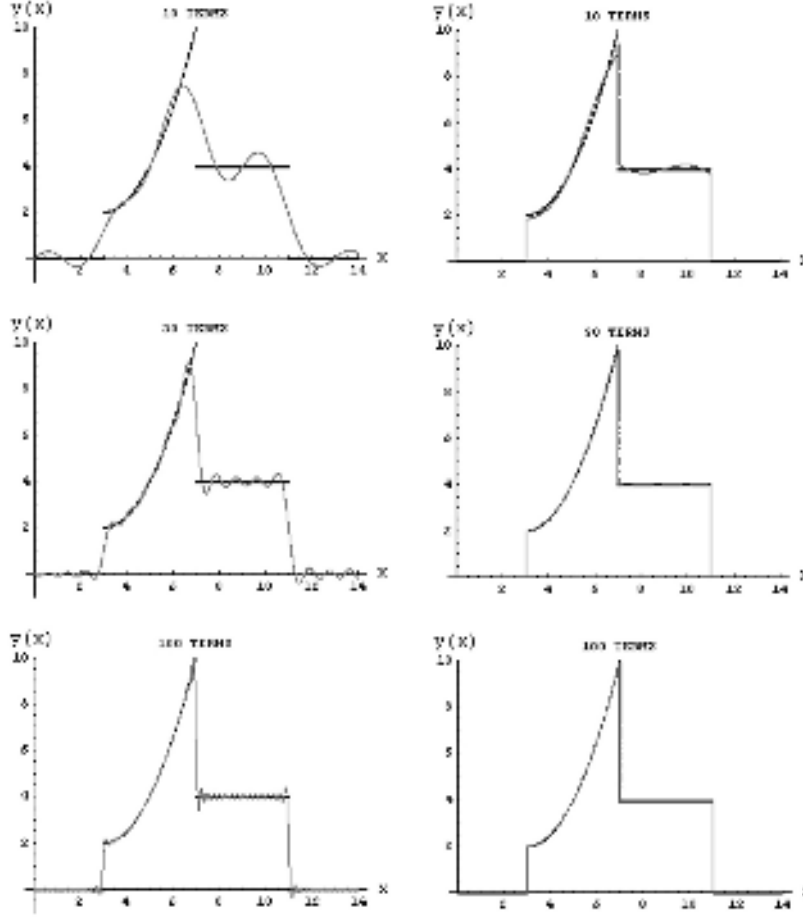


Fig.4. Fourier series expansion plots superimposed on the plot of the original function . On the left side: plots of the *direct* expansion of the *original discontinuous* function. On the right side: plots of Eqs. (11), the expansion of the *continuous parametric* representation of the function.

17.4 Results

Thus, the following pair of equations constitute the Fourier sine series expansion of the *parametric* representation of the original function:

$$y(u) = \sum_{n=1}^{\infty} b_n \sin \frac{n\pi u}{L} \quad (11a)$$

$$\begin{aligned} x(u) = & [h(u,0) - h(u,2)](3) + [h(u,2) - h(u,6)](u+1) \\ & + [h(u,6) - h(u,12)](7) + [h(u,12) - h(u,16)](u-5) \\ & + [h(u,16) - h(u,20)](11) + [h(u,20)](u-9) \end{aligned} \quad (11b)$$

Eq. (11b) is the same as Eq. (8a), of course.

The plots of the direct expansion in Fourier series of the original function, Fig. 4, left side, show greater amplitude in the oscillations next to the jump discontinuities which means that these are the spurious oscillations characteristic of the Gibbs phenomenon. All three curves deviate considerably from the original PCF, as expected. As it is well known, the error will never be smaller than about 9% no matter how many terms are included in the series expansion [9].

In the plots of the expansion in Fourier series of the parametric, continuous representation, Eqs. (11), Fig.4, right side, the plot for 10 terms of the series is the only one that shows any significant deviation from the original PCF but the oscillations in each component curve have constant amplitudes, so, it must be concluded, that *no Gibbs phenomenon is present*. The curves for 30 and 100 terms of the series practically coincide with the original PCF except for the vertical line segments.

Furthermore, it is obvious that considerably fewer terms are required for convergence in the Fourier series expansion of the linked parametric representation of the PCF than in the conventional Fourier series expansion.

But it is possible to reconvert the parametric representation into a single (nonparametric) equation in terms of the independent variable x .

From Fig. 3a and Eq.(8a), for the first component function, in the interval, $2 \leq u < 6$, we have: $u = x - 1$; for the second component function, in the interval, $12 \leq u < 16$, we have: $u = x + 5$. Substituting these relations into equation (11a) yields:

$$\begin{aligned} y(x)_{DF} = & \sum_{n=1}^{\infty} b_n \left\{ -[h(x,3) - h(x,7)] \sin \frac{n\pi(x-1)}{L} \right. \\ & \left. + [h(x,7) - h(x,11)] \sin \frac{n\pi(x+5)}{L} \right\} \end{aligned} \quad (12)$$

which is the single equation representing the Fourier sine series expansion of the given function in terms of the independent variable. The plots of Eq. (12) are exactly the same as those of the right side of Fig. 4 except for the fact that there are no vertical links.

17.5 Conclusions

A simple procedure has been established to represent a function with jump discontinuities by continuous, closed and exact parametric equations. This is accomplished by closing the gaps of the jump discontinuities by means of vertical lines which may be established analytically by a specific parameterization of the function. The function in terms of the parameter is free of discontinuities. Since in the series expansion of the function, the jump discontinuities give rise to the Gibbs phenomenon, eliminating them prevents the Gibbs phenomenon from arising.

Furthermore, the parametric Fourier series expansion converges with significantly fewer terms than those required by the conventional, direct Fourier series representation.

It is worth emphasizing that, since in the proposed methodology there is no reconstruction of the series, the series expansion is, obviously, also free of the distortions introduced by the reconstruction methods such as: convergence to a deviated value of the function and decreased steepness at the points of discontinuity. To be sure, in the proposed procedure the series converges to the value of the original function and the rise, or fall, at the points of the jump discontinuities is vertical.

References

1. Boyd JP (2005) Trouble with the Gegenbauer reconstruction for defeating Gibbs phenomenon in the diagonal limit of Gegenbauer polynomial approximations. *J Comp Phys* 204: 253-264
2. Brezinski C (2004) Extrapolation algorithms for filtering series of functions, and treating the Gibbs phenomenon. *Num Algorithms* 36:309-329
- 3.. Chicurel-Uziel E (2006) Representation of rotational mechanical parts by parametric equations. *WSEAS Trans on Math* 5:540-548
4. Chicurel-Uziel E (2007) Dirac delta representation by exact parametric equations. Application to impulsive vibration systems. *J Sound and Vib*, 305:134–150
5. Gottlieb D, Shu CW (1997) On the Gibbs phenomenon and its resolution. *SIAM Rev* 39:644-668

6. Greene N (2008) Fourier series of orthogonal polynomials. Proc Am Conf on Appl Math, WSEAS, pp 247-251
7. Greene N (2008) A wavelet-based method for overcoming the Gibbs phenomenon. Proc of the Am Conf on Appl Math, WSEAS, pp 408-412
8. Greene N (2009) Contributions to the resolution of the Gibbs phenomenon. Proc Int Conf on Electronic, Sign Proc and Control, WSEAS, pp 261-265
9. Hesthaven JS, Kaber SM, Lurati L (2006) Padé-Legendre Interpolants for Gibbs Reconstruction. J of Sci Comp 28:337-359
10. Hoskins R.F (1979) Generalized Functions. Ellis Horwood Ltd. Wiley, Chichester, W. Sussex, England, p 42
11. Jerri AJ (2000) Lanczos-like σ -factors for reducing the Gibbs phenomenon in general orthogonal expansions and other representations. J Comp Analysis and Applica 2:111-127
12. Lanczos C (1956) Applied Analysis. Prentice-Hall, Englewood Cliffs, NJ, USA, pp 32-36, 225-229
13. Rodrigues J, Pai KR. (2009) A new model of sharp transition FIR filter. Proc 4th WSEAS International Conference on Electronic, Sign Proc and Control, article No 33
14. Shizgal BD, Jung Jae-Hun (2003) Towards the resolution of the Gibbs Phenomena. J Comp and Appl Math 161:41-65
15. Yun BI, Rim KS (2009) Construction of Lanczos type filters for the Fourier series approximation. Appl Numl Math 59:280-300

Chapter 18

Multi Party Computation using Infinite Product Series over Distributed Data

Abdolreza Rasouli, Mohd Aizaini Maarof, Mahboubeh Shamsi

Faculty of Computer Science & Information System, University
Technology Malaysia, Skudai, Johor, Malaysia – 81310
rs.reza@gmail.com, aizaini@utm.my, mahboubeshamsi@yahoo.com

Abstract. The potency of data mining tools to discover useful rules among large data inspires the companies to share their data to be mined. The big problem of these companies is trusting to a third party miner. Currently, the issue of Multi Party Computation (MPC) tries to solve this conflict using cryptographic tool. In this paper, we propose a secure summation algorithm for online transactions, where the users will send the data to the miner, stepwise. The algorithm emerges the excessively useful response time, so the execution time of summation for 1000 users' data is only 0.9s.

Keywords. Multi Party Computation, Privacy Preserving, Secure Counting Algorithm, Distributed Data

18.1 Introduction

Today, the large amount of data will be used among the large number of companies. There exist a lot of valuable patterns and roles which are hided between wooded data. Data mining tools has been developed to discover these worth facts. It is sophisticated where the data has been distributed among number of parties because of privacy of their sensitive data. Each company leans to gain the advantage of data mining tools, but he does not propend to share his sensitive data. Multi Party Computation (MPS) has been solved this conflict. MPC ensures the privacy preserving of data

using cryptographic tools as well as precious aggregation over the distributed data.

In this paper, we propose a cryptographic method that allows a Miner system to compute aggregation of respondents' data. As we mentioned before, it is important that content of data should not be revealed. In this algorithm, each client only sends a message to Miner system and there is no need to any inner interaction between clients. The algorithm ensures that no extra information is revealed to Miner System except the summation of respondents' data.

Our algorithm could be used as basic fundamental for data mining tools. It is also useful for implementing an E-Voting system or calculation of web pages poll result.

First, we will briefly introduce some related work and basic cryptographic method in section 1.2, respectively and then we will fully explain the base algorithm in section 1.3. We explain the problem and proposed algorithm in section 1.4 and we represent our result in section 1.5. The paper will ends with a discussion about the earned result and precise conclusion in section 1.6.

18.2 Background and Related Work

The secure multi-party computation also known as (MPC) is one of the main results of the theory of cryptography. First, Yao's [2] introduced the multi-party computation and nowadays many authors have attend many optimizations and extensions to the basic concept, for two main branch; the two-party (2PC) and the multi-party (MPC) [6,7,9,10,14,15]. Most of recently papers on secure multi-party computation area have been focused on theory of multi-party computation and there is no much applicable implementing of MPC, although, in the few last year some practical implementation of multi-party computation has been appeared [1,3,11,12,21].

There exist many algorithm and techniques for secure multi-party computation. We have focused on more practical and high speed algorithms which have been published.

Secure multi-party computation essentially comes in two flavors [16]. The first approach is typically based upon secret sharing and operates on an arithmetic circuit representation of the computed function, such as in the BGW (Ben-Or, Goldwasser and Wigderson) or CCD (Chaum, Crepeau and Damgård) protocols [5,13]. This approach is usually applied when there is an honest majority among the participants (which can only exist if more than two parties participate in the protocol). An alternative approach

represents the function as a binary circuit. This approach was used in the original two-party garbled circuit construction of Yao [2], and in the GMW (Goldreich, Micali and Wigderson) multi-party protocol [7].

Our proposed algorithm is secure within the semi-honest model. The algorithm is secure under Diffie-Hellman DDH assumption and uses ElGamal encryption for increasing robustness and speed.

18.2.1 ElGamal Cryptosystem

The ElGamal cryptosystem is a part of public encryption systems. The public key is (h, G, q, g) where G is a cyclic group of order q with the generator g , $h=g^x$ and x is the private key which is randomly chosen from $[1, q]$. All computation in the ElGamal scheme is done in the group G .

Under the public key (h, G, q, g) , the ciphertext of a message m (which is the representation of an element of G) is encrypted as $E(m)=(c_1, c_2)$ where $c_1=m.h^r$, $c_2=g^r$ and r is randomly chosen from $[1, q]$. To decrypt the ciphertext (c_1, c_2) with the private key x , the plaintext message m can be decrypted as $m=c_1(c_2^x)^{-1}$. It clearly is true because

$$c_1(c_2^x)^{-1} = m.h^r(g^{rx})^{-1} = m.h^r(h^r)^{-1} = m \quad (1)$$

ElGamal encryption is semantically secure under the Decisional Diffie-Hellman (DDH) assumption [4]. One family in which DDH is believed to be intractable is the quadratic residue subgroup Q_p of Z_p^* where p, q are two primes and $p=2q+1$.

In the ElGamal encryption, one cleartext has many possible ciphertexts because of the random value r . ElGamal supports rerandomization: a new ciphertext $E'(m)$ of m can be computed from a ciphertext $E(m)=(c_1, c_2)$ as $E'(m)=(c_1.h^{r'}, c_2.g^{r'})$ where r' is randomly chosen from $[1, q]$.

18.2.2 Secret Sharing

Secret sharing is the method of sharing a secret by multiple parties, so that no one and no party know the secret, but the secret could be constructed by combining some parties' shares.

For example, in a two-party case, Alice and Bob share a value x modulo some appropriate value N , in such a way that Alice holds a , Bob holds b , and x is equal to $(a+b) \bmod N$. This is called additive secret sharing. An important property of this kind of secret sharing is that if Alice and Bob have shares of a and b , then they can each locally add their shares modulo N to obtain shares of $a+b$.

Shamir secret sharing is a threshold scheme [19]. In Shamir secret sharing, there are N parties and a polynomial P of degree $k-1$ such that $P(0)=x$ where x is a secret. Each of the N parties holds a point in the polynomial P . Because k points (x_i, y_i) ($1 \leq i \leq k$) uniquely define a polynomial P of degree $k-1$, a subset of at least k parties can reconstruct the secret x . But, fewer than k parties cannot construct the secret x . This scheme is also called (N, k) Shamir secret sharing.

18.3 Aggregation Algorithm

In our scenario, there are n clients, so we call them C_1, \dots, C_n ; respectively. Each client owns his private data d_i . The aim of the Miner System is to calculate the sum $d = \sum_{i=1}^n d_i$ with ensuring the privacy of d_i .

In our model, because of its practicability, the clients do not need to know about other clients and they never communicate themselves. So there is no communication channel between different clients. Moreover, each client only sends one encrypted message to the Miner. So, they do not need multi round interaction between clients and the Miner System.

The implemented algorithm [17] is based on the homomorphism property of mentioned ElGamal encryption [8,20]. The DDH assumption and the ElGamal cryptosystem ensure the privacy of the algorithm. The algorithm also uses the exponentiation's mathematical properties for converting multiplication to desired sums. We also use modular arithmetic operation to speed up the computing time of big prime numbers. It is surely affect on algorithm time.

Let G be a group where $(|G|=q$ for a large prime $q)$, and let g be a generator of G . The group G is assumed for all computations in this paper. Suppose that each respondent C_i has two pairs of keys: $((x_i \bmod q), (X_i \bmod q = (g^{x_i} \bmod q)))$, $((y_i \bmod q), (Y_i \bmod q = (g^{y_i} \bmod q)))$. We also define

$$(X) \quad (2)$$

$$(Y) \quad (3)$$

The values x_i and y_i are private keys; X_i and Y_i are public keys. All respondents need to calculate the values of X and Y from public keys.

First, each respondent C_i try to encrypt his value d_i using his private key x_i , y_i and shared public key X , Y in ElGamal encryption system as described below

$$E(d_i) = \begin{cases} (m_i \bmod q) = (g^{d_i} \bmod q) \\ (h_i \bmod q) = (Y^{x_i} \bmod q) \end{cases} \quad (4)$$

Then all clients send their encrypted message to Miner. The Miner System gathers all encrypted data together and computes m, h as:

$$m = \prod_{i=1}^n (m_i \bmod q), h \quad (5)$$

The Miner will use m, h for decrypting the d as sum of d_i . Then the Miner tries to find correct d between all possible values. It means that the miner tries to calculate $(g^d \bmod p)$ for all possible of d values. This stage is more time consuming step and will continue until there exist any d as

$$(m \bmod q) = \quad (6)$$

This value is the desired summation of respondents' votes. Mathematic demonstration is shown in [18]. The Miner cannot take discrete logarithms; the Miner must use trial and error to learn d . The time consumption parameter of algorithm returns to the range of possible values of d . In case of Boolean votes the range of d is the number of respondent's n .

18.4 Problem Explanation

Base on the mentioned algorithm, before the process starts, the miner should ensure that all clients are ready to send their data and the Miner knows the number of clients. It means that the system is semi-offline system. This condition avoids us to use the algorithm in real world online problems like as E-Voting or web page's poll result. Because in this cases, the clients are not ready at the beginning time and the time of user's appearance is optional. The miner does not have any knowledge which how many clients could join to system and when they will appear. So the Miner could not able to compute the secret shared between the clients. As we mentioned before, the secret shared between the users is the product of their public keys as we called X, Y . All clients should be present their public keys to compute the shared secret key X, Y before aggregation started. Moreover, all users should send their data; therefore the Miner could compute the summation.

So, in most online cases the algorithm will fail and is inapplicable. Due to the Miner, likewise, clients cannot assess the product of user's public keys that do not join yet.

In this paper, infinite product series have been used as shared secret key instead of X, Y . The final amounts of these series do not change with increasing the number of elements. An infinite product series define as

$$A = \prod_{i=1}^{\infty} \alpha_i \text{ like as } \prod_{i=1}^{\infty} \frac{1}{e} \left(\frac{1}{2i+1} \right)^{2i+\frac{1}{2}} \quad (7)$$

18.4.1 Proposed Algorithm

The proposed algorithm also is based on the homomorphism property of mentioned ElGamal encryption. We note that the security of ElGamal encryption depends on new random values being used for each encryption. Same as before, each client C_i has two pairs of keys: $((x_i \bmod q), (X_i \bmod_q = (g^{x_i} \bmod_q)))$, $((y_i \bmod_q), (Y_i \bmod_q = (g^{y_i} \bmod_q)))$ as private and public keys, respectively. Remember that the x_i and y_i values cannot be reused in different uses of the running algorithm. Here, we are not able to compute the X, Y . So we use two infinite series Ω, Ψ instead of X, Y .

$$(\Omega_i \bmod_q) = (g^{\omega_i} \bmod_q), (\Omega \bmod_q) = \prod_i \quad (8)$$

$$(\Psi_i \bmod_q) = (g^{\psi_i} \bmod_q), (\Psi \bmod_q) = \prod_i \quad (9)$$

The client C_i encrypt his value d_i in ElGamal encryption system using Ω, Ψ as described below

$$E(d_i) = \begin{cases} (m_i \bmod_q) = (g^{d_i} \bmod_q) \\ (h_i \bmod_q) = (\Psi^{x_i}) \end{cases} \quad (10)$$

and send his encrypted message in assition to his public keys X_i, Y_i to the Miner. The Miner System computes m, h, X, Y as before:

$$m = \prod_{i=1}^n (m_i \bmod_q), h \quad (11)$$

$$X = \prod_{i=1}^n (X_i \bmod_q), Y \quad (12)$$

The Miner tries to find desired d to satisfy below

$$(m \bmod_q), (X^{\sum_{i=1}^n \psi_i}) = (g^d \bmod_q), (h \bmod_q) \quad (13)$$

The desired d is the summation of clients' data.

18.4.2 Algorithm Demonstration

In this section, we prove that the algorithm represented in 4.1 correctly computes the summation of respondents' data. Suppose that the Miner finds a d so

$$(m \bmod_q), (X^{\sum_{i=1}^n \psi_i}) = (g^d \bmod_q), (h \bmod_q), (Y^{\sum_{i=1}^n \omega_i}) \quad (13^*)$$

We will show that $d = \sum_{i=1}^n$.

$$\begin{aligned} (m \bmod_q), (X^{\sum_{i=1}^n \psi_i}) &= (g^d \bmod_q), (h \bmod_q) \\ \Rightarrow (\prod_{i=1}^n (m_i \bmod_q)), (X^{\sum_{i=1}^n \psi_i}) &= (g^d \bmod_q), (\prod_{i=1}^n (h_i \bmod_q)), (Y^{\sum_{i=1}^n \omega_i}) \\ \Rightarrow (g^d \bmod_q) &= \frac{(\prod_{i=1}^n (m_i \bmod_q)) \cdot (X^{\sum_{i=1}^n \psi_i})}{(\prod_{i=1}^n (h_i \bmod_q)) \cdot (Y^{\sum_{i=1}^n \omega_i})} = \frac{(\prod_{i=1}^n (g^{d_i} \bmod_q) \cdot (g^{y_i} \bmod_q)) \cdot (X^{\sum_{i=1}^n \psi_i})}{(\prod_{i=1}^n (\varphi^{x_i} \bmod_q)) \cdot (Y^{\sum_{i=1}^n \omega_i})} \\ &= \frac{(\prod_{i=1}^n (g^{d_i} \bmod_q) \cdot \prod_{i=1}^n (g^{y_i} \bmod_q)) \cdot (X^{\sum_{i=1}^n \psi_i})}{(\prod_{i=1}^n (\varphi^{x_i} \bmod_q)) \cdot (Y^{\sum_{i=1}^n \omega_i})} = \frac{\prod_{i=1}^n (g^{d_i} \bmod_q) \cdot \prod_{i=1}^n (g^{y_i} \bmod_q)}{\prod_{i=1}^n (\varphi^{x_i} \bmod_q)} \cdot \frac{(X^{\sum_{i=1}^n \psi_i})}{(Y^{\sum_{i=1}^n \omega_i})} \\ &= g^{\sum_{i=1}^n (d_i \bmod_q)} \cdot \frac{X^{\sum_{i=1}^n \psi_i}}{Y^{\sum_{i=1}^n \omega_i}} \\ &= g^{\sum_{i=1}^n (d_i \bmod_q)} \cdot \frac{(\prod_{i=1}^n (g^{d_i} \bmod_q))^{\sum_{j=1}^n y_j}}{(\prod_{i=1}^n (\varphi^{x_i} \bmod_q))^{\sum_{j=1}^n x_j}} \\ &= g^{\sum_{i=1}^n (d_i \bmod_q)} \cdot \frac{(\prod_{i=1}^n (g^{d_i} \bmod_q))^{\sum_{j=1}^n y_j}}{(\prod_{i=1}^n (g^{y_i} \bmod_q))^{\sum_{j=1}^n x_j}} \\ &= g^{\sum_{i=1}^n (d_i \bmod_q)} \cdot \frac{g^{\sum_{i=1}^n (d_i \bmod_q) \sum_{j=1}^n y_j}}{g^{\sum_{i=1}^n (y_i \bmod_q) \sum_{j=1}^n x_j}} \\ &= g^{\sum_{i=1}^n (d_i \bmod_q)} \cdot \frac{g^{\sum_{j=1}^n (y_j \bmod_q) \sum_{i=1}^n \omega_i}}{g^{\sum_{j=1}^n (x_j \bmod_q) \sum_{i=1}^n \psi_i}} \\ &= g^{\sum_{i=1}^n (d_i \bmod_q)} \cdot \frac{(\prod_{i=1}^n (g^{y_i} \bmod_q))^{\sum_{i=1}^n \omega_i}}{(\prod_{i=1}^n (g^{x_i} \bmod_q))^{\sum_{i=1}^n \psi_i}} \\ &= g^{\sum_{i=1}^n (d_i \bmod_q)} \cdot \frac{(\prod_{i=1}^n (g^{y_i} \bmod_q))^{\sum_{i=1}^n \omega_i}}{(\prod_{i=1}^n (g^{x_i} \bmod_q))^{\sum_{i=1}^n \psi_i}} \\ &= g^{\sum_{i=1}^n (d_i \bmod_q)} \cdot \frac{Y^{\sum_{i=1}^n \omega_i}}{X^{\sum_{i=1}^n \psi_i}} \\ \Rightarrow (g^d \bmod_q) &= g^{\sum_{i=1}^n (d_i \bmod_q)} \therefore d = \sum_{i=1}^n \end{aligned}$$

18.5 Experimental Result

We implemented our algorithm in Delphi. All cryptographic operations use the OpenSSL and FGBigInt libraries. The OpenSSL library is accessible at www.openssl.org website. The IIS HTTP Server has been used for network simulation under windows Vista on a PC with a 2.4GHz processor and 2GB memory. We choose the 512 bits as the length of each cryptographic key. The main parameters of experiment are: Number of Clients and the range of d . Moreover, the time of computing d_i in each respondent will affect on final time. The key generation also is time consuming process, but these parameters are ignored in our experimental result. Because, as we mentioned before, these values can be precomputed offline before the protocol starts.

In our result, we propose the time consumption of algorithm as the main factor. The algorithm's stages are:

- Sending Encrypted message to Miner
- Computing m, h, X, Y on Miner
- Finding desired d across its' possible values

We define two different phase of result. In first phase, each client owns a Boolean data. In sooth, we are counting the users' data. This case is similar to E-Voting systems that each client is free to select an option and the Miner aims to know how many users do vote the idea. The final result of phase 1 is shown in Fig. 1.1. In this experiment, the respondents send the Boolean data to Miner. In this condition, the range of d is equal to number of clients. We use 50, 100, 200, 500 and 1000 users in our experiment and the earned time is base on the average of five algorithm runs. As you can see, the time offers a linear behavior related to number of users. For example, Miner computation takes 952 milliseconds for 1000 voters.

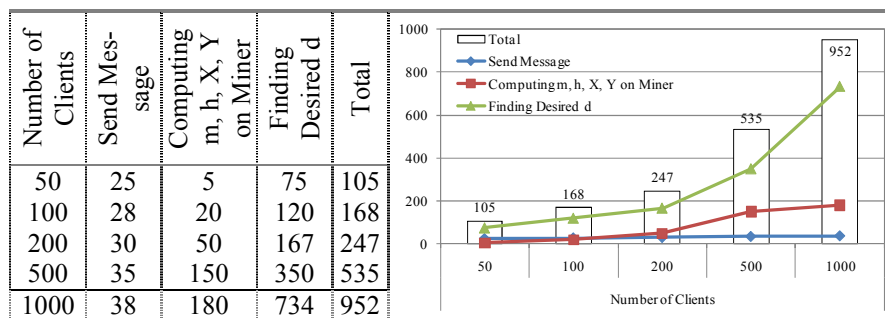


Fig. 1.1. Total time of Protocol related to Number of respondents

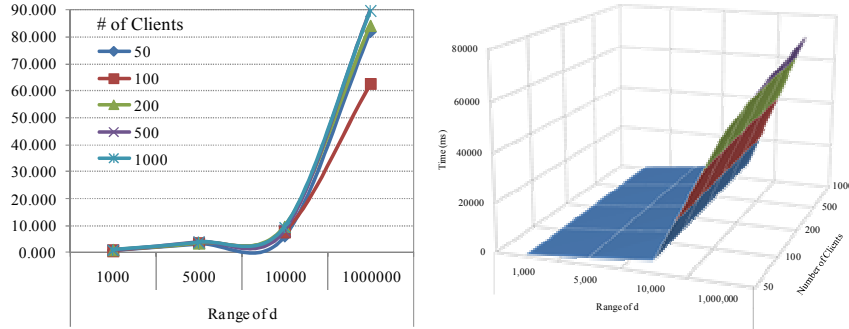


Fig. 1.2. a) Total time related to Number of Voters and Range of d- b) 3d view

In second phase, we have focused on non Boolean cases. In this case, the time is highly related to range of d . The range of d is not a simple parameter and return to many conditions. The Fig. 1.2 shows the time of algorithm execution related to number of clients and range of d . We earn the algorithm takes less than 0.5 second for 1000 voters with d in range of 1 to 1,000,000.

18.6 Discussion and Conclusion

In this paper, we proposed a secure aggregation algorithm. Our proposed algorithm ensures the confidentiality of sensitive respondents' data. Furthermore, it does not need communication channels between different respondents or multi round interaction between any respondent and the Miner Systems.

The mathematical demonstration ensures the accuracy of data's aggregation. On the other hand, The DDH assumption and the ElGamal cryptosystem ensure the privacy of the algorithm and respondents' data, so the Miner system can not reveal respondents' data. Our experimental results also show significantly desirable response time. The time increasing is derived from high security of algorithm, because the Miner could not able to decrypt the message, so the Miner should use trial and error to find the result. The algorithm can also be used for any model enabled by counting values. Our both theoretical analysis and proof in addition to experimental results show that the algorithm is very efficient and runs in desirable time.

References

1. Amirbekyan A, Estivil-Castro V (2007) Privacy-Preserving Regression Algorithms. In: Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization, vol., Beijing, China
2. Andrew Chi-Chih Y (1986) How to generate and exchange secrets. In: Proceedings of the 27th Annual Symposium on Foundations of Computer Science, vol. IEEE Computer Society,
3. Bogetoft P, Damgård I, Jakobsen T, Nielsen K, Pagter J, Toft T (2006) A Practical Implementation of Secure Auctions Based on Multiparty Integer Computation. In: Financial Cryptography and Data Security, vol., pp 142-147
4. Boneh D (1998) The decision Diffie-Hellman problem. Lecture Notes in Computer Science
5. David C, Claude C, peau, Ivan D (1988) Multiparty unconditionally secure protocols. In: Proceedings of the twentieth annual ACM symposium on Theory of computing, vol. ACM, Chicago, Illinois, United States
6. Goldreich O (2004) Foundations of Cryptography, vol 2. Cambridge Univ. Press,
7. Goldreich O, Micali S, Wigderson A (1987) How to play ANY mental game. In: Proceedings of the nineteenth annual ACM symposium on Theory of computing, vol. ACM, New York, New York, United States
8. Hirt M, Sako K (2000) Efficient receipt-free voting based on homomorphic encryption. In Advances in Cryptology - Proceedings of EUROCRYPT volume 1807 of Lecture Notes in Computer Science
9. Jarecki S, Shmatikov V (2007) Efficient Two-Party Secure Computation on Committed Inputs. EUROCRYPT
10. Lindell Y, Pinkas B (2007) An Efficient Protocol for Secure Two-Party Computation in the Presence of Malicious Adversaries. EUROCRYPT 52-79
11. Ma J, Sivakumar K (2005) Privacy-Preserving Bayesian Network Parameter Learning. In: 4th WSEAS Int. Conf. on Computational Intelligence, Man-Machine Systems and Cybernetics, vol., Miami, Florida, USA, pp 46-51
12. Ma J, Sivakumar K (2006) A Framework of Privacy-Preserving Bayesian Network Parameter Learning using Post Randomization. WSEAS Transaction of Information Science and Applications 3(1)
13. Michael B-O, Shafi G, Avi W (1988) Completeness theorems for non-cryptographic fault-tolerant distributed computation. In: Proceedings of the twentieth annual ACM symposium on Theory of computing, vol. ACM, Chicago, Illinois, United States
14. Mishra DK, Chandwani M (2007) Extended Protocol for Secure Multiparty Computation using Ambiguous Identity. WSEAS Transaction on Computer Research 2(2)
15. Mohassel P, Franklin M (2006) Efficiency Tradeoffs for Malicious Two-Party Computation. In: Public Key Cryptography - PKC 2006, vol., pp 458-473
16. Pinkas B, Schneider T, Smart N, Williams S (2009) Secure Two-Party Computation Is Practical. In: Advances in Cryptology – ASIACRYPT 2009,

- vol., pp 250-267
17. Rasouli A, Hosseinkhani J, Shamsi M, Harouni M (2010) A Robust and High Speed E-Voting Algorithm Using ElGammel CryptoSystem. In: The 2nd International Conference on Computer and Automation Engineering, vol., Singapore
 18. Rasouli A, Maarof MAB, Shamsi M (2010) A New Multi Party Aggregation Algorithm Using Infinite Product Series. In: The 9th WSEAS International Conference on Applications of Computer Engineering, vol., Penang, Malaysia
 19. Shamir A (1979) How to share a secret. *Communications of the ACM*:22(11) 612–613
 20. Wang M-N, Yen S-M, Wu C-D, Lin C-T (2006) Cryptanalysis on an Elgamal-like cryptosystem for encrypting large messages. In: *Proceedings of the 6th WSEAS International Conference on Applied Informatics and Communications*, vol., Elounda, Greece
 21. Yehuda L, Benny P, Nigel PS (2008) Implementing Two-Party Computation Efficiently with Security Against Malicious Adversaries. In: *Proceedings of the 6th international conference on Security and Cryptography for Networks*, vol. Springer-Verlag, Amalfi, Italy

Chapter 19

A Method of Extended Linearization for Polynomial Periodic and Autonomous Systems

V. G. Melnikov

Department of Theoretical and Applied Mechanics, The National Research University of Information Technologies, Mechanics and Optics,
Sablinskaya 14, 197101, St.- Petersburg, Russia, melnikov@mail.ifmo.ru

Abstract. This paper presents a new method of making an approximately equivalent extended linear systems for autonomous nonlinear systems with polynomial structure. This method is based on the method of extended model. The Chebyshev economization is used to obtain higher accuracy. The method is extended on polynomial systems with periodic coefficients. Two examples are given to illustrate the developed method.

Keywords. Chebyshev economization, Chebyshev polynomials, telescoping power series, nonlinear systems, approximation, extended systems, autonomous systems, periodic systems.

19.1 Introduction

The problem of has received much attention in recent years [1-12]. In this paper we consider a system of autonomous dynamic equations in the normal form with a polynomial right side [1],[5],[6]. We apply a method of extending of an object by introduction of a countable set of additional phase coordinates, that transforms the nonlinear equations of perturbed motion into the linear infinite system of equations [2], [3], and then we apply Chebyshev economization method [4], [7] to obtain higher accuracy of the finite linear extended system.

19.2 Initial and extended systems

Suppose a nonlinear system can be defined or approximated up to desired accuracy as below:

$$\frac{dx_i}{dt} = \sum_{j=1}^n a_i^j x_j + \sum_{|\nu|=2}^m x_\nu a_i^\nu, \quad x_j(t_0) = x_{j0} \quad (1)$$

where a_{ij} are constant coefficients, $\nu = (\nu_1, \dots, \nu_n)$ are vector indexes, $x_\nu = x_1^{\nu_1} \dots x_n^{\nu_n}$ are monomials of the degree $|\nu| = \nu_1 + \dots + \nu_n$, a_i^ν are constant coefficients with vector indexes ν , $m = 2n$, $i = 1, \dots, n$, $e_j \equiv [0, \dots, 0, 1, 0, \dots, 0]$ - unit indexes, $a_i^j \equiv a_i^{e_j}$ - constant coefficients of the linear forms.

By definition put

$$\bar{D} = \{x = [x_1, \dots, x_n] \in \mathbb{R}^n, \quad |x_j| \leq r, \quad i = 1, \dots, n\} \quad (2)$$

where \bar{D} is a finite region of the state-space. We assume that the nonlinear terms in (1) are small $a_i^\nu = O(\varepsilon)$, and that we have applied a linear change of independent and dependent variables to normalize the region \bar{D} ,

As it is known [2], [3] systems with polynomial nonlinearity of the form (1) can be transformed to the infinite set of linear equations using the object extension method.

Let us introduce the modified method. First we add a finite set of additional variables

$$x_\nu = x_1^{\nu_1} \dots x_n^{\nu_n}, \quad |\nu| = 2, 3, \dots, m, \quad m = 2n. \quad (3)$$

Note that each homogeneous form of the power $|\nu| = k$ includes $N_k = ((k+1)(k+2)(k+n-1)) / ((n-1)!)$ monomials. As a whole we add $N = N_2 + N_3 + \dots + N_m$ new variables. Let us redenote these variables as follows

$$\{x_\nu, |\nu| = 2, 3, \dots, m\} \Rightarrow [x_{n+1}, x_{n+2}, \dots, x_{n+N}]. \quad (4)$$

Differentiating (3), taking into account (1) and using vector indexes we obtain

$$\dot{x}_\nu = \sum_i \nu_i \dot{x}_i x_{\nu-e_i} = \sum_i \nu_i x_{\nu-e_i} \left(\sum_j x_j a_i^j + \sum_{\nu'} x_{\nu'} a_i^{\nu'} \right) \quad (5)$$

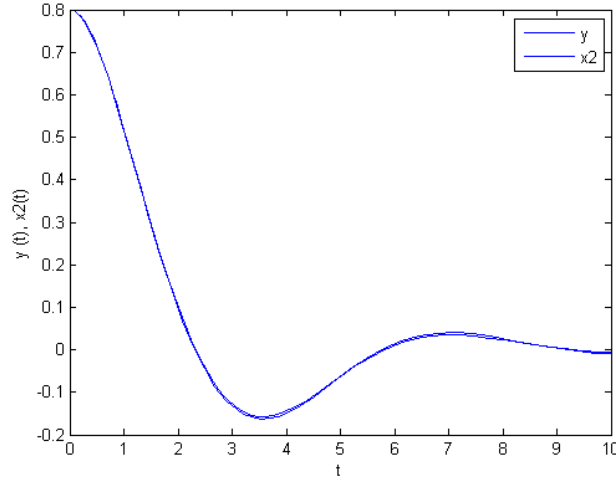


Fig. 1.1. Solutions of the nonlinear and extended linear systems

By $X^{(2,m)}$ denote the polynomials with terms of the power $2, 3, \dots, m$ and by $X^{(m+1, 2m-1)}$ denote the remainder polynomials. We obtain

$$\dot{x}_v = \sum_{i,j} v_i a_i^j x_{v-e_i+e_j} + X^{(2,m)} + X_v^{(m+1, 2m-1)}. \quad (6)$$

Since every term of the remainder polynomials has at least one variable to the power $k \geq 3$, we can apply a Chebyshev economization method [4], [7] as follows:

$$\begin{aligned} x_s^3 &\approx \frac{3}{4}r^2 x_s, & \delta^{(3)} &\leq \frac{1}{4}r^3, & r &\leq 1 \\ x_s^4 &\approx r^2 x_s^2 - \frac{1}{8}r^4, & \delta^{(4)} &\leq \frac{1}{8}r^4, \\ x_s^5 &\approx \frac{5}{4}r^2 x_s^3 - \frac{5}{16}x_s r^4, & \delta^{(5)} &\leq \frac{1}{16}r^5, \end{aligned} \quad (7)$$

where δ is the economization error. Note that for some cases it is more convenient to use the alternative approximations $x_s^4 = x_s x_s^3 \approx \frac{3}{4}r^2 x_s^2$, although they have higher error $\delta = \delta^{(3)}$.

Then using this method repeatedly, we finally get the following approximation of the remainder polynomials $X_v^{(m+1, 2m-1)}$ by polynomials $\tilde{X}_v^{(2,m)}$ and additional linear and constant terms

$$X_v^{(m+1, 2m-1)} \approx c_v + \sum_i x_i b_v^i + \tilde{X}_v^{(2, m)} \quad (8)$$

Substituting (8) for (6), we get the extended linear system in a matrix form

$$\begin{aligned} \frac{dy}{dt} &= yB + C, \quad B = [b_{ij}]_1^{n+N}, \quad C = [0, 0, \dots, c_{n+1}, \dots, c_{n+N}], \quad y \in \tilde{D}, \\ \tilde{D} &= \{y = [x_1, \dots, x_{n+N}]: |x_i| \leq r, \quad i = 1, 2, \dots, n, \quad |x_{n+1}| \leq r^{|v|}\} \end{aligned} \quad (9)$$

with the initial conditions

$$x_i(t_0) = x_{i0}, \quad i = 1, \dots, n, \quad (x_v(t_0) = x_{v0} = x_{10}^{v_1} \dots x_{10}^{v_n}, \quad |v| = 2, 3, \dots, m) \quad (10)$$

Note that in (9) we neglected small quantities. Therefore the solutions of the extended system (9)-(10) will satisfy (1) with some residual error.

19.3 Transformation of polynomial nonlinear systems with periodic coefficients

Suppose a nonlinear system with periodic coefficients is defined as below:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n-2} A_i^j x_j + \sum_{|\mu|=2}^m \tilde{x}_\mu A_i^\mu, \quad x_j(t_0) = x_{j0}, \quad j = 1, \dots, n-2 \quad (11)$$

where $\mu = (\mu_1, \dots, \mu_{n-2})$, $|\mu| = \mu_1 + \dots + \mu_{n-2}$, $\tilde{x}_\mu = x_1^{\mu_1} \dots x_{n-2}^{\mu_{n-2}}$, $m \geq 2n$.

Let the coefficients of the polynomials in (11) be 2π -periodic functions of time and let these coefficients be defined as power polynomials of the two trigonometric functions $x_{n-1} = r \cos t$, $x_n = r \sin t$:

$$\begin{aligned} A_i^j &= a_i^j + \sum_{\substack{|k|=1 \\ m-|\mu|}}^{m-1} \xi_k A_i^{jk}, \quad \xi_k = x_{n-1}^{k_{n-1}} x_n^{k_n}, \quad k = (k_{n-1}, k_n); \quad i, j = 1, \dots, n-2 \\ A_i^\mu &= a_i^\mu + \sum_{|k|=1}^{m-1} \xi_k A_i^{\mu k}, \quad |k| = k_{n-1} + k_n \end{aligned} \quad (12)$$

For instance, it could be partial sums of Fourier series, that approximate 2π -periodic functions (12). By definition put

$$\bar{D}_1 = \{\tilde{x} = [x_1, \dots, x_{n-2}] \in \mathbb{R}^{n-2}, \quad |x_i| \leq r \leq 1, \quad i = 1, \dots, n-2\}, \quad t \geq 0$$

where \bar{D}_1 is a finite region of the state-space, where this system is defined

Substituting (12) for (11), we get the system (1)-(2) with the last two equations

$$\frac{dx_{n-1}}{dt} = -x_n, \quad \frac{dx_n}{dt} = x_{n-1}, \quad x_{n-1}(0) = r, \quad x_n(0) = 0 \quad (13)$$

We obtain the united monomials of the form $x_\nu = x_\mu \xi_k = x_1^{\mu_1} \dots x_{n-2}^{\mu_{n-2}} x_{n-1}^{\mu_{n-1}} x_n^{\mu_n} \equiv x_1^{\nu_1} \dots x_n^{\nu_n}$. Finally, using the developed method, we get the extended linear autonomous system (9). As a result, we can estimate the solution of the initial periodic system (11), (12) by the solution of the extended linear system (9), at (13). Note that two equations (13) are linear.

19.4 Examples

Consider the equation of holonomic systems with stationary nonlinear resistance:

$$\ddot{y} + 2n\dot{y} + k^2 y + a\dot{y}^3 = 0, \quad \bar{D} = [(\dot{y}, y) : |\dot{y}| \leq r, \quad y \leq r, \quad r < 1] \subset R^2,$$

or

$$\dot{x}_1 = -(2nx_1 + k^2 x_2) - ax_1^3, \quad \dot{x}_2 = x_1, \quad \bar{D} = [(x_1, x_2) : |x_{12}| \leq r] \quad (14)$$

Then $X = [x_1, \dots, x_6]$ is the extended phase row vector of the system and $X(0) = [x_{10}, \dots, x_{60}]$ is its initial phase, where

$$x_3 = x_1^3, \quad x_4 = x_1^2 x_2, \quad x_5 = x_1 x_2^2, \quad x_6 = x_2^3, \quad x_1(0) = x_{10}, \quad x_2(0) = x_{20}, \\ x_3(0) = x_{10}^3, \quad x_4(0) = x_{10}^2 x_{20}, \quad x_5(0) = x_{10} x_{20}^2, \quad x_6(0) = x_{20}^3$$

We apply the following economizations for additional monomials:

$$x_s^5 \approx \frac{5}{4} r^2 x_s^3 - \frac{5}{16} r^4 x_s, \quad x_s^4 x_j \approx r^2 x_s^2 x_j - \frac{1}{8} r^4 x_j, \quad x_s^3 x_j^2 \approx \frac{3}{4} r^2 x_s x_j^2, \quad (15)$$

where $j=1,2$. For example here is how we get the first additional equation using (14) and (15)

$$\begin{aligned}\dot{x}_3 &= -3x_1^2(2nx_1 + k^2x_2) - 3ax_1^5 = -6nx_3 - 3k^2x_4 - 3a\left(\frac{5}{4}r^2x_3 - \frac{5}{16}r^4x_1\right) \Rightarrow \\ \Rightarrow \dot{x}_3 &= XB_3, \quad B_3 = [b, 0, -c, -3k^2, 0, 0]^T, \quad b = \frac{15}{16}ar^4, \quad c = 6n + \frac{15}{4}ar^2.\end{aligned}$$

Finally, we get the extended linear system $\dot{X} = XB$, where B be as follows:

$$B = \begin{bmatrix} -2n & 1 & b & 0 & 0 & 0 \\ -k^2 & 0 & 0 & e & 0 & 0 \\ -a & 0 & -c & 1 & 0 & 0 \\ 0 & 0 & -3k^2 & -g & 2 & 0 \\ 0 & 0 & 0 & -2k^2 & -h & 3 \\ 0 & 0 & 0 & 0 & -k^2 & 0 \end{bmatrix},$$

where $e = ar^4 / 4$, $g = 4n + 2ar^2$, $h = 2n + 3ar^2 / 4$.

Now we can estimate the stability of the initial nonlinear system by the eigenvalue spectrum of the extended system, i.e. by the six roots of the characteristic equation for B . For example, for $r = 0.8$, $n = 0.1$, $k = 1$, $a = 0.8$ we get $\lambda = [-0.156 \pm 0.988i, -1.167 \pm 0.936i, -1.041 \pm 2.736i]$. The error and the initial conditins are defined as follows: $\Delta_1 = x_1 - y$, $\Delta_2 = x_2 - y$, $\dot{y}_0 = x_{10}$, $y_0 = x_{20} \Leftrightarrow \Delta_1(0) = \Delta_2(0) = 0$.

Two plots are shown in Fig. 1:1 one - of the solution of initial nonlinear system ($y(t)$ at $y(0) = 0.8$, $\dot{y}(0) = 0$, $t \in [0, 10]$) and another one - of the solution of the extended linear system $x_2(t)$ with the corresponding initial conditions. As it could be seen, these two graphs almost merged. That proofs a good accuracy of the proposed linearization method.

Now consider another example - Van der Pol equation

$$\ddot{x} + \epsilon(x^2 - 1)\dot{x} + x = 0 \quad (16)$$

Substituting $\sqrt{k}x_2$ for x and $\sqrt{k}x_1$ for \dot{x} in (16) we obtain

$$\begin{aligned}\dot{x}_1 &= \epsilon x_1 - x_2 - \mu x_1 x_2^2 \\ \dot{x}_2 &= x_1, \quad |x_1| \leq 1, \quad |x_2| \leq 1, \quad \mu = \epsilon k\end{aligned} \quad (17)$$

Let us introduce additional variables as follows

$$x_3 = x_1 x_2^2, \quad x_4 = x_1^2 x_2, \quad x_5 = x_1^3, \quad x_6 = x_2^3 \quad (18)$$

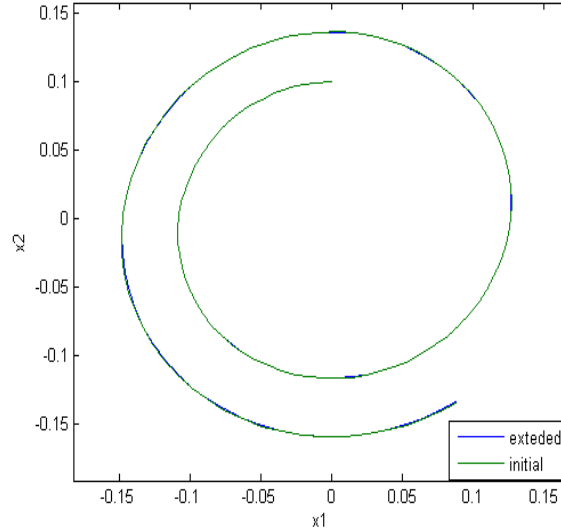


Fig. 1.2. Van der Pol equation solution and the extended linear system solution on the phase plane

Adding (18) to (17) and taking into account Chebyshev economization (19)

$$\mu x_1 x_2^4 \approx \mu x_1 (x_2^2 - \frac{1}{8}), \quad \mu x_1^2 x_2^3 \approx \frac{3}{4} \mu x_1^2 x_2, \quad \mu x_1^3 x_2^2 \approx \frac{3}{4} \mu x_1 x_2^2 \quad (19)$$

we obtain

$$\begin{aligned} \dot{x}_1 &= \epsilon x_1 - x_2 - \mu x_3, & \dot{x}_2 &= x_1, & \dot{x}_3 &= \frac{1}{8} \mu x_1 - (\mu - \epsilon) x_3 + 2x_4 - x_6 \\ \dot{x}_4 &= -2x_3 - \left(\frac{3\mu}{2} - 2\epsilon \right) x_4 + x_5, & \dot{x}_5 &= -\frac{9}{4} \mu x_3 - 3x_4 + 3\epsilon x_5, & \dot{x}_6 &= 3x_3 \end{aligned} \quad (20)$$

where $\mu = \epsilon k$. The initial conditions for this system are

$$\begin{aligned} x_1(0) &= x_{10}, & x_2(0) &= x_{20}, & x_3(0) &= x_{10} x_{20}^2 \\ x_4(0) &= x_{10}^2 x_{20}, & x_5(0) &= x_{10}^3, & x_6(0) &= x_{20}^3 \end{aligned} \quad (21)$$

The results of comparative simulation of the initial nonlinear system (16) and the extended linear system (20)-(21) at $x_0 = 0.1, \dot{x}_0 = 0, k = 1, \epsilon = 0.1$ are shown in (Fig.1.2). Two graphs almost merged in one, so we achieve a high accuracy of the approximation. Note

that for the phase trajectory in outside of the periodic cycle we need the extended linear system, constructed for the increased area of linearization.

19.5 Conclusions

In this paper the new method of making an approximately equivalent extended linear systems for autonomous nonlinear systems with polynomial structure is presented. It is based on the method of extended model. The Chebyshev economization is used to obtain higher accuracy. The example illustrates the developed method and shows good accuracy of the approximation. Using the developed method we can estimate the stability and transient time of the nonlinear system (1) by the spectrum of eigenvalues of the matrix B of the extended linear system. Reliable estimates of stability of motion can be obtained by the method constructing a linear inhomogeneous differential inequality for positive definite quadratic Lyapunov function for the extended system variables, including error estimates economization.

Acknowledgement. The research was supported by the grant of the Russian Foundation for Basic Research (RFBR).

References

1. Lyapunov AM (1992) The general problem of the stability of motion, Taylor&Francis, London
2. Babadzanjan LK. (2010) A Method of additional variables. Vestnik of St. Petersburg State University. Applied mechanics, computer science, control processes, 10: 3--11.
3. Tartakovsky VA (1950) Explicit formulas for local expansions of differential equations solutions. Reports of the Russian Academy of Sciences. 72(5): 853-856.
4. Lustig SR, Rastogi S, Wagner N (1995) Telescoping fast multipole methods using Chebyshev economization, J. of Comp. Phys., 22(2): 317-322
5. Matrosov VM (2001) Vector Lyapunov functions method: nonlinear analysis of dynamical properties. Fizmatlit, Moscow
6. Vasilyev SN, Kozlov RI, Ulyanov SA (2009) Coordinate transformations and other models of dynamic systems analysis. Proc. Institute of Mathematics and Mechanics UB RAS. Ekaterinburg, IMM UB RAS:1-18.

- 7 Melnikov VG (2005) Chebyshev economization in Poincare-Dulac transformations of nonlinear systems. *Nonlinear Analysis, Elsevier*, 63 (5-7): e1351 - e1355.
- 8 Yacine Z, Benfdila A, Djennoune S. (2008) Model order reduction of nonlinear systems via the trajectory piecewise linear approach application to transmission lines circuit, 8th WSEAS Int. Conf. AIC'08, Rhodes, Greece, August 20-22, 370-375
- 9 Boumhidi S (2002) A combined adaptive PID and Chebyshev series to control a class of MIMO non-linear systems, 2002 WSEAS Int.Conf. ISPR '02, 1271-1279
- 10 Bekiarski A, Lungu M, Lungu R, Calbureanu M. (2010) The use of the hierarchical structured dynamic inversion to the aircrafts lateral movement, 14th WSEAS International Conference on systems pp 66-71
11. Melnikov VG (2010) About root-clustering in sophisticated regions, 14th WSEAS International Conference Latest Trends on Systems, 1, pp. 297-300.
12. Melnikov VG (2010) Chebyshev economization in transformations of nonlinear systems with polynomial structure, 14th WSEAS International Conference Latest Trends on Systems, 1, pp.301-303.

Chapter 20

Power Transmission Line Magnetic Fields for different Phase Arrangements

L. Neamț¹, L. E. Petrean,¹O. Chiver,¹Z. Erdei¹

¹Electrical Engineering Department, North University of Baia Mare, 430083 – Baia Mare, Maramureș, Dr. Victor Babeș street, no. 62/A, Romania, liviu_neamt@ubm.ro

Abstract. The influence of phase-sequence orientation for transposed and untransposed phasing overhead power lines on surrounding magnetic field is studied based on Finite Element Method analysis. All available combinations for double circuit overhead power line phase arrangements were considered. The soil under the lines and a simplified model of human body complete the analyzed configurations. The magnetic field distributions and strengths and the outlined The paper presents the Finite element analysis of the magnetic field around a 110 KV, double circuit overhead power line for different phase arrangements. have been evaluated in each case, the results demonstrating that the is an important factor that influences the electromagnetic field distribution around the transmission line and some recommendations. Also some aspects concerning the human exposure are pointed, using the same computer based simulation on a simplified.

Keywords. Electromagnetic compatibility, Overhead power lines, Electromagnetic fields, Finite element method, Human exposure.

20.1 Introduction

The electromagnetic fields (EMFs) have effects on humans. At high field levels, bigger than we usually meet in the environment, the effects are well

understood and there are exposure guidelines in place to protect against them. Over the past 20 years, scientists have linked exposure to everyday levels of EMFs with various health problems, ranging from headaches to Alzheimer's disease. The most persistent of these suggestions relates to childhood leukaemia. But the evidence is not straightforward, [1].

ICNIRP 1998, [2], guidelines specify, for occupational exposures at power frequencies, the basic restriction of 10 mA/m^2 . For the general public, is applied an extra factor of 5, giving a basic restriction of 2 mA/m^2 . They also give investigation reference levels for EMFs: $500 \text{ } \mu\text{T}$ and 10 kV/m for workers and $100 \text{ } \mu\text{T}$ and 5 kV/m for the public. The fields required to produce the basic restriction are higher.

One of the most import sources of EMFs are the overhead power lines due to their high levels of current and voltages. There are a variety of ways to reduce the EMFs around lines, some imply expensive measures, but there are others costless, simple and efficient. One option is outlined in this paper, namely phase arrangements.

For the calculation of EMFs produced by overhead power lines, the Finite Element Method (FEM) was used. FEM is the most popular, precise and developed tools for numerical analysis in engineering. It provides easy way to accomplish the non linearity and the complicated structure of the materials, great accuracy of the simulation, reduced costs, and speed of analysis.

20.2 Problem formulation

In this work a double circuit line 110 KV at a load of 500 A with architecture dictated by steel pylons SN 110.252 type is considered [3]. In Fig. 1.1 is shown the geometry of the line:

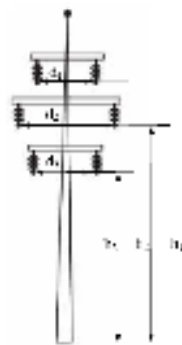


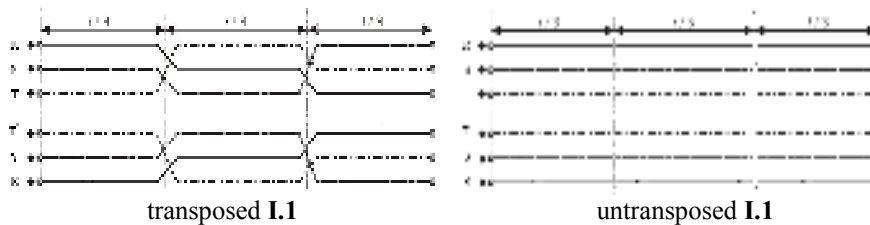
Fig. 1.1. The 110 KV double circuit overhead line geometry

The Romanian standards in overhead power lines call for transposed phasing only for long lines. This measure is due to the conflict between equal impedances (equal voltages drop) of each phase and the exploitation safety and simplicity. For double circuit lines, the transposing is applied both for the phases of each circuit and for circuits, one to another. To achieve the same impedances, the phases have to be placed alternately in each possible location on the total length of the line. The problem is more complex if the line is with double circuit. So each phase has to take hold these location but not neglecting the entire circuit mutual influences. Possible solutions imply to create 3 or more equal parts with transposed phases. The most common principle is to put one's phases of each circuit in corresponding position and the other 2 in inverse disposal. In these cases there are 3 possibilities of phase architectures, shown in Table 1.1 and named I.1, I.2, I.3. Other ways are to avoid any correspondence between phases, II.1 and II.2, or to transposing only the phases from circuits, both having the same arrangements, III. For untransposed phases it can be take any combinations corresponding to $1/3$ line length, from Table 1.1 applied to entire line.

Table 1.1. Transposing principle

	Circuit											
	1	2	1	2	1	2	1	2	1	2	1	2
1/3 line length	R	T	R	R	R	S	R	T	R	S	R	R
	S	S	S	T	S	R	S	R	S	T	S	S
	T	R	T	S	T	T	T	S	T	R	T	T
1/3 line length	T	S	T	T	T	R	T	S	T	R	T	T
	R	R	R	S	R	T	R	T	R	S	R	R
	S	T	S	R	S	S	S	R	S	T	S	S
1/3 line length	S	R	S	S	S	T	S	R	S	T	S	S
	T	T	T	S	T	S	T	S	T	R	T	T
	R	S	R	T	R	R	R	T	R	S	R	R
Name	I.1		I.2		I.3		II.1		II.2		III	

In Fig. 1.2 are shown some examples of phases transposing, e.g. I.1, III and untransposed I.1 and III:



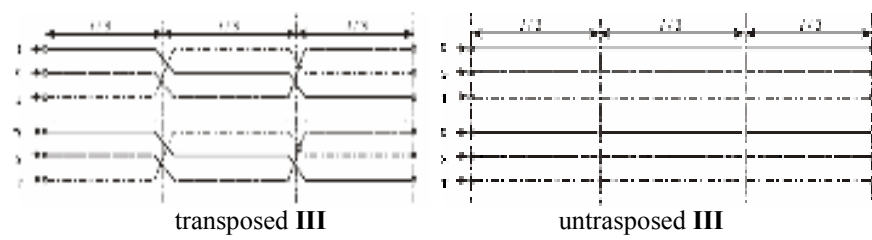


Fig. 1.2. Phase transposing examples of double circuit overhead line

20.3 Finite Element Method solutions

The Finite Element Method analyses were done using David Meeker Finite Element Method Magnetics FEMM® for 2D configurations. In FEMM® the mathematical model of time harmonic magnetic problems is defined as follows [4]: the electric field intensity, E , and the current density, J , obey the constitutive relationship:

$$J = \sigma \cdot E \quad (1.1)$$

Substituting the vector potential, A , form of magnetic flux density, B , into Faraday's Law:

$$\nabla \times E + \frac{\partial B}{\partial t} = 0 \quad (1.2)$$

yields:

$$\nabla \times E = -\nabla \times \frac{\partial A}{\partial t} \quad (1.3)$$

In the case of 2-D problems, (1.3) can be integrated to give:

$$E = -\frac{\partial A}{\partial t} - \text{grad}V \quad (1.4)$$

and the constitutive relationship, (1.1) goes to:

$$J = -\sigma \cdot \frac{\partial A}{\partial t} - \sigma \cdot \text{grad}V \quad (1.5)$$

Substituting into Maxwell-Ampere law, written in terms of vector potential:

$$\nabla \times \left(\frac{1}{\mu(B)} \nabla \times A \right) = J \quad (1.6)$$

yields the partial differential equation:

$$\nabla \times \left(\frac{1}{\mu(B)} \nabla \times A \right) = -\sigma \cdot \frac{\partial A}{\partial t} + J_{src} - \sigma \cdot gradV \quad (1.7)$$

where J_{src} represents the applied currents sources.

FEMM® considers (1.7) for the case in which the field is oscillating at one fixed frequency. For this case, a phasor transformation yields a steady-state equation that is solved for the amplitude and phase of A . This transformation is:

$$A = \text{Re}[a(\cos \omega t + j \cdot \sin \omega t)] = \text{Re}(a \cdot e^{j \cdot \omega t}) \quad (1.8)$$

Substituting into (1.7) and dividing out the complex exponential term yields the equation that FEMM® solves for harmonic magnetic problems:

$$\nabla \times \left(\frac{1}{\mu(B)} \nabla \times a \right) = -j \cdot \omega \cdot \sigma \cdot a + \bar{J}_{src} - \sigma \cdot gradV \quad (1.9)$$

All phase arrangements from Table 1.1 were considered in the 110 KV double circuit overhead line geometry, showed in Figure 1, where $d_1 = 6$ m, $d_2 = 8$ m, $d_3 = 6$ m, $h_1 = 20$ m, $h_2 = 16$ m, $h_3 = 12$ m. The 500 A current trough phases is closed to maximum load for 300/50 mm² Aluminium-steel reinforced conductor type. The load is considered symmetrically, and then the currents of each phase have the expressions:

$$\begin{aligned} i_R[A] &= 500 \cdot \sqrt{2} \\ i_S[A] &= 500 \cdot \sqrt{2} \cdot \left(-\frac{1}{2} - j \cdot \frac{\sqrt{3}}{2} \right) \\ i_T[A] &= 500 \cdot \sqrt{2} \cdot \left(-\frac{1}{2} + j \cdot \frac{\sqrt{3}}{2} \right) \end{aligned} \quad (1.10)$$

The resulted configurations including the soil under the line, having a 0.05 S/m conductivity and a simplified model of human body (a homogeneous 0.2 m radius sphere with a conductivity of 0.2 S/m [1]) placed at 1.6 m above ground are analyzed on FEMM®.

20.4 Results

The time-varying magnetic field effect on humans is caused by generates circulating tissue currents, which are basic restrictions in ICNIRP guidelines. Because these currents are very hard to be measured, the value of the magnetic fields is considered as reference levels. The resulted values of magnetic flux density around the overhead double circuit line, for all phase arrangements, corresponding with notations from Table 1, were plotted in Fig. 1.3. It is obviously that I.3 and III arrangements create more intense field under the line, than the others. Also a symmetric distribution can be observed in 4 analyzed configurations. So some explanatory inferences can be drawn, but a final clarification need values to be compared.

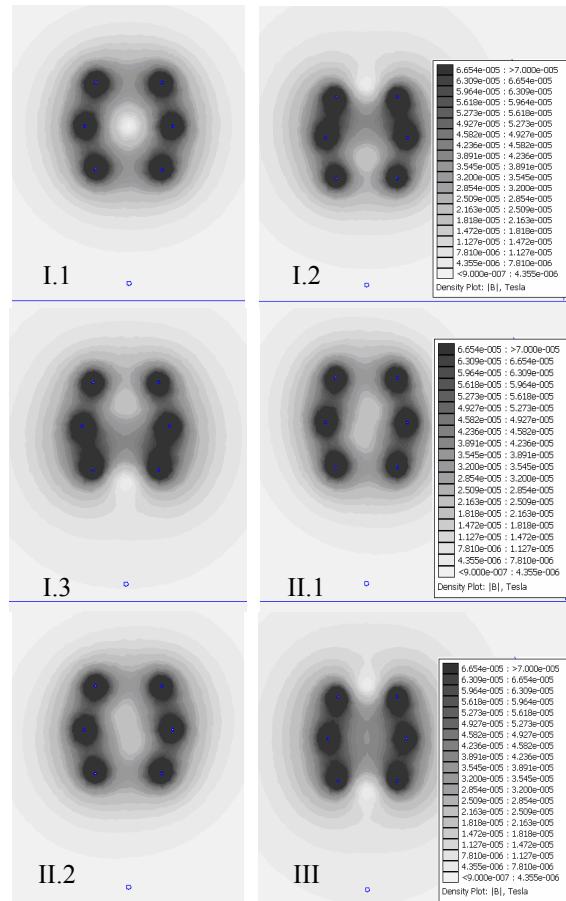
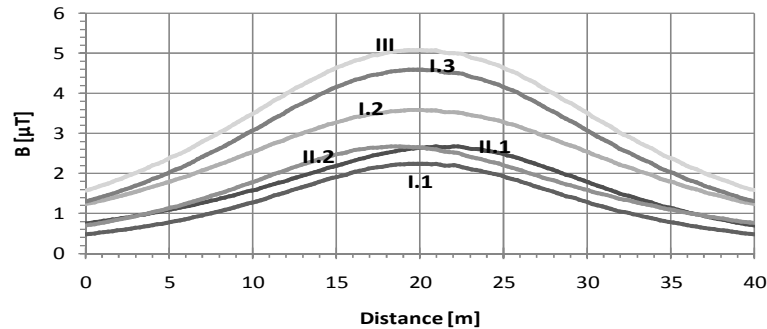
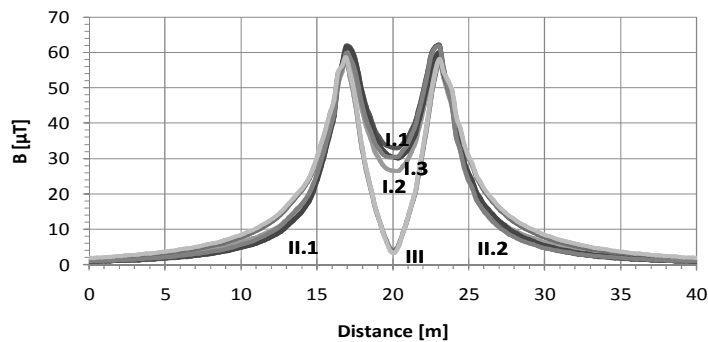


Fig. 1.3. Magnetic flux density distributions around line

For deeper investigations, the graphical variations of magnetic flux density at 1 m above the ground, Fig. 1.4, respectively at 1 m under the lowest conductor, Fig. 1.5, are very useful instruments:

**Fig. 1.4.** Magnetic flux density at 1 m above the ground

Again I.3 and III reveal that they are the most powerful sources of EMFs, followed by I.2 and II.1 and II.2. It is easily seen that without any other measure to limit the EMFs, but the phase arrangements, the field can be reduced twice.

**Fig. 1.5.** Magnetic flux density at 1 m under the lowest conductor

A proper phase arrangement do not guaranties a smaller value in conductor vicinity. It ensures a faster depreciation of field with increasing of distance from center line.

To link the field with the effects on humans, a simplified model is introduced under the overhead power line. The induced currents are computed via FEMM® in the body and also in the soil. As results were

graph the modulus of total induced current density, for all phase arrangements and having on the horizontal axis the vertical coordinate, see Fig. 1.6. The coordinate between 0 and 0.4 m means the human model, between 0.4 m and 1.8 m is the air space (where the induced current is obviously 0) and over 1.8 m is the soil.

Induced peripheral current density in human body has the maximum value for III phase arrangement, i.e. 0.06185 mA/mm². For homogeneous sphere in its centre the current is zero. The value is lower than public reference value of ICNIRP. In the soil for same power line configuration maximum induced current density is 0.79473 mA/mm². The current in the ground is evaluated until 10 m below surface where it has a diminution until 0.33941 mA/mm².

It has to be mentioned that in FEMM® all values computed in time harmonic regime are maximum values and not the rms values. Interpreting the EMFs effect on biodiversity by using these values offer a safety comparison, but in case of exceeding the imposed values, the rms can simply be computed via $\sqrt{2}$ factor

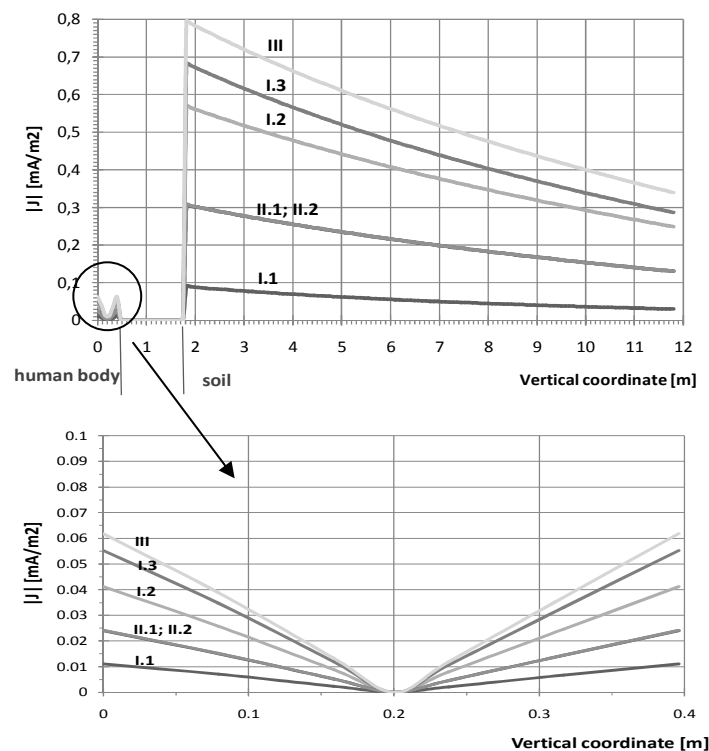


Fig. 1.6. Induced current densities in body and in soil

20.5 Conclusions

This paper has studied magnetic field distribution resulting from 6 typical phase-sequence orientations of 110 KV double circuit overhead power lines supported on steel pylons SN 110.252 type. Every line type can be considered with or without transposed phases trough entire length, see Figure 2. The 2D time harmonic FEM analyses were perform to reach the magnetic field around lines and the induced currents in a simplified human model.

All the computations were verified by implementing in FEMM® environment the problems with the results reported in [1], [5]. Also the specific values obtaining here were compared using interpolation, with results of overhead power lines EMFs evaluations [3], [6].

It was clearly outlined that the phase-sequence orientation is one of the key factors that influence magnetic field distribution in electric power transmission and distribution lines. The magnetic field strengths can be reduced either underneath the lines or far from the lines by changing phase arrangement

The magnetic flux density at height of 1 m above ground level underneath a 110 KV double circuit power lines with rms phase currents of 500 A does not exceed 5.1 μT , which is 19 times lower than ICNIRP public reference value. At 1 m below the lowest conductor the magnetic flux density is under 64 μT .

Induced peripheral current density in human body has the maximum value of 0.06185 mA/mm^2 , which is also a value below the public reference value of ICNIRP. In the soil maximum induced current density is 0.79473 mA/mm^2 .

Even the values are not exceeding the imposed values for public exposure, a costless, simple and efficient measure to reduce the double circuit overhead power lines EMFs is to use the phase arrangements (in transposed or untransposed phases mode) named in this paper I.1, see Fig. 1.7 and Table 1. This ensure the minimum field strength at public level access (≤ 2.5 m above ground) and of course in direct relation the minimum induced current in exposed body. The worst phase sequence orientation, having the same distribution of phases for both circuits, Fig. 1.7, is very used in distribution system, especially for short lines length. The ratio between magnetic flux densities corresponding to these 2 arrangements is 0.43.

Because the sources of magnetic field are the currents trough conductors and not the voltages, these results have applications at any higher and

lower voltage double circuit overhead lines. For different pylon types the same algorithm can be applied and a solution will be distinguished.

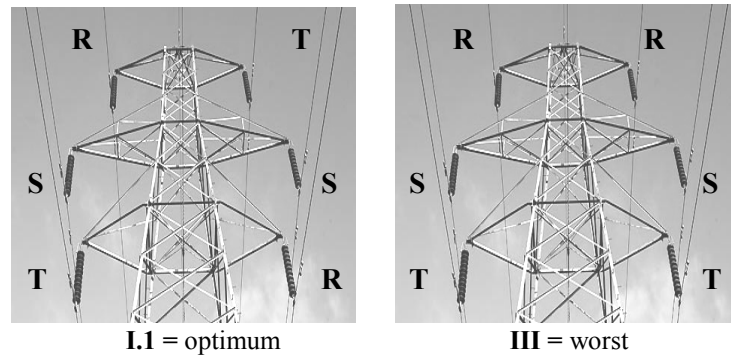


Fig. 1.6. Optimum and worst phase arrangements

References

1. <http://www.emfs.info/>
2. International Commission on Non-ionizing Radiation Protection (1998), Guidelines for limiting exposure to time-varying electric, magnetic and electromagnetic fields up to 300 GHz. Health Physics 74: 494-522
3. Popadiuc S, Popa B (2006) Romania Compliance with the EU Requirements in Environmental Impact of HV Installations. Proceedings of the 4th WSEAS Int. Conf. on Heat transfer, thermal engineering and environment, Elounda, Greece, August 21-23: 49-54
4. Meeker D (2006) Finite Element Method Magnetic, Users Manual. <http://femm.info>
5. Pao-La-Or P, Isaramongkolrak A, Kulworawanichpong T (2008) Electromagnetic Field Distribution of Phase-sequence Orientation of a Double-Circuit Power Transmission Line Based on Finite Element Method. WSEAS Transactions on Power Systems 10, 3: 653-663
6. Study of Influence of an Overhead Ground Wire on Electric Fields around the HV Power Transmission Line Using 2D and 3D Finite Element Method PAOLA-OR*, A. ISARAMONGKOLRAK, and T. KULWORAWANICHPONG WSEAS TRANSACTIONS on POWER SYSTEMS Issue 11, Volume 3, November 2008
7. Regulations for protection against electric and magnetic fields and optimum solution for the development of 150kV transmission lines in suburban regions D. TSANAKAS E. MIMOS A. TZINEVRAKIS Proceedings of the 2006 IASME/WSEAS International Conference on Energy & Environmental Systems, Chalkida, Greece, May 8-10, 2006 (pp237-242)

8. MARILENA UNGUREANU ADRIAN CONSTANTIN RUSU Electromagnetic environment of the overhead transmission lines in low and high frequency Proc. of the 3rd IASME/WSEAS Int. Conf. on Energy, Environment, Ecosystems and Sustainable Development, Agios Nikolaos, Greece, July 24-26, 2007

Chapter 21

An Ellipse Framework for Precise Iris Edge Detection

Mahboubeh Shamsi, Puteh Bt Saad, Subariah BT Ibrahim, Abdolreza Rasouli

Faculty of Computer Science & Information System, University
Technology Malaysia, Skudai, Johor, Malaysia – 81310
mahboubeshamsi@yahoo.com, drputeh@gmail.com, subariah@utm.my,
rs.reza@gmail.com

Abstract. Iris segmentation is the principal task of Iris based biometric identification systems. Always the low contrast between pupil and iris always affects on accuracy of detecting boundary between them. In order to increasing the accuracy of boundary detecting, we propose a new technique using a difference function and a factor matrix. We also enhance the technique to detect the pupil and iris as ellipse instead of circle. Experiments show that the proposed technique can segment the iris region and pupil region precisely. Based on our result, 99.34% of eyes have been segmented accurately in 1.24s averagely.

Keywords. Daugman's method, Average Square Shrinking, Difference Function, Contour Factor Matrix

21.1 Introduction

Iris is the most reliable biometric in secure transaction proposals. Iris in an eye image is situated between sclera and pupil. Before iris can be utilized for a specific application, it has to be localized first. Iris localization is a challenging ordeal due to several reasons such as occlusions to the presence of eyelids and eyebrows and also due to the uneven texture

contrast [12; 19]. The textural contrast between sclera and iris is high; conversely the textural contrast between iris and pupil is low [6-8]. The problem is further aggravated with the presence of light reflection in the pupil. Hence, an accurate algorithm is desired to detect the subtle difference between the two regions. In this paper, we propose a new algorithm to capture a maximum difference value of both, inner and outer iris boundaries.

This paper next describes related work, followed by new algorithm in section 1.3. The paper then continues on result in section 1.4. The paper ends with a conclusion in section 1.5.

21.2 Related Work

Iris is between sclera and pupil. Sclera region and iris region is easier to differentiate due to higher texture contour contrast between the two regions. However, there is a lower texture contrast between iris region and pupil region. Hence, it is difficult to automatically detect the edge between iris and pupil [9].

Daugman [9] uses a differential operator for locating the circular iris, pupil regions and the arcs of the upper and lower eyelids. The differential operator is defined as

$$\max_{(r, x_0, y_0)} \left| G_s(r) * \frac{\partial}{\partial r} \oint_{r, x_0, y_0} \frac{I(x, y)}{2\pi r} d_s \right| \quad (1)$$

Where $I(x, y)$ is the gray level of image in pixel (x, y) , $G_{s(r)}$ is Gaussian smoothing filter, s is the counter of circle represented by (x_0, y_0) as center and r as radius. The operator searches for the circular path where there is maximum change in pixel values, by varying the radius and centre x and y position of the circular contour. However, if there is noise in the image, the algorithm can fail, such as reflection. When the iris is dark and the image is under natural light, the contrast between iris and pupil is low and it makes the segmentation process more difficult [4].

An automatic segmentation algorithm based on the circular Hough transform is employed by Wildes *et al.* [3; 18]. They also make use of the parabolic Hough transform to detect the eyelids, approximating the upper and lower eyelids with parabolic arcs, which are represented as;

$$-(x - h_j)\sin\theta_j + (y - k_j)\cos\theta_j)^2 = a_j((x - h_j)\cos\theta_j + (y - k_j)\sin\theta_j) \quad (2)$$

Where a_j is the controls of the curvature, (h_j, k_j) is the peak of the

parabola and θ_j is the angle of rotation relative to the x axis. For edge detection in this method we need to choose threshold values.

Camus and Wildes [3] use similar method to Daugman's method [5]. Their algorithm, finds three circumference parameters (centre (x, y) and radius z) by maximizing the following function

$$C = \sum_{\theta=1}^n \left((n-1) \|g_{\theta,r}\| - \sum_{\phi=\theta+1}^n \|g_{\theta,r} - g_{\phi,r}\| - \frac{I_{\theta,r}}{n} \right) \quad (3)$$

Where n is the total number of directions and $I_{\theta,r}$ and $g_{\theta,r}$ are the image intensity and derivatives with respect to the radius in the polar coordinate system respectively.

The performance of algorithm degraded for noisy iris images and when there is reflection in the image.

Daugman's operator is based on the fact that the illumination difference between inside and outside of pixels in iris edge circle is maximum [16]. It means the difference values of pixel's gray level in iris circle are higher than any other circles in image. This fact is based on color of iris and color of sclera.

We are not able to calculate Daugman Operator for all feasible circles of an image. Therefore, we should restrict the space of potential circles. Many researchers assume that the center of iris is near the center of image. But in many cases, the center of iris does not fit to the center of image. Also they find a range of radius which is based on the size of image.

In our first paper, we proposed a new Average Square Shrinking (ASS) Approach [14] for initializing the range of potential centers. Therefore, we restricted the algorithm to a range of (x, y) as potential centers and a range of r as potential radiuses. The ASS approach is based on this fact that all eyes' pupils are black and therefore the center of iris must be black. Gaussian Blur or other smoothing method is applied to find dark integrated pixels in image processing methods. Therefore we used the darkest place to detect the center of the iris. We had broken the image to some small squares. Each square in a source image will be converted into one pixel in shrunken image during ASS process. The size of square and the stages of shrinking are related to shrinking (smoothing) factor S_f and number of shrinking stages N respectively. The values of all pixels inside the square will be averaged in the shrunken image. The darkest pixel (x_0, y_0) in the last image is the pupil center. The range of $[x_0 \pm S_f] \times [y_0 \pm S_f]$ will be used for potential centers of Daugman operator.

We had converted the Daugman's environmental integral to a discreet summation of a simple difference function on circle's contour to be computable by computer programming. The difference function is

$$diff(x, y) = I(x + \Delta_\alpha, y + \Delta_\alpha) - I(x - \Delta_\alpha, y - \Delta_\alpha) \quad (4)$$

We showed that the difference function also should be computed approximately, because we had only integer values for coordinates (x, y) and we could not calculate exact difference values. We had only the values of top, bottom, left, right and diagonal pixels. Due to this, we converted the Daugman operator as follows:

$$\begin{aligned} & \max_{(x_c, y_c, r)} \sum_{j=1}^{CS} diff(x_j, y_j) \quad \forall (x_c, y_c) \in \text{potential centers} \& r \in \text{potential radius} \quad (5) \\ & x_j = x_c + r \cdot \cos(\alpha_j), y_j = y_c + r \cdot \sin(\alpha_j), \alpha_j = 2\pi * j / CS \\ & diff(x_j, y_j) = I_1 + I_2 + I_3 + I_4 \\ & I_1 = (I(x_j + 1, y) - I(x_j - 1, y)) \cdot \cos(\alpha_j), I_2 = (I(x, y_j + 1) - I(x, y_j - 1)) \cdot \sin(\alpha_j) \\ & I_3 = (I(x_j + 1, y_j + 1) - I(x_j - 1, y_j - 1)) \cdot \sin(45 + \alpha_j) \\ & I_4 = (I(x_j + 1, y_j - 1) - I(x_j - 1, y_j + 1)) \cdot \cos(45 + \alpha_j) \end{aligned}$$

However, our previous method worked very well, but it did not compute the difference value correctly. The performance of our algorithm degraded for low texture contrast eye's image.

21.3 Proposed Algorithm

We use the Average Square Shrinking Process to find the potential centers and the estimated range of radius. Then, we apply the Daugman operator for improving the iris center and radius. The higher value of Daugman operator corresponded to the exact center and radius of iris. These steps will be continued iteratively on higher shrunk image to find the final center and radius.

21.3.1 Ellipse Segmentation

The experiment was conducted on many eyes' images; it has been observed that so many irises are not exactly circle. So an ellipse view will improve the accuracy of iris segmentation. We have made a slight modification to Daugman's operator to make it desirable for ellipse function. The new operator is:

$$\max_{(r_a, r_b, x_0, y_0)} \left| G_\delta(r) * \frac{\partial}{\partial r} \oint_{r_a, r_b, x_0, y_0} \frac{I(x, y)}{2\pi \sqrt{\frac{1}{2}(r_a^2 + r_b^2)}} d_s \right| \quad (6)$$

Two parameters r_a , r_b have been replaced by radius parameter r . The new operator is able to look for any feasible ellipse around the iris to find the best match.

21.3.2 Improving Difference function

The Daugman Operator tries to compute the difference gray level value between inside and outside pixels of iris circle. Daugman demonstrated that this value is the greatest one between all feasible circles of an eye image [17].

There are discrepancies in the difference function created in our first algorithm as described next. The difference function computes the difference value by subtracting the opposite pixels. These opposite pixels are right and left pixels, top and bottom and two pair of diagonal pixels. Fig 1.1.a shows the problem of this assumption that should be removed. The outside and inside pixels of difference function are colored in Fig 1.1.a. The gray rectangles are outside pixels set and hatched rectangles are inside pixels. Let us highlight the problem by an example. If you focus on Octad₅, the gray rectangles are really inside the circle, whereas they are assumed as outside pixels in difference function.

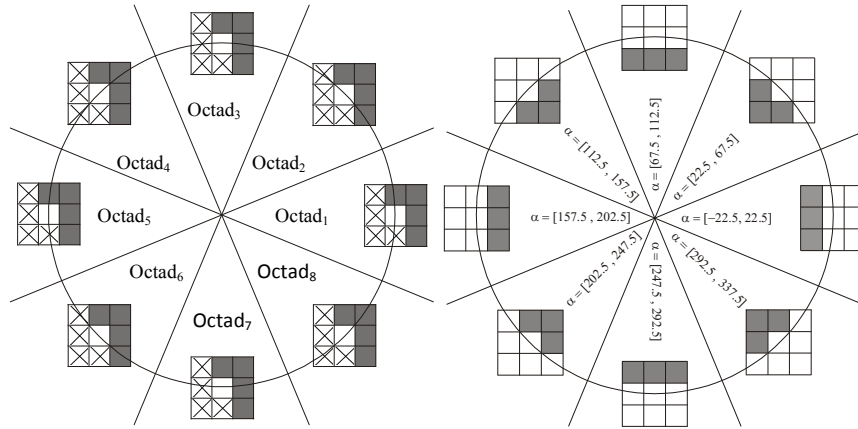


Fig. 1.1. a) The Wrong Pixel Setting. b) Correct Pixel Setting

The correct adjustment of outside and inside pixels around the circle's contour is shown in Fig 1.1.b.

We have only eight pixels around the main pixel. So we divide the circle to eight regions, because the outside and inside states of pixel in each region are the same. This fact is shown in Fig 1.1.b. Each region covers $2\pi/8$ or $\pi/4$ which is equal to 45° . First the center degree of each region is

computed, because each region corresponds to its center. The center of each region is represented as θ_j with 22.5° offset as follows:

$$\theta_j = \left(\left\lfloor \frac{\alpha_j + 22.5}{45} \right\rfloor \times 45 \right) \quad (7)$$

We proposed a new factor matrix with 3×3 dimension. The matrix is multiplied by the neighbors of the main pixel correspondingly. The factor matrix values are selected from (-1, 0, 1). The multiplication factors for outside pixel is 1 and for contour and inside pixels are 0 and -1, respectively. By adding the values of pixels after applying the factors, we obtain the real difference between outside and inside pixels. The factor matrix is:

$$M = \begin{bmatrix} \sin\left(\left(\theta_j - 45\right) \times \left(\frac{2\pi}{360}\right)\right) & \sin\left(\theta_j \times \left(\frac{2\pi}{360}\right)\right) & \sin\left(\left(\theta_j + 45\right) \times \left(\frac{2\pi}{360}\right)\right) \\ \sin\left(\left(\theta_j - 90\right) \times \left(\frac{2\pi}{360}\right)\right) & 0 & \sin\left(\left(\theta_j + 90\right) \times \left(\frac{2\pi}{360}\right)\right) \\ \sin\left(\left(\theta_j - 135\right) \times \left(\frac{2\pi}{360}\right)\right) & \sin\left(\left(-\theta_j\right) \times \left(\frac{2\pi}{360}\right)\right) & \sin\left(\left(\theta_j + 135\right) \times \left(\frac{2\pi}{360}\right)\right) \end{bmatrix} \quad (8)$$

Some examples of factor matrix are shown below.

$$\alpha = [-22.5, 22.5] \Rightarrow M = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \alpha = [22.5, 67.5] \Rightarrow M = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix}, \alpha = [67.5, 112.5] \Rightarrow M = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

The factor matrix ensures the correct setting for outside and inside pixels as shown in Fig 1.1.b. So the new difference function, $\text{diff}(x, y)$ is:

$$\max_{(x_c, y_c, r_a, r_b)} \sum_{j=1}^{CS} \text{diff}(x_j, y_j) \quad \forall (x_c, y_c) \in \text{potential centers} \& r_a, r_b \in \text{potential radius} \quad (9)$$

$$x_j = x_c + r_a \cdot \cos(\alpha_j), y_j = y_c + r_b \cdot \sin(\alpha_j), \alpha_j = 360 * j / CS, \theta_j = \left(\left\lfloor \frac{\alpha_j + 22.5}{45} \right\rfloor \times 45 \right)$$

$$\text{diff}(x_j, y_j) = \left| \sum_{\tilde{i}=1}^3 \sum_{\tilde{j}=1}^3 \tilde{I}_{\tilde{i}, \tilde{j}} \right|, \tilde{I} = M \circ \begin{bmatrix} I(x_j - 1, y_j - 1) & I(x_j, y_j - 1) & I(x_j + 1, y_j - 1) \\ I(x_j - 1, y_j) & I(x_j, y_j) & I(x_j + 1, y_j) \\ I(x_j - 1, y_j + 1) & I(x_j, y_j + 1) & I(x_j + 1, y_j + 1) \end{bmatrix}$$

We delete the number of stages parameter N that is required to be input manually. This is accomplished by adopting a threshold for the size of the last shrunk image. The image is shrunk until the size of image is less than the threshold. Consequently, the range of radius parameter can also be ignored, since in an eye image with width 10 pixels, the radius can be estimated by 2 pixels.

21.4 Experimental Result

The algorithm is developed using Delphi programming language. It is tested on 2.4 GHz CPU with Windows Vista and 2 GB Ram. Two famous iris databases have been selected for experiments. CASIA-IrisV3 [1] includes three subsets which are labeled as CASIA-IrisV3-Interval, CASIA-IrisV3-Lamp, CASIA-IrisV3-Twins. CASIA-IrisV3 contains a total of 22,051 iris images from more than 700 subjects. All iris images are 8 bit gray-level JPEG files, collected under near infrared illumination. MMU iris database [2] contributes a total number of 450 iris images. Subjects come from Asia, Middle East, Africa and Europe. Each of them contributes 5 iris images for each eye.

The maximum value of difference function for edge between sclera and iris (outer boundary) and for edge between iris and pupil (inner boundary), are computed using the previous algorithm and the current algorithm. The results are depicted in Fig 1.2, 1.3. In addition, we investigate the effect of circle sample (CS) on the maximum value of difference function. It is noticed that, as the circle sample increases, so does the maximum value of difference function. Hence, the circle sample is relatively proportional to maximum value of difference function. The results show that with the low value of CS both algorithms display a similar performance and there is no significant variation between difference values. However, by increasing CS, the distinction of difference value on the new algorithm will be extremely improved. The amount of circle contour sample (CS) versus detection accuracy and time consuming has been studied in [14; 15]. With choosing correct amount of CS, the manual inference of user decreases.

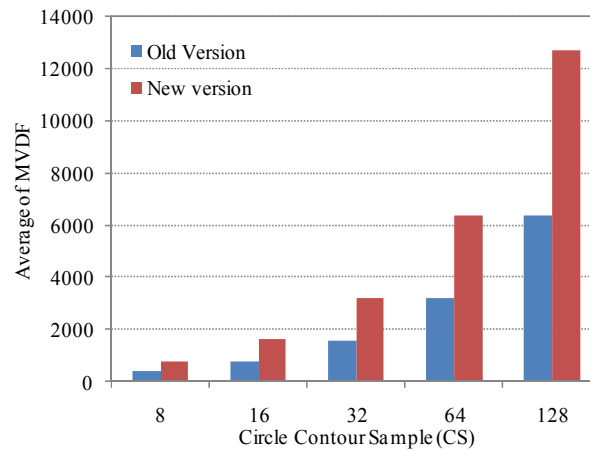


Fig. 1.2. Maximum Value Related to CS for Iris Circle

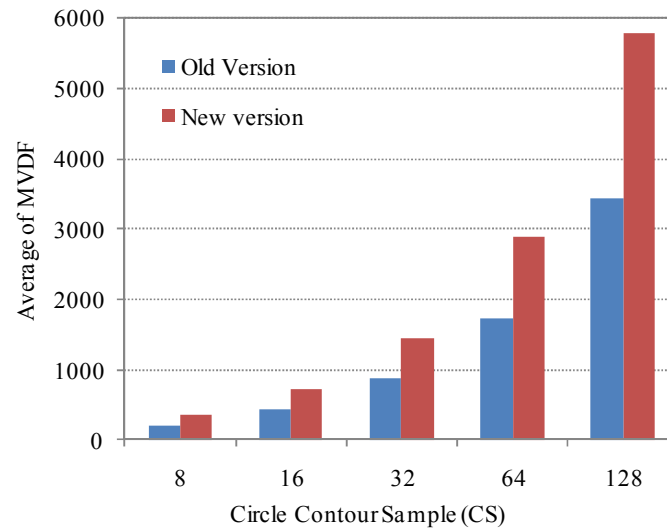


Fig. 1.3. Maximum Value Related to CS for Pupil Circle

It is noticed that the maximum difference value is doubly increased in the new algorithm for both outer and inner iris boundaries when compared to our previous algorithm. The maximum difference value is bigger for outer boundaries as compared to inner boundaries. This is due to the high texture contrast between sclera and iris. Conversely, the small difference value for inner boundary is due to the low texture contrast between iris and pupil.

The result of ellipse operator has been shown in Fig 1.4. It is disclosed that the ellipse operator redounded to better segmentation and accuracy has been extremely increased. But, due to increasing one more radius parameter, we expect that the processing time increases, respectively. So we compare the performance and processing time of our new algorithm with other algorithms obtained from [13]. Table 1 shows the boundary detection rate of various algorithms in comparison to our algorithm.

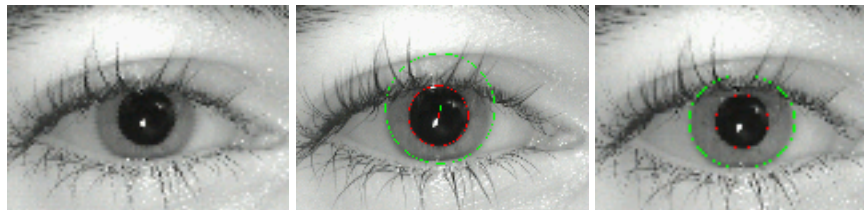


Fig. 1.4. The Original image and result of Ellipse operator vs. Circle operator

Table 1. Detection Rate of Proposed Algorithms

Algorithm	Time (s)	Detect Rate
Daugman [11]	5.36	98.58%
Daugman [10]	0.984	54.44%
Wildes [11]	6.34	99.82%
Wildes [10]	1.35	86.49%
Masek [10]	7.5	83.92%
Proposed Algorithm	1.24	99.34%

With regards to Table 1, it is noticed that our new algorithm shows the highest rate of iris boundaries detection related to its execution time in comparison to existing algorithms. Based on these findings, we are confident that the subsequent step of feature extraction will aid us to produce good quality textural features for further analysis.

21.5 Conclusion

The detection of iris boundaries is a difficult task in iris recognition systems. This is due to the low texture contrast between pupil region and iris region. In this paper, we hybrid summation function and factor matrix to be able to detect the iris boundaries. Both theoretically and our experimental results show that the proposed new algorithm strongly improved the Daugman Operator Difference Function and the detection accuracy has been significantly improved by new algorithm. The segmentation rate using this algorithm is 99.34%.

In future, we shall improve the texture contrast on iris boundaries by applying the high contrast image processing technique to obtain better results.

Acknowledgment. This project is supported by Ministry of Science, Technology and Innovation (MOSTI) through the E-Science Fund grant titled Iris Cryptography for Identity Document. We would like to thank to Madam Nazeema Abd Rahim for all her assistance in this project.

References

1. CASIA iris image database. In: <http://www.sinobiometrics.com>, vol. Chinese Academy of Sciences Institute of Automation,
2. MMU iris image database, In: <http://pesona.mmu.edu.my/~ccteo>, vol.

- Multimedia University,
3. Camus TA, Wildes R (2004) Reliable and fast eye finding in closeup images. IEEE 16th Int. Conf. on Pattern Recognition, Quebec, Canada:389-394
4. Daugman JG (2002) How iris recognition works. Proceedings of 2002 International Conference on Image Processing
5. Daugman JG, Anderson R, Hao F (2006) Combining Crypto with Biometrics Effectively. IEEE Trans. on Computers 55(9)
6. El-taweel GS, Helmy AK (2008) Efficient iris recognition scheme based on difference of filters. WSEAS Trans. Comp. Res. 3(3):152-161
7. Fenghua W, Jiuqiang H, Xianghua Y (2007) Iris recognition based on multialgorithmic fusion. WSEAS Trans. Info. Sci. and App. 4(12):1415-1421
8. Kocer HE, Novruz A (2008) An efficient iris recognition system based on modular neural networks. In: Proceedings of the 9th WSEAS International Conference on Neural Networks, vol., Sofia, Bulgaria
9. Mazur J (2007) Fast Algorithm for Iris Detection. Advances in Biometrics:858-867
10. Narote SP, Narote AS, Waghmare LM, Gaikwad AN (2006) An Automated Segmentation Method For Iris Recognition. In: TENCON 2006. 2006 IEEE Region 10 Conference, vol., pp 1-4
11. Nguyen Van H, Hakil K (2008) A Novel Circle Detection Method for Iris Segmentation. In: Proceedings of the 2008 Congress on Image and Signal Processing, Vol. 3 - Volume 03, vol. IEEE Computer Society,
12. Park J-h, Son B, Lee Y (2004) Iris Tilt Correctness in Iris Recognition System. WSEAS Trans. Info. Sci. and App. 1(1):139-144
13. Proenca H, Alexandre LA (2006) Iris segmentation methodology for non-cooperative recognition. IEEE Proceedings on Vision, Image and Signal Processing 153(2):199-205
14. Shamsi M, Saad PB, Ibrahim SB, Rasouli A (2009) Fast Algorithm for Iris Localization Using Daugman Circular Integro Differential Operator. International Conference of Soft Computing and Pattern Recognition, SoCPaR09
15. Shamsi M, Saad PB, Ibrahim SB, Rasouli A (2010) Iris Boundary Detection Using A Novel Algorithm. In: The 9th WSEAS International Conference on Applications of Computer Engineering (ACE'10) - (ISI, Elsevier Indexed), vol. WSEAS, Penang, Malaysia
16. Shamsi M, Saad PB, Rasouli A (2008) Iris Image Localization using Binning Approach. Proceedings of 2008 Student Conference on Research and Development (SCoReD 2008)
17. Shamsi M, Saad PB, Rasouli A (2008) Iris Segmentation and Normalization Approach. Journal of Information Technology (Universiti Teknologi Malaysia) 19(2)
18. Wildes R, Asmuth J, Green G, Hsu S (2004) A system for automated iris recognition. Proceedings IEEE Workshop on Applications of Computer Vision
19. Yuen CT, Rizon M, Shazri M (2009) Real-Time Detection of Face and Iris. WSEAS TRANSACTIONS on SIGNAL PROCESSING 5(6)

Chapter 22

Adaptive, Model-Following Chaos Suppression and Synchronization

J. Kabzinski

Institute of Automatic Control, Technical University of Lodz,
Stefanowskiego 18/22 90 924 Lodz, Poland, jacek.kabzinski@p.lodz.pl

Abstract. We present an innovative adaptive, model following control technique for single-input nonlinear systems and apply it to control a chaotic system with unknown parameters. The same idea is used to solve three important problems: suppression of chaos, anticontrol of chaos and chaos synchronization. We propose to use ‘approximators’ linearly parameterized by adaptive parameters. ‘Approximators’ may be constructed as a neural network, fuzzy system or may be built using a simplified description of the unknown part of the plant. We illustrate the proposed approach with control of Duffing oscillator problem.

Keywords. Adaptive control, model-following control, chaotic systems, Duffing oscillator.

22.1 Introduction

Recently, as more and more knowledge is gained about the nature of chaos and number of possible applications of chaotic system increases, scientific interests are directed to problem of controlling a chaotic system [1]. In the beginning, the goal of controlling chaos was to eliminate the harmful chaotic motion, usually by steering the trajectory of a chaotic system to a periodic one. This operation is usually called ‘suppression of chaos’ [1]. Another control problem is to synchronize two chaotic systems (‘chaos

synchronization'). It is utilized in applications in digital communication, and save data transition [3]. Since lots of applications of chaos are found significant attention has been focused on problem of chaos generation and a new notion 'chaotification' was proposed. One of possible approaches is to control a system (linear or nonlinear, chaotic or free of chaos) to follow chaotic movement of a 'master' system ('anticontrol of chaos'). All these control tasks (suppression of chaos, chaos synchronization and anticontrol of chaos) may be consider as a nonlinear tracking or model following problem, and so several nonlinear control techniques were used to find a solution [1,2]. Integrator backstepping and adaptive integrator backstepping were found as useful and effective [4,5,7].

In this contribution we propose a model-following adaptive technique based on utilization of linearly parameterized 'approximators' of plant nonlinearities. The proposed method is applied for suppression, synchronization and anticontrol of chaos in Duffing oscillator – a two-dimensional chaotic forced system wildly discussed in the literature of chaos research [1,4,5,8]. We assume that parameters of the controlled system are unknown. The results may be easily transit to another chaotic systems and used in many applications [9,10].

22.2 Adaptive model-following control

Motivated by a structure of many chaotic systems we consider a single input n -dimensional system in a canonical form described by:

$$\begin{aligned}\dot{x}_1 &= x_2 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= f(x, u) + F(x)\end{aligned}\quad (1.1)$$

where $x=[x_1, \dots, x_n]^T$ are state variables u – control input, $f(x, u)$ describes the known part of the plant and $F(x)$ – the unknown one. The desired behavior of the system is described by a model with state $x_d=[x_{d1}, \dots, x_{dn}]^T$ and control u_d :

$$\begin{aligned}\dot{x}_{d1} &= x_{d2} \\ &\vdots \\ \dot{x}_{d(n-1)} &= x_{dn} \\ \dot{x}_{dn} &= f_d(x_d, u_d)\end{aligned}\quad (1.2)$$

The control goal is to make the state variables of the plant (1.1) to follow the state of the model (1.2) in sense of upper ultimate boundness of the error. The first aim is to capture the unknown function $F(x)$ in the plant description. We accomplish it constructing a special system to approximate $F(x)$ over a compact domain.

22.2.1 Nonlinear ‘approximators’

Consider a nonlinear function F and assume that on a compact subset D_F of its domain we are able to approximate F by a model described by:

$$\hat{F} = \hat{A}_F^T \xi_F(x_F) \quad (1.3)$$

where \hat{A}_F^T is a vector of parameters to be changed during adaptation and ξ_F is a nonlinear mapping of inputs $x_F \in D_F$. At this point it is not necessary to precise number of parameters, but of course dimensions of \hat{A}_F^T and ξ_F agree. An ‘approximator’ of the presented type may be constructed as an artificial neural network, fuzzy system or may be build using a simplified description of F . The nature of the ‘approximator’ is not important as long as it possesses the structure defined by (1.3) and we may assume existence of ‘optimal’ parameters:

$$A_F^* = \arg \min \left\{ \sup_{x_F \in D_F} |F(x_F) - \hat{A}_F^T \xi_F(x_F)| \right\} \quad (1.4)$$

The model corresponding to the ‘optimal’ parameters will be denoted by:

$$F^* = A_F^{*T} \xi_F(x_F) \quad (1.5)$$

We assume that the upper bound for the modeling error exists:

$$\forall x_F \in D_F \quad |\varepsilon_F| = |F(x_F) - A_F^{*T} \xi_F(x_F)| \leq \varepsilon_{Fmax}. \quad (1.6)$$

Finally the difference between ‘optimal’ and actual parameters will be denoted by:

$$\tilde{A}_F = A_F^* - \hat{A}_F. \quad (1.7)$$

22.2.2 Model-following by ‘approximate system’

Consider a dynamical system utilizing the ‘approximator’ constructed for $F(x)$ – the unknown function in the plant description (1.1):

$$\begin{aligned}\dot{x}_{a1} &= x_{a2} \\ &\vdots \\ \dot{x}_{a(n-1)} &= x_{an} \\ \dot{x}_{an} &= f(x, u) + \hat{F}(x) + \alpha^T(x - x_a).\end{aligned}\quad (1.8)$$

The design parameters $\alpha^T = [\alpha_0, \dots, \alpha_{n-1}]$ will be defined in the next section. If we apply control law satisfying condition:

$$f(x, u) + \hat{F}(x) + \alpha^T(x - x_a) - f_d(x_d, u_d) = -\beta^T(x_a - x_d) \quad (1.9)$$

we will obtain

$$\begin{aligned}\dot{x}_{a1} - \dot{x}_{d1} &= x_{a2} - x_{d2} \\ &\vdots \\ \dot{x}_{a(n-1)} - \dot{x}_{d(n-1)} &= x_{an} - x_{dn} \\ \dot{x}_{an} - \dot{x}_{dn} &= -\beta^T(x_a - x_d),\end{aligned}\quad (1.10)$$

so design parameters $\beta^T = [\beta_0, \dots, \beta_{n-1}]$ will be used to decide speed of convergence $x_a \rightarrow x_d$, as the characteristic equation of (1.10) is $s^n + \beta_{n-1}s^{n-1} + \beta_1s + \beta_0 = 0$.

Equation (1.9) gives open expression for u if $f(x, u)$ is linear in u with invertible gain $b(x)$:

$$f(x, u) = f_1(x) + b(x)u \quad (1.11)$$

In this case:

$$u = \frac{1}{b(x)} [-f_1(x) - \hat{F}(x) - \alpha^T(x - x_a) + f_d(x_d, u_d) - \beta^T(x_a - x_d)] \quad (1.12)$$

22.2.3 Adaptation

Adaptive laws will be used to ensure that $x \rightarrow x_a$.

After denoting $e = x - x_a$, under condition (1.9) and taking into account (1.6) and (1.7) we get:

$$\frac{d}{dt} \begin{bmatrix} e_1 \\ \vdots \\ e_{n-1} \\ e_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\alpha_0 & -\alpha_1 & \cdots & -\alpha_{n-1} \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_{n-1} \\ e_n \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \tilde{A}_F^T \xi_F + \varepsilon_F \end{bmatrix} \quad (1.13)$$

Let us choose such $\alpha^T = [\alpha_0, \dots, \alpha_{n-1}]$, that $E = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ -\alpha_0 & -\alpha_1 & \cdots & -\alpha_{n-1} \end{bmatrix}$ is stable and let us solve Lyapunov equation:

$$PE + E^T P = -Q \quad (1.14)$$

with a positive definite matrix Q . As E is stable P is positive definite. We

choose Lyapunov function:

$$V = \frac{1}{2} [e^T P e + \tilde{A}_F^T \Gamma^{-1} \tilde{A}_F] \quad (1.15)$$

where Γ^{-1} is positive definite design parameter. After calculation of Lyapunov function system derivative we get:

$$\dot{V} = -\frac{1}{2} e^T Q e + (e^T p_n) (\tilde{A}_F^T \xi_F + \varepsilon_F) + \tilde{A}_F^T \Gamma^{-1} \frac{d}{dt} \tilde{A}_F \quad (1.16)$$

where p_n is the last column of P . Choosing adaptation law

$$\frac{d}{dt} \tilde{A}_F = -\Gamma [(e^T p_n) \xi_F - \sigma \hat{A}_F] \Leftrightarrow \frac{d}{dt} \hat{A}_F = \Gamma [(e^T p_n) \xi_F - \sigma \hat{A}_F] \quad (1.17)$$

with a small positive σ we are able to prove that \dot{V} is negative outside the set

$$\|e\|^2 < \varrho = \frac{1/2 \|p_n\|^2 \varepsilon_{Fmax}^2 + \sigma \|A_F^*\|^2}{\lambda_{min}(Q) - 1/2}, \quad (1.18)$$

where $\lambda_{\min}(Q)$ is the minimal eigenvalue of matrix Q and so the bound q decreases with increasing eigenvalues of Q and increases with growing ε_{Fmax} .

According to the well known generalization of Lyapunov theorem due to LaSalle and Lefschetz [6] the tracking errors e_1 and e_2 are uniformly ultimately bounded. Adaptation laws (1.17) describe a response of stable linear system to a bounded input, so adaptive parameters are bounded.

We are able to modify the tracking errors by proper choice of design parameters. Adaptation speed is influenced mostly by Γ and final parameters value by σ .

We have to stress that the number of adaptive parameters is not decided yet. If approximators are more accurate the tracking error bound will be smaller and the control effort will be lower. Increasing the number of adaptive parameters results usually in improvement of model accuracy, but each parameter means one more differential adaptive law to be solved on-line, so we should compromise between these criteria. Sometime, if fast adaptation is possible just one adaptive parameter in a model is enough.

22.3 Duffing oscillator

We consider a second-order chaotic system described by the Duffing's equation:

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -px_2 - p_1x_1 - p_2x_1^3 + q\cos(\omega t) + u\end{aligned}\tag{1.19}$$

where t is time, ω - external force frequency, q - it's amplitude, p, p_1, p_2 are real constants. The control variable is u . For $p_1 > 0$ Duffing oscillator can be interpreted as a forced oscillator with a spring, whose restoring force is $R = -p_1x_1 - p_2x_1^3$. When $p_2 > 0$, we have a 'hardening spring', and when $p_2 < 0$ - a 'softening spring' although this interpretation is valid only for small x . For $p_1 < 0$, the Duffing oscillator describes the dynamics of a point mass in a double well potential, and it can be a model of a periodically forced steel beam deflected toward the two magnets. Solutions of (1.19) may display complex phenomena, including various limit cycles and chaotic behaviors. The uncontrolled system has different chaotic trajectories for different values of q , as it is demonstrated in fig. 1.1 and 1.2.

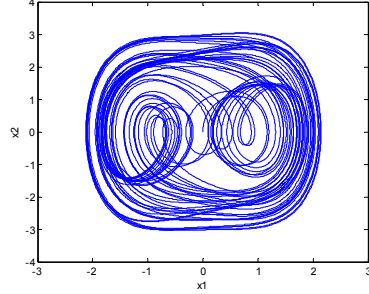


Fig. 1.1. Duffing oscillator phase-plot. Chaotic behavior, $p=0.4$, $p_1=-1.1$, $p_2=1$, $q=2.1$, $\omega=1.8$, $u=0$

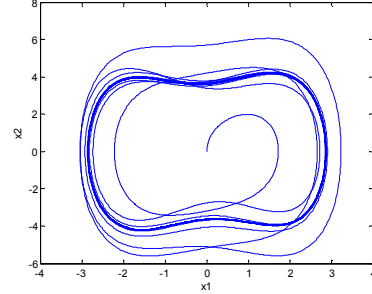


Fig. 1.2. Duffing oscillator phase-plot. Limit cycle, $p=0.4$, $p_1=-1.1$, $p_2=1$, $q=4$, $\omega=1.8$, $u=0$

For numerical experiments let us consider Duffing oscillator with parameters $p=0.4$, $p_1=-1.1$, $p_2=1$, $q=2.1$, $\omega=1.8$. The uncontrolled system demonstrates chaotic behavior (fig. 1.1). We assume that all system parameters are not known exactly. The unknown part of the plant is described by:

$$F(x) = -px_2 - p_1x_1 - p_2 \quad (1.20)$$

while the known one is just $f(x, u) = u$. We consider two possible approximators for F . The first one with one adaptive parameter:

$$\hat{F} = \hat{A}_F^T \xi_F \quad \hat{A}_F^T = a \quad (1.21)$$

$$\xi_F = -p_0x_2 - p_{10}x_1 - p_{20}x_1^3 + q_0 \cos(\omega t)$$

and the second with four adaptive parameters:

$$\hat{F} = \hat{A}_F^T \xi_F \quad \hat{A}_F^T = [a_1 \quad \dots \quad a_4] \quad (1.22)$$

$$\xi_F = [-p_0x_2, -p_{10}x_1, -p_{20}x_1^3, q_0 \cos(\omega t)]$$

where $*_0$ denotes nominal, known value of the parameter which differs up to 20% from the actual one.

During all numerical experiments parameters β_0, β_1 responsible for speed of convergence of $x_a \rightarrow x_d$ were chosen to place eigenvalues of linear system described in equation (1.10) in -5 and -8. Parameters α_0, α_1 responsible for speed of convergence of $x \rightarrow x_a$ were chosen to place eigenvalues of linear system described in equation (1.13) (matrix E) in -10 and -12. Matrix Q in Lyapunov equation was $Q = 10 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The remaining parameters were $\sigma = 0.01$ and $\Gamma = 1$ or 5.

The first control task is to synchronize the plant output x with a limit cycle described by:

$$\begin{aligned}\dot{x}_{1d} &= x_{2d} \\ \dot{x}_{2d} &= f_d(x_d, u_d) = -2.5^2 x_{1d}\end{aligned}\quad (1.23)$$

Tracking error is kept into the bound, which may be controlled by design parameters. All adaptive parameters remain bounded. The number of adaptive parameters is not a critical factor for tracking accuracy – using one parameter we were able to obtain better results than for four parameters. System performance is illustrated in fig. 1.3-1.5.

During the second test Duffing oscillator with parameters $p=0.4$, $p_1=-1.1$, $p_2=1$, $q=4$, $\omega=1.8$, (slave), which uncontrolled converges to a limit cycle (fig. 1.2) will be forced to follow (master) chaotic Duffing system with parameters $p=0.4$, $p_1=-1.1$, $p_2=1$, $q=2.1$, $\omega=1.8$, $u=0$, (fig. 1.1) – it is chaos anticontrol problem. Control performance is illustrated in fig. 1.6,1.7,1.8. Again we notice that using one parameter we were able to obtain better or similar results than applying four parameters.

In the third experiment two first state variables of 3-dimensional chaotic Amedeo system (master):

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= -a_0 x_1 - a_1 x_2 - a_2 x_3 + a_3 x_3^3\end{aligned}\quad (1.25)$$

with parameters $a_0=-5.5$, $a_1=3.5$, $a_2=1$, $a_3=-1$, $x_1(0)=3$ will generate desired trajectories for chaotic Duffing system with parameters $p=0.4$, $p_1=-1.1$, $p_2=1$, $q=2.1$, $\omega=1.8$, $u=0$, $b=0.8(1+0.2\cos 2t)$ (slave). Results are presented in fig. 1.9,1.10. Again tracking performance is satisfactory and adaptive parameters remain bounded. In this case tuning the system with four adaptive parameters was significantly more difficult than with one adaptive parameter.

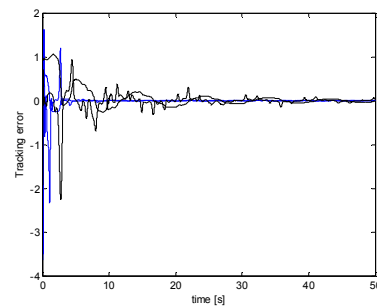
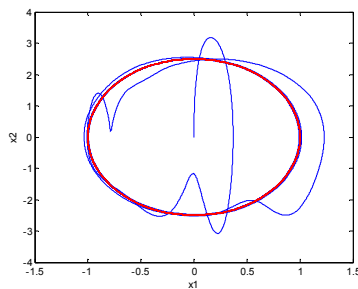


Fig. 1.3. Phase plot of controlled Duffing **Fig. 1.4.** Tracking error during suppress-

oscillator. One adaptive parameter. *Red* – desired, *blue* – Duffing oscillator.

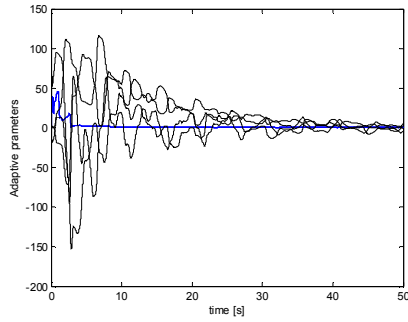


Fig. 1.5. Adaptive parameters during suppression of chaos. *Blue* – 1 adaptive parameter, *black* – 4 adaptive parameters.

sion of chaos. *Blue* – 1 adaptive parameter, *black* – 4 adaptive parameters.

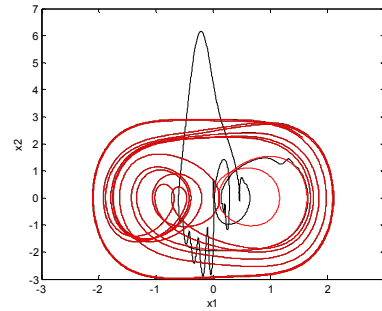


Fig. 1.6. Chaos anti-control of two Duffing oscillators. One adaptive parameter. *Red* – master, *black* – slave

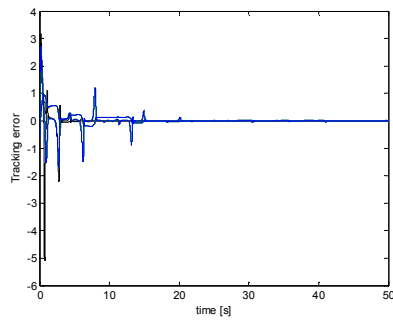


Fig. 1.7. Tracking error during chaos anti-control. *Blue* – 4 adaptive parameters, *black* – 1 adaptive parameter.

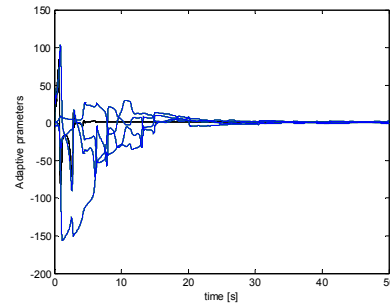


Fig. 1.8. Adaptive parameters during chaos anti-control. *Blue* – 4 adaptive parameters, *black* – 1 adaptive parameter.

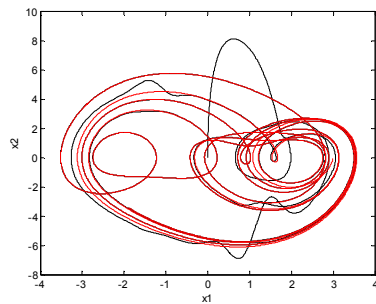


Fig. 1.9. Synchronization of Ameodo (*red*) and Duffing (*black*) chaotic systems. One adaptive parameter

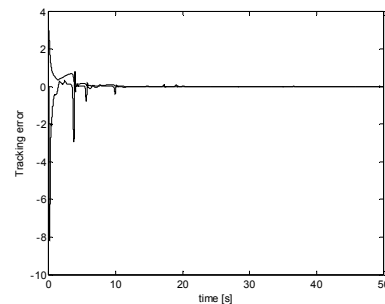


Fig. 1.10. Tracking error during synchronization of Ameodo and Duffing chaotic systems. One adaptive parameter.

22.4 Conclusions

We present an innovative adaptive, model following control scheme and apply it to control a chaotic system with unknown parameters. Although presented controller derivation techniques are general or may be easily generalized to a wide class of nonlinear plants we illustrate them with control of Duffing oscillator. The same approach is used to solve three important problems:

- chaos stabilization to a given orbit or point (suppression of chaos),
- anticontrol of chaos,
- chaos synchronization.

The main idea of adaptive control was to use ‘approximators’ linearly parameterized by adaptive parameters. Such approximator may be constructed as an artificial neural network, fuzzy system or may be build using a simplified description of the unknown part of the plant as it was done in this study.

It is important that number of adaptive parameters is not stiffly imposed by the design procedure and for all tested control problems one adaptive parameter for one approximate model was sufficient. It makes the controller simpler and easier applicable.

Comparing the presented approach with the adaptive backstepping technique developed in [7] we find several similarities, although backstepping controllers were slightly easier to tune.

References

1. Huaguang Zhang, Derrong Liu, Zhilian Wang (2009) Controlling Chaos – Supression, Synchronization and Chaotification, Springer Verlag
2. Anaya GF, Lucatero CR, Flores-Godoy JJ, Miranda-Reyes C, (2010) Preservation of Synchronization in Dynamical Systems via Lyapunov Methods, WSEAS Trans. on Circuits and Systems, vol. 9, pp 248-257
3. Lin CF, Chung CH, Chen ZL, Song CJ, Wang ZX, (2008) A Chaos-based Unequal Encryption Mechanism in Wireless Telemedicine with Error Decryption, WSEAS Trans. on Systems, vol. 7 pp 49-55
4. Harb AM, Zaher AA, Al-Qaisia AA, (2007) Recursive Backstepping Control of Chaotic Duffing Oscillators, Chaos, Solitons & Fractals, 34, pp 639-645
5. Yeoung-Jeu Sun (2009) Robust Tracking Control of Uncertain Duffing-Holmes Control Systems, Chaos, Solitons & Fractals, 40, pp1282-1287
6. Krstic M, Kanellakopoulos I, Kokotovic PV, (1995) Nonlinear and Adaptive Control Design, New York: Wiley

7. Kabziński J, (2010) Adaptive Control of Duffing Oscillator with Unknown Input Gain, Proc. of 15th International Conference on Methods and Models in Automation and Robotics
8. Jeng JD, Kang Y, Chang YP, (2008) An Alternative Poincaré Section for Steady-State Responses and Bifurcations of a Duffing-Van der Pol Oscillator, WSEAS Trans. on Systems, vol. 7 pp 748-757
9. Jabli N, Khammari H, Minouni MF, Dhifaoui R, (2010) Bifurcation and chaos phenomena appearing in induction motor under variation of PI controller parameters, WSEAS Trans. on Systems, vol. 9 pp 784-793
10. Negoitescu D, Lascu D, Popescu V, Ivan C, (2008) Bifurcation and Chaotic Aspects in Peak Current Controlled Buck-Boost Converters, WSEAS Trans. on Circuits and Systems, vol. 7, pp 688-697

Author Index

A

Alexandrov V. V., 53
Alexandrova T. B., 53
Ali Ahmed A., 163
Ariff Baharudin M., 163
Audenaert A., 21

B.

Bessmeltsev M. V., 32
Bognar G., 143

C.

Castillo-Quiroz G., 53
Chavez-Bosquez O., 132
Chicurel-Uziel E., 184
Chiver O., 214
Choudhuri A., 109
Ciupala L., 174
Ciurea E., 174

D.

Dullaert W., 21

E.

Erdei Z., 214

F.

Fisal N., 163

G.

Ghandriz R., 71
Gomez Ramos J. L., 132
Goudarzi M., 71

H.

Hafizah S., 163
Hajek P., 88

I.

Iqbal Faruque M. R., 61
Islam M. T., 61

K.

Kabzinski J., 235
Kamilah S., 163
Kuleshov A., 99

M.

Maarof M. A., 194
McTavish D. J., 109
Melnikov V. G., 122, 205
Misran N., 61

N.

Neamt L., 214
Neubauer A., 10

O.

Okouneva G., 109
Olej V., 88
Ortega A., 53

P.

Parpalea M. M., 42
Parpalea M. N., 42
Peremans H., 21
Petrean L. E., 214
Pozos-Parra P., 132

R.

Rasouli A., 194, 225
Reniers G., 21

S.

Saad P. B., 225
Saleem K., 163
Sangeorzan L., 42

Shamsi M., 194, 225

Shulenina N., 53

Sidorenko G., 53

Sohrab S. H., 143

Soto E., 53

Strbac P., 153

Subariah I. B., 225

T.

Tuba M., 153

V.

Vega R., 53