RECENT ADVANCES in MATHEMATICS

Proceedings of the 2015 International Conference on Pure Mathematics, Applied Mathematics and Computational Methods (PMAMCM 2015)

Zakynthos Island, Greece
July 16-20, 2015

RECENT ADVANCES in MATHEMATICS

Proceedings of the 2015 International Conference on Pure Mathematics, Applied Mathematics and Computational Methods (PMAMCM 2015)

Zakynthos Island, Greece
July 16-20, 2015

All papers of the present volume were peer reviewed by no less than two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.

Series: Mathematics and Computers in Science and Engineering Series | 48

ISSN: 2227-4588
ISBN: 978-1-61804-323-8

RECENT ADVANCES in MATHEMATICS

Proceedings of the 2015 International Conference on Pure Mathematics, Applied Mathematics and Computational Methods (PMAMCM 2015)

Zakynthos Island, Greece
July 16-20, 2015

Organizing Committee

Editor:
Prof. Imre J. Rudas, Obuda University, Hungary

Associate Editor:
Prof. Valery. V. Kozlov

Program Committee:
Prof. Ferhan M. Atici, Western KentuckyUniversity, Bowling Green, KY 42101, USA
Prof. Ravi P. Agarwal, Texas A&M University - Kingsville, Kingsville, TX, USA
Prof. Martin Bohner, Missouri University of Science and Technology, Rolla, Missouri, USA
Prof. Dashan Fan, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
Prof. Paolo Marcellini. University of Firenze, Firenze, Italy
Prof. Xiaodong Yan, University of Connecticut, Connecticut, USA
Prof. Ming Mei, McGill University, Montreal, Quebec, Canada
Prof. Enrique Llorens, University of Valencia, Valencia, Spain
Prof. Yuriy V. Rogovchenko, University of Agder, Kristiansand, Norway
Prof. Yong Hong Wu, Curtin University of Technology, Perth, WA, Australia
Prof. Angelo Favini, University of Bologna, Bologna, Italy
Prof. Andrew Pickering, Universidad Rey Juan Carlos, Mostoles, Madrid, Spain
Prof. Guozhen Lu, Wayne state university, Detroit, MI 48202, USA
Prof. Gerd Teschke, Hochschule Neubrandenburg - University of Applied Sciences, Germany
Prof. Michel Chipot, University of Zurich, Switzerland
Prof. Juan Carlos Cortes Lopez, Universidad Politecnica de Valencia, Spain
Prof. Julian Lopez-Gomez, Universitad Complutense de Madrid, Madrid, Spain
Prof. Jozef Banas, Rzeszow University of Technology, Rzeszow, Poland
Prof. Ivan G. Avramidi, New Mexico Tech, Socorro, New Mexico, USA
Prof. Kevin R. Payne, Universita' degli Studi di Milano, Milan, Italy
Prof. Juan Pablo Rincon-Zapatero, Universidad Carlos III De Madrid, Madrid, Spain
Prof. Valery Y. Glizer, ORT Braude College, Karmiel, Israel
Prof. Norio Yoshida, University of Toyama, Toyama, Japan
Prof. Feliz Minhos, Universidade de Evora, Evora, Portugal
Prof. Mihai Mihailescu, University of Craiova, Craiova, Romania
Prof. Lucas Jodar, Universitat Politecnica de Valencia, Valencia, Spain
Prof. Dumitru Baleanu, Cankaya University, Ankara, Turkey
Prof. Jianming Zhan, Hubei University for Nationalities, Enshi, Hubei Province, China
Prof. Zhenya Yan, Institute of Systems Science, AMSS, Chinese Academy of Sciences, Beijing, China
Prof. Nasser-Eddine Mohamed Ali Tatar, King Fahd University of Petroleum and Mineral, Dhahran, S Arabia
Prof. Jianqing Chen, Fujian Normal University, Cangshan, Fuzhou, Fujian, China
Prof. Josef Diblik, Brno University of Technology, Brno, Czech Republic
Prof. Stanislaw Migorski, Jagiellonian University in Krakow, Krakow, Poland
Prof. Qing-Wen Wang, Shanghai University, Shanghai, China
Prof. Luis Castro, University of Aveiro, Aveiro, Portugal
Prof. Alberto Fiorenza, Universita' di Napoli "Federico II", Napoli (Naples), Italy
Prof. Patricia J. Y. Wong, Nanyang Technological University, Singapore
Prof. Salvatore A. Marano, Universita degli Studi di Catania, Catania, Italy
Prof. Sung Guen Kim, Kyungpook National University, Daegu, South Korea
Prof. Maria Alessandra Ragusa, Universita di Catania, Catania, Italy
Prof. Gerassimos Barbatis, University of Athens, Athens, Greece
Prof. Jinde Cao, Distinguished Prof., Southeast University, Nanjing 210096, China
Prof. Kailash C. Patidar, University of the Western Cape, 7535 Bellville, South Africa
Prof. Mitsuharu Otani, Waseda University, Japan
Prof. Luigi Rodino, University of Torino, Torino, Italy

Additional Reviewers

Bazil Taha Ahmed	Universidad Autonoma de Madrid, Spain
James Vance	The University of Virginia's College at Wise, VA, USA
Sorinel Oprisan	College of Charleston, CA, USA
M. Javed Khan	Tuskegee University, AL, USA
Jon Burley	Michigan State University, MI, USA
Xiang Bai	Huazhong University of Science and Technology, China
Hessam Ghasemnejad	Kingston University London, UK
Angel F. Tenorio	Universidad Pablo de Olavide, Spain
Yamagishi Hiromitsu	Ehime University, Japan
Imre Rudas	Obuda University, Budapest, Hungary
Takuya Yamano	Kanagawa University, Japan
Abelha Antonio	Universidade do Minho, Portugal
Andrey Dmitriev	Russian Academy of Sciences, Russia
Valeri Mladenov	Technical University of Sofia, Bulgaria
Francesco Zirilli	Sapienza Universita di Roma, Italy
Ole Christian Boe	Norwegian Military Academy, Norway
Masaji Tanaka	Okayama University of Science, Japan
Jose Flores	The University of South Dakota, SD, USA
Kazuhiko Natori	Toho University, Japan
Matthias Buyle	Artesis Hogeschool Antwerpen, Belgium
Frederic Kuznik	National Institute of Applied Sciences, Lyon, France
Minhui Yan	Shanghai Maritime University, China
Eleazar Jimenez Serrano	Kyushu University, Japan
Konstantin Volkov	Kingston University London, UK
Miguel Carriegos	Universidad de Leon, Spain
Zhong-Jie Han	Tianjin University, China
Francesco Rotondo	Polytechnic of Bari University, Italy
George Barreto	Pontificia Universidad Javeriana, Colombia
Moran Wang	Tsinghua University, China
Alejandro Fuentes-Penna	Universidad Autónoma del Estado de Hidalgo, Mexico
Shinji Osada	Gifu University School of Medicine, Japan
Kei Eguchi	Fukuoka Institute of Technology, Japan
Philippe Dondon	Institut polytechnique de Bordeaux, France
Dmitrijs Serdjuks	Riga Technical University, Latvia
Deolinda Rasteiro	Coimbra Institute of Engineering, Portugal
Stavros Ponis	National Technical University of Athens, Greece
Tetsuya Shimamura	Saitama University, Japan
João Bastos	Instituto Superior de Engenharia do Porto, Portugal
Genqi Xu	Tianjin University, China
Santoso Wibowo	CQ University, Australia
Tetsuya Yoshida	Hokkaido University, Japan
José Carlos Metrôlho	Instituto Politecnico de Castelo Branco, Portugal

Table of Contents

Plenary Lecture 1

Evolutionary Algorithms - Cybernetic Overview

Professor Ivan Zelinka
Faculty of Electrical Engineering and Computer Science
Technical University of Ostrava (VSB-TU)
Czech Republic
E-mail: ivan.zelinka@vsb.cz

Abstract: Proposed tutorial is focused on mutual intersection of two interesting fields of research i.e. evolutionary algorithms and complex system dynamics and structure. It discusses recent progress on evolutionary algorithms that can be considered like a dynamical complex system with inherent nonlinear dynamics and feedback loop. This dynamics can generate different kind of behavior including chaotic one and can be visualized as a complex geometrical structure. Basics of deterministic chaos will be explained in order to better understand proposed topic, such as universal features of that kind of behavior are explained, quantifying chaotic, period doubling, intermittence, chaotic transients and crises.

With conjunction on this will be demonstrated fruitful intersection of complex dynamics, structures and evolutionary algorithms, that are discussed from a few points of views (chaos control, chaotic behavior of evolutionary dynamics...) including evolutionary dynamics conversion into complex network that can be analyzed and controlled by means of classical as well as modern evolutionary methods.

Further will be explained and demonstrated that evolutionary algorithm can be understand also as a nonlinear feedback loop system that can be controlled and thus performance of the evolutionary algorithm can be controlled too.

Brief Biography of the Speaker: Ivan Zelinka (born in 1965, ivanzelinka.eu) is currently associated with the Technical University of Ostrava (VSB-TU), Faculty of Electrical Engineering and Computer Science. He graduated consequently at the Technical University in Brno (1995 - MSc.), UTB in Zlin (2001 - Ph.D.) and again at Technical University in Brno (2004 - Assoc. Prof.) and VSB-TU (2010 - Professor).

Prof. Zelinka is responsible supervisor of grant research of Czech grant agency GA?R named a) Highly Scalable Parallel and Distributed Methods of Data Processing in E-science (focused on astroinformatics), b) Softcomputing methods in control, c) Control Algorithm Design by Means of Evolutionary Approach, and co-supervisor of grant FRV? - Laboratory of parallel computing. He was also working on numerous grants and two EU projects as member of team (FP5 - RESTORM) and supervisor (FP7 - PROMOEVO) of the Czech team. He is also head of research team NAVY http://navy.cs.vsb.cz/.

Prof. Zelinka was awarded by Siemens Award for his Ph.D. thesis, as well as by journal Software news for his book about artificial intelligence. He is a member of the British Computer Society, Machine Intelligence Research Labs (MIR Labs - http://www.mirlabs.org/czech.php), IEEE (committee of Czech section of Computational Intelligence), a few international program committees of various conferences, and three international journals. He is also the founder and editor-in-chief of a new book series entitled Emergence, Complexity and Computation (Springer series 10624, see also www.ecc-book.eu).

A priori hybrid mesh adaptation for turbo-machinery CFD

François Guibault
École Polytechnique de Montréal
Montréal, Canada, H3R 2A5
Email: francois.guibault@polymtl.ca

Abstract—This paper presents a fully automatic approach to generate anisotropic adapted hybrid meshes for high Reynolds number, turbulent flow simulations in radial turbo-machinery cascades. The meshes are composed of hexahedral mesh layers near boundaries, combined to prisms in the remainder of the computational domain. A Riemannian metric is constructed to control element size and orientation in the domain, based on distance from boundaries and obstacles. The boundary layer mesh thickness is automatically adjusted to account for the presence of neighbouring objects and domain boundaries, and geometric and topological mesh modification operations are used to control element density, shape and orientation. This mesh generation methodology produces meshes that suitable for advanced computational fluid dynamics simulation.

Keywords—mesh adaptation, CFD, Riemannian metric, high Reynolds-number flow, anisotropic mesh, turbo-machinery

I. Introduction

Mesh generation still constitutes one of the most important and time consuming steps in simulation-based engineering, and particularly in computational fluid dynamics (CFD) simulations. The density and quality of the meshes largely determine the accuracy of the numerical methods, both in finite-volume and finite element methods [1]. This is particularly the case for high Reynolds-number, turbulent flow simulations, where element quality and grading directly influence simulation accuracy. Mesh generation also constitutes one of the most human-intensive tasks in the simulation process [2]. There is therefore a need to provide CFD analysts with more automatic and robust mesh generation approaches that allow to automatically cope with geometric domain complexities and construct meshes with guaranteed quality.

This paper presents an automatic and robust approach to construct high quality meshes for turbo-machinery flow simulation applications. The proposed approach is based on the combination of automatically determined geometric domain characteristics and user-specified concentration functions, which are merged in a single mesh control function expressed as a metric tensor. This metric tensor, defined on an initial background mesh, is used to iteratively modify the mesh using geometric and topological modification operations so that it correspond to the metric specification [3]. This paper is structured as follows. First the overall mesh generation process in described, and the initial mesh generation method is introduced. The formal definition of the metric tensor is then given. The algorithms used to determine a metric, based on geometric domain characteristics and user specifications, is then presented. The next section introduces the mesh adaptation operations used to iteratively modify the mesh. The paper then presents some results and conclusions.

II. Global mesh generation process

Algorithm 1 presents the global mesh generation process. Each step of this algorithm is described in a subsequent section.

Algorithm 1 Global mesh generation process.
Input: set of blade profiles Γ_i, blade angles α_i, inlet radius R_i, outlet radius R_o, machine pitch θ_p.
Output: adapted 3D mesh

```
1: Generate initial mesh
   A: Construct a 2D boundary representation of the computational domain
   B: Decompose the computational domain into zones
   C: Construct an initial in each zone
2: Construct a metric tensor
3: Adapt the unstructured mesh
4: Extrude the hybrid mesh
```

III. Initial mesh generation

The first step in the global mesh generation process consists in generating an initial hybrid mesh which will then be adapted according to user specifications and geometric domain characteristics.

A. Computational domain representation

The initial mesh generation process starts by automatically reconstructing a computational domain based on the description of blade profiles, blade angles and machine pitch.

In order to reduce computational cost, only a fraction of the total 360° domain is simulated. The fraction of the domain is determined based on the machine pitch, provided as an input to the mesh generation process. Figure 1 illustrates a 2-stage cascade typical of hydraulic turbo-machines. Inlet and outlet boundary curves are automatically determined based on inlet radius R_i and outlet radius R_o. Periodic boundary curves are automatically computed based on the position and opening angle of the each blade, so that the periodic boundary lie the furthest from all blades, while maintaining periodicity and a fixed pitch angle θ_p. Figure 2 illustrates the computation

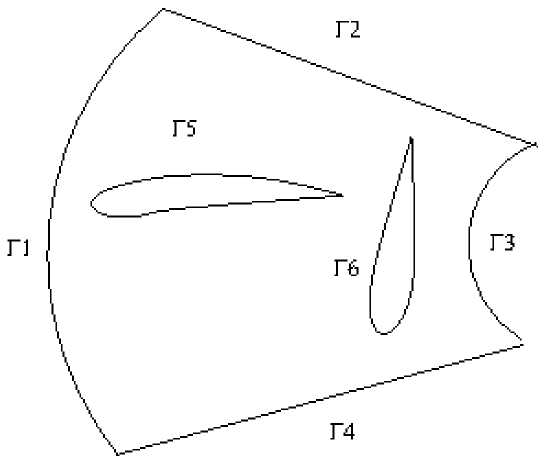


Fig. 1. Reconstructed computational domain

process applied to determine the position of the periodic boundary in the presence of narrow passage between blades. The main branch of a medial axis curve connecting the inlet to the outlet is constructed, and in this process, distance information for each boundary edge to all other adjacent boundaries is computed and stored.

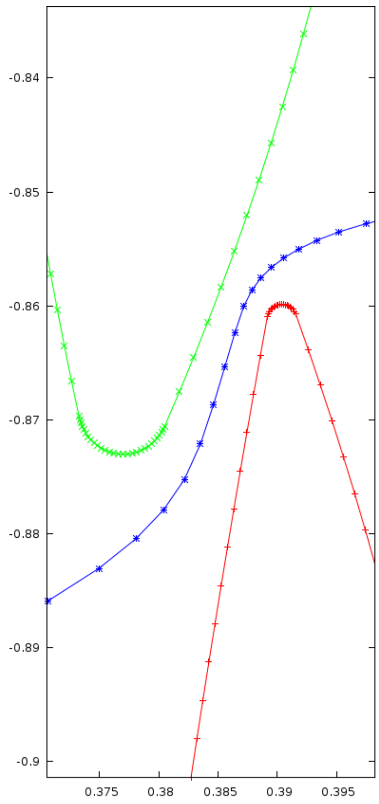


Fig. 2. Medial axis-based periodic boundary

B. Domain decomposition

The initial mesh generation is based on an automatic decomposition of the computational domain, which yields near-wall regions to be meshed using a structured approach, and a remainder domain to be meshed using an unstructured approach. Given a 2D computational domain Ω and a set of curves Γ_i representing voids in Ω, the domain decomposition step consists in constructing a near-wall region around each curve corresponding to a solid wall, by computing an offset curve $\tilde{\Gamma}_i$, as illustrated in Fig. 3. The remainder of the computational domain, consisting in the restriction of $\Omega \backslash \bigcup_i \tilde{\Gamma}_i$, forms the core domain where an unstructured mesh is constructed.

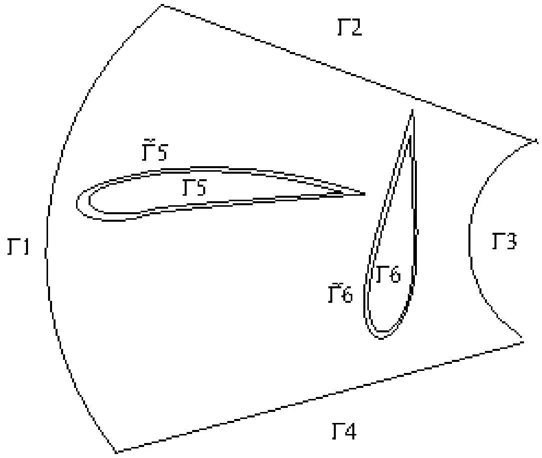


Fig. 3. Offset curves used for domain decomposition

In the case of a blunt shape at the blade trailing edge, a special treatment is applied in the blocking procedure to improve future mesh orthogonality, as illustrated in Fig. 4.

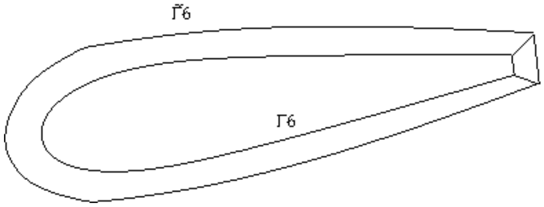


Fig. 4. Blunt trailing edge domain decomposition

C. Initial mesh generation

The near-wall region between each curve Γ_i and its corresponding offset curve $\tilde{\Gamma}_i$ forms a region in which a structured mesh is constructed. The remainder of the computational domain is meshed using a Delaunay-based mesh generation algorithm [4]. Both meshes are conforming at the common zone boundary, along the offset curve $\tilde{\Gamma}_i$, as illustrated in Fig. 5. This provides the initial mesh which will be adapted.

IV. Riemannian Metric

Controlling mesh adaptation through metric tensors provides several advantages, among which 1) being independent of the equations being solved, 2) being usable with different types of elements, 3) allowing to control simultaneously desired size, shape and orientation of elements, and 4) allowing to include

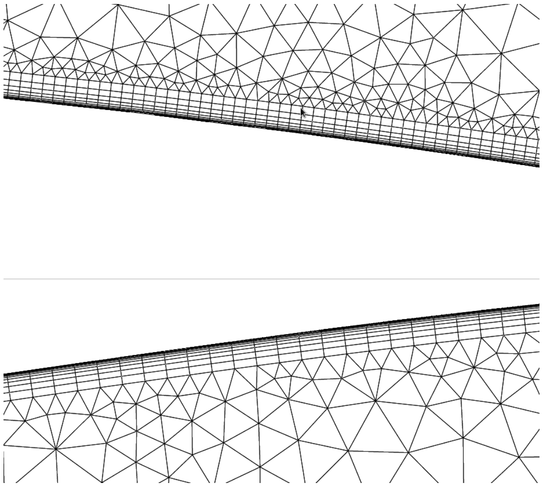


Fig. 5. Initial conforming hybrid mesh

constraints related to domain geometry. In 2D, a metric tensor is defined as:

$$M_S = R\Lambda R^{-1} = [\vec{e_1}\vec{e_2}] \left[\begin{array}{cc} h_1^{-2} & 0 \\ 0 & h_2^{-2} \end{array} \right] \left[\begin{array}{c} \vec{e_1^T} \\ \vec{e_2^T} \end{array} \right]$$

where vectors $\vec{e_1}$ and $\vec{e_2}$ are the eigenvectors of M_S, which correspond to two prescribed directions, and the quantities h_1 et h_2 are the inverse of square roots of the eigenvalues, which correspond to two prescribed sizes along vectors $\vec{e_1}$ and $\vec{e_2}$. The eigenvalues of M_S are always real positive numbers since M_S must be a positive definite matrix. A measure of length between two points A and B is defined in the metric space of M_S by the expression

$$l_{AB}^{M_S} = \int_0^1 \sqrt{(\vec{P_B} - \vec{P_A})^T M_S(\vec{P_t})(\vec{P_B} - \vec{P_A})}\, dt, \qquad (1)$$

with

$$\vec{P_t} = \vec{P_A} + t(\vec{P_B} - \vec{P_A}).$$

The goal of metric-based mesh adaptation can thereby be formulated as a process which generates of modifies a mesh so that all edges be of unit length, when measured in the space induced by the specified metric M_S.

V. Metric specification

In the present adaptation process, the target metric is obtained as the combination of information from three distinct sources, namely:

- The distance from a solid boundary,
- The distance of a given boundary edge to distinct nearby obstacles,
- Other user-specified information.

In the vicinity of a solid boundary, a structured mesh layer is constructed through an explicit blocking procedure described above. The unstructured triangular mesh connected to the structured boundary layer mesh must both connect conformally to the structured mesh, and present element grading and stretching characteristics that are compatible with the elements in the boundary layer mesh. To achieve a compatible mesh density and stretching close to the boundary mesh, a local metric M_G is computed, based on the element characteristics of the structured mesh along the offset curve.

The second source of information used to construct the target metric comes from the distance characteristics computed while determining the position of the periodic boundary interface, as described in section III-A. The minimum distance of each boundary edge to all other obstacles and domain boundaries is computed and stored. This information is used to determine a target element size on the boundary edges, which is propagated inside the computational domain, in the presence of nearby obstacles. The local distance information is therefore converted into a second metric field, M_D, which contains the local distance of a point in the computational domain to nearby obstacles and boundaries.

A third source of metric information M_U may come from the user, who has the possibility of specifying local refinement regions in the domain. These regions may take the shape of circles or rectangles, which are linked to specific blade positions, such as the blade leading or trailing edge. This feature provides a natural mechanism for the user to concentrate the mesh, for instance, in the blade wake, or at the blade leading edge, as illustrated in Fig. 6.

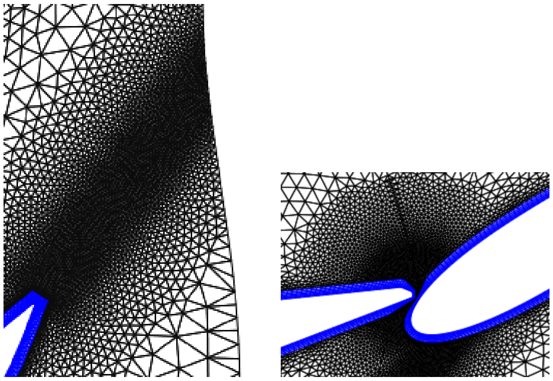


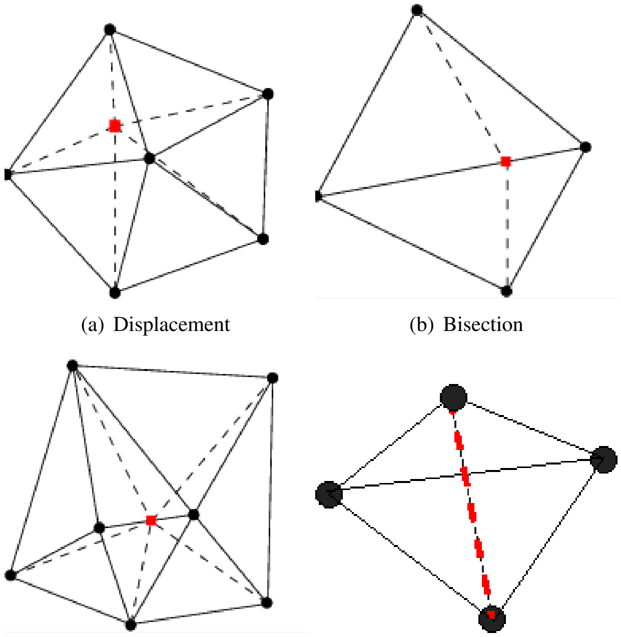
Fig. 6. User specified concentration regions

The target metric used for adaptation is therefore obtained as the tensor combination of the three metrics: $M_S = M_G \cap M_D \cap M_U$.

VI. Metric-based mesh adaptation

For triangular elements, four adaptation techniques are employed, namely vertex displacement, edge refinement, edge coarsening and edge swapping (Fig. 7). Hence, both geometric and topological modifications are applied to the original mesh to produce an adapted mesh.

The adaptation algorithm is an iterative scheme which tries to satisfy at best the specified metric. The different operations mentioned above are successively combined until the mesh satisfies the metric up to some level of conformity or up to a specified number of iterations. The order in which each operation is applied during one adaptation iteration was



(a) Displacement

(b) Bisection

(c) Coarsening

(d) Swap

Fig. 7. Mesh adaptation operations

determined experimentally in order to optimize the overall algorithm performance.

VII. Mesh extrusion

The final step in the generation of a tridimensional mesh consists in extruding the adapted 2D mesh in the third dimension. Generally, radial turbo-machines consist of a main flow region enclosed between two parallel surfaces, which connect to inlet and outlet regions where the machine height varies. It is therefore often necessary to proceed with a variable height extrusion to accommodate these extension regions.

VIII. Results

Figure 8 illustrates the final result of the mesh adaptation process in the presence of a complex periodic interface. This figure also illustrates the rectangular-shaped user defined refinement regions in the wake of the blades, which are automatically reproduced across the periodic boundary.

IX. Conclusion

This paper has presented a robust mesh generation methodology for high Reynolds number flow simulation applications. The methodology, based on a priori mesh adaptation, is very general and may be applied in a wide variety of applications. The recourse to Riemannian metric specifications provides a very flexible and highly configurable mechanism to control of the mesh density and stretching. Several other applications of this methodology are currently being investigated.

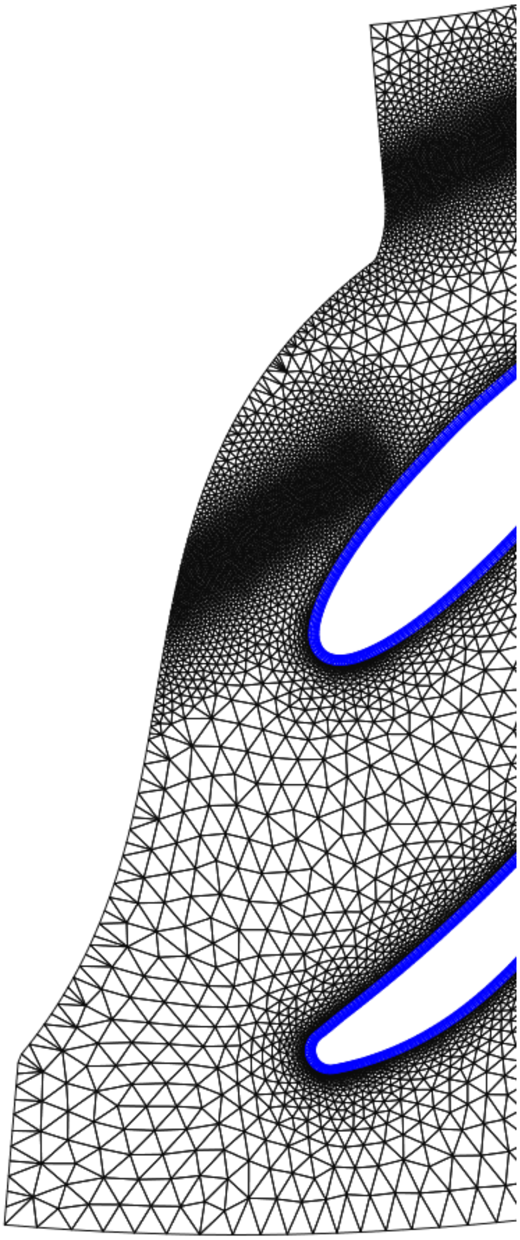


Fig. 8. 2-stage adapted mesh

References

[1] E. F. D'Azevedo and R. B. Simpson, "On optimal triangular meshes for minimizing the gradient error," *Numer. Math.*, vol. 59, no. 4, pp. 321–348, 1991.

[2] J. A. Samareh, "Geometry and grid/mesh generation issues for CFD and CSM shape optimization," vol. 6, no. 1, pp. 21–32, 2005.

[3] P. Labb, J. Dompierre, M.-G. Vallet, and F. Guibault, "Verification of a three-dimensional anisotropic adaptation method by local mesh modification," *International Journal for Numerical Methods in Engineering*, vol. 88, no. 4, pp. 350–369, 2011.

[4] P.-L. George and H. Borouchaki, *Delaunay Triangulation and Meshing. Applications to Finite Elements*. Paris: Hermès, 1998.

A Simple Dual Method for Optimal Allocation of Total Network Resources

I.V. Konnov, A.Yu. Kashuba, E. Laitinen

***Abstract*—We consider a general problem of optimal allocation of a homogeneous resource (bandwidth) in a wireless communication network, which is decomposed into several zones (clusters). Due to instability of users requirements, the fixed network resource volume may be not sufficient in some time periods, hence the network manager can buy additional volumes of this resource. This approach leads to a constrained convex optimization problem. We suggest the dual Lagrangian method to be applied to a selected constraints. This enables us to replace the initial problem with one-dimensional dual one. We consider the case of the affine cost (utility) functions, when each calculation of the value of the dual function requires solution of a special linear programming problem. The results of the numerical experiments confirm the preferences of the new method over the previous ones.**

***Index Terms*—Resource allocation, wireless networks, bandwidth, zonal network, dual Lagrange method, linear programming.**

I. Introduction

THe current development of telecommunication systems creates a number of new challenges of efficient management mechanisms involving various aspects. One of them is the efficient allocation of limited communication networks resources. In fact, despite the existence of powerful processing and transmission devices, increasing demand of different communication services and its variability in time, place, and quality, leads to serious congestion effects and inefficient utilization of significant network resources (e.g., bandwidth and batteries capacity), especially in wireless telecommunication networks. This situation forces one to replace the fixed allocation rules with more flexible mechanisms; see e.g. [1]–[4]. Naturally, treatment of these very complicated systems is often based on a proper decomposition/clustering approach, which can involve zonal, time, frequency and other decomposition procedures for nodes/units; see e.g. [5], [6]. In [7], [8], several optimal resource allocation problems in telecommunication networks and proper decomposition based methods were suggested. A further development of these models, where a system manager can utilize additional external resources for satisfying current users resource requirements, was considered in [9]. We note that such a strategy is rather typical for contemporary wireless communication networks, where WiFi or femtocell communication services are utilized in addition to the usual network resources; see e.g. [10]. A decomposition method for solution of the arising optimization problem was also suggested in [9]. It was based on an explicit volume resource allocation procedure and gave a multi-level iterative procedure. In this paper, we consider some other approach to enhance the performance of the solution method. It consists in utilization of the Lagrangian multipliers only for the total resource bound, which yields an one-dimensional dual optimization problem. We consider the case of the affine cost (utility) functions, when each calculation of the value of the dual function requires solution of a special linear programming problem. The results of the numerical experiments confirms the preferences of the new method over the previous ones.

II. Problem description

Let us consider a network with nodes (attributed to users), which is divided into n zones (clusters) within some fixed time period. For the k-th zone ($k = 1, \dots, n$), I_k denotes the index set of nodes (currently) located in this zone, b_k is the maximal fixed resource value. The network manager satisfies users resource requirements in the k-th zone by allocation of the own (inner) resource value $x_k \in [0, b_k]$ and also by taking the external resource value $z_k \in [0, c_k]$. Clearly, these values require proper maintenance expenses $f_k(x_k)$ and side payments $h_k(z_k)$ for each $k = 1, \dots, n$. We suppose also that there exists the upper bound B for the total amount of the inner resource of the network. Next, if the i-th user receives the resource amount y_i with the upper bound a_i, then he/she pays the charge $\varphi_i(y_i)$. The problem of the network manager is to find an optimal allocation of the resource among the zones and can be written as follows:

$$\max_{(x,y,z)\in W, \sum\limits_{k=1}^{n} x_k \le B} \to \mu(x, y, z) \tag{1}$$

where

$$\mu(x, y, z) = \sum_{k=1}^{n} \left[\sum_{i\in I_k} \varphi_i(y_i) - f_k(x_k) - h_k(z_k) \right], \tag{2}$$

and

$$W = \left\{ (x, y, z) \;\middle|\; \begin{array}{c} \sum_{i\in I_k} y_i = x_k + z_k,\ 0 \le y_i \le a_i,\ i \in I_k, \\ 0 \le x_k \le b_k,\ 0 \le z_k \le c_k,\ k = 1, \dots, n \end{array} \right\}. \tag{3}$$

In what follows we assume all the functions $\varphi_i(y_i)$, $f_k(x_k)$, and $h_k(z_k)$ are affine, i.e.

$$\begin{array}{rcl} \varphi_i(y_i) &=& \alpha'_i y_i + \alpha''_i,\ \alpha'_i > 0,\ i \in I_k,\ k = 1, \dots, n, \\ f_k(x_k) &=& \beta'_k x_k + \beta''_k,\ \beta'_k > 0,\ k = 1, \dots, n, \\ h_k(z_k) &=& \gamma'_k z_k + \gamma''_k,\ \gamma'_k > 0,\ k = 1, \dots, n. \end{array}$$

I.V. Konnov is with the Department of System Analysis and Information Technologies, Kazan Federal University, ul. Kremlevskaya, 18, Kazan 420008, Russia. E-mail: konn-igor@ya.ru

A.Yu. Kashuba is with LLC "AST Povolzhye", ul.Sibirskiy trakt, 34A, Kazan, 420029, Russia. E-mail: leksser@rambler.ru

E. Laitinen is with the Department of Mathematical Sciences, University of Oulu, Oulu, Finland. E-mail: erkki.laitinen@oulu.fi

III. Solution method

Let us define the Lagrange function of problem (1)–(3) as follows:

$$L(x,u,z,\lambda) = \mu(x,y,z) - \lambda\left(\sum_{k=1}^{n} x_k - B\right).$$

We utilize the Lagrangian multiplier λ only for the total resource bound. We can now replace problem (1)–(3) with its one-dimensional dual:

$$\min_{\lambda \geq 0} \rightarrow \psi(\lambda), \tag{4}$$

where

$$\begin{aligned}\psi(\lambda) &= \max_{(x,y,z)\,\in\,W} L(x,y,z,\lambda) = \lambda B \\ &+ \max_{(x,y,z)\,\in\,W} \sum_{k=1}^{n}\left[\sum_{i\in I_k}\varphi_i(y_i) - (f_k(x_k)+\lambda x_k) - h_k(z_k)\right]\end{aligned}$$

Its solution can be found by one of well-known single-dimensional optimization problem.

In order to calculate the value of $\psi(\lambda)$ we have to solve the inner problem:

$$\max \rightarrow \sum_{k=1}^{n}\left[\sum_{i\in I_k}\varphi_i(y_i) - (f_k(x_k)+\lambda x_k) - h_k(z_k)\right]$$

subject to

$$\begin{aligned}&\sum_{i\in I_k} y_i = x_k + z_k,\ 0 \leq y_i \leq a_i,\ i \in I_k,\\ &0 \leq x_k \leq b_k,\ 0 \leq z_k \leq c_k,\ k = 1,\dots,n.\end{aligned}$$

Obviously, this problem decomposes into n independent zonal linear programming problems

$$\max \rightarrow \left[\sum_{i\in I_k}\varphi_i(y_i) - (f_k(x_k)+\lambda x_k) - h_k(z_k)\right], \tag{5}$$

$$\sum_{i\in I_k} y_i = x_k + z_k,\ 0 \leq y_i \leq a_i,\ i \in I_k, \tag{6}$$

$$0 \leq x_k \leq b_k,\ 0 \leq z_k \leq c_k, \tag{7}$$

for $k = 1,\dots,n$. Note that the cost function in (5) is rewritten as

$$\sum_{i\in I_k} \alpha_i' y_i - (\beta_k' + \lambda)x_k - \gamma_k' z_k.$$

It follows that we can find very easily an exact solution of each of problems (5)–(7) in a finite number of iterations by a simple ordering algorithm.

This approach gives an alternative to the method from [9], which involved explicit marginal profit values for each zone depending on its resource allocation share. Then one can also replace the initial problem (1)–(3) with a sequence of one-dimensional ones, but that requires a multi-level iterative procedure with concordance of accuracies at each level. Therefore, our new approach simplifies essentially that from [9].

TABLE I
Results of testing with $J = 510$, $n = 70$, $\delta = 10^{-2}$

ε	N_ε	T_ε (DML)	T_ε (SDM)
10^{-1}	20	3.3907	0.0050
10^{-2}	24	3.9427	0.0038
10^{-3}	29	4.9633	0.0043
10^{-4}	34	5.7347	0.0057

TABLE II
Results of testing with $n = 70$, $\varepsilon = 10^{-2}$, $\delta = 10^{-2}$

J	N_ε	T_ε (DML)	T_ε (DMLA)	T_ε (SDM)
210	24	1.7453	1.2240	0.0009
310	24	2.4480	1.7967	0.0025
410	24	3.1980	2.3910	0.0028
510	24	3.9427	2.9007	0.0038
610	24	4.6097	3.4167	0.0038
710	24	5.3070	3.9220	0.0040
810	24	6.0260	4.4427	0.0031
910	24	6.9170	4.9533	0.0047
1010	24	7.4843	5.4797	0.0047

IV. Numerical experiments

In order to evaluate the performance of the new method denoted as (SDM) and to compare it with that from [9] denoted as (DML) we made a number of computational experiments.

We utilized the golden section method for solving the single-dimensional optimization problems. The methods were implemented in C++ with a PC with the following facilities: Intel(R) Core(TM) i7-4500, CPU 1.80 GHz, RAM 6 Gb.

The initial intervals for choosing the dual variable λ (and the additional dual variables in (DML)) were taken as [0,1000]. The initial intervals for choosing the zonal allocation shares u_k in (DML) were taken as $[0, R]$ with $R = B + \sum_{k=1}^{n} c_k$, B was chosen to be 1000. Values of b_k and c_k were chosen by trigonometric functions in $[1, 11]$, values of a_i were chosen by trigonometric functions in $[1, 2]$. The number of zones was varied from 5 to 105, the number of users was varied from 210 to 1010. Users were distributed in zones either uniformly or according to the normal distribution. The processor time and number of iterations, which were necessary to find an approximate solution of problem (4) within the same accuracy, were not significantly different for these two cases of distributions.

Further we report the results of tests, which include the time and number of iterations needed to find a solution of problem (4) within some accuracies. Let ε and δ denote the desired accuracy of finding a solution of problem (4) and solutions of auxiliary inner problems in (DML). Let J denote the total number of users, N_ε the number of upper iterations in λ, T_ε the total processor time in seconds. For the same accuracy, both the methods gave the same numbers of upper iterations, so that the main difference was in the processor time. The results of computations are given in Tables I–III. We inserted also the results for (DML) with adaptive strategy of choosing the inner accuracies. We named by (DMLA) this version of the method. In Table I, we vary the accuracy ε, in Tables II and III we vary the total number of users and the number of zones, respectively. From the results we can conclude that the

TABLE III
RESULTS OF TESTING WITH $J = 510$, $\varepsilon = 10^{-2}$, $\delta = 10^{-2}$

n	N_ε	T_ε (DML)	T_ε (DMLA)	T_ε (SDM)
5	24	3.5730	2.6517	0.0032
15	24	3.6200	2.6877	0.0019
25	24	3.6927	2.7240	0.0016
35	24	3.7500	2.7917	0.0013
45	24	3.7970	2.7970	0.0034
55	24	3.8487	2.8383	0.0034
65	24	3.9480	2.8857	0.0044
75	24	3.9740	2.9167	0.0047
85	24	4.0210	2.9530	0.0038
95	24	4.1720	3.0260	0.0035
105	24	4.2187	3.0467	0.0053

new method (SDM) has the significant preference over that in [9], which enables us to apply (SDM) for online solution of these resource allocation problems.

V. CONCLUSION

In this work, we considered a problem of managing limited resources in a zonal wireless communication network and gave its constrained convex optimization problem formulation. We proposed a new dual Lagrangian method, which was applied to the case of the affine cost (utility) functions. The results of the numerical experiments confirmed the preferences of this method over the previous ones.

ACKNOWLEDGMENT

In this work, the first and second authors were supported by the RFBR grant, project No. 13-01-00029a. Also, the first and third authors were supported by grant No. 276064 from Academy of Finland.

REFERENCES

[1] Stańczak, S., Wiczanowski, M., Boche, H.: Resource Allocation in Wireless Networks. Theory and Algorithms. Springer, Berlin (2006)

[2] Wyglinski, A.M., Nekovee, M., Hou, Y.T.: (eds.) Cognitive Radio Communications and Networks: Principles and Practice. Elsevier, Amsterdam (2010)

[3] Zhao, Q., Sadler, B.: A survey of dynamic spectrum access. IEEE Commun. Mag., vol. 24, pp.79–89 (2009)

[4] Rao, K.R., Bojkovic, Z.S., Bakmaz, B.M.: Wireless Multimedia Communication Systems: Design, Analysis, and Implementation. CRC Press, Boca Raton (2014)

[5] Chen, Y., Liestman, A.L.: A zonal algorithm for clustering ad hoc networks. Int. J. Found. of Computer Sci., vol.14, pp.305–322 (2003)

[6] Rohloff, K., Ye, J., Loyall, J., Schantz, R.: A hierarchical control system for dynamic resource management. Proc. of the 12th IEEE Real-Time and Embed. Technol. and Appl. Symposium (RTAS 2006). Work in Progress Symposium, San Jose, (2006)

[7] Konnov, I.V., Kashina, O.A., Laitinen, E.: Optimisation problems for control of distributed resources. Int. J. Model., Ident. and Contr., vol.14, pp.65–72 (2011)

[8] Konnov, I.V., Kashina, O.A., Laitinen, E.: Two-level decomposition method for resource allocation in telecommunication network. Int. J. Dig. Inf. Wirel. Comm., vol.2, pp.150–155 (2012)

[9] Konnov, I.V., Laitinen, E., Kashuba, A.: Optimization of zonal allocation of total network resources. Proc. of the 11th International Conference ”Applied Computing 2014”, Porto, pp.244–248 (2014).

[10] Gao, L., Iosifidis, G., Huang, J., Tassiulas, L.: Economics of mobile data offloading. IEEE Conference on Computer Communications, pp. 351–356 (2013)

Ph.D Erkki Laitinen is a university lecturer at the Department of Mathematical Sciences of University of Oulu, and an adjunct professor of Computer Science at the University of Jyväskylä, Finland. His research interests include numerical analysis, optimization and optimal control. He is active in promoting these techniques in practical problem solving in engineering, manufacturing, and industrial process optimization. He has published more than hundred peer reviewed scientific papers in international journals and conferences. He has participated in several applied projects dealing with optimization and control of production processes or wireless telecommunication systems.

D.Sc. Igor Konnov is a professor at the Department of System Analysis and Information Technologies, Kazan Federal University, Kazan, Russia. His research topics include theory, methods, and applications of nonsmooth optimization, equilibrium problems, and variational inequalities. He has published five books and more than 230 peer reviewed scientific papers in these fields.

Aleksey Kashuba is a leading engineer at the LLC ”AST Povolzhye”, and a junior researcher fellow at the Scientific Research Laboratory ”Computational Technologies and Computer Modeling” of Kazan Federal University. His research interests include numerical analysis, optimization and system analysis.

Multiple-Layer Parking with Screening

Sjoert Fleurke and Aernout C. D. van Enter

***Abstract*—**In this article a multilayer parking system with screening of size $n = 3$ is studied with a focus on the time-dependent particle density. We prove that the asymptotic limit of the particle density increases from an average density of $1/3$ on the first layer to the value of $(10 - \sqrt{5})/19 \approx 0.4086$ in higher layers.

***Keywords*—**Multi-layer car parking, Particle deposition.

I. Introduction

SUPPOSE we have a lattice $\mathcal{L}(x, r)$ consisting of sites (x, r) with positions $x \in \{-2, -1, 0, 1, 2\}$ and heights r $\in \mathbb{N}$. At each position particles arrive according to independent Poisson processes $N_t(x)$. We impose boundary conditions $N_t(-2) = N_t(2) = 0$. The particles pile up across the layers but they are not allowed to "interfere" with particles earlier deposited in neighboring sites at the same layer. In other words, the horizontal distance between two particles has to be at least 2. Furthermore, in this model the particles are not allowed to pass earlier deposited particles. As a consequence a new particle is always deposited in the layer above the highest layer that rejected it. This model property is sometimes called "screening" (see Fig. 1).

Our model can be formulated more precisely in the following way.

1) The state-space is $\mathcal{F} \coloneqq (\mathcal{L}, \mathbb{N}^+)^{\{0,1\}}$.

2) The process $\kappa_t(x, r) = 1$ if there is a particle at (x, r) at time t and 0 otherwise.

3) When a particle arrives at site x at time t, it will be deposited at $h_t(x) \coloneqq 1 + \max\{r: \exists_{y \in N_x}, \kappa_t(y, r) = 1\}$, where neighborhood set N_x consists of site x and the sites with distance 1 from it.

The density $\rho_t(x, r)$ of a site at $(x, r) \in \mathcal{L}$ is defined as the expectation of the occupancy of that site at time t, i.e. $\rho_t(x, r) = E\kappa_t(x, r)$. The end-density of a site is $\rho_\infty(x, r)$.

The majority of the existing literature in which discrete parking is analytically treated is about monolayer models [1, 2, 3], while most literature about multi-layer models is based on simulations [4, 5]. However, recently there is some interest in analytical results on multilayer parking models. In [6], for example, it was shown that in an infinite parking system the second layer has a higher end-density than the first layer for models both with and without screening. Analytical formulas for the time-dependent densities were derived for small parking systems without screening in [7] and [8]. It was found that the end-density in the case of a system of size three tends to exactly $\frac{1}{2}$ for high layers. It is conjectured that the same counts for bigger parking systems. In [9] density formulas are calculated for the model with screening in the case of infinite-sized regular and random trees. Contrary to the model without screening the layer densities turn out to decrease with the layer number. In [10] it was proven that the end-density of an infinite parking system with screening tends to a value that lies between $\frac{1}{k^*}$ and $\frac{1}{2}$, where k^* is such that $\left(2\frac{e}{k^*}\right)^{k^*} - e = 0$, which means that $0.232 < \rho < 0.500$. To our knowledge a precise value for the end-densities of a system with screening of any size is yet to be found.

In this paper we continue the work on calculating the particle densities in a small multi-layer parking model. We hope our result will lead to further insights also in systems with bigger sizes and systems with neighborhoods of cardinality greater than 2.

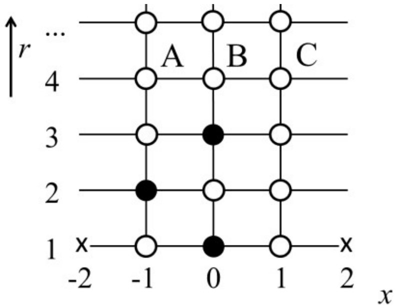


Fig. 1: Parking on a lattice with 3 positions where parking is allowed. In this picture 3 particles have arrived at positions $x_1 = 0, x_2 = -1$, and $x_3 = 0$. The next particle will be deposited at position A, B, or C depending on its x-position. Note that in the case of the parking model without screening a particle at position 1 would not be deposited at C(1,4) but at position $(x, r) = (1, 2)$. The '×' symbols at positions -2 and 2 are indicating that at those positions no

This work was supported by the Radio Communications Agency, Netherlands.

S. R. Fleurke is with the Radiocommunications Agency Netherlands, Postbus 450, 9700 AL, Groningen, The Netherlands. (corresponding author: sjoert.fleurke@agentschaptelecom.nl).

A. C. D. van Enter is with the Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Nijenborgh 9, 9747 AG, Groningen, The Netherlands. (e-mail: a.c.d.van.enter@rug.nl).

particles are dropped during the process.

II. Time-Dependent Particle Densities

This section is dedicated to the calculation of the time-dependent density of vertex $(0, r)$ at time t. The result is displayed in the following theorem.

Theorem 1: Consider parking with screening on a lattice $\mathcal{L}(x, r)$ consisting of sites (x, r) with positions $x \in \{-2, -1, 0, 1, 2\}$ and heights $r \in \mathbb{N}$ and boundary conditions $N_t(-2) = N_t(2) = 0$. The time-dependent particle density at layer r is given by:

$$\rho_t(0, r) = \sum_{i=1}^{r} \sum_{d_1+d_2+\cdots+d_i=r} \prod_{j=1}^{i} \left[\frac{2}{3} \sum_{k=0}^{d_j-2} \binom{d_j+k-1}{k} \left(\frac{1}{3}\right)^{d_j+k-1} + \right. \qquad (1)$$

$$\left. + \binom{2d_j-2}{d_j-1} \left(\frac{1}{3}\right)^{2d_j-1} \right] \left(1 - e^{-t} \sum\nolimits_{l=0}^{i} \frac{t^l}{l!}\right)$$

Using Theorem 1 we find the following densities for the first 4 layers:

$$\rho_t(0, 1) = \frac{1}{3} - \left(\frac{1}{3} + \frac{1}{3} t\right) e^{-t} \qquad (2)$$

$$\rho_t(0, 2) = \frac{11}{27} - \left(\frac{11}{27} + \frac{11}{27} t + \frac{1}{9} \frac{t^2}{2}\right) e^{-t} \qquad (3)$$

$$\rho_t(0, 3) = \frac{11}{27} - \left(\frac{11}{27} + \frac{11}{27} t + \frac{19}{81} \frac{t^2}{2!} + \frac{1}{27} \frac{t^3}{3!}\right) e^{-t} \qquad (4)$$

$$\rho_t(0, 4) = \frac{893}{2187} - \left(\frac{893}{2187} + \frac{893}{2187} t + \frac{229}{729} \frac{t^2}{2!} + + \frac{1}{27} \frac{t^3}{3!} + \frac{1}{81} \frac{t^4}{4!}\right) e^{-t} \qquad (5)$$

Theorem 1 can be proven using the fact that the process has a renewal structure. Every particle arrival at the center vertex $x = 0$ counts as a renewal. Between every arrival at the center there have been zero or more arrivals at the neighboring sites. The vertical distance between two consecutive particles arriving in the center is thus determined by the maximum of the number of arriving particles at the neighboring sites. More precisely, if t_n denotes the arrival time of the nth particle at the center and $N_{t_1, t_2}(x)$ denotes the number of arrivals at x between time t_1 and t_2, then the vertical distance between the nth and the (n+1)st consecutively arrived particles at $x = 0$ is equal to

$$\psi_n := 1 + \max\{N_{t_{n-1}, t_n}(-1), N_{t_{n-1}, t_n}(1)\}. \qquad (6)$$

This means that, for example, the probability that the first center particle is deposited at height $r = 5$, is equal to $P(\psi_1 = 5) = P\left(\max\{N_{0, t_1}(-1), N_{0, t_1}(1)\} = 4\right)$.

More generally we can write:

$$\rho_t(0, r) = P\left(\sum_{j=1}^{N_t(0)} \psi_j = r\right) \qquad (7)$$

This leads to:

$$\rho_t(0, r) = \int_0^t \sum_{i=1}^{r} P\left(\sum_{j=1}^{N_t(0)} \psi_j = r \middle| N_t(0) = i\right) \cdot P(N_t(0) = i) du \qquad (8)$$

Note that the $(\psi_n)_{n \in \mathbb{N}^+}$ are independently and identically distributed. Also remember that $N_t(0)$ is Poisson distributed, so that (4) can be rewritten as

$$\rho_t(0, r) = \int_0^t \sum_{i=1}^{r} \sum_{d_1+d_2+\cdots+d_i=r} P\left(\bigcap_{j=1}^{i} \psi_j = d_j\right) \cdot e^{-u} \frac{u^i}{i!} du \qquad (9)$$

$$= \int_0^t \sum_{i=1}^{r} \sum_{d_1+d_2+\cdots+d_i=r} \prod_{j=1}^{i} P(\psi_j = d_j)\, e^{-u} \frac{u^i}{i!} du \qquad (10)$$

$$= \sum_{i=1}^{r} \sum_{d_1+d_2+\cdots+d_i=r} \prod_{j=1}^{i} P(\psi_j = d_j) \int_0^t e^{-u} \frac{u^i}{i!} du \qquad (11)$$

$$= \sum_{i=1}^{r} \sum_{d_1+d_2+\cdots+d_i=r} \prod_{j=1}^{i} P\left(\psi_j = d_j\right) \cdot \left(1 - e^{-t} \sum\nolimits_{k=0}^{i} \frac{t^k}{k!}\right) \qquad (12)$$

To complete the result we need to calculate the distribution of the stochastic variable ψ_n .

Lemma 1: The distribution of ψ_n is given by:

$$P(\psi_n = d) = \frac{2}{3} \sum\nolimits_{k=0}^{d-2} \binom{d+k-1}{k} \left(\frac{1}{3}\right)^{d+k-1} \qquad (13)$$

$$+\binom{2d-2}{d-1}\left(\frac{1}{3}\right)^{2d-1}$$

Proof: For this proof we use expression (6). We can calculate the distribution of $S_T = \max\{N_T(-1), N_T(1)\}$, i.e. the maximum of the number of particles that arrive at position $x = -1$ and $x = 1$ in a period of time of length T.

$$\begin{aligned}
&P(S_T = d) = \\
&= \int_0^\infty P(\max\{N_T(-1), N_T(1)\} = d | T = u) \\
&\qquad \cdot P(T = u)du \\
&= \int_0^\infty P(N_T(-1) = d \ \cap N_T(1) < d \, | T = u)e^{-u}du \\
&+ \int_0^\infty P(N_T(-1) < d \ \cap N_T(1) = d \, | T = u)e^{-u}du \\
&+ \int_0^\infty P(N_T(-1) = N_T(1) = d \, | T = u)e^{-u}du \\
&= \int_0^\infty 2P(N_T(-1) = d \cap N_T(1) < d \, | T = u)e^{-u}du \\
&+ \int_0^\infty [P(N_T(1) = d \, | T = u)]^2 \, e^{-u}du \qquad (14) \\
&= 2\int_0^\infty e^{-u}\frac{u^d}{d!}\sum_{k=0}^{d-1} e^{-u}\frac{u^d}{k!}e^{-u}du \\
&+ \int_0^\infty \left[e^{-u}\frac{u^d}{d!}\right]^2 e^{-u}\,du \\
&= 2\sum\nolimits_{k=0}^{x-1}\frac{(d+k)!}{d!\,k!}\left(\frac{1}{3}\right)^{d+k+1} \\
&\cdot \int_0^\infty \frac{3^{2d+1}}{(d+k)!}u^{(d+k+1)-1}e^{-3u}du \\
&+ \frac{(2d)!}{d!\,d!}\left(\frac{1}{3}\right)^{2d+1}\int_0^\infty \frac{3^{2d+1}}{(2d)!}u^{(2d+1)-1}e^{-3u}du \\
&= \frac{2}{3}\sum\nolimits_{k=0}^{d-1}\binom{d+k}{k}\left(\frac{1}{3}\right)^{d+k} + \binom{2d}{d}\left(\frac{1}{3}\right)^{2d+1}
\end{aligned}$$

Now we take $\psi_n = 1 + S_T$ to establish our result of Lemma 1. Finally, combining Lemma 1 with equation (12) yields the formula of Theorem 1.

In Figure 2 developments in time of particle densities are displayed for several layers. It is interesting to see that in higher layers the probability of a particle hit at the center site tends to be (slightly) bigger than in lower layers.
The opposite phenomenon was observed in larger systems in [9]. In an infinite-sized system the density decreases when parking with screening is conducted.

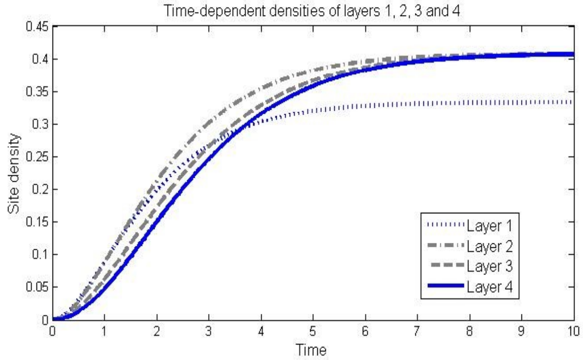


Fig. 2: Particle densities at the sites (0,r) as a function of time for r is 1, 2, 3 and 4 according to Theorem 1 and formulas (2), (3), (4) and (5) in particular.

III. Layer-dependent End-densities

With the above result it is straightforward to find the end-density for any layer. Take formula (1) and then let $t \rightarrow \infty$. This immediately yields:

$$\rho_\infty(0,r) = \sum_{i=1}^{r}\ \sum_{d_1+d_2+\cdots+d_i=r}$$

$$\prod_{j=1}^{i}\left[\frac{2}{3}\sum\nolimits_{k=0}^{d_j-2}\binom{d_j+k-1}{k}\left(\frac{1}{3}\right)^{d_j+k-1} + \right. \qquad (15)$$

$$\left. +\binom{2d_j-2}{d_j-1}\left(\frac{1}{3}\right)^{2d_j-1}\right]$$

In Figure 3 the end-densities are displayed for the first 4 layers based on formula (15) which yields $\frac{1}{3}, \frac{11}{27}, \frac{11}{27}$, and $\frac{893}{2187}$ respectively. These numbers have been confirmed by simulations.

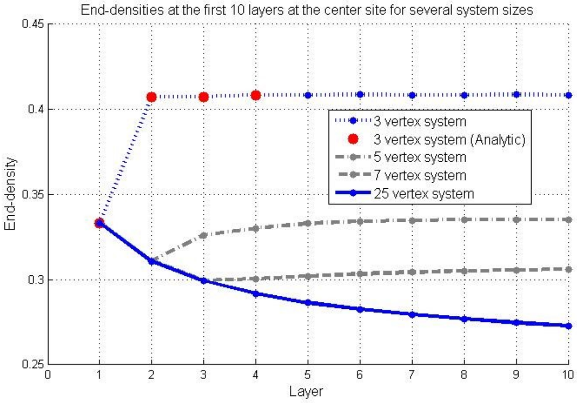


Fig. 3: End-densities as a function of the layer number based on formula (15) and simulation results. For some other bigger systems the end-densities are plotted as well. It can be seen that whereas in the 3-vertex the end-densities grow with the layer number this effect disappears in bigger systems. It is conjectured that for an infinite-sized system the end-densities of the center site decrease monotonically with the layer number.

IV. End-Density for High Layers

As can be seen in Figure 3 the end-density grows with the layer number. A natural and interesting question is what the exact limiting value for this end-density could be. The purpose of this section is to prove the following theorem.

Theorem 2. The average end-density of the center vertex ultimately tends to the value

$$p_0 := \lim_{r\to\infty} \rho_\infty(0,r) = \frac{10-\sqrt{5}}{19} \approx 0.408628 \qquad (16)$$

The same holds for the neighboring sites: $p_{-1} = p_1 = p_0 = (10-\sqrt{5})/19$.

One may attempt to obtain this limit from formula (15) by taking $r \to \infty$ but this seems rather difficult. Fortunately, there is an alternative, and perhaps more interesting, approach which starts with the observation that the average end-density is the inverse of 1 plus the average number of vertical consecutive empty positions. If, for example, the average run of empty sites between two occupied sites were 4, it would follow that the average occupancy is 1/(4+1) = 0.20. In other words, the following formula holds:

$$\lim_{r\to\infty} \rho_\infty(0,r) = \frac{1}{1+EX} \qquad (17)$$

Our efforts are now concentrated on calculating the value of EX. We will prove the following lemma.

Lemma 2. The expected run of empty center sites for high layers is given by:

$$\mathrm{EX} = 1 + \frac{1}{5}\sqrt{5} \qquad (18)$$

Proof: If the number of particles arrived at the border sites is known, we can calculate the expected run length EX as follows:

$$E(X|N=n) = \sum_{i=0}^{n} \max(i, n-i) \binom{n}{i} \left(\frac{1}{2}\right)^{n} \qquad (19)$$

If n is an even number this can be written as:

$$\sum_{i=0}^{n/2} (n-i)\binom{n}{i}\left(\frac{1}{2}\right)^{n} + \sum_{i=\frac{n}{2}+1}^{n} i\binom{n}{i}\left(\frac{1}{2}\right)^{n} \qquad (20)$$

while if n is an odd number it can be written as:

$$\sum_{i=0}^{(n-1)/2} (n-i)\binom{n}{i}\left(\frac{1}{2}\right)^{n} + \sum_{i=\frac{n-1}{2}+1}^{n} i\binom{n}{i}\left(\frac{1}{2}\right)^{n} \qquad (21)$$

So, we continue with splitting the cases where n is even and n is odd. When n is an even number we find with some algebraic manipulation:

$$\begin{aligned} E(X|N=2k) &= \sum_{i=0}^{2k} \max(i, 2k-i)\binom{2k}{i}\left(\frac{1}{2}\right)^{2k} \\ &= \sum_{i=0}^{k} (2k-i)\binom{2k}{i}\left(\frac{1}{2}\right)^{2k} + \sum_{i=k+1}^{2k} i\binom{2k}{i}\left(\frac{1}{2}\right)^{2k} \\ &= \left(\frac{1}{2}\right)^{2k} 2k\left[\sum_{i=0}^{k}\binom{2k-1}{i} + \sum_{i=k}^{2k-1}\binom{2k-1}{i}\right] \\ &= \left(\frac{1}{2}\right)^{2k} 2k\left[2^{2k-1} + \binom{2k-1}{k}\right] \\ &= k + k\binom{2k-1}{k}\left(\frac{1}{2}\right)^{2k-1} \\ &= k + k\binom{2k}{k}\left(\frac{1}{2}\right)^{2k} \end{aligned} \qquad (22)$$

Likewise, for the odd case we find:

$$\begin{aligned} &E(X|N=2k+1) \\ &= \left(\frac{2k+1}{2}\right) + \left(\frac{2k+1}{2}\right)\binom{2k}{k}\left(\frac{1}{2}\right)^{2k} \end{aligned} \qquad (23)$$

With this result we can calculate EX:

$$\begin{aligned} \mathrm{EX} &= \sum_{n=0}^{\infty} E(X|N=n)\,P(N=n) \\ &= \sum_{k=0}^{\infty} [E(X|N=2k)\,P(N=2k) \\ &+ E(X|N=2k+1)\,P(N=2k+1)] \end{aligned} \qquad (24)$$

But first we must calculate P(N = n), the probability that during a run of empty center sites a total of n particles were dropped in the border vertices. This is given by

$$\mathrm{P(N=n)} = \frac{1}{3}\left(\frac{2}{3}\right)^{n} \qquad (25)$$

which may be regarded trivial but can easily be calculated as follows. Let T denote the time between two droppings in the center. Then we have

$$\begin{aligned} \mathrm{P(N=n)} &= \int_0^\infty P(N_T = n|T=t)P(T=t)dt \\ &= \int_0^\infty \frac{(2t)^n}{n!} e^{-2t} e^{-t} dt \\ &= \frac{2^n}{3^{n+1}} \int_0^\infty \frac{3^{n+1}}{n!} t^{(n+1)-1} e^{-3t} dt \\ &= \frac{1}{3}\left(\frac{2}{3}\right)^{n} \end{aligned} \qquad (26)$$

Continuing with Equation (24) we can now write:

$$\mathrm{EX} = \sum_{\mathrm{n}=0}^{\infty} \mathrm{E(X|N=n)\;P(N=n)} \qquad (27)$$

$$
\begin{aligned}
&= \sum_{k=0}^{\infty}\Bigg[\left(k + k\binom{2k}{k}\left(\frac{1}{2}\right)^{2k}\right)\frac{1}{3}\left(\frac{2}{3}\right)^{2k} \\
&+ \left(\left(\frac{2k+1}{2}\right) + \left(\frac{2k+1}{2}\right)\binom{2k}{k}\left(\frac{1}{2}\right)^{2k}\right)\frac{1}{3}\left(\frac{2}{3}\right)^{2k+1}\Bigg] \\
&= \sum_{k=0}^{\infty}\Bigg[\left(k + \frac{2}{3}\frac{2k+1}{2}\right)\frac{1}{3}\left(\frac{2}{3}\right)^{2k} \\
&+ \frac{1}{3}\,k\binom{2k}{k}\left(\frac{1}{3}\right)^{2k} + \frac{2}{9}\left(\frac{2k+1}{2}\right)\binom{2k}{k}\left(\frac{1}{3}\right)^{2k}\Bigg] \\
&= \frac{1}{9}\sum_{k=0}^{\infty}\left(\frac{4}{9}\right)^{k} + \frac{5}{9}\sum_{k=0}^{\infty}k\left(\frac{4}{9}\right)^{k} \\
&+ \frac{1}{3}\sum_{k=0}^{\infty}k\binom{2k}{k}\left(\frac{1}{9}\right)^{k} + \frac{2}{9}\sum_{k=0}^{\infty}k\binom{2k}{k}\left(\frac{1}{9}\right)^{k} \\
&+ \frac{1}{9}\sum_{k=0}^{\infty}\binom{2k}{k}\left(\frac{1}{9}\right)^{k} \\
&= \frac{1}{5} + \frac{4}{5} + \frac{5}{9}\frac{2/9}{(5/9)^{3/2}} + \frac{1}{9}\frac{3}{\sqrt{5}} = 1 + \frac{1}{\sqrt{5}}
\end{aligned}
$$

where we used the identities (with $x = 1/9$)

$$\sum_{k=0}^{\infty}\binom{2k}{k}x^{k} = \frac{1}{\sqrt{1-4x}} \qquad (28)$$

and

$$\sum_{k=0}^{\infty}k\binom{2k}{k}x^{k} = \frac{2x}{\sqrt{(1-4x)^{3}}} \qquad (29)$$

where (28) can be easily checked by writing down the Taylor series expansion for $(1-4x)^{-1/2}$ while (29) follows immediately from (28) by differentiation with respect to x.

The first part of Theorem 2 follows now from Formula (17) and Lemma 2.

$$\lim_{r\to\infty}\rho_{\infty}(0,r) = \frac{1}{1+EX} = \frac{1}{2+\frac{1}{5}\sqrt{5}} = \frac{10-\sqrt{5}}{19} \qquad (30)$$

The proof of the second part of the theorem, i.e. that the limiting densities for the position left and right to the center are identical to the center, goes as follows. The expected total number of arrivals on one of the borders is equal to the number of arrivals in the center. This means that on average between two arrivals at position 0 there will be an arrival on the border site too. The distance between two particles which arrive at the center is on average EX. At the border site we expect one particle too. Therefore, the expected end-density on a border site must be $1/(1+\text{EX})$. This is indeed the same density as in the center.

References

[1] A. Rényi, "On a One-dimensional Problem Concerning Random Space-filling," *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 3, pp. 109–127, 1958.

[2] R. Cohen, H. Reiss, "Kinetics of Reactant Isolation I. One-Dimensional Problems," *J. Chem. Phys.,* vol. 38, no. 3, pp. 680–691, 1963.

[3] H. G. Dehling, S. R. Fleurke, and C. Külske, "Parking on a Random Tree," *J. Stat. Phys.*, vol. 133, no. 1, pp. 151–157, 2008.

[4] P. Nielaba, V. Privman, "Multilayer Adsorption with Increasing Layer Coverage," *Phys. Rev. A*, vol. 45, pp. 6099–6102, 1992.

[5] H. G. Dehling, S. R. Fleurke, "The Sequential Frequency Assignment Process," *Proc. 12th WSEAS Internat. Conf. on Appl. Math.,* 2007, pp. 280–285.

[6] S. R. Fleurke, C. Külske, "A Second-row Parking Paradox," *J. Stat. Phys.*, vol. 136, no. 2, pp. 285–295, 2009.

[7] S. R. Fleurke , A. C. D. van Enter, "Analytical Results for a Small Multilayer Parking System," in *Computational Problems in Engineering: Lecture Notes in Electrical Engineering*, vol. 307, V. Mladenov, N. Mastorakis, Eds. Springer International Publishing, 2014, ch. 4.

[8] S. R. Fleurke, *Multilayer Particle Deposition Models*. Saarbrücken: VDM Verlag Dr. Muller, 2011, p. 33.

[9] S. R. Fleurke, C. Külske, "Multilayer Parking with Screening on a Random Tree," *J. Stat. Phys.*, vol. 139, no. 3, pp. 417–431, 2010.

[10] T. S. Mountford, A. Sudbury, "Deposition processes with Hardcore Behavior," *J. Stat. Phys.*, vol. 146, no. 4, pp. 687-700, 2012.

Data interpolation with applications via probabilistic distribution and nodes combination

Dariusz J. Jakóbczak

Abstract**—**Proposed method, called Probabilistic Nodes Combination (PNC), is the method of 2D curve modeling and handwriting identification by using the set of key points. Nodes are treated as characteristic points of signature or handwriting for modeling and writer recognition. Identification of handwritten letters or symbols need modeling and the model of each individual symbol or character is built by a choice of probability distribution function and nodes combination. PNC modeling via nodes combination and parameter γ as probability distribution function enables curve parameterization and interpolation for each specific letter or symbol. Two-dimensional curve is modeled and interpolated via nodes combination and different functions as continuous probability distribution functions: polynomial, sine, cosine, tangent, cotangent, logarithm, exponent, arc sin, arc cos, arc tan, arc cot or power function.

Keywords**—** handwriting identification, shape modeling, curve interpolation, PNC method, nodes combination, probabilistic modeling.

I. Introduction

Handwriting identification and writer verification are still the open questions in artificial intelligence and computer vision. Handwriting based author recognition offers a huge number of significant implementations which make it an important research area in pattern recognition [1]. There are so many possibilities and applications of the recognition algorithms that implemented methods have to be concerned on a single problem. Handwriting and signature identification represents such a significant problem. In the case of writer recognition, described in this paper, each person is represented by the set of modeled letters or symbols. The sketch of proposed method consists of three steps: first handwritten letter or symbol must be modeled by a curve, then compared with unknown letter and finally there is a decision of identification. Author recognition of handwriting and signature is based on the choice of key points and curve modeling. Reconstructed curve does not have to be smooth in the nodes because a writer does not think about smoothing during the handwriting. Curve interpolation in handwriting identification is not only a pure mathematical problem but important task in pattern recognition and artificial intelligence such as: biometric recognition [2-4], personalized handwriting recognition [5], automatic forensic document examination [6,7], classification of ancient manuscripts [8]. Also writer recognition in monolingual handwritten texts is an extensive area of study and the methods independent from the language are well-seen. Proposed method represents language-independent and text-independent approach because it identifies the author via a single letter or symbol from the sample. This novel method is also applicable to short handwritten text.

Writer recognition methods in the recent years are going to various directions: writer recognition using multi-script handwritten texts [9], introduction of new features [10], combining different types of features [3], studying the sensitivity of character size on writer identification [11], investigating writer identification in multi-script environments [9], impact of ruling lines on writer identification [12], model perturbed handwriting [13], methods based on run-length features [14,3], the edge-direction and edge-hinge features [2], a combination of codebook and visual features extracted from chain code and polygonized representation of contours [15], the autoregressive coefficients [9], codebook and efficient code extraction methods [16], texture analysis with Gabor filters and extracting features [17], using Hidden Markov Model [18-20] or Gaussian Mixture Model [1]. But no method is dealing with writer identification via curve modeling or interpolation and points comparing as it is presented in this paper.

The author wants to approach a problem of curve interpolation [21-23] and shape modeling [24] by characteristic points in handwriting identification. Proposed method relies on nodes combination and functional modeling of curve points situated between the basic set of key points. The functions that are used in calculations represent whole family of elementary functions with inverse functions: polynomials, trigonometric, cyclometric, logarithmic, exponential and power function. These functions are treated as probability distribution functions in the range [0;1]. Nowadays methods apply mainly polynomial functions, for example Bernstein polynomials in Bezier curves, splines and NURBS [25]. But Bezier curves do not represent the interpolation method and cannot be used for example in signature and handwriting modeling with characteristic points (nodes). Numerical methods for data interpolation are based on polynomial or trigonometric functions, for example Lagrange, Newton, Aitken and Hermite methods. These methods have some weak sides [26] and are not sufficient for curve interpolation in the situations when the curve cannot be build by polynomials or trigonometric functions. Proposed 2D curve interpolation is the functional modeling via any elementary

Dariusz Jacek Jakóbczak, Department of Electronics and Computer Science, Technical University of Koszalin, Sniadeckich 2, 75-453 Koszalin, Poland
dariusz.jakobczak@tu.koszalin.pl

functions and it helps us to fit the curve during handwriting identification.

This paper presents novel Probabilistic Nodes Combination (PNC) method of curve interpolation and takes up PNC method of two-dimensional curve modeling via the examples using the family of Hurwitz-Radon matrices (MHR method) [27], but not only (other nodes combinations). The method of PNC requires minimal assumptions: the only information about a curve is the set of at least two nodes. Proposed PNC method is applied in handwriting identification via different coefficients: polynomial, sinusoidal, cosinusoidal, tangent, cotangent, logarithmic, exponential, arc sin, arc cos, arc tan, arc cot or power. Function for PNC calculations is chosen individually at each modeling and it represents probability distribution function of parameter $\alpha \in [0;1]$ for every point situated between two successive interpolation knots. PNC method uses nodes of the curve $p_i = (x_i,y_i) \in \boldsymbol{R}^2$, $i = 1,2,\ldots n$:

1. PNC needs 2 knots or more ($n \geq 2$);
2. If first node and last node are the same ($p_1 = p_n$), then curve is closed (contour);
3. For more precise modeling knots ought to be settled at key points of the curve, for example local minimum or maximum and at least one node between two successive local extrema.

Condition 3 means for example the highest point of the curve in a particular orientation, convexity changing or curvature extrema. The goal of this paper is to answer the question: how to model a handwritten letter or symbol by a set of knots [28]?

II. Probabilistic CURVE Interpolation

The method of PNC is computing points between two successive nodes of the curve: calculated points are interpolated and parameterized for real number $\alpha \in [0;1]$ in the range of two successive nodes. PNC method uses the combinations of nodes $p_1=(x_1,y_1)$, $p_2=(x_2,y_2)$,…, $p_n=(x_n,y_n)$ as $h(p_1,p_2,\ldots,p_m)$ and $m = 1,2,\ldots n$ to interpolate second coordinate y for first coordinate $c = \alpha \cdot x_i + (1-\alpha)\cdot x_{i+1}$, $i = 1,2,\ldots n-1$:

$$y(c) = \gamma \cdot y_i + (1-\gamma) y_{i+1} + \gamma(1-\gamma) \cdot h(p_1, p_2, ..., p_m)\,, \qquad (1)$$

$\alpha \in [0;1]$, $\gamma = F(\alpha) \in [0;1]$.

Here are the examples of h computed for MHR method [29]:

$$h(p_1, p_2) = \frac{y_1}{x_1} x_2 + \frac{y_2}{x_2} x_1 \qquad (2)$$

or

$$h(p_1, p_2, p_3, p_4) = \frac{1}{x_1^{\,2} + x_3^{\,2}} (x_1 x_2 y_1 + x_2 x_3 y_3 + x_3 x_4 y_1 - x_1 x_4 y_3) + \\ + \frac{1}{x_2^{\,2} + x_4^{\,2}} (x_1 x_2 y_2 + x_1 x_4 y_4 + x_3 x_4 y_2 - x_2 x_3 y_4)\,.$$

The examples of other nodes combinations:

$$h(p_1, p_2) = \frac{y_1 x_2}{x_1 y_2} + \frac{y_2 x_1}{x_2 y_1}$$

or

$$h(p_1, p_2) = \frac{y_1 x_2}{y_2} + \frac{y_2 x_1}{y_1}$$

or

$$h(p_1, p_2) = x_1 y_1 + x_2 y_2$$

or

$$h(p_1, p_2) = x_1 x_2 + y_1 y_2$$

or

$$h(p_1, p_2, ..., p_m) = 0$$

or

$$h(p_1) = x_1 y_1$$

or others. Nodes combination is chosen individually for each curve. Formula (1) represents curve parameterization as $\alpha \in [0;1]$:

$x(\alpha) = \alpha \cdot x_i + (1-\alpha)\cdot x_{i+1}$

and

$$y(\alpha) = F(\alpha) \cdot y_i + (1 - F(\alpha)) y_{i+1} + F(\alpha)(1 - F(\alpha)) \cdot h(p_1, p_2, ..., p_m)$$

$$y(\alpha) = F(\alpha) \cdot (y_i - y_{i+1} + (1 - F(\alpha)) \cdot h(p_1, p_2, ..., p_m)) + y_{i+1}\,.$$

Proposed parameterization gives us the infinite number of possibilities for curve calculations (determined by choice of F and h) as there is the infinite number of human signatures, handwritten letters and symbols. Nodes combination is the individual feature of each modeled curve (for example a handwritten letter or signature). Coefficient $\gamma = F(\alpha)$ and nodes combination h are key factors in PNC curve interpolation and shape modeling.

A. *Distribution Functions in PNC Modeling*

Points settled between the nodes are computed using PNC method. Each real number $c \in [a;b]$ is calculated by a convex combination $c = \alpha \cdot a + (1 - \alpha) \cdot b$ for

$$\alpha = \frac{b-c}{b-a} \in [0;1].$$

Key question is dealing with coefficient γ in (1). The simplest way of PNC calculation means $h = 0$ and $\gamma = \alpha$ (basic probability distribution). Then PNC represents a linear interpolation. MHR method [30] is not a linear interpolation. MHR [31] is the example of PNC modeling. Each interpolation requires specific distribution of parameter α and γ (1) depends on parameter $\alpha \in [0;1]$:

$\gamma = F(\alpha)$, F:[0;1]→[0;1], $F(0) = 0$, $F(1) = 1$

and F is strictly monotonic. Coefficient γ is calculated using different functions (polynomials, power functions, sine, cosine, tangent, cotangent, logarithm, exponent, arc sin, arc cos, arc tan or arc cot, also inverse functions) and choice of function is connected with initial requirements and curve specifications. Different values of coefficient γ are connected with applied functions $F(\alpha)$. These functions $\gamma = F(\alpha)$ represent the examples of probability distribution functions for random variable $\alpha \in [0;1]$ and real number $s > 0$:

$\gamma=\alpha^s$, $\gamma=sin(\alpha^s\cdot\pi/2)$, $\gamma=sin^s(\alpha\cdot\pi/2)$, $\gamma=1\text{-}cos(\alpha^s\cdot\pi/2)$, $\gamma=1\text{-}cos^s(\alpha\cdot\pi/2)$, $\gamma=tan(\alpha^s\cdot\pi/4)$, $\gamma=tan^s(\alpha\cdot\pi/4)$, $\gamma=log_2(\alpha^s+1)$, $\gamma=log_2^{\,s}(\alpha+1)$, $\gamma=(2^\alpha-1)^s$, $\gamma=2/\pi\cdot arcsin(\alpha^s)$, $\gamma=(2/\pi\cdot arcsin\alpha)^s$, $\gamma=1\text{-}2/\pi\cdot arccos(\alpha^s)$, $\gamma=1\text{-}(2/\pi\cdot arccos\alpha)^s$, $\gamma=4/\pi\cdot arctan(\alpha^s)$, $\gamma=(4/\pi\cdot arctan\alpha)^s$, $\gamma=ctg(\pi/2-\alpha^s\cdot\pi/4)$, $\gamma=ctg^s(\pi/2\text{-}\alpha\cdot\pi/4)$,

$\gamma=2-4/\pi\cdot arcctg(\alpha^s)$, $\gamma=(2-4/\pi\cdot arcctg\alpha)^s$.

Functions above, used in γ calculations, are strictly monotonic for random variable $\alpha\in[0;1]$ as $\gamma = F(\alpha)$ is probability distribution function. Also inverse functions $F^{-1}(\alpha)$ are appropriate for γ calculations. Choice of function and value s depends on curve specifications and individual requirements. Considering nowadays used probability distribution functions for random variable $\alpha\in[0;1]$ - one distribution is dealing with the range [0;1]: beta distribution. Probability density function f for random variable $\alpha\in[0;1]$ is:

$$f(\alpha)=c\cdot\alpha^s\cdot(1-\alpha)^r\ ,\ s\geq 0, r\geq 0. \qquad (3)$$

When r = 0 probability density function (3) represents $f(\alpha)=c\cdot\alpha^s$ and then probability distribution function F is like $f(\alpha)=3\alpha^2$ and $\gamma = \alpha^3$. If s and r are positive integer numbers then γ is the polynomial, for example $f(\alpha)=6\alpha(1-\alpha)$ and $\gamma = 3\alpha^2-2\alpha^3$. Beta distribution gives us coefficient γ in (1) as polynomial because of interdependence between probability density f and distribution F functions:

$$f(\alpha)=F'(\alpha)\ ,\ F(\alpha)=\int_0^\alpha f(t)dt\cdot \qquad (4)$$

For example (4): $f(\alpha)=\alpha\cdot e^\alpha$ and $\gamma=F(\alpha)=(\alpha-1)e^\alpha+1\cdot$

What is very important in PNC method: two curves (for example a handwritten letter or signature) may have the same set of nodes but different h or γ results in different interpolations (Fig.6-14).

Algorithm of PNC interpolation and modeling (1) looks as follows:

Step 1: Choice of knots p_i at key points.

Step 2: Choice of nodes combination $h(p_1,p_2,\ldots,p_m)$.

Step 3: Choice of distribution $\gamma = F(\alpha)$.

Step 4: Determining values of α: α = 0.1, 0.2…0.9 (nine points) or 0.01, 0.02…0.99 (99 points) or others.

Step 5: The computations (1).

These five steps can be treated as the algorithm of PNC method of curve modeling and interpolation (1).

Curve interpolation has to implement the coefficients γ. Each strictly monotonic function F between points (0;0) and (1;1) can be used in PNC interpolation.

III. Data Modeling and Recognition

PNC method enables signature and handwriting recognition. This process of recognition consists of three parts:

1. Modeling – choice of nodes combination and probabilistic distribution function (1) for known signature or handwritten letters;
2. Unknown writer - choice of characteristic points (nodes) for unknown signature or handwritten word and the coefficients of points between nodes;
3. Decision of recognition - comparing the results of PNC interpolation for known models with coordinates of unknown text.

A. Modeling – the Basis of Patterns

Known letters or symbols ought to be modeled by the choice of nodes, determining specific nodes combination and characteristic probabilistic distribution function. For example a handwritten word or signature "*rw*" may look different for persons A, B or others. How to model "*rw*" for some persons via PNC method? Each model has to be described by the set of nodes for letters "*r*" and "*w*", nodes combination h and a function $\gamma=F(\alpha)$ for each letter. Less complicated models can take $h(p_1,p_2,\ldots,p_m)$ = 0 and then the formula of interpolation (1) looks as follows:

$y(c)=\gamma\cdot y_i+(1-\gamma)y_{i+1}\cdot$

It is linear interpolation for basic probability distribution ($\gamma = \alpha$). How first letter "*r*" is modeled in three versions for nodes combination h = 0 and α=0.1,0.2…0.9? Of course α is a random variable and $\alpha\in[0;1]$.

Person A

Nodes (1;3), (3;1), (5;3), (7;3) and $\gamma = F(\alpha) = \alpha^2$:

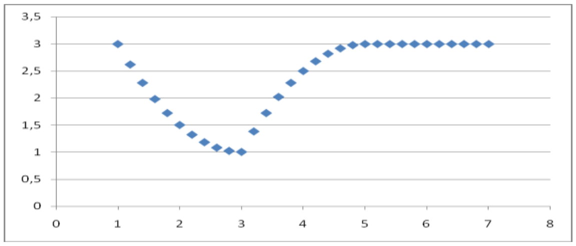


Fig. 1. PNC modeling for nine reconstructed points between nodes.

Person B

Nodes (1;3), (3;1), (5;3), (7;2) and $\gamma = F(\alpha) = \alpha^2$:

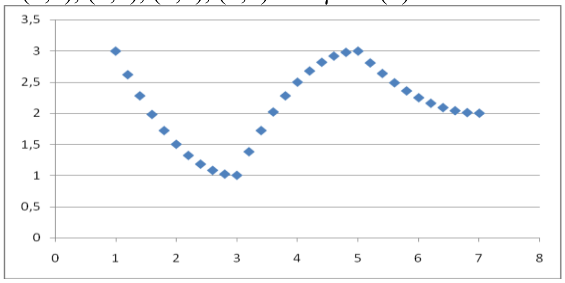


Fig. 2. PNC modeling of letter "*r*" with four nodes.

Person C

Nodes (1;3), (3;1), (5;3), (7;4) and $\gamma = F(\alpha) = \alpha^3$:

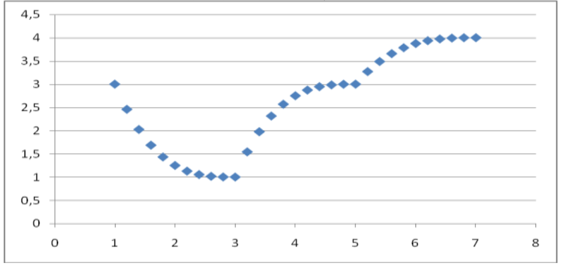


Fig. 3. PNC modeling of handwritten letter "*r*".

These three versions of letter "*r*" (Fig.1-3) with nodes combination h = 0 differ at fourth node and probability distribution functions $\gamma = F(\alpha)$. Much more possibilities of modeling are connected with a choice of nodes combination

$h(p_1,p_2,\ldots,p_m)$. MHR method [32] uses the combination (2) with good features because of orthogonal rows and columns at Hurwitz-Radon family of matrices:

$$h(p_i, p_{i+1}) = \frac{y_i}{x_i} x_{i+1} + \frac{y_{i+1}}{x_{i+1}} x_i$$

and then (1)

$$y(c) = \gamma \cdot y_i + (1-\gamma) y_{i+1} + \gamma(1-\gamma) \cdot h(p_i, p_{i+1},) \cdot$$

Here are two examples of PNC modeling with MHR combination (2).

Person D

Nodes (1;3), (3;1), (5;3) and $\gamma = F(\alpha) = \alpha^2$:

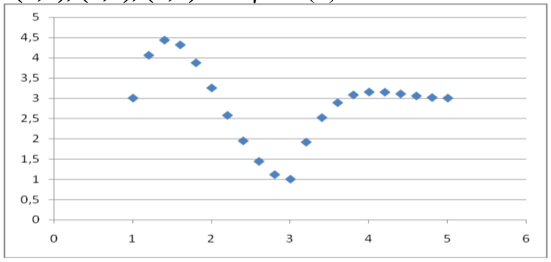


Fig. 4. PNC modeling of letter "*r*" with three nodes.

Person E

Nodes (1;3), (3;1), (5;3) and $\gamma = F(\alpha) = \alpha^{1.5}$:

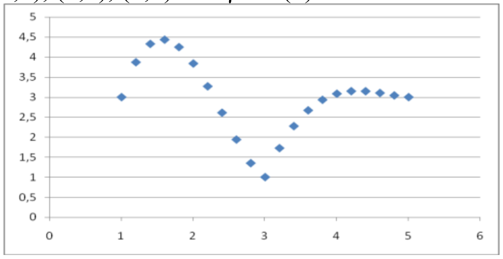


Fig. 5. PNC modeling of handwritten letter "*r*".

Fig.1-5 show modeling of letter "*r*". Now let us consider a letter "*w*" with nodes combination $h = 0$.

Person A

Nodes (2;2), (3;1), (4;2), (5;1), (6;2) and $\gamma = F(\alpha) = (5^\alpha - 1)/4$:

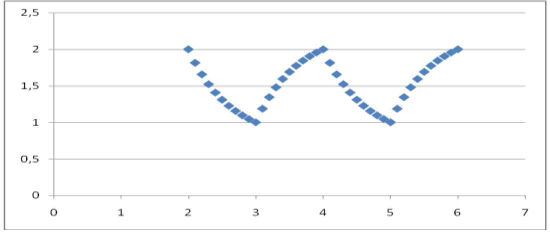


Fig. 6. PNC modeling for nine reconstructed points between nodes.

Person B

Nodes (2;2), (3;1), (4;2), (5;1), (6;2) and $\gamma = F(\alpha) = sin(\alpha\cdot\pi/2)$:

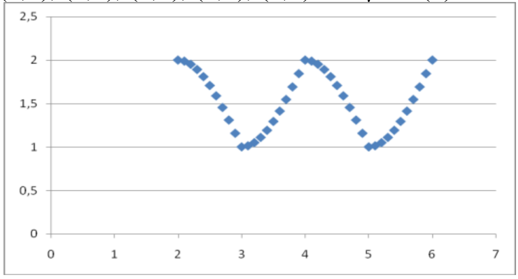


Fig. 7. PNC modeling of letter "*w*" with five nodes.

Person C

Nodes (2;2), (3;1), (4;2), (5;1), (6;2) and $\gamma = F(\alpha) = sin^{3.5}(\alpha\cdot\pi/2)$:

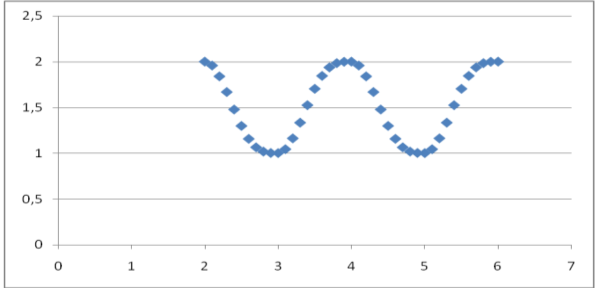


Fig. 8. PNC modeling of handwritten letter "*w*".

These three versions of letter "*w*" (Fig.6-8) with nodes combination $h = 0$ and the same nodes differ only at probability distribution functions $\gamma = F(\alpha)$. Fig.9 is the example of nodes combination h (2) from MHR method:

Person D

Nodes (2;2), (3;1), (4;1), (5;1), (6;2) and $\gamma = F(\alpha) = 2^\alpha - 1$:

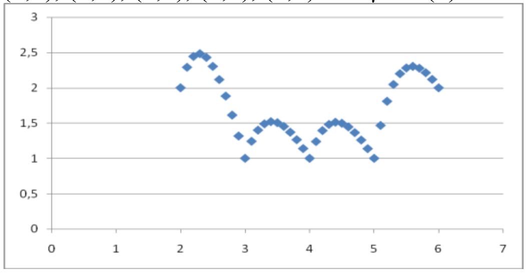


Fig. 9. PNC modeling for nine reconstructed points between nodes.

Examples above have one function $\gamma = F(\alpha)$ and one combination h for all ranges between nodes. But it is possible to create a model with functions $\gamma_i = F_i(\alpha)$ and combinations h_i individually for a range of nodes $(p_i;p_{i+1})$. It enables very precise modeling of handwritten symbol between each successive pair of nodes.

Each person has its own characteristic and individual handwritten letters, numbers or other marks. The range of coefficients x has to be the same for all models because of comparing appropriate coordinates y. Every letter is modeled by PNC via three factors: the set of nodes, probability distribution function $\gamma = F(\alpha)$ and nodes combination h. These three factors are chosen individually for each letter, therefore this information about modeled letters seems to be enough for specific PNC curve interpolation, comparing and handwriting identification. Function γ is selected via the analysis of points between nodes and we may assume $h = 0$ at the beginning. What is very important - PNC modeling is independent of the language or a kind of symbol (letters, numbers or others). One person may have several patterns for one handwritten letter. Summarize: every person has the basis of patterns for each handwritten letter or symbol, described by the set of nodes, probability distribution function $\gamma = F(\alpha)$ and nodes combination h. Whole basis of patterns consists of models $\boldsymbol{S_j}$ for $j = 0,1,2,3\ldots K$.

B. Unknown Author – Points of Handwritten Text

Choice of characteristic points (nodes) for unknown letter or handwritten symbol is a crucial factor in object recognition. The range of coefficients x has to be the same like the x range in the basis of patterns. Knots of the curve (opened or closed) ought to be settled at key points, for example local minimum or maximum (the highest point of the curve in a particular orientation), convexity changing or curvature maximum and at least one node between two successive key points. When the nodes are fixed, each coordinate of every chosen point on the curve (x_0^c,y_0^c), (x_1^c,y_1^c),…, (x_M^c,y_M^c) is accessible to be used for comparing with the models. Then probability distribution function $\gamma = F(\alpha)$ and nodes combination h have to be taken from the basis of modeled letters to calculate appropriate second coordinates $y_i^{(j)}$ of the pattern $\boldsymbol{S_j}$ for first coordinates x_i^c, $i = 0,1,\ldots,M$. After interpolation it is possible to compare given handwritten symbol with a letter in the basis of patterns.

C. Recognition – the Writer

Comparing the results of PNC interpolation for required second coordinates of a model in the basis of patterns with points on the curve (x_0^c,y_0^c), (x_1^c,y_1^c),…, (x_M^c,y_M^c), we can say if the letter or symbol is written by person A, B or another. The comparison and decision of recognition [33] is done via minimal distance criterion. Curve points of unknown handwritten symbol are: (x_0^c,y_0^c), (x_1^c,y_1^c),…, (x_M^c,y_M^c). The criterion of recognition for models $\boldsymbol{S_j} = \{(x_0^c,y_0^{(j)}), (x_1^c,y_1^{(j)}),\ldots, (x_M^c,y_M^{(j)})\}$, j=0,1,2,3…K is given as:

$$\sum_{i=0}^{M}\left|y_i^{\,c} - y_i^{\,(j)}\right| \to \min .$$

Minimal distance criterion helps us to fix a candidate for unknown writer as a person from the model $\boldsymbol{S_j}$.

IV. CONCLUSION

The method of Probabilistic Nodes Combination (PNC) enables interpolation and modeling of two-dimensional curves [34] using nodes combinations and different coefficients γ: polynomial, sinusoidal, cosinusoidal, tangent, cotangent, logarithmic, exponential, arc sin, arc cos, arc tan, arc cot or power function, also inverse functions. Function for γ calculations is chosen individually at each curve modeling and it is treated as probability distribution function: γ depends on initial requirements and curve specifications. PNC method leads to curve interpolation as handwriting or signature identification via discrete set of fixed knots. PNC makes possible the combination of two important problems: interpolation and modeling in a matter of writer identification. Main features of PNC method are:

a) the smaller distance between knots the better;
b) calculations for coordinates close to zero and near by extremum require more attention because of importance of these points;
c) PNC interpolation develops a linear interpolation into other functions as probability distribution functions;
d) PNC is a generalization of MHR method via different nodes combinations;
e) interpolation of L points is connected with the computational cost of rank $O(L)$ as in MHR method;
f) nodes combination and coefficient γ are crucial in the process of curve probabilistic parameterization and interpolation: they are computed individually for a single curve.

Future works are going to: application of PNC method in signature and handwriting recognition, choice and features of nodes combinations and coefficient γ, implementation of PNC in computer vision and artificial intelligence: shape geometry, contour modelling, object recognition and curve parameterization.

REFERENCES

[1] Schlapbach, A., Bunke, H.: Off-line writer identification using Gaussian mixture models. In: International Conference on Pattern Recognition, pp. 992–995 (2006).

[2] Bulacu, M., Schomaker, L.: Text-independent writer identification and verification using textural and allographic features. IEEE Trans. Pattern Anal. Mach. Intell. 29 (4), 701–717 (2007).

[3] Djeddi, C., Souici-Meslati, L.: A texture based approach for Arabic writer identification and verification. In: International Conference on Machine and Web Intelligence, pp. 115–120 (2010).

[4] Djeddi, C., Souici-Meslati, L.: Artificial immune recognition system for Arabic writer identification. In: International Symposium on Innovation in Information and Communication Technology, pp. 159–165 (2011).

[5] Nosary, A., Heutte, L., Paquet, T.: Unsupervised writer adaption applied to handwritten text recognition. Pattern Recogn. Lett. 37 (2), 385–388 (2004).

[6] Van, E.M., Vuurpijl, L., Franke, K., Schomaker, L.: The WANDA measurement tool for forensic document examination. J. Forensic Doc. Exam. 16, 103–118 (2005).

[7] Schomaker, L., Franke, K., Bulacu, M.: Using codebooks of fragmented connected-component contours in forensic and historic writer identification. Pattern Recogn. Lett. 28 (6), 719–727 (2007).

[8] Siddiqi, I., Cloppet, F., Vincent, N.: Contour based features for the classification of ancient manuscripts. In: Conference of the International Graphonomics Society, pp. 226–229 (2009).

[9] Garain, U., Paquet, T.: Off-line multi-script writer identification using AR coefficients. In: International Conference on Document Analysis and Recognition, pp. 991–995 (2009).

[10] Bulacu, M., Schomaker, L., Brink, A.: Text-independent writer identification and verification on off-line Arabic handwriting. In: International Conference on Document Analysis and Recognition, pp. 769–773 (2007).

[11] Ozaki, M., Adachi, Y., Ishii, N.: Examination of effects of character size on accuracy of writer recognition by new local arc method. In: International Conference on Knowledge-Based Intelligent Information and Engineering Systems, pp.1170–1175 (2006).

[12] Chen, J., Lopresti, D., Kavallieratou, E.: The impact of ruling lines on writer identification. In: International Conference on Frontiers in Handwriting Recognition, pp. 439–444 (2010).

[13] Chen, J., Cheng, W., Lopresti, D.: Using perturbed handwriting to support writer identification in the presence of severe data constraints. In: Document Recognition and Retrieval, pp. 1–10 (2011).

[14] Galloway, M.M.: Texture analysis using gray level run lengths. Comput. Graphics Image Process. 4 (2), 172–179 (1975).

[15] Siddiqi, I., Vincent, N.: Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recogn. Lett. 43 (11), 3853–3865 (2010).

[16] Ghiasi, G., Safabakhsh, R.: Offline text-independent writer identification using codebook and efficient code extraction methods. Image and Vision Computing 31, 379–391 (2013).

[17] Shahabinejad, F., Rahmati, M.: A new method for writer identification and verification based on Farsi/Arabic handwritten texts, Ninth

International Conference on Document Analysis and Recognition (ICDAR 2007), pp. 829–833 (2007).

[18] Schlapbach, A., Bunke, H.: A writer identification and verification system using HMM based recognizers, Pattern Anal. Appl. 10, 33–43 (2007).

[19] Schlapbach, A., Bunke, H.: Using HMM based recognizers for writer identification and verification, 9th Int. Workshop on Frontiers in Handwriting Recognition, pp. 167–172 (2004).

[20] Marti, U.-V., Bunke, H.: The IAM-database: an English sentence database for offline handwriting recognition, Int. J. Doc. Anal. Recognit. 5, 39–46 (2002).

[21] Collins II, G.W.: Fundamental Numerical Methods and Data Analysis. Case Western Reserve University (2003).

[22] Chapra, S.C.: Applied Numerical Methods. McGraw-Hill (2012).

[23] Ralston, A., Rabinowitz, P.: A First Course in Numerical Analysis – Second Edition. Dover Publications, New York (2001).

[24] Zhang, D., Lu, G.: Review of Shape Representation and Description Techniques. Pattern Recognition 1(37), 1-19 (2004).

[25] Schumaker, L.L.: Spline Functions: Basic Theory. Cambridge Mathematical Library (2007).

[26] Dahlquist, G., Bjoerck, A.: Numerical Methods. Prentice Hall, New York (1974).

[27] Jakóbczak, D.: 2D and 3D Image Modeling Using Hurwitz-Radon Matrices. Polish Journal of Environmental Studies 4A(16), 104-107 (2007).

[28] Jakóbczak, D.: Shape Representation and Shape Coefficients via Method of Hurwitz-Radon Matrices. Lecture Notes in Computer Science 6374 (Computer Vision and Graphics: Proc. ICCVG 2010, Part I), Springer-Verlag Berlin Heidelberg, 411-419 (2010).

[29] Jakóbczak, D.: Curve Interpolation Using Hurwitz-Radon Matrices. Polish Journal of Environmental Studies 3B(18), 126-130 (2009).

[30] Jakóbczak, D.: Application of Hurwitz-Radon Matrices in Shape Representation. In: Banaszak, Z., Świć, A. (eds.) Applied Computer Science: Modelling of Production Processes 1(6), pp. 63-74. Lublin University of Technology Press, Lublin (2010).

[31] Jakóbczak, D.: Object Modeling Using Method of Hurwitz-Radon Matrices of Rank k. In: Wolski, W., Borawski, M. (eds.) Computer Graphics: Selected Issues, pp. 79-90. University of Szczecin Press, Szczecin (2010).

[32] Jakóbczak, D.: Implementation of Hurwitz-Radon Matrices in Shape Representation. In: Choraś, R.S. (ed.) Advances in Intelligent and Soft Computing 84, Image Processing and Communications: Challenges 2, pp. 39-50. Springer-Verlag, Berlin Heidelberg (2010).

[33] Jakóbczak, D.: Object Recognition via Contour Points Reconstruction Using Hurwitz-Radon Matrices. In: Józefczyk, J., Orski, D. (eds.) Knowledge-Based Intelligent System Advancements: Systemic and Cybernetic Approaches, pp. 87-107. IGI Global, Hershey PA, USA (2011).

[34] Jakóbczak, D.: Curve Parameterization and Curvature via Method of Hurwitz-Radon Matrices. Image Processing & Communications- An International Journal 1-2(16), 49-56 (2011).

Dariusz Jacek Jakóbczak was born in Koszalin, Poland, on December 30, 1965. He graduated in mathematics (numerical methods and programming) from the University of Gdansk, Poland in 1990. He received the Ph.D. degree in 2007 in computer science from the Polish – Japanese Institute of Information Technology, Warsaw, Poland.

From 1991 to 1994 he was a civilian programmer in the High Military School in Koszalin. He was a teacher of mathematics and computer science in the Private Economic School in Koszalin from 1995 to 1999. Since March 1998 he has worked in the Department of Electronics and Computer Science, Technical University of Koszalin, Poland and since October 2007 he has been an Assistant Professor in the Chair of Computer Science and Management in this department. His research interests connect mathematics with computer science and include computer vision, artificial intelligence, shape representation, curve interpolation, contour reconstruction and geometric modeling, numerical methods, probabilistic methods, game theory, operational research and discrete mathematics.

Polar, Spherical and Orthogonal Space Subdivisions for an Algorithm Acceleration: O(1) Point-in-Polygon/Polyhedron Test

Vaclav Skala

Abstract**—**Acceleration of algorithms is becoming a crucial problem, if larger data sets are to be processed. Evaluation of algorithms is mostly done by using computational geometry approach and evaluation of computational complexity. However in today's engineering problems this approach does not respect that number of processed items is always limited and a significant role plays also speed of read/write operations. One general method how to speed up an algorithm is application of space subdivision technique and usually the orthogonal space subdivision is used. In this paper non-orthogonal subdivisions are described. The proposed approach can significantly improve memory consumption and run-time complexity. The proposed modified space subdivision techniques are demonstrated on two simple problems Point-in-Convex Polygon and Point-in-Convex Polyhedron tests.

Keywords**—**containment test, orthogonal space subdivision, polar space subdivision, spherical space subdivision, point-in-polygon, point-in-polyhedron.

I. Introduction

ALGORITHM'S evaluation is quite complicated issue. One approach is based on the Computational Geometry (CG) approach, which deals with the complexity issues expressed by $O(g(N))$ terms, where $g(N)$ is a function, typically. $g(N) = N \lg N$ etc. and N is a number of processed items. It should be noted that $O(N \lg N)$ actually means that the "measured" complexity will be probably something like

$$O(N \lg N) = C_1 N + C_2 \lg N + C_3 \lg^2 N + \cdots + C_k \lg N$$

where constants $C_1, C_2, \dots, C_{k-1}$ might .be very high. The CG approach evaluates the complexity for $N \to \infty$, which is not a typical engineering problem where algorithms have to optimize for $N \in \langle n_1, n_2 \rangle$. Also the CG approach does not handle problems related to read/write time complexity, caching and speed of the data transfer itself etc. All that mean, that the algorithm with a better computational complexity in the CG sense might be actually slower in real engineering applications for the given interval of N, i.e. $N \in \langle n_1, n_2 \rangle$.

Algorithms do have their "optimal" complexity, which is actually the lowest complexity without preprocessing. Of course, for some specific data sets the algorithm might be actually faster.

To speed-up computation a parallel processing can be used. However, it should be noted that parallel processing is actually a "brute force" approach. If we have, e.g. 95% of the code, which can be made in parallel, the final speed-up for an infinite number of processors according to the Amdahl's law will be 20, i.e. the final computation will be 20 times faster but we spend also *infinite resources*!

In the case that many items are processed, the constant part can be preprocessed to speed-up the actual run-time. Space subdivision technique is one of the mostly used approaches beside of the parallel processing.

II. Space Subdivision

Space subdivision techniques are used in many algorithms across all computational fields. In the following basic types of space subdivision techniques are presented with their fundamental properties.

A. Orthogonal space subdivision

Orthogonal space subdivision is the simplest and mostly used subdivision techniques. For simplicity, let us consider the E^2 case. If we know the Axis Aligned Bounding Box (AABB) of the data set, the AABB is split regularly in one axis, resp. in two axes to smaller cells, where the data, resp. their references are to be stored. No we are getting M, resp. $M \times M$, cells, where N items are stored. It means that each item has to be examined and determined a cell, to where the data item is to be stored. So the preprocessing is of the $O(MN)$, resp. $O(M^2 N)$, complexity. In geometric problems an object can interfere with more cells. In the E^3 case the preprocessing is $O(M^3 N)$ complexity.

However, there is a possibility how to decrease the memory complexity to $O(N\,dM)$, where d is a dimensionality. However, this technique slightly extend the run-time, details can be found in [4].

In spite of the simplicity, there are some significant disadvantages, especially:

- in the interval of data is unknown, all data have to be read and AABB has to be found
- if large data are to be processed, the preprocessing time is very high due to reading data from an external memory

This work was supported by the Ministry of Education of the Czech Republic, project No.LH12181.

Vaclav Skala is with the University of West Bohemia, Faculty of Applied Sciences, Department .of Computer Science and Engineering, Center of Computer Graphics and Visualization, Univerzitni 8, CZ 30614 Plzen, Czech Republic (http://www.VaclavSkala.eu, http://Graphics.zcu.cz)..

- if only one axis is subdivided, the preprocessing and run-time is quite data sensitive
- it is easy to show that in many cases irregular splitting to cells is needed

However there are some simple problems like

- point-in-convex polygon or point-in-convex polyhedron tests
- convex hull construction in E^2, resp. E^3

The orthogonal space subdivision causes some problems in preprocessing due to increase of preprocessing complexity with some influences to the run-time computations.

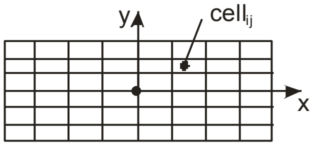


Fig.1 Orthogonal space subdivision

in the case of a convex polygon, the space subdivision is strongly dependent on the polygon rotation, too.

B. Polar space subdivision

Polar space subdivisions are used in related problems as well. Usually the space is split to regular angular sectors. It means that if the point coordinates are given in the Cartesian coordinates, that an angle and radius has to computed using formulas:

$$r = \sqrt{(x - x_A)^2 + (y - y_A)^2}$$
$$\varphi = \arctan 2(x, y)$$

where (x_A, y_A) is a reference point of the "virtual" origin.

However those formulae do have some weak points, especially:

- radius r has to computed in double precision if coordinates are given is a single precision
- function $\arctan 2(x, y)$ is in *principle imprecise* and giving the relevant angle $\varphi \in \langle 0, 2\pi)$

In the case of the polar space subdivision the computation of r, φ is still quite simple in the comparison with the spherical space subdivision.

C. Spherical space subdivision

In the case of E^3 the situations seems to be very similar, however, the formulae are more complex and usually the following formulae are used:

$$r = \sqrt{x^2 + y^2 + z^2}$$
$$\theta = \cos^{-1}\left(\frac{z}{\sqrt{x^2 + y^2 + z^2}}\right)$$
$$\varphi = \tan^{-1}\left(\frac{y}{x}\right)$$

In this case the imprecision of computation is quite high and may lead to wrong selection a cell in the space subdivision.

D. Cylindrical space subdivision

In the case of the cylindrical space subdivision the situation is simple and is made as a simple modification of the polar space subdivision and subdivision has to be made for the z coordinate, too.

III. Proposed modifications

If the space subdivision is to be effective not only in the preprocessing stage but also during the run-time, all computations have to be as simple as possible. As space subdivision is not actually strictly defined method, it might be imprecise, but reliable for the given purpose giving correct answers all the time.

A. Modified Polar Space Subdivision

Modification of the original polar space subdivision is quite simple as we do not need strictly regular subdivision, but subdivision close to regular with simple computation. It can be seen that the function $tan\,\varphi$ is "nearly" linear for $\varphi \in \langle 0, \pi/_4 \rangle$, so the edges of the "virtual" square can be split to segments with the same Δx, or Δy in order to simplify computation of the angular section in which the given point$\boldsymbol{x}$ lies. If $\Delta x = 1$ and $\Delta y = 1$

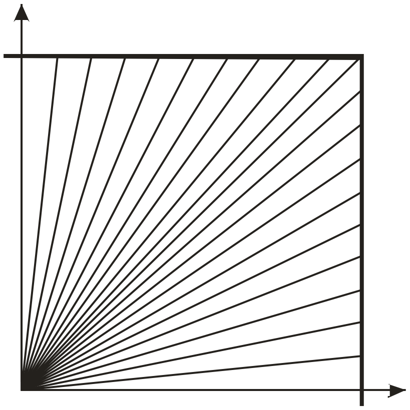


Fig.2 "Virtual"square - 1st quadrant split irregularly

The relevant index i of a segment in the case of the 1st octant for the given point $\boldsymbol{x} = \langle x, y \rangle$ is computed as:

$$i = \frac{y}{x} * m$$

where: m is e number of segments on the "virtual" square edge. Of course, if the point is in the 2nd octant the index has to be computed as:

$$i = \frac{x}{y} * m$$

in order to keep high precision. Other octants are computed similarly.

It can be seen that the computation is much simpler than in the original polar space subdivision approach. Now, there is a question is a similar approach can be applied in the E^3 case, i.e. for the spherical space subdivision, as well.

B. Modified Spherical Space Subdivision

The modification of the polar space subdivision is quite simple as instead of edges split to segments, the faces of a "virtual" cube are split to cells. The advantage of this proposed approach is its simplicity of the index of a cell computation as Δx, Δy, Δz are constant. Indexes of the relevant cells are again quite simple as in principle:

$$i = \frac{y}{x} * m \qquad j = \frac{z}{x} * m$$

and similarly for all faces of the "virtual" cube.

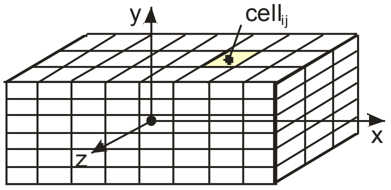


Fig.3 "Virtual" cube splitting

In the E^3 case, there is a small complication with indexes as we have to map indexes of cells of the "virtual" cube surface to two dimensional data structure.

IV. Point-in-Convex polygon with $O(1)$ complexity

Containment test Point-in-Convex Polygon is a fundamental test in many applications. It is quite a simple test and two algorithms are mostly used. The first one with $O(N)$ complexity is extremely simple to implement. It relies on the edges orientation, e.g. all normal vectors of edges are consistently oriented "out" or "in". This algorithm uses actually a separation by half-planes.

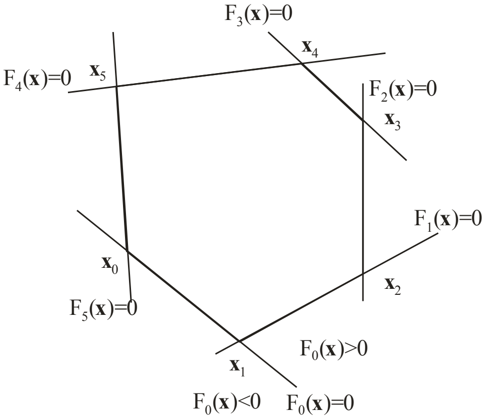


Fig.4 Point-in-Convex Polygon test with $O(N)$ complexity

The second one with $O(\lg N)$ complexity is relatively simple to implement. It relies on the edges orientation as well, but uses vertices ordering, i.e. clockwise or anticlockwise, and binary subdivision on "axis of indexes", i.e. indexes of vertices are binary subdivided, see [12] for details.

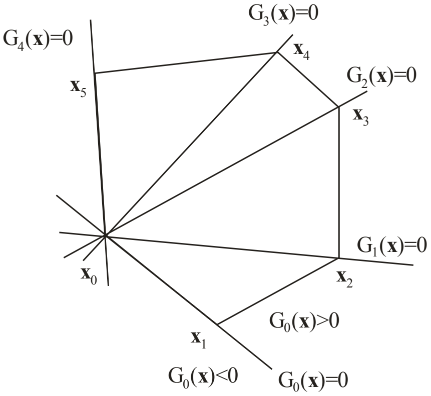


Fig.5 Point-in-Convex Polygon test with $O(\lg N)$ complexity

If the preprocessing can be used, e.g. if many points against a convex constant polygon are to be tested, an algorithm based on one dimensional space subdivision in y axis has been derived recently, Fig.6, and was described in [12]. However this approach has the following problems:

- interval of y axis of all vertices has to be known,
- the memory consumption is highly dependent on geometrical properties, as the width of a slab must be smaller than the smallest edge length in the y coordinate, Fig.7.

It can be seen that the function $\tan\varphi$ is "close" to linear for $\varphi \in \langle 0, \pi/4 \rangle$. In the case of space subdivision we actually do not need *exact regularity*, but "close enough" is acceptable, especially in the case that it leads to some positives.

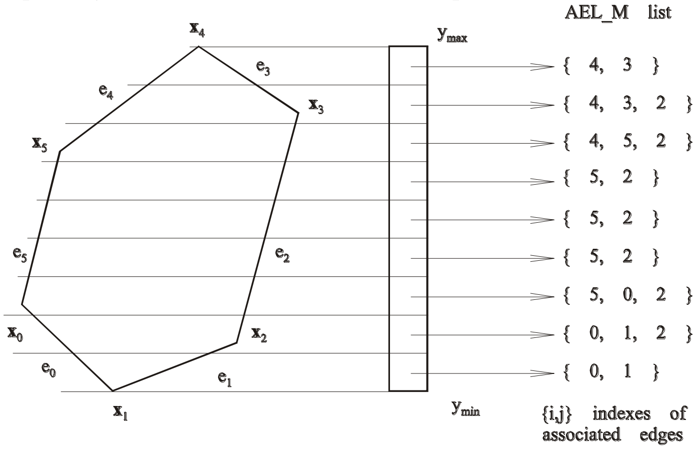


Fig.6 Point-in-Convex Polygon test with $O(1)$ complexity

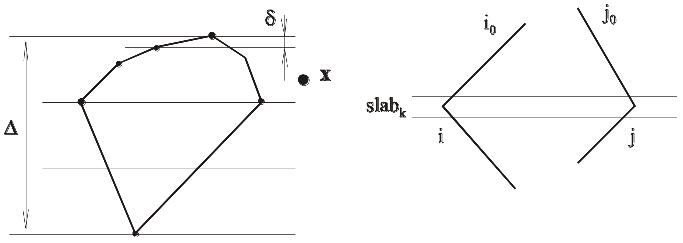


Fig.7 Influence of geometrical properties

Let us consider a convex polygon and let us made a "virtual" square containing the given convex polygon, see Fig.8, for a simplicity (this assumption is not a critical one).

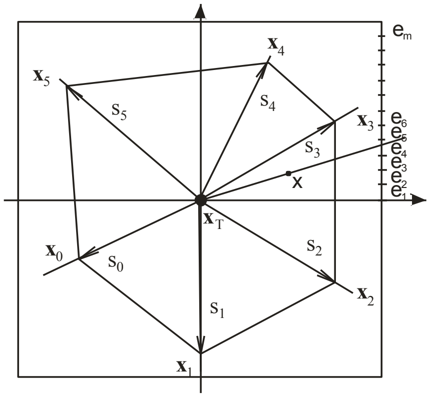


Fig.7 "Virtual" square and the given polygon

Now for the each section on the square edge the relevant edge of the given convex polyhedron can be determined. Due to the limited "resolution", i.e. length of a section, two edges have to be considered if the angular section contains a vertex.

Now, for the given point $\boldsymbol{x}$ the relevant angular section is determined and simple test determine if the point $\boldsymbol{x}$ is inside or outside to the given convex polygon. As the test is based on a half-plane separation test, only 2 addition and 2 multiplication operations are needed.

It can be seen that the test is clearly of $O(1)$ run-time complexity. It can be proved that the section length is determined by the shortest edge length of the given convex polygon.

It should be noted that the proposed algorithm does not need to know AABB box as the "reference" point $\boldsymbol{x}_T$, i.e. the center of the "virtual square" can be computed as:

$$\boldsymbol{x}_T = \left(\boldsymbol{x}_0 + \boldsymbol{x}_{\lfloor N/2 \rfloor}\right)/2$$

The size of the "virtual square" is not actually related to the size of the convex polygon.

V. Point-in-Convex polyhedron with $O(1)$ complexity

The modification for the E^3 case, i.e. for the case of convex polyhedron, the modification is now straightforward. However, there is e a little bit more complex test as we have to test the given point $\boldsymbol{x}$ against planes of the given convex polyhedron incidenting with the relevant spatial angular segment, which is actually of a "pyramid" shape, Fig.3.

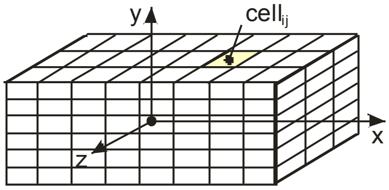


Fig.8 "Virtual" cube for the given polyhedron

Of course, there, there is a legitimate question. How the proposed approach will handle non-self intersecting convex polygon and polyhedron. It seems to that in the non-convex case, the proposed approach is applicable as well, however the expected complexity is of $O(H)$, where H is a number of edges, resp. faces incidenting with the angular segment.

VI. Conclusion

In this paper new modification of space subdivision techniques based on polar and spherical subdivisions are presented. The proposed approaches are simple to implement and they are robust as well. As a direct consequence algorithms for point-in-convex polygon and point-in-convex polyhedron with $O(1)$ run-time complexity using polar or spherical space subdivision in the preprocessing were briefly described. The algorithms are convenient for application in cases when many points are tested and the given polygon, resp. polyhedron, is constant. It is expected, that the proposed modified polar and spherical space subdivision is widely applicable in many geometry related problems, as well.

Acknowledgment

The author would like to thank to colleagues at the University of West Bohemia in Plzen and VSB-Technical University, Ostrava for fruitful discussions and to anonymous reviewers for their comments and hints which helped to improve the manuscript significantly.

References

[1] Haines,E.: Point in polygon strategies. Graphics Gems IV, Ed. Heckbert,P., Academic Press, pp.24-46,1994.

[2] Gombosi,M., Zalik,B.: Point-in-polygon tests for geometric buffers. Comp.& Geosciences 31 (10), 1201–1212, 2005

[3] Hormann,K, Agathos,A.: The point in polygon problem for arbitrary polygons, Computational Geometry: Theory and Applications, 2001,20(3),pp.131-144, 2001

[4] Huang,C.W.,Shih,T.Y.: On the complexity of point-in-polygon algorithms, Comp.& Geosciences 23(1),109–118, 1997.

[5] Jime´nez, J.J., Feito, F.R., Segura, R.J.: A new hierarchical triangle-based point-in-polygon data structure. Comp.&Geosciences 35 (9), 1843–1853, 2009

[6] Jiménez JJ, Feito FR, Segura RJ. Robust and optimized algorithms for the point-in-polygon inclusion test without pre-processing. Computer Graphics Forum, 28(8), pp.2264-2274, 2009

[7] Lane,J., Megedson,B., Rarick,M.: An efficient Point in Polyhedron Algorithm, Computer Vision, Graphics and Image Processing, Vol.26, pp.118-125,1984

[8] Li,J., Wang.W.,Wu,E.: Point-in-polygon tests by convex decomposition, Comp.&Graphics, 31, (4), pp.636-648, 2007

[9] Li,J., Wang,W.-C.: Point-in-polygon test method based on center points of grid, Ruan Jian Xue Bao/Journal of Software, Vol.23, No.9, 2012, pp. 2481-2488, 2012

[10] Martinez,F., Rueda,A.J., Feito,F.R.: The multi-L-REP decomposition and its application to a point-in-polygon inclusion test. Comp.&Graphics 30 (6), 947–958, 2006.

[11] Skala,V.: Algorithms for Line and Plane Intersection with a Convex Polyhedron with O(sqrt(N)) Complexity in E3, SIGGRAPH Asia 2014, ISBN:978-1-4503-2792-3, Shenzen, China, 2014

[12] Skala,V.: Trading Time for Space: an O(1) Average time Algorithm for Point-in-Polygon Location Problem. Theoretical Fiction or Practical Usage? Machine Graphics and Vision, Vol.5., No.3., pp. 483-494, 1996

[13] Skala,V.: Line Clipping in E2 with O(1) Processing Complexity, Comp.&Graphics, Vol.20, No.4, pp.523-530, 1996.

[14] Skala,V.: Memory Saving Technique for Space Subdivision Technique, Machine Graphics and Vision, Vol.2, No.3, pp.237-250, , ISSN 1230-0535, 1993.

[15] Solomon,K.: An Efficient Point-in-Polygon Algorithm, Comp.&Geosciences, Vol.4, pp.173-178,1978

[16] Wang,W.C., Li,J., Wu,E.H.: 2D point-in-polygon test by classifying edges into layers, Comp.& Graphics, Vol.29, No.3., pp.427-439, 2005

[17] Yang S, Yong JH, Sun J, Gu H, Paul JC. A point-in-polygon method based on a quasi-closest point. Computers & Geosciences, 36(2):205-213, 2010

Vaclav Skala is a Full professor of Computer Science at the University of West Bohemia, Plzen, Czech Republic where he is a full professor of Computer Science. He is the Head of the Center of Computer Graphics and Visualization since 1996.

Vaclav Skala is a member of editorial board of The Visual Computer (Springer), Computers and Graphics (Elsevier), Machine Graphics and Vision (Polish Academy of Sciences), The International Journal of Virtual Reality (IPI Press, USA) and the Editor in Chief of the Journal of WSCG. He has been a member of several international program committees of prestigious conferences and workshops. He is a member of Eurographics Association and he became a Fellow of the Eurographics Association in 2010.

Vaclav Skala has published over 200 research papers in scientific journal and at international research conferences. His current research interests are computer graphics, visualization and mathematics, especially geometrical algebra, algorithms and data structures.

Details can be found at http://www.VaclavSkala.eu

Cluster flow modeling on multi-lane supporters

Pavel Strusinskiy, *Fellow, MADI*

***Abstract*—The cluster flow model on multi-lane supporters is researched. The scheme of clusters movement on multi-lane supporter is described. The system of cluster flow parameters is obtained. The problem statements are formulated, theoretical results are obtained and proved. The software, created by author, reconstructs movement of clusters on two-lane closed supporter (*belt*) and determines the flow characteristics.**

***Index Terms*—Cluster, NODE-model, totally-connected flow, multi-lane.**

I. Introduction

A. Definition of the cluster

Cluster is a state movement of two or more particles (cars, cells and etc.), distributed uniformly in car-following model. In cluster flow model cluster is a moving rectangle, model of moving batch of particles, distributed uniformly, with constant velocity v on the lane with length d in the flow with density y (Fig. 1). Then M is a cluster mass. The first definition of the cluster was given in [2]. The velocity is a function of flow density: $f(y)$[1],[3],[4]. For example,

$$v(y) = v_0 \cdot (\frac{y_{max} - y}{y_{max}})^{\alpha} \tag{1}$$

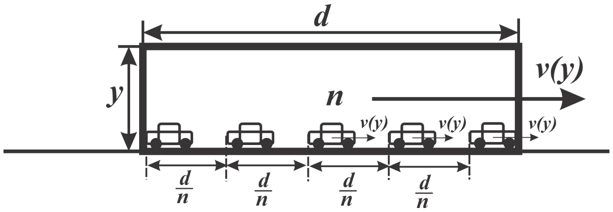


Fig. 1. Cluster

B. Definition of supporter

The Ring is a closed lane with length d_0 (Fig. 2). Clusters are moving on the it one by one and don't change the lane.

The Belt is a system of n parallel rings (Fig. 3). Clusters are moving on the it one by one and can change their lane.

II. Cluster flow on multi-lane supporter

There are two lanes for movement $X^{(1)}$ and $X^{(2)}$ in the same direction from left to right. On the lane $X^{(1)}$ there are two clusters: $[x_1, x_2]$ and $[x_3, x_4]$, $x_2 < x_3$, with densities $y_1 < y_2$ and velocities $v_2 > v_1$. Cluster $[x_1, x_2]$ is the *fast* cluster or is the *overtaking* cluster. Other cluster is the *slow* cluster. After some period of time the fast cluster overtakes the slow one, $x_2(t_0) = x_3(t_0)$ (Fig. 5).

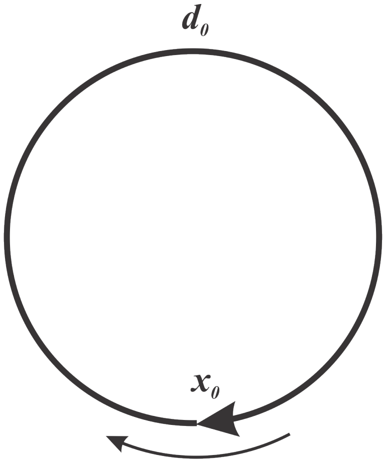


Fig. 2. Ring

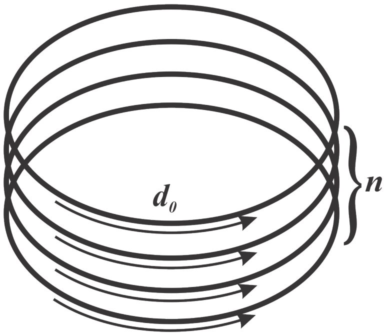


Fig. 3. Belt

There are 2 scenarios of clusters behavior: cluster consecutively [5] or synchronously changes it's lane.

A. Consecutive change of the lane

1) if on the neighbor lane $X^{(2)}$ there is a coordinate x_2, that is not blocked by other clusters, then the fast cluster overtakes the slow one through the coordinate x_2;
2) if on the neighbor lane $X^{(2)}$ there is a *blocking* cluster with parameters y_3, d_3, v_3, that moves through the coordinate x_2 and blocks it, then the fast cluster moves behind the slow one without interaction with the velocity v_1;
3) when the neighbor lane $X^{(2)}$ will be free (the blocking cluster will leave the coordinate x_2), the fast cluster will overtake the slow one through the coordinate x_2 on the lane $X^{(2)}$;
4) the fast cluster can divide into parts, if it won't finish the overtake process completely, before the moment, when

P. Strusinskiy is with the Department of High Mathematics, Moscow Automobile and Road State Technical University (MADI), Moscow, Leningradskiy prosp. 64, 125319 Russia e-mail: perssot@gmail.com

the coordinate x_2 is been blocked by another blocking cluster.

B. Synchronous change of the lane

1) if on the neighbor lane $X^{(2)}$ there is a segment $[x_1, x_2]$, that is not blocked by other clusters, then the fast cluster overtakes the slow one through the segment $[x_1, x_2]$;
2) if on the neighbor lane $X^{(2)}$ there is a blocking cluster with parameters y_3, d_3, v_3, that moves through the segment $[x_1, x_2]$ and blocks it partly or completely, then the fast cluster moves behind the slow one without interaction with the velocity v_1;
3) when the neighbor lane $X^{(2)}$ will be free (the blocking cluster will leave the segment $[x_1, x_2]$ completely), the fast cluster will overtake the slow one through the segment $[x_1, x_2]$ on the lane $X^{(2)}$;
4) the fast cluster can't divide into parts in this scenario, because it changes the lane completely at once, when the segment $[x_1, x_2]$ is free.

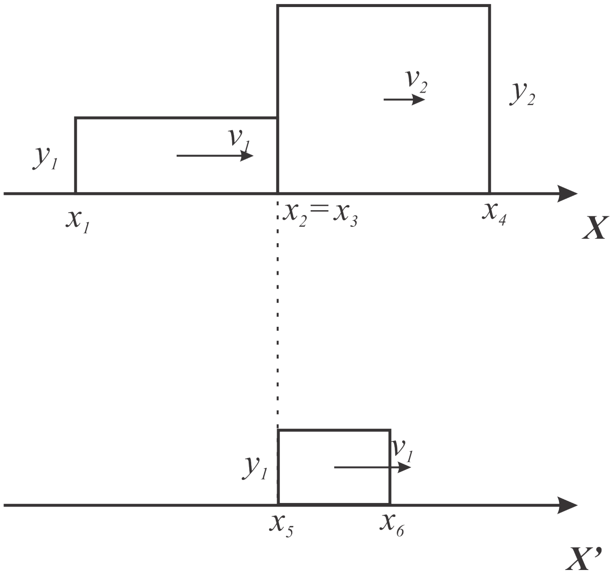


Fig. 4. Overtake process

III. Problem Statement

A. Problem with scenario 2.1

There are two lanes $X^{(1)}$ and $X^{(2)}$ both with length d_0. On the lane $X^{(1)}$ there are n *slow* clusters with the velocities v_1, the densities y_1 and the lengths d_1. On the lane $X^{(2)}$ there are m *fast* clusters with the velocities $v_2 > v_1$, the densities y_2 and the lengths d_2. Slow and fast clusters are moving only on their initial lanes. The slow clusters are moving on the same distance l_1 between each other, the distance between fast clusters is equal to l_2. On both lanes $X^{(1)}$ and $X^{(2)}$ there is 1 *special* cluster with the velocity $v_3 > v_2$, the density y_3 and the length d_3. This cluster is the fastest and it will move through other clusters, using the **scenario 2.1**.

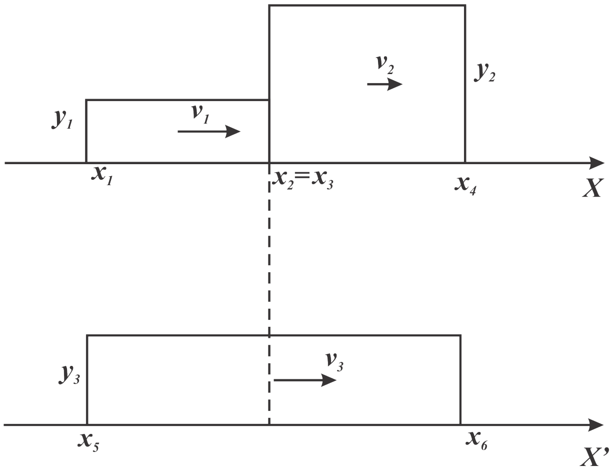


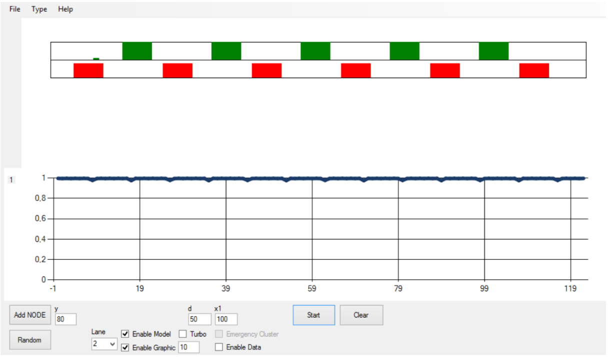
Fig. 5. Fast cluster behind the slow one

B. Problem with scenario 2.2

There are two lanes $X^{(1)}$ and $X^{(2)}$ both with length d_0. On the lane $X^{(1)}$ there are n *slow* clusters with the velocities v_1, the densities y_1 and the lengths d_1. On the lane $X^{(2)}$ there are m *fast* clusters with the velocities $v_2 > v_1$, the densities y_2 and the lengths d_2. Slow and fast clusters are moving only on their initial lanes. The slow clusters are moving on the same distance l_1 between each other, the distance between fast clusters is equal to l_2. On both lanes $X^{(1)}$ and $X^{(2)}$ there is 1 *special* cluster with the velocity $v_3 > v_2$, the density y_3 and the length d_3. This cluster is the fastest and it will move through other clusters, using the **scenario 2.2**.

C. Numerical parameters

1) The velocity of the special cluster v_3;
2) configuration of the clusters.



y 80

d 50

x1 100

Lane 2

☑ Enable Model ☐ Turbo ☐ Emergency Cluster

☑ Enable Graphic 10 ☐ Enable Data

Fig. 6. The problem statement in the program

IV. Computer Simulation

The developed *NODE model* software allows to create a simulation of cluster movement on the *two-lane Belt*.

A. Data storage about clusters

1) Cluster $i = 1..n$ has three parameters: the density of cluster $y_{1..n}$, the length $d_{1..n}$ and the location of the cluster on the lane $x0_{1..n}$;

2) after pressing the button Add the data about clusters is stored in the one-dimensional array with type $TRoad$. With each button press the number of the clusters n is increasing, the computer adds the data about cluster in the array;
3) every element of the array $TRoad$ has data about number of clusters on the lane, the lengths, densities, velocities and the locations of the each cluster on the lane;
4) the computer calculates the velocity of the each cluster by the equation: $v_i = v_0 \cdot (1 - \frac{y_i}{y_{max}})$.

B. Visualization of the clusters and lanes

1) In program, the lane $j = 1..m$ is a rectangle with height y_{max} and length d_0. The number of lanes m is defined manually in program;
2) computer creates m elements in array $TRoad$, one for each lane;
3) the data about clusters can be specified manually or automatically in program;
4) computer creates a cluster like a filled rectangle with height y and length d.

C. Manual location of clusters

1) After the required data about cluster is selected and confirmed and the button Add is pressed, the clusters data will be added to array $TRoad$ and the cluster will be showed on the lane with the selected number in the program;
2) the location of the cluster $x_{1..n}$ is defined by equation: $x_i = x_{i-1} + d_i + x0_i$. If this is the first cluster, then $x_i = x0_i$.
3) there is condition for clusters: $(x_i + d_i) < d_0$. If this condition is not satisfied, this means, that the cluster is located outside the lane, and it won't be added to array and won't be showed on the lane.

D. Automatical location of clusters

This option is required for uniform distribution of clusters on the lane for the experiment.

1) For automatical location of clusters the maximal values of cluster density, length and its location are required:$y_{up}, d_{up}, x0_{up}$. There is an option to fix this values if the same density, length or location of the clusters are required;
2) after pressing the $Random$ button the program starts the automatical location of the clusters;
3) for each i cluster the computer automatically sets the density as follows: $y_i = Random(0, y_{up})$, or $y_i = y_{up}$ if the value y_{up} was fixed. Function $Random(a, b)$ returns random value between a and b if: $a > b$;
4) similarly to previous item computer sets the length and the location of the cluster. For the value $x0_i$ there is additional condition: $(x_i + d_i) < d_0$. If this condition is not satisfied, then computer restarts the selection process of the value $x0_i$.

E. Clusters movement

1) If the button $Start$ is pressed, then the computer starts the clusters movement;
2) the program starts the Timer, that repeats the next operation with the time interval t_0:
   a) the location coordinate x_i of the i cluster changes to value:$x_i = x_i + v_i$;
   b) if $x_i > d_0$, then $x_i = d_0 - x_i$;
   c) after this computer clears the picture with lanes and clusters and shows it again with new locations.
3) if the $Stop$ button is pressed, then the computer stops the Timer.

F. Process of changing lanes

1) During the clusters movement computer examines the following condition for each i cluster: $x_i + d_i = x_{i+1}$;
2) if this condition is not satisfied, then the cluster continue to move;
3) otherwise:
   a) the velocity of the i cluster is: $v_i = v_{i+1}$;
   b) during the next step of the Timer on the lane $j + 1$ the cluster with the number $k = n + 1$, density $y_k = y_i$, length $d_k = v_i - v_{i+1}$ and with the location $x_k = x_{i+1}$ is created. The velocity of this cluster is $v_k = v_i$;
   c) every Timer step the clusters length d_k will increase by the value: $v_i - v_{i+1}$, the length of the cluster d_i will decrease by the value: $v_i - v_{i+1}$.
4) if the length of the i cluster $d_i = 0$, then this cluster will be deleted for array $TRoad$, the number of cluster on the lane j will decrease by 1.

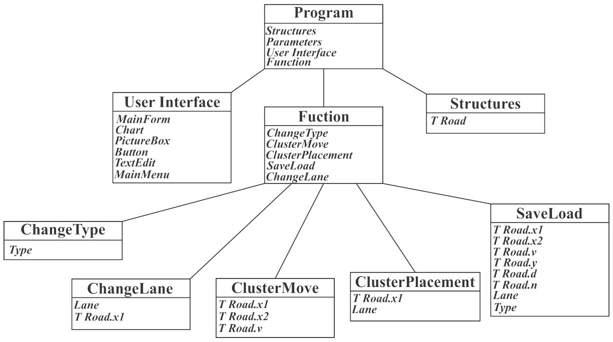


Fig. 7. The scheme of the program

V. Results

A. Synergy in scenario 2.2

Let be $v_1 = 0$, then $y_1 = y_{max}$, and:

$$\Delta v = v_3 - v_2, \tag{2}$$

where Δv is a velocity of the special cluster relative to the velocity of the fast cluster.

Statement 1: Let be $d_3 < min(l_1, l_2)$, $d_2 < l_1$. If the following condition is satisfied:

$$\frac{v_3}{v_2} > \frac{l_1}{l_1 - (d_2 + d_3)} > 0, \tag{3}$$

then the special cluster moves with the velocity: v_3.

Proof: Let be:

$$t_1 = \frac{d_2 + d_3}{\Delta v}, t_2 = \frac{l_1 - (d_2 + d_3)}{v_2}, \tag{4}$$

where t_1 is a time, during that the special cluster overtakes the fast cluster, t_2 is a time, during that the fast cluster blocks the lane $X^{(2)}$, and after this moment the special cluster can not overtake the fast cluster. Then, if the special cluster needs to overtake the fast one, the following condition must be satisfied: $0 < t_1 < t_2$. After setting the values t_1 and t_2 in this condition:

$$\frac{d_2 + d_3}{\Delta v} < \frac{l_1 - (d_2 + d_3)}{v_2}$$

$$\frac{d_2 + d_3}{l_1 - (d_2 + d_3)} < \frac{\Delta v}{v_2}$$

$$-\frac{l_1 - (d_2 + d_3) - l_1}{l_1 - (d_2 + d_3)} < \frac{v_3 - v_2}{v_2}$$

$$-1 + \frac{l_1}{l_1 - (d_2 + d_3)} < \frac{v_3}{v_2} - 1$$

$$\frac{l_1}{l_1 - (d_2 + d_3)} < \frac{v_3}{v_2}. \tag{5}$$

The statement 1 is proved.

Statement 2: Let $d_3 > l_1$, $d_3 < l_2$, then for any value of the fast cluster length the velocity of the special cluster is: $v_3 = v_2$.

Proof: In this statement the special cluster can not change the lane to $X^{(1)}$, because there is not enough space for it: $d_3 > l_1$, so it can not overtake the fast cluster and the special cluster will move behind the fast cluster with the velocity v_2.

Statement 3: Let $d_3 < l_1$, $d_3 > l_2$, then for any value of the fast cluster length the velocity of the special cluster is: $v_3 = v_1 = 0$.

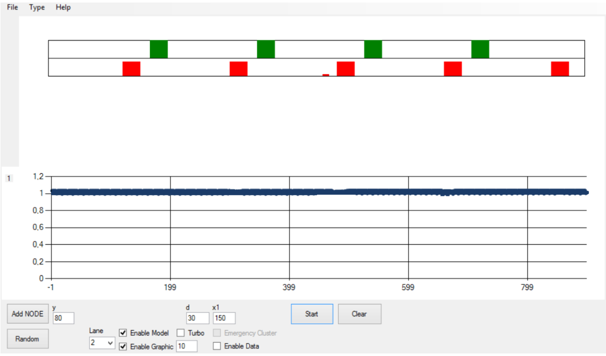
Proof: Similarly to proof of the statement 2, the special cluster can not change the lane to $X^{(2)}$, because there is not enough space for it: $d_3 > l_2$, so it can not overtake the slow cluster and it will move behind the slow cluster with the velocity v_1.

Statement 4: Let $d_3 < l_1, d_3 < l_2, d_2 > l_1$, then the velocity of the special cluster is: $v_3 = v_2$.

Proof: In this case, where $d_2 > l_1$ the special cluster can not overtake the fast cluster, because there is not enough time and the inequality 3 is not satisfied, so the special cluster will move behind the same fast cluster.

B. Conclusions

The problem statement for scenario 2.2 is researched and the following results are obtained: for any initial configuration of the clusters after some moment of time there is synergy in their movement. The parameters, that can affect the result of the cluster movement are l_1, l_2, d_2, d_3.



y: 80
d: 30
x1: 150
Lane: 2
☑ Enable Model
☐ Turbo
☐ Emergency Cluster
☑ Enable Graphic 10
☐ Enable Data

Fig. 8. Results

C. Synergy in scenario 2.1

Let be the special on the lane $X_{(2)}$, its length $d_3 > l_1$ and $d_3 > l_2$, the length of the fast cluster $d_2 < l_1$ and:

$$\Delta l = l_1 - d_2, \Delta v_1 = v_3 - v_1, \Delta v_2 = v_3 - v_2, \Delta v_3 = v_2 - v_1 \tag{6}$$

All clusters are moving, and the value Δv_1 is a velocity of the special cluster relative to the velocity of the slow cluster, the value Δv_2 is a velocity of the special cluster relative to the velocity of the fast cluster, the value Δv_3 is a velocity of the fast cluster relative to the velocity of the slow cluster. Let define two variables:

$$t_1 = \frac{\Delta l}{\Delta v_3} \tag{7}$$

where t_1 is a time, during that the fast cluster blocks the lane $X^{(2)}$, and after this moment the special cluster can not overtake the fast cluster,

$$t_2 = \frac{l_1}{\Delta v_1} \tag{8}$$

where t_2 is a time, during that the minimal part of the special cluster can overtake the fast cluster.

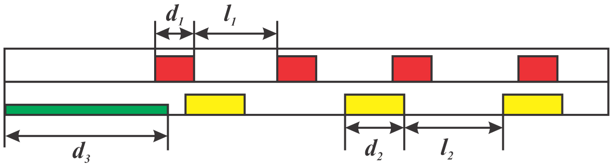


Fig. 9. Problem statement in scenario 2.1

Statement 1: If $t_1 < t_2$, then the segments with length l_1 between slow clusters will be filled by parts of the special cluster with the length:

$$\Delta d = (\Delta v_1 - \Delta v_3) \cdot t_2, \tag{9}$$

that will move with the velocities v_1. The rest of the special cluster will move behind the fast cluster with the velocity v_2.

Proof: In this statement if $t_1 < t_2$, then it means that the special cluster doesn't have enough time to overtake the fast cluster. One part of the special cluster with the length Δd will move on the lane $X_{(1)}$, filling the segments l_1 between the slow clusters and after some moment of time will return on the lane $X_{(2)}$. The rest of the special cluster will move behind the fast cluster.

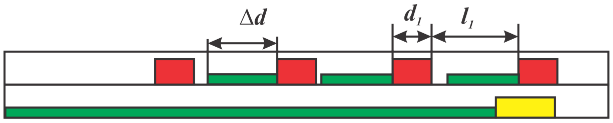


Fig. 10. The statement 1

Statement 2: If $t_1 > t_2$, then the special cluster will divide on segments with the lengths:

$$\Delta d_1 = (v_3 - \Delta v_1) \cdot (t_1 - t_2) \tag{10}$$

that will move with the velocities v_3, and will overtake other clusters.

Proof: If $t_1 > t_2$, then it means that the special cluster has enough time overtake the fast cluster. One part of the special cluster with length Δd will move behind the slow cluster on the lane $X_{(1)}$ and will wait until the lane $X_{(2)}$ will be free and it will return on this lane. The other part of the special cluster with the length Δd_1, that has overtaken the fast cluster successfully, will continue to move with the velocity v_3 till the next fast cluster. The rest of the special cluster will move behind the fast cluster and will wait till next chance to overtake it. The special cluster will fill some segments between slow clusters. After some moment of time the whole special cluster will divide into segments with length Δd_1, that will move and will overtake other clusters.

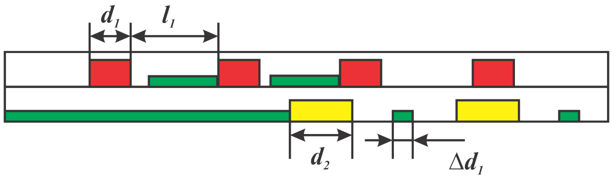


Fig. 11. The statement 2

D. Conclusions

The problem statement for scenario 2.1 is researched and the following results are obtained: for several initial configurations of the clusters after some moment of time there is synergy in their movement. The parameters, that can affect the result of the cluster movement are $l_1, l_2, l_3, d_1, d_2, d_3$.

VI. Future works

Further there are extension by increasing the number of cluster types and increasing the number of lanes, inclusion of the possibility of changing lanes by clusters not only for overtake but for ”free” the lane or for ”occupy” the necessary lane. For example, all slow clusters occupy the right lane and all fast clusters occupy the left lane.

VII. Conclusion

In the article the cluster flow model on the two-lane Belt is considered, problem statements and results for there types of clusters are formulated. Statements and proofs of this model are obtained. Newly developed software, that simulates cluster movement on the two-lane Belt, proved some of this results by simulation.

Acknowledgment

The author would like to thank his supervisor prof. Buslaev A.P. for discussion and problem statement.

References

[1] Buslaev A.P., Tatashev A.G., Yashina M.V., Cluster flow models and properties of appropriate dynamic systems, Journal of Applied Functional Analysis, Vol.8, 2012, pp 54-76.

[2] Bugaev A.S, Buslaev A.P., Kozlov V.V., Yashina M.V., Distributed problems of monitoring and modern approaches to traffic modeling, 14th International IEEE conference on intelligent transportation systems (ITSC-2011), Washington, USA, 2011, pp 477-481.

[3] Kozlov V.V., Buslaev A.P., Metropolis traffic modeling: from intelligent monitoring through physical representation to mathematical problems, International conference on computational and mathematical methods in science and engineering, Vol. 1, 2012, pp 750-756.

[4] Buslaev A.P., Strusinskiy P.M., Computer simulation analysis of cluster model of totally-connected flows on the chain mail, New results in dependability and computer systems, Springer, 2012, pp 63-71.

[5] Strusinskiy P.M., On cluster flow models on multi-lane supporters, Proceedings of the 14th International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 14, pp 1208-1218.

[6] Buslaev A.P., Yashina M.V., Cluster flow of totally-connected flow with local information, International conference on computational and mathematical methods in science and engineering (CMMSE), Vol. 1, 2012, pp 225-232.

The linear complexity over $\mathbb{F}_2$ and $\mathbb{F}_p$ of binary sequences of length $4p$ with optimal autocorrelation

Vladimir Edemskiy

***Abstract*—We derive the linear complexity of binary sequences of period $4p$ with optimal autocorrelation value/magnitude over the finite fields of order two and p. These sequenced are constructed by cyclotomic classes of orders two and four.**

***Index Terms*—Binary sequences, linear complexity, finite field**

I. Introduction

THE autocorrelation is an important measure of pseudo-random sequence for their application in code-division multiple access systems, spread spectrum communication systems, radar systems and so on [7]. An important problem in sequence design is to find sequences with optimal autocorrelation. In their paper, Arasu et al. [1] investigated almost difference sets and constructed new classes of binary sequences of period $4n$ with optimal autocorrelation $\{0, -4\}$. These sequences have also been referred to as interleaved sequences. Sun et al. presented another construction method of binary sequences of period $4p$ with optimal autocorrelation magnitude as well [13].

The linear complexity is another important characteristic of pseudo-random sequence significant for cryptographic applications [4]. It may be defined as the length of the shortest linear feedback shift register that is capable of generating the sequence. The feedback function of this shift register can be deduced from knowledge of just $2L$ consecutive digits of the sequence. Thus, it is reasonable to suggest that "good" sequences have $L > N/2$ (where N denotes the period of the sequence) [10]. The linear complexity of interleaved sequences over the finite field of order two was investigated in series of papers [11], [14], [15], [6] (see also references therein). Also, the linear complexity of Legendre sequences and other cyclotomic sequences of length p was derived in [2], [3] over the finite field $\mathbb{F}_p$.

In this paper we derive the linear complexity over $\mathbb{F}_2$ and $\mathbb{F}_p$ of binary sequences from [1], and over $\mathbb{F}_p$ from [13].

II. Preliminaries

First, we briefly repeat the basic definitions from [1] and some general information.

The ring residue classes $\mathbb{Z}_{4n} \cong \mathbb{Z}_4 \times \mathbb{Z}_n$ relative to isomorphism $\phi(a) = (a \bmod 4, a \bmod n)$ [9]. In [1] a new family of binary sequences with optimal autocorrelation defined as

$$s_i = \begin{cases} 1, & \text{if } i \mod 4n \in C; \\ 0, & \text{if } i \mod 4n \notin C, \end{cases} \qquad (1)$$

for

$$C = \phi^{-1}\big(\{0\} \times C_0 \cup \{1\} \times (C_0 - \delta)^* \cup \\ \{2\} \times C_0^* \cup \{3\} \times (C_0 - \delta)^*\big)$$

where C_0 is the support of binary sequences of period n with optimal autocorrelation, C_0^* and $(C_0 - \delta)^*$ denote the complements of C_0 and $(C_0 - \delta)$ in $\mathbb{Z}_n$, respectively; $0 \le \delta \le n - 1$.

In this paper we consider only the case when $n = p$ where p is an odd prime. One more approach to design of binary sequences of period $4p$ with optimal autocorrelation magnitude was suggested in [13] (see section IV).

It is well known [4] that if $\{s_i\}$ is a binary sequence with period N, then the minimal polynomial $m(x)$ and the linear complexity L of this sequence is defined by

$$m(x) = (x^N - 1)/\gcd\big(x^N - 1, S(x)\big), \\ L = N - \deg\gcd\big(x^N - 1, S(x)\big),$$

where $S(x) = s_0 + s_1 x + ... + s_{N-1}x^{N-1}$.

In our case $N = 4p$, hence over $\mathbb{F}_2$ we have

$$L = 4p - \deg\gcd\left((x^p - 1)^4, S(x)\right). \qquad (2)$$

We use Günther-Blahut theorem to calculate the linear complexity over $\mathbb{F}_p$ of $\{s_i\}$ with a period $4p$ (see, for example [12]). But first we need to prove intermediate lemmas.

Let G be a subset of the residue class ring $\mathbb{Z}_{4p}$, and b be an element of $\mathbb{Z}$. Define

$$bG = \{ba \mod 4p | a \in G\}, \\ G + b = \{(a + b) \mod 4p | a \in G\}.$$

Let C_1 be a compliment of C_0 in $\mathbb{Z}_p^*$. Then Lemma 1 follows from our definitions.

Lemma 1:

(i) $\phi^{-1}\big(\{0\} \times C_m + kp\big) = \phi^{-1}\big(\{kp \bmod 4\} \times C_m\big)$, $m = 0, 1; k = 0, \dots, 3$;

(ii) $\phi^{-1}\big(\{1\} \times (C_0 - \delta)^*\big) =$
$\big(\phi^{-1}\big(\{(1 + \delta)p \bmod 4\} \times C_1\big) - \delta\big) \cup \{(3 - \delta)p - \delta\}$;

(iii) $\{3\} \times (C_0 - \delta)^* =$
$\big(\phi^{-1}\big(\{(\delta - 1)p \bmod 4\} \times C_1\big) - \delta\big) \cup \{(1 - \delta)p - \delta\}$.

V. Edemskiy is with the Department of Applied Mathematics and Information Science, Novgorod State University, Veliky Novgorod, Russia, 173003 e-mail: Vladimir.Edemsky@novsu.ru.

This work was supported by the Ministry of Education and Science of Russia as a part of state-sponsored project no 1.949.2014/K.

By definition, put $B_m = \{0\} \times C_m, m = 0, 1$. Let us introduce the auxiliary polynomials $E_m(x) = \sum_{i \in B_m} x^i$.

Lemma 2: Let $\{s_i\}$ be defined by (1). Then

$$S(x) \equiv \big(E_0(x)+x^{(1+\delta)p-\delta}E_1(x)+x^{(3-\delta)p-\delta}+x^{2p}E_1(x)+ x^{2p}+x^{(\delta-1)p-\delta}E_1(x)+x^{(1-\delta)p-\delta}\big)(\mathrm{mod}(x^{4p}-1)). \quad (3)$$

III. The linear complexity of sequences obtained from Legendre sequences

In this section we consider sequences with optimal autocorrelation value obtained from Legendre sequences [1].

Let p be a prime of the form $p \equiv 1(\bmod\ d)$, where d is an even integer, and let $R = (p-1)/d$, and g be a primitive root modulo p [9]. Then the integers $\bmod\ p$ can be partitioned into d cosets $H_i, 0 \le i \le d-1$, each containing R elements, such that H_0 contains the dth power residues $\bmod\ p$, and the remaining H_i are formed from $g^i H_0$. Cosets H_i are also called the cyclotomic classes of order d with respect to p [4].

Let $d = 2, p \equiv 3(\bmod\ 4)$. The Legendre sequence $l = \{l_i\}$ of period p is defined by

$$l_i = \begin{cases} 1, & \text{if } i \in H_0, \\ 0, & \text{otherwise .} \end{cases}$$

It is well known that Legendre binary sequences have optimal autocorrelation value if $p \equiv 3(\mathrm{mod} 4)$.

Theorem 3: Let C_0 be the support of a Legendre sequence, and let $\{s_i\}$ be defined by (1). Then the linear complexity over $\mathbb{F}_2$ of $\{s_i\}$ is equal to

$$L = \begin{cases} 2p+2, & \text{if } p \equiv \pm 3(\mathrm{mod}\ 8) \text{ and } \delta \neq 0, \\ p+3, & \text{if } p \equiv \pm 1(\mathrm{mod}\ 8) \text{ and } \delta \neq 0, \\ & \text{or} p \equiv \pm 3(\mathrm{mod}\ 8) \text{ and } \delta = 0, \\ (p+7)/4, & \text{if } p \equiv \pm 3(\mathrm{mod}\ 8) \text{ and } \delta \neq 0. \end{cases}$$

Proof: Using (3) we obtain $S(x) = E_0(x) + E_1(x) + 1 + (x^{2p}+1)(x^{p-\delta}+1)(E_1(x)+1)$. From our definitions it follows that $E_0(1)+E_1(1)+1 = 1$ and $E_0(x)+E_1(x)+1 = ((x^p-1)/(x-1))^4$. Hence

$$\gcd\big((x^p-1)^4, S(x)\big) = ((x^p-1)/(x-1))^2 \gcd\big((x^p-1)^2, E_1(x)+1\big) \quad (4)$$

for $\delta \neq 0$ and

$$\gcd\big((x^p-1)^4, S(x)\big) = ((x^p-1)/(x-1))^3 \gcd\left((x^p-1), E_1(x)+1\right) \quad (5)$$

for $\delta = 0$.

The linear complexity of Legendre sequences was investigated in [5]. Since $E_1(x) = \left(\sum_{i \in H_1} x^i\right)^4$, by [5] we have

$$\deg\gcd\left(x^p-1, E_1(x)+1\right) = \begin{cases} 0, & \text{if } p \equiv \pm 3(\mathrm{mod}\ 8), \\ (p-1)/2, & \text{if } p \equiv \pm 1(\mathrm{mod}\ 8). \end{cases}$$

The conclusions of this theorem then follows from (4), (5) and (2). ■

Now, we derive the linear complexity of these sequences over $\mathbb{F}_p$.

Theorem 4: Let C_0 be the support of the Legendre sequence, and let $\{s_i\}$ be defined by (1). Then the linear complexity over $\mathbb{F}_p$ of $\{s_i\}$ is equal to $4p$.

Proof: Let β be a primitive root 4th power of unity in an extension of $\mathbb{F}_p$. Since $E_0(\beta) = E_1(\beta) = (p-1)/2$, it follows from (3) that $S(1) \neq 0$, $S(-1) \neq 0$, $S(\beta) \neq 0$, $S(-\beta) \neq 0$. So, the statement of Theorem 4 follows from Günther-Blahut theorem. ■

The results of computing the linear complexity by Berlekamp-Massey algorithm when $p = 7, 11, 19, 23, 31, 43, 47, \ldots$ confirm Theorems 3 and 4.

In the conclusion section we note that if $\{s_i\}$ equals product the binary sequence of length 4 with ideal autocorrelation and the Legendre sequence, i.e., $C = \phi^{-1}\left(\{0,1,2\} \times C_0 \cup \{3\} \times (C_1 \cup \{0\})\right)$ then $S(x) \equiv (1 + x^p + x^{2p})E_0(x) + x^{3p}E_1(x) + x^{3p}(\bmod (x^{4p}-1))$. In this case $S(\beta^i) \neq 0, i = 0, 1, 2, 4$ and $L = 4p$ over $\mathbb{F}_p$. But over $\mathbb{F}_2$ we have $S(x) \equiv (x^p-1)^3 E_0(x) + x^{3p}\left((x^p-1)/(x-1)\right)^4 (\bmod (x^{4p}-1))$. Hence, here $L \le p+3$.

IV. The linear complexity of sequences obtained from biquadratic residues

In this section we consider sequences with optimal autocorrelation magnitude to present in [13]. Let $d = 4$, $p \equiv 5(\mathrm{mod}\ 8)$, and has a quadratic partition of the form $p = x^2 + 4$. Here x is an integer and $x \equiv 1(\mathrm{mod}\ 4)$.

We derive the linear complexity of sequence defined by

$$s_i = \begin{cases} 1, & \text{if } i \mod 4p \in C; \\ 0, & \text{if } i \mod 4p \notin C, \end{cases} \quad (6)$$

for

$$C = \phi^{-1}\big(\{0,1\} \times (H_0 \cup H_1) \cup \{2\} \times (H_0 \cup H_3) \cup \{3\} \times (H_1 \cup H_2) \cup \{0,2\}\big).$$

By definition, put $D_m = \phi^{-1}\left(\{0\} \times H_m\right)$. Let us introduce the auxiliary polynomials $F_m(x) = \sum_{i \in D_m} x^i, m = 0, 1, 2, 3$. Hence, in this case by Lemma 1 for $S(x) = \sum_{i \in C} x^i$ we obtain that

$$S(x) \equiv T(x)(\mathrm{mod}(x^{4p}-1)) \quad (7)$$

where

$$T(x) = (1 + x^p + x^{2p})F_0(x) + (1 + x^p + x^{3p})F_1(x) + x^{2p}F_3(x) + x^{3p}F_2(x) + 1 + x^{2p}.$$

In the following subsection we prove a few propositions about $F_m(x)$ and $F_m^{(n)}(x)$.

A. Auxiliary lemmas

Let us introduce the auxiliary polynomials $T_{m,0}(x) = F_m(x)$, $T_{m,1}(x) = xF_m^{'}(x)$ and $T_{m,n}(x) = xT_{n-1}^{'}(x), n = 2, 3, \ldots$. Then $T_{m,n}(x) = \sum_{i \in D_m} i^n x^i$ and

$$x^n F_m^{(n)}(x) = T_{m,n}(x) - \sum_{j=1}^{n-1} a_{j,n}(x) F_m^{(j)}(x), \quad (8)$$

where $a_{j,n}(x)$ are polynomials.

Lemma 5: Let β be a primitive root of 4th power of unity in $\mathbb{F}_p$, $1 \le n \le p-1$ and $1 \le m \le d-1$. Then

$$T_{m,n}(\beta^j) = \begin{cases} 0, & \text{if } n \not\equiv 0(\mathrm{mod}(p-1)/4), \\ g^{mn}(p-1)/4, & \text{if } n \equiv 0(\mathrm{mod}(p-1)/4). \end{cases}$$

Proof: By definitions of D_m and $T_{m,n}$ we have that $T_{m,n}(\beta^j) = \sum_{i \in D_m} i^n, n = 1, 2, \ldots$. If $n \equiv 0(\mathrm{mod}(p-1)/4)$ then $\sum_{i \in D_m} i^n = g^{mn}(p-1)/4$. Suppose $n \not\equiv 0(\mathrm{mod}\ (p-1)/4)$; denote $\sum_{i \in D_0} i^n$ by B. Since $g \bmod p$ is a primitive root modulo p, it follows that

$$0 = \sum_{j=1}^{p-1} j^n = \sum_{t=0}^{3} \sum_{i \in H_t} i^n = B + g^n B + g^{2n} B + g^{3n} B = B(g^{4n}-1)/(g^n-1).$$

Hence, $B = 0$ and $\sum_{i \in D_m} i^n = 0$. ■

Corollary 6: If $1 \le n < (p-1)/4$ then $F_m^{(n)}(\beta^j) = 0, j = 0, 1, 2, 3$.

Theorem 7: Let the balanced binary sequences $\{s_i\}$ be defined by (6). Then $L = (11p+5)/4$.

Proof: If $p \equiv 1(\mathrm{mod}\ 4)$ then β belongs to $\mathbb{F}_p$. Hence, without loss of generality, we can assume that $\beta = g^{(p-1)/4}$. By G$\ddot{u}$nther-Blahut theorem, to prove the assertion it suffices to find for $S(x)$ the multiplicity of roots $\pm 1, \pm\beta$. By definition and (7) we have $S(1) = S(\beta) = S(-\beta) = 0, S(-1) = 2$.

Let $(p-1)/4 = 1 + 2r, r \in \mathbb{N}$. Further, since

$$T^{(n)}(x) = (1+x^p+x^{2p})F_0^{(n)}(x) + (1+x^p+x^{3p})F_1^{(n)}(x) + x^{3p}F_2^{(n)}(x) + x^{2p}F_3^{(n)}(x). \quad (9)$$

by Corollary 6, (8), and Lemma 5 it follows that:

(i) $T^{(n)}(1) = 0$ for $1 \le n \le (p-5)/4$ and $T^{((p-1)/4)}(1) = (1+\beta)(p-1)/2 \ne 0$;

(ii) $T^{(n)}(\beta) = 0$ for $1 \le n \le (p-5)/4$ and $T^{((p-1)/4)}(\beta) = \beta(1+\beta^{2r})(p-1)/2$;

(iii) $T^{(n)}(-\beta) = 0$ for $1 \le n \le (p-5)/4$ and $T^{((p-1)/4)}(-\beta) = -\beta(1-\beta^{2r})(p-1)/2$.

So, the multiplicity of root 1 is equal to $(p-1)/4$. The multiplicity of roots $\beta, -\beta$ depends from r. This require the additional study.

Using (8) we obtain from (9) that

$$x^n T^{(n)}(x) = (1+x^p+x^{2p})T_{m,1}(x) + (1+x^p+x^{3p})T_{m,2}(x) + x^{3p}T_{m,2}(x) + x^{2p}T_{m,3}(x) - x^n \sum_{j=1}^{n-1} a_{j,n}(x)T^{(j)}(x).$$

Hence, if $T^{((p-1)/4)}(\pm\beta) = 0$ then $T^{(j)}(\pm\beta) = 0$ for $(p-1)/4 < j < (p-1)/2$ by Lemma 5.

Further, by (9) and Lemma 5 $T^{((p-1)/2)}(\pm\beta) = 0$ and $T^{(3(p-1)/4)}(\pm\beta) = \pm\beta(1 \mp \beta^{2r})(p-1)/2$.

From this we can establish that the multiplicity of root β is equal to $(p-1)/4$ for $r \equiv 0(\mathrm{mod}\ 2)$ and $3(p-1)/4$ for $r \equiv 1(\mathrm{mod}\ 2)$; also the multiplicity of root $-\beta$ is equal to $3(p-1)/4$ for $r \equiv 0(\mathrm{mod}\ 2)$ and $(p-1)/4$ for $r \equiv 1(\mathrm{mod}\ 2)$. This completes the proof of Theorem 7. ■

The results of computing the linear complexity by Berlekamp-Massey algorithm when $p = 5, 13, 29, 53, 173, 229, 293, \ldots$ confirm Theorem 7.

V. Conclusion

In this paper we examine the linear complexity of binary sequences with period $4p$ constructed on cyclotomic classes. First, we derive the linear complexity of binary sequences obtained from Legendre sequences. These sequences of length $4p$ with optimal autocorrelation were constructed by method proposed by Arasu et al. [1]. Second, we investigate the linear complexity of binary sequences of length $4p$ with optimal autocorrelation obtained from the cyclotomic classes of order four [13].

We determine the parameters of sequences with optimal autocorrelation values and high linear complexity.

March 12, 2015

References

[1] K.T. Arasu, C. Ding, T. Hellesenh, P. V. Kumar, H.M. Martinsen. "Almost difference sets and their sequences with optimal autocorrelation". *IEEE transactions on information theory.* vol. 47, 7, pp. 2934-2943, 2001.

[2] H. Aly, A. Winterhof. "On the k-error linear complexity over F_p of Legendre and Sidelnikov sequences". *Des Codes Crypt.,* vol.40, pp. 369-374, 2006.

[3] H. Aly, W. Meidl,A. Winterhof. "On the k-Error Linear Complexity of Cyclotomic sequences". *J. Math. Crypt.*, vol.1, pp.1-14, 2007.

[4] T.W. Cusick, C. Ding, A. Renvall. *Stream Ciphers and Number Theory*, North-Holland Publishing Co., Amsterdam (1998)

[5] C. Ding, T. Helleseth, W. Shan. "On the Linear Complexity of Legendre Sequences", *IEEE Trans. Info Theory,* vol. IT-44, pp. 1276 - 1278, 1998.

[6] V. Edemskiy. "On the linear complexity of interleaved binary sequences of period 4p obtained from Hall sequences or Legendre and Hall sequences". *Electronics Letters.*, vol.50. Issue 8, p. 604-605, 2014.

[7] S.W. Golomb, G. Gong. *Signal Design for Good Correlation: For Wireless Communications, Cryptography and Radar Applications.* Cambridge University Press (2005)

[8] M. Hall.*Combinatorial Theory*, Wiley, New York (1975)

[9] K. Ireland, M. Rosen. *A Classical Introduction to Modern Number Theory*, Springer, Berlin (1982)

[10] R. Lidl, H. Niederreiter. *Finite Fields.* Addison-Wesley (1983).

[11] N. Li, X. Tang."On the linear complexity of binary sequences of period $4N$ with optimal autocorrelation value/magnitude", *IEEE Trans. Inf. Theory*, vol.57, pp. 7597-7604.

[12] J.L. Massey, S. Serconek. "Linear complexity of periodic Sequences: A General Theory". *Journal of Complexity* . Lecture Notes in Computer Science. pp. 358–371, 1996.

[13] Y. Sun, H. Shen. "New Binary Sequences of Length 4p with Optimal Autocorrelation Magnitude". *Ars Combinatoria (A Canadian Journal of Combinatorics)*, Vol. LXXXIX (89), pp. 255-262, 2008.

[14] Q. Wang, X. N. Du. "The linear complexity of binary sequences with optimal autocorrelation", *IEEE Trans. Inf. Theory*, vol. 56, pp. 6388-6397, 2010.

[15] H. Xiong, L. Qu, C. Li, S. Fu. "Linear complexity of binary sequences with interleaved structure", *IET Communications*, vol. 7(15), pp. 1688-1696, 2013.

Design optimization of sandwich structure subjected to maximum displacement criterion

E. Kormanikova and K. Kotrasova

***Abstract*—**The paper deals with design optimization of sandwich structure made of laminate outer layers and PUR foam core. The thickness of outer layers with the known fibre orientation angle of individual laminae, referred to as the thickness variable, will be used as design variable. The optimization problem with displacement constraint will be formulated to minimize the weight of sandwich with laminate outer layers. The design is optimized using continuous design variable.

***Keywords*—**Design optimization, displacement criterion, sandwich structure, thickness design variable, weight objective function.

I. Introduction

ASYMETRIC sandwich plate to the mid-plane has the objective function of minimizing the weight function. As design variable is the thickness of outer layers and is computed by using the Sequential Linear Programming method. Within this method the Modified Feasible Direction method was used.

The typical sandwich structure compounds of three layers. The outer layers are made of a material that has high strength (fiber reinforced laminates), which can transfer axial forces and bending moments, while the core is made of lightweight materials such as foam, alder wood etc. The material used in sandwich core must be resistant to compression and capable of transmitting shear [1].

The design optimization problem of current interest is the minimization of the weight function for a sandwich composite plate. This is a design optimization problem which optimizes the thickness of the composite laminae to give the minimum weight. Of greater interest to current study are the works on the design optimization of composite plates where the laminae thicknesses are taken as the design variables.

The optimization of a composite plate is important analysis for design of structures ranging from aircrafts to civil engineering structures.

II. Effective Moduli of Composites

If the composite has periodic microstructure, then Fourier series can be used to estimate all the components of the stiffness tensor of a composite. Explicit formulas for a composite reinforced by long circular cylindrical fibres, which are periodically arranged in a square array (Fig. 1) are written in the following way.

Because the microstructure has square symmetry, the stiffness tensor has six unique coefficients given by [2]

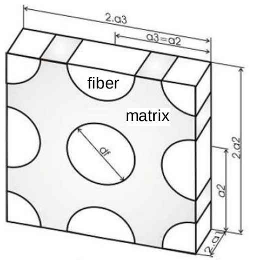


Fig. 1 periodic square microstructure model

$$C_{11}=\lambda^{(m)}+2\mu^{(m)}-\frac{\xi}{D}\left(\begin{array}{l}\frac{S_3^2}{\mu^{(m)2}}-\frac{2S_6S_3}{\mu^{(m)2}g}-\frac{aS_3}{\mu^{(m)}c}+\frac{S_6^2-S_7^2}{\mu^{(m)2}g^2}+\\+\frac{aS_6+bS_7}{\mu^{(m)}gc}+\frac{a^2-b^2}{4c^2}\end{array}\right) \tag{1}$$

$$C_{12}=\lambda^{(m)}+\frac{\xi}{D}b\left(\frac{S_3}{2c\mu^{(m)}}-\frac{S_6-S_7}{2c\mu^{(m)}g}-\frac{a+b}{4c^2}\right) \tag{2}$$

$$C_{22}=\lambda^{(m)}+2\mu^{(m)}-\frac{\xi}{D}\left(-\frac{aS_3}{2\mu^{(m)}c}+\frac{aS_6}{2\mu^{(m)}gc}+\frac{a^2-b^2}{4c^2}\right) \tag{3}$$

$$C_{66}=\mu^{(m)}-\xi\left(-\frac{S_3}{\mu^{(m)}}+\left(\mu^{(m)}-\mu^{(f)}\right)^{-1}\right)^{-1} \tag{4}$$

$$C_{23}=\lambda^{(m)}+\frac{\xi}{D}\left(\frac{aS_7}{2\mu^{(m)}gc}-\frac{ba+b^2}{4c^2}\right) \tag{5}$$

$$C_{44}=\mu^{(m)}-\xi\left(\frac{2S_3}{\mu^{(m)}}+\left(\mu^{(m)}-\mu^{(f)}\right)^{-1}+\frac{4S_7}{\mu^{(m)}\left(2-2\nu^{(m)}\right)}\right)^{-1} \tag{6}$$

where

This work was supported by the Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under Project VEGA 1/0477/15.

Kamila Kotrasova is with the Department of Structural Mechanics, Institute of Structural Engineering, The Technical University of Kosice, Faculty of Civil Engineering, Vysokoskolska 4, 042 00 Kosice, Slovak Republic (corresponding author to provide phone: +421 55 6024294; e-mail: kamila.kotrasova@tuke.sk).

Eva Kormanikova is with the Department of Structural Mechanics, Institute of Structural Engineering, The Technical University of Kosice, Faculty of Civil Engineering, Vysokoskolska 4, 042 00 Kosice, Slovak Republic (e-mail: eva.kormanikova@tuke.sk).

$$D=\frac{aS_3^2}{2\mu^{(m)2}c}-\frac{aS_6S_3}{\mu^{(m)2}gc}+\frac{a\left(S_6^2-S_7^2\right)}{2\mu^{(m)2}g^2c}+\frac{S_3\left(b^2-a^2\right)}{2\mu^{(m)}c^2}+ \\ +\frac{S_6\left(a^2-b^2\right)+S_7\left(ab+b^2\right)}{2\mu^{(m)}gc^2}+\frac{\left(a^3-2b^3-3ab^2\right)}{8c^3} \qquad (7)$$

$$a=\mu^{(f)}-\mu^{(m)}-2\mu^{(f)}\nu^{(m)}+2\mu^{(m)}\nu^{(f)} \qquad (8)$$

$$b=-\mu^{(m)}\nu^{(m)}+\mu^{(f)}\nu^{(f)}+2\mu^{(m)}\nu^{(m)}\nu^{(f)}-2\mu^{(f)}\nu^{(f)}\nu^{(m)} \qquad (9)$$

$$c=\left(\mu^{(m)}-\mu^{(f)}\right)\begin{pmatrix}\mu^{(f)}-\mu^{(m)}+\mu^{(f)}\nu^{(f)}-\mu^{(m)}\nu^{(m)}+2\mu^{(m)}\nu^{(f)}- \\ -2\mu^{(f)}\nu^{(m)}+2\mu^{(m)}\nu^{(f)}\nu^{(m)}-2\mu^{(f)}\nu^{(m)}\nu^{(f)}\end{pmatrix} \qquad (10)$$

$$g=\left(2-2\nu^{(m)}\right) \qquad (11)$$

Assuming the fibre and matrix are both isotropic, Lame constants of both materials are obtained by

$$\lambda=\frac{E}{(1+\nu)(1-2\nu)} \qquad \mu=G\,. \qquad (12)$$

For a composite reinforced by long circular cylindrical fibres, periodically arranged in a square array (Fig. 1) the constants S_i, i = 3, 6, 7 are given as follows [2]

$$S_3=0{,}49247-0{,}47603\xi-0{,}02748\xi^2$$

$$S_6=0{,}36844-0{,}14944\xi-0{,}27152\xi^2$$

$$S_7=0{,}12346-0{,}32035\xi+0{,}23517\xi^2\,. \qquad (13)$$

Further alternative is the periodic microstructure with square arrangement of fibers in the representative volume element (RVE) solved using FEM (Fig. 2).

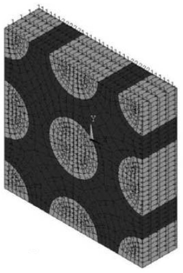


Fig. 2 periodic square microstructure FEA model

The elastic properties of the homogenized material can be computed by [2]

$$E_1=C_{11}-2C_{12}^2/\left(C_{22}+C_{23}\right) \qquad (14)$$

$$\nu_{12}=C_{12}/\left(C_{22}+C_{23}\right) \qquad (15)$$

$$E_2=\left(C_{11}\left(C_{22}+C_{23}\right)-2C_{12}^2\right)\left(C_{22}-C_{23}\right)/\left(C_{11}C_{22}-C_{12}^2\right) \qquad (16)$$

$$G_{12}=C_{44} \quad \nu_{23}=\left(C_{11}C_{23}-C_{12}^2\right)/\left(C_{11}C_{22}-C_{12}^2\right) \quad G_{23}=C_{66}\,. \qquad (17)$$

III. Sandwich Plates with Laminate Faces

A sandwich can be defined as a special laminate with three layers. The thin cover sheets, i.e. the layers 1 and 3, have the thicknesses h_1 for the lower skin and h_3 for the upper skin. The thickness of the core is h_2. In a general case h_1 does not have to be equal to h_3, but in the most important practical case of symmetric sandwiches $h_1 = h_3$.

Most sandwich structures can be modelled and analyzed using the shear deformation theory for laminate plates [3]. The in-plane resultants ***N*** for sandwiches are defined by

$$\boldsymbol{N}=\int_{-\left(\frac{1}{2}h_2+h_1\right)}^{-\frac{1}{2}h_2}\boldsymbol{\sigma}dz+\int_{\frac{1}{2}h_2}^{\frac{1}{2}h_2+h_3}\boldsymbol{\sigma}dz\,. \qquad (18)$$

The moment resultants are defined by

$$\boldsymbol{M}=\int_{-\left(\frac{1}{2}h_2+h_1\right)}^{-\frac{1}{2}h_2}\boldsymbol{\sigma}zdz+\int_{\frac{1}{2}h_2}^{\frac{1}{2}h_2+h_3}\boldsymbol{\sigma}zdz \qquad (19)$$

and the transverse shear force by

$$\boldsymbol{V}=\int_{-\frac{1}{2}h_2}^{\frac{1}{2}h_2}\boldsymbol{\tau}dz\,. \qquad (20)$$

For the resultants ***N*** and ***M*** the integration is carried out over the sheets only and for the transverse shear force over the core.

The constitutive equations for a sandwich are written in the hypermatrix form

$$\begin{pmatrix}\mathbf{N}\\ \mathbf{M}\\ \mathbf{V}\end{pmatrix}=\begin{pmatrix}\mathbf{A} & \mathbf{B} & \mathbf{0}\\ \mathbf{C} & \mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{A}^s\end{pmatrix}\begin{pmatrix}\boldsymbol{\varepsilon}_\mathbf{m}^\mathbf{0}\\ \boldsymbol{\kappa}\\ \boldsymbol{\gamma}\end{pmatrix}, \qquad (21)$$

with stiffness coefficients

$$A_{ij}=A_{ij}^{(1)}+A_{ij}^{(3)}, \qquad B_{ij}=\frac{1}{2}h^{(2)}\left(A_{ij}^{(3)}-A_{ij}^{(1)}\right),$$

$$C_{ij}=C_{ij}^{(1)}+C_{ij}^{(3)}, \qquad D_{ij}=\frac{1}{2}h^{(2)}\left(C_{ij}^{(3)}-C_{ij}^{(1)}\right),$$

$$A_{ij}^s=E_{ij}^s h^{(2)};\ i,j = 4, 5, \qquad (22)$$

where E_{ij}^s are the transverse shear moduli of the core.

From equilibrium equations results the set of five differential equations correspond with five partial differential equilibrium equations. For the symmetric sandwich plate element with laminated faces gilt

$$A_{11}^{(1)}\frac{\partial^2u_1}{\partial x^2}+2A_{14}^{(1)}\frac{\partial^2u_1}{\partial x\partial y}+A_{44}^{(1)}\frac{\partial^2u_1}{\partial y^2}+A_{14}^{(1)}\frac{\partial^2v_1}{\partial x^2}+\left(A_{12}^{(1)}+A_{44}^{(1)}\right)\frac{\partial^2u_1}{\partial x\partial y}+ \\ +A_{24}^{(1)}\frac{\partial^2v_1}{\partial y^2}-E_{55}\left(\frac{\partial\overline{w}}{\partial x}-\psi\right)-E_{56}\left(\frac{\partial\overline{w}}{\partial y}-\varphi\right)=0 \qquad (23)$$

$$A_{22}^{(1)}\frac{\partial^2 v_1}{\partial y^2}+\left(A_{12}^{(1)}+A_{44}^{(1)}\right)\frac{\partial^2 u_1}{\partial x\partial y}+A_{24}^{(1)}\frac{\partial^2 u_1}{\partial y^2}+A_{44}^{(1)}\frac{\partial^2 v_1}{\partial x^2}+2A_{24}^{(1)}\frac{\partial^2 u_1}{\partial x\partial y}+$$
$$+A_{22}^{(1)}\frac{\partial^2 v_1}{\partial y^2}-E_{56}\left(\frac{\partial \overline{w}}{\partial x}-\psi\right)-E_{66}\left(\frac{\partial \overline{w}}{\partial y}-\varphi\right)=0 \qquad (24)$$

$$A_{11}^{(3)}\frac{\partial^2 u_3}{\partial x^2}+2A_{14}^{(3)}\frac{\partial^2 u_3}{\partial x\partial y}+A_{44}^{(3)}\frac{\partial^2 u_3}{\partial y^2}+A_{14}^{(3)}\frac{\partial^2 v_3}{\partial x^2}+\left(A_{12}^{(3)}+A_{44}^{(3)}\right)\frac{\partial^2 u_3}{\partial x\partial y}+$$
$$+A_{24}^{(3)}\frac{\partial^2 v_3}{\partial y^2}+E_{55}\left(\frac{\partial \overline{w}}{\partial x}-\psi\right)+E_{56}\left(\frac{\partial \overline{w}}{\partial y}-\varphi\right)=0 \qquad (25)$$

$$A_{22}^{(3)}\frac{\partial^2 v_3}{\partial y^2}+\left(A_{12}^{(3)}+A_{44}^{(3)}\right)\frac{\partial^2 u_3}{\partial x\partial y}+A_{24}^{(3)}\frac{\partial^2 u_3}{\partial y^2}+A_{44}^{(3)}\frac{\partial^2 v_3}{\partial x^2}+2A_{24}^{(3)}\frac{\partial^2 u_3}{\partial x\partial y}+$$
$$+A_{22}^{(3)}\frac{\partial^2 v_3}{\partial y^2}+E_{56}\left(\frac{\partial \overline{w}}{\partial x}-\psi\right)+E_{66}\left(\frac{\partial \overline{w}}{\partial y}-\varphi\right)=0 \qquad (26)$$

$$D_{11}\frac{\partial^4 w}{\partial x^4}+4D_{14}\frac{\partial^4 w}{\partial x^3\partial y}+2\left(D_{12}+2D_{44}\right)\frac{\partial^4 w}{\partial x^2\partial y^2}+4D_{24}\frac{\partial^4 w}{\partial x\partial y^3}+$$
$$+D_{22}\frac{\partial^4 w}{\partial y^4}-\overline{A}_{55}\left(\frac{\partial^2 \overline{w}}{\partial x^2}-\frac{\partial \psi}{\partial x}\right)-\overline{A}_{56}\left(\frac{\partial^2 \overline{w}}{\partial x\partial y}-\frac{\partial \varphi}{\partial x}\right)-$$
$$-\overline{A}_{65}\left(\frac{\partial^2 \overline{w}}{\partial x\partial y}-\frac{\partial \psi}{\partial y}\right)-\overline{A}_{66}\left(\frac{\partial^2 \overline{w}}{\partial y^2}-\frac{\partial \varphi}{\partial y}\right)=p\,. \qquad (27)$$

The solving of unknown functions $u_1(x,y)$, $u_3(x,y)$, $v_1(x,y)$, $v_3(x,y)$, $w(x,y)$ have to perform the boundary conditions for each boundary. Consistent with the eight order set of differential equations four boundary conditions must be prescribed for each edge of the plate. The classical boundary conditions

N_n or u, N_{nt} or v, M_n or $\partial w/\partial n$,

$$V_n=Q_n+\frac{\partial M_{nt}}{\partial t}\quad \text{or} \quad w \qquad (28)$$

must be specified. The subscripts n and t in the boundary conditions above denote the coordinates normal and tangential to the boundary. It is well known that in the classical plate theory the boundary cannot responded separately to the shear force resultant Q_n and the twisting moment M_{nt} but only to the effective or Kirchhoff shear force resultant. Equations (28) may be used to represent any form of simple edge conditions, e.g. clamped, simply supported and free.

If the sandwich layers are symmetrical to the mid-plane, for the simplified form of equations (28), the boundary conditions are

Simply supported edge: $w=0,\quad M_n=0$,

Clamped edge: $w=0,\quad \partial w/\partial n=0$,

Free edge: $M_n=0,\quad V_n=0$.

We have used the finite element method for solving the problem. The continuum was divided to the finite number of rectangular finite plate elements.

IV. Design Optimization

Before starting the topic of design optimization, it is important to distinguish between analysis and design. Analysis is the process of determining the response of a specified system to its environment. Design is the actual process of defining the system. Analysis is therefore a subset of design.

Engineering design is an iterative process. The design is continuously modified until it meets evaluation and acceptance criteria set by the engineer. Mathematical and empirical formulas and experience have been useful in the traditional design processes to verify the adequacy of designs. A fully automated design optimization is used when engineers are trying to modify a design which level of complexity exceeds their ability to make appropriate changes. It is not surprising that even what might appear as extremely simple design task may easily be a real challenge to the designer during the decision-making process.

Design optimization refers to the automated redesign process that attempts to minimize an objective function subject to limits or constraints on the response by using a rational mathematical approach to yield improved designs.

A feasible design is a design that satisfies all of the constraints. A feasible design may not be optimal. An optimum design is defined as a point in the design space for which the objective function is minimized or maximized and the design is feasible. The process of design optimization can be pictorially represented as shown in the Figure 3.

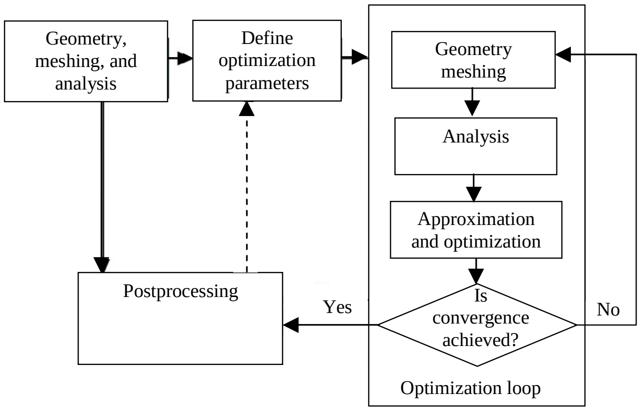

Geometry, meshing, and analysis
Define optimization parameters
Geometry meshing
Analysis
Approximation and optimization
Is convergence achieved?
Yes
No
Postprocessing
Optimization loop

Fig. 3 design optimization process

The optimization process is applied to the approximate problem represented by the polynomial approximation. The coefficients of the polynomial function are determined by the least squares regression.

For regression analysis the singular value decomposition is used. When the objective function and constraints are approximated and their gradients with respect to the design variables are calculated based on chosen approximation, it is possible to solve the approximate optimization problem.

One of the algorithms used in the optimization module is called the Modified Feasible Direction method (MFD). The solving process is iterated until convergence is achieved.

It is important to distinguish the iteration inside the approximate optimization from the loop in the overall optimization process. Figure 4 shows the iterative process within the optimization loop.

Using the modified feasible direction method (MFD) [4] the solving process is iterated until convergence is achieved:

1. $q = 0,\ X^q = X^m$.
2. $q = q+1$.
3. Evaluate objective function and constraints.
4. Identify critical and potentially critical constraints $\overline{N}_c$.
5. Calculate gradient of objective function $\nabla F(X_i)$ and constraints $\nabla g_k(X_i)$, where $k = 1,2,...,\overline{N}_c$.
6. Find a usable-feasible search direction S^q.
7. Perform a one-dimensional search $X^q = X^{q-1} + \alpha S^q$.
8. Check convergence. If satisfied, make $X^{m+1} = X^q$. Otherwise, go to 2.
9. $X^{m+1} = X^q$.

Convergence of MFD to the optimum is checked by criteria of maximum iterations and criteria changes of objective function. Besides the previously mentioned criteria, the Kuhn-Tucker conditions necessary for optimality must be satisfied.

The other algorithm for solving the nonlinear approximate optimization problem is called the Sequential Linear Programming method (SLP). The iterative process of SLP within each optimization loop is shown below:

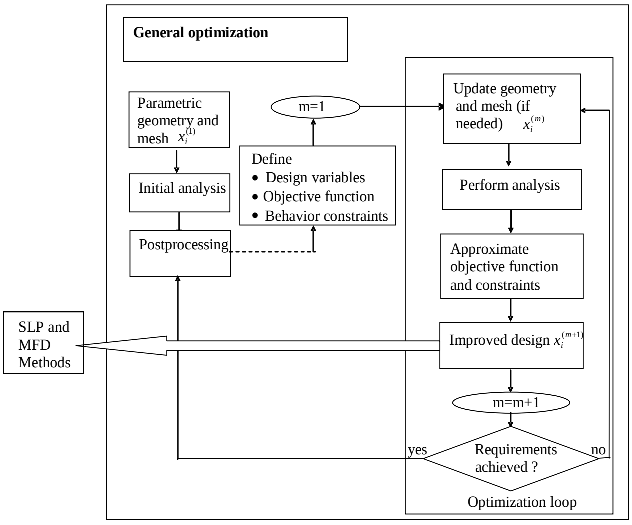


Fig. 4 general optimization process

1. p=0, X^p=X^m.
2. p=p+1.
3. Linearize the problem at X^{p-1} by creating a first order Taylor Series expansion of the objective function and retained constraints

$$F(X) = F(X^{p-1}) + \nabla F(X^{p-1})(X - X^{p-1})$$

$$g(X) = g(X^{p-1}) + \nabla g(X^{p-1})(X - X^{p-1}).$$

4. Use this approximation of optimization instead of the original nonlinear functions:

Maximize: $F(X)$

Subject to: $g(X) \leq 0$ and $\overline{X}_i^L \leq X_i \leq \overline{X}_i^U$.

5. Find an improved design X^p (using the Modified Feasible Direction method).
6. Check feasibility and convergence. If both of them are satisfying, go to 7. Otherwise, go to step 2.
7. $X^{m+1} = X^p$.

Using the SLP method the solving process is iterated until convergence is achieved. Convergence or termination checks are performed at the end of each optimization loop in general optimization. The optimization process continues until either convergence or termination occurs.

The process may be terminated before convergence in two cases:

- the number of design sets so far exceeds the maximum number of optimization loops,
- if the initial design is infeasible and the allowed number of consecutive infeasible designs has been exceeded.

The optimization problem is considered converged if all of the following conditions are satisfied:

1. the current design is feasible,
2. changes in the objective function F:

- the difference between the current value and the best design so far is less than the tolerance τ_F

$$\left|F_{\text{current}} - F_{\text{best}}\right| \leq \tau_F \tag{29}$$

- the difference between the current value and the previous design is less than the tolerance

$$\left|F_{\text{current}} - F_{\text{current}-1}\right| \leq \tau_F \tag{30}$$

- the differences between the current value and two previous designs are less than the tolerance

$$\left|F_{\text{current}} - F_{\text{current}-1}\right| \leq \tau_F \tag{31}$$

$$\left|F_{\text{current}} - F_{\text{current}-2}\right| \leq \tau_F \tag{32}$$

3. changes in the design variables Xi:

- the difference between the current value of each design variable and the best design so far is less than the respective tolerance τ^i

$$\left|X^i_{\text{current}} - F^i_{\text{best}}\right| \leq \tau^i \tag{33}$$

- the difference between the current value of each design variable and the previous design is less than the respective tolerance

$$\left|X^i_{\text{current}} - F^i_{\text{current}-1}\right| \leq \tau^i \tag{34}$$

- the differences between the current value of each design variable and two previous designs are less than the respective tolerance

$$\left|X^i_{\text{current}} - F^i_{\text{current}-1}\right| \leq \tau^i \tag{35}$$

$$\left|X^i_{\text{current}} - F^i_{\text{current}-2}\right| \leq \tau^i \tag{36}$$

V. Modeling of Sandwich Plates and Numerical Solution

For the numerical solution the simply supported panel with laminate facings was used [4, 5]. Panel length is 3750 mm, nominal width is 1000 mm. Thickness of the facings is h_1=h_3 and core is h_2=0.1m (Fig. 5). On the panel affects uniform static wind load with intensity of 2 kPa in the bending plane.

The laminate Carbon/epoxy facings are composed of eight identical thickness layers of a symmetric laminate $[0/\pm 45/90]_s$.

It was considered the carbon fibres in epoxy matrix, while unidirectional laminate layer has characteristics:

E_f = 230 GPa; E_m = 3 GPa; ν_f = 0.2; ν_m = 0.3; V_f = 0.6; ρ_k = 1580 kg/m^3.

Sandwich core, consisting of PUR foam, has material constants: E_{PUR} = 16 MPa; ν_{PUR} = 0.3; ρ_{PUR} = 150 kg/m^3.

Laminate properties were determined by homogenization techniques [6, 13]. Computational program MATLAB was used to calculate the effective material properties of laminate facings. Numerical experiments were conducted through the COSMOS/M program. STAR module for solving linear static was used for calculations. There were used finite elements of type SHELL4L. These are the 4-node multi-layer quadrilateral elements with membrane and bending response; can be enter up to fifty layers.

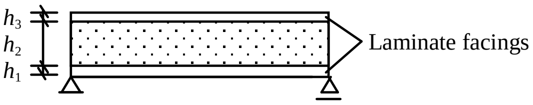


Fig. 5 scheme of sandwich structure

VI. Results

The design optimization problem can be written as follows:

$$F(\boldsymbol{X}) = G(h_1) \rightarrow \min \quad [\mathrm{N}]$$

$$1 \cdot 10^{-4} \le h_1 \le 0.01 \quad [\mathrm{m}]$$

$$0 \le w \le 0.0375 \quad [\mathrm{m}]$$

The initial values and bounds of design variables, constraints and the objective function are shown in the Table 1.

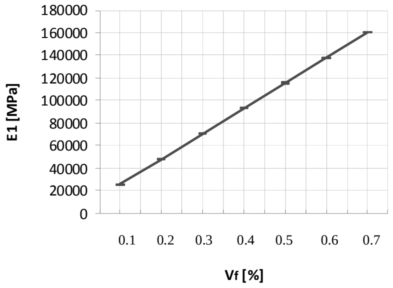


Fig. 6 longitudinal modulus E_1 versus fiber volume fraction

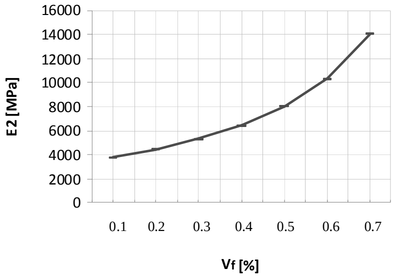


Fig. 7 transversal modulus E_2 versus fiber volume fraction

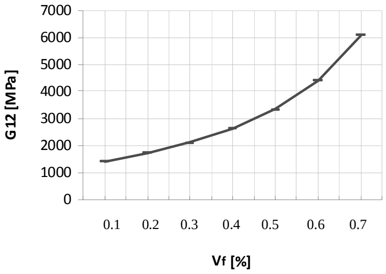


Fig. 8 shear modulus G_{12} versus fiber volume fraction

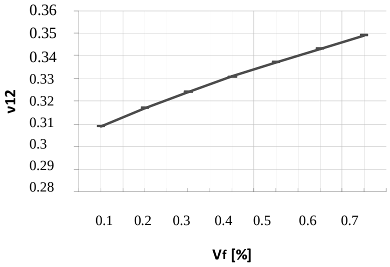


Fig. 9 Poisson ratio ν_{12} versus fiber volume fraction

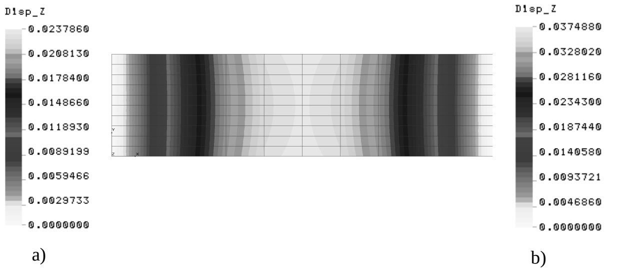


Fig. 10 deflections w before a) and after b) the optimization process

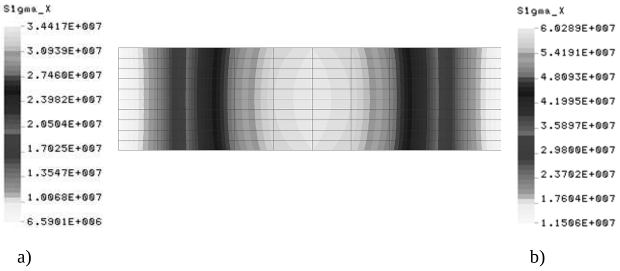


Fig. 11 effective stresses σ_x at the bottom of first layer before a) and after b) the optimization process

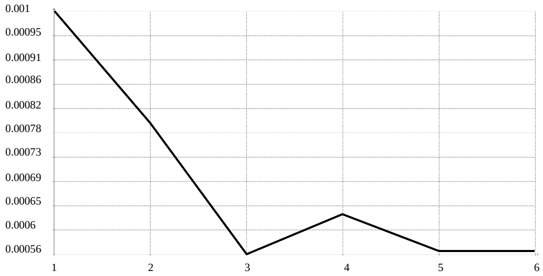


Fig. 12 variation of design variable h_1 [m] during the optimization process

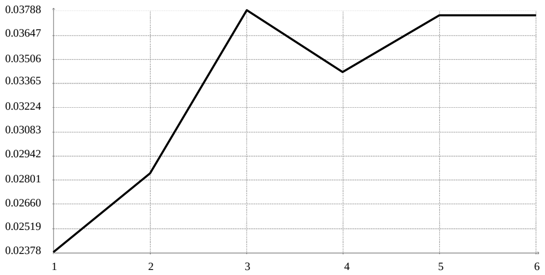


Fig. 13 variation of constraint values *w* [m] during the optimization process

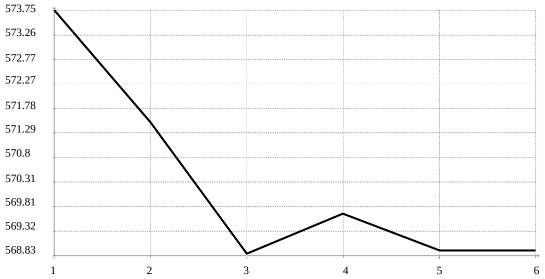


Fig. 14 variation of objective function values *G* [N] during the optimization process

Table 1 summary of results

Optimization parameters		Initial values	Final values	Tolerance τ
Design variable	h_1 [m]	0.001	$5.683 \cdot 10^{-4}$	$1 \cdot 10^{-5}$
Objective function	G [N]	573.75	568.893	$1 \cdot 10^{-3}$
Constraint	w [m]	0.02378	0.0375	$1 \cdot 10^{-4}$

VII. Discussions and Conclusion

The homogenization techniques applied for periodical RVE was used to get the material characteristics [7] of outer laminate layers of sandwich structure (Figs. 6-9).

The first order shear laminate theory was used by the FEM analysis of the problem [8-13]. The problem was formulated as a minimum weight of simply supported rectangular sandwich plate subject to deflection constraint in the middle of the plate. Design variable was thickness of sandwich layers. The optimal problem was solved using SLP and MFD method [15] with maximum 70 iterations in each own optimization loop. The main optimization process was finished after 6 iterations in general optimization loop. In the Figs. 10 and 11 are shown deflections and stresses σ_x before and after optimization process, respectively. In the Figs. 12-14 there are depicted variations of design variable, constraint and objective function during the optimization process, respectively. Initial and final values of optimization process are shown in the Table 1. There was not taken into account a hygroscopic effect of environment. Only static analysis under mechanical loading there was performed. It was designed the optimized thickness of laminate layer $h_1 = 0.0006$ m. The total thickness of sandwich plate is $h = 0.1012$ m.

Acknowledgment

This work was supported by the Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under Project VEGA 1/0477/15.

References

[1] H. Altenbach, J. Altenbach, “W. Kissing, Structural analysis of laminate and sandwich beams and plates,” Lublin, 2001.

[2] E. J. Barbero, “Finite element analysis of composite materials,” CRC Press, USA, ISBN-13: 978-1-4200-5433-0, 2007.

[3] Z. Gürdal, R.T. Haftka, P. Hajela, “Design and Optimization of Laminated Composite Materials,” J. Wiley & Sons, 1999.

[4] E. Kormanikova, I. Mamuzic, “Buckling analysis of a laminate plate,” Metalurgija. Vol. 47, no. 2 (2008), pp. 129-132. - ISSN 0543-5846.

[5] M. Mihalikova [et al.], “Influence of strain rate on automotive steel sheet breaking,” Chemické listy. Vol. 105, no. 17 (2011), pp. 836-837. - ISSN 0009-2770.

[6] Sykora, M. Sejnoha, J. Sejnoha, “Homogenization of coupled heat and moisture transport in masonry structures including interfaces,” Applied Mathematics and Computation, 219 (13), pp. 7275-7285, 2013.

[7] S. Harabinova, E. Panulinova, “Properties of Aggregates of Steel-Making Slag,” GeoConference on Energy and Clean Technologies: conference proceedings, Albena, Bulgaria – Sofia, Volume 2, 2014, pp. 199-202.

[8] N. Jendzelovsky, “Analysis of the 3D state of stress of a concrete beam,” Advanced Materials Research, Volume 969, pp. 45-50, 2014.

[9] K. Tvrda, “Probability and sensitivity analysis of plate,” Applied Mechanics and Materials, Volume 617, pp. 193-196, 2014.

[10] M. Krejsa, P. Janas, I. Yilmaz, M. Marschalko, T. Bouchal, “The use of the direct optimized probabilistic calculation method in design of bolt reinforcement for underground and mining workings,” The Scientific World Journal, Volume 2013, Article number 267593, 2013.

[11] M. Zmindak, Z. Pelagic, M. Bvoc, “Analysis of high velocity impact on composite structures,” Applied Mechanics and Materials, Volume 617, pp. 104-109, 2014.

[12] J. Kralik, “Optimal design of npp containment protection against fuel container drop”, Advanced Materials Research, Volume 688, pp. 213-221, 2013.

[13] J. Melcer, G. Lajcakova, “Comparison of finite element and classical computing models of reinforcement pavement,” Advanced Materials Research, Volume 969, pp. 85-88, 2014.

[14] M. Sejnoha, J. Zeman, “Micromechanical modeling of imperfect textile composites,” (2008) *International Journal of Engineering Science*, 46 (6), pp. 513-526.

[15] E. Kormanikova, and I. Mamuzic, “Optimization of laminates subjected to failure criterion,” *Metalurgija*, vol. 50 (1), pp. 41-44, 2011.

E. Kormanikova graduated at the Technical University of Košice, Civil Engineering Faculty, study program - Building Construction. After finishing the university she started to work at the Technical University of Košice, Civil Engineering Faculty, Department of Structural Mechanics. Since 2009 she has worked at Civil Engineering Faculty, Technical University of Košice, study program - Theory and Design of Engineering Structures, as associate professor. Her research topic is design and optimization of structural elements and structures made of composite materials.

K. Kotrasova graduated at the Technical University of Košice, Civil Engineering Faculty, study program - Building Construction. After finishing of the university she started to work at RCB in Spišská Nová Ves as designer and then at the Technical University of Košice, Faculty of Mechanical Engineering, study program - Applied Mechanics. The research topics: seismic design of liquid storage ground-supported tanks, interaction problems of fluid, solid and subsoil.

Seismic behavior of fluid flows fully coupled with rectangular tank

K. Kotrasova and E. Kormanikova

Abstract**—**Liquid-containing tanks are used to store variety of liquids. This paper provides theoretic background for specification of impulsive and convective action of fluid in liquid storage rectangular container. Numerical model of tank seismic response - the endlessly long shipping channel was obtained by using of Finite Element Method (FEM), Arbitrary Lagrangian Eulerian (ALE), Fluid Structure Interactions (FSI) formulation in software ADINA. It was considered the horizontal ground motion of the earthquake in Loma Prieta.

Keywords**—**arbitrary Lagrangian-Eulerian formulation, earthquake, finite element method, Fluid-structure interaction, tank

I. Introduction

LIQUID-containing rectangular tanks are used to store variety of liquids, e.g. water for drinking and fire fighting, petroleum, oil, liquefied natural gas, chemical fluids, and wastes of different forms. Therefore, this type of structures must show satisfactory performance, especially, during earthquakes.

In particular, the analysis and design of liquid storage tanks against earthquake/induced action has been the subject of numerous analytical, numerical, and experimental works.

Numerous studies have been carries out about seismic behavior of ground-level cylindrical tanks. However, the conditions are not the same for underground tanks, rectangular tanks, and elevated tanks.

The seismic analysis and design of liquid storage tanks is, due to the high complexity of the problem, in fact, really complicated task. Number of particular problems should be taken into account, for example: dynamic interaction between contained fluid and tank, sloshing motion of the contained fluid; and dynamic interaction between tank and sub-soil. Those belong to wide range of so called fluid structure interactions (FSI). The knowledge of pressures acting onto walls and the bottom of containers, pressures in solid of tanks, liquid surface sloshing process and maximal height of liquid's wave during an earthquake plays essential role in reliable and durable design of earthquake resistance structure/facility - tanks. The analysis of a coupled multi-physics system is frequently required today to understand the behaviour of the system. In particular, the analysis of problems that involve fluid flows interacting with solids or structures is increasingly needed in diverse applications including ground-supported tanks used to store a variety of liquids.

To model the behavior of solid media the Lagrangian formulation of motion is employed, whereas, for a fluid flow analysis the Eulerian formulation is usually used since it is of interest to know the behavior of the fluid. However, when considering a fluid flow interacting with a solid medium and with free surface, the fluid domain changes as a function of time, and an arbitrary Lagrangian-Eulerian (ALE) description of the Eulerian and Lagrangian descriptions. We will concentrate in thes paper on the analysis of fluid flows that can deform. The flow equations are modeled using the Navier-Stokes equations of motion, and the constitutive relations of the structure are assumed to be either linear. [3-6,13]

II. Mechanical model

The dynamic analysis of a liquid - filled tank may be carried out using the concept of generalized single - degree - of freedom (SDOF) systems representing the impulsive and convective modes of vibration of the tank - liquid system as shown in Fig. 1. For practical applications, only the first convective modes of vibration need to be considered in the analysis, mechanical model. The impulsive mass of liquid m_i is rigidly attached to tank wall at height h_i. Similarly convective mass m_{cn} is attached to the tank wall at height h_{cn} by a spring of stiffness k_{cn}. The mass, height and natural period of each SDOF system are obtained by the methods described in [20].

$$m_i = m 2\gamma \sum_{n=0}^{\infty} \frac{I_1(\nu_n/\gamma)}{\nu_n^3 I_1^{'}(\nu_n/\gamma)}, \tag{1}$$

$$h_i = H \frac{\displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n I_1(\nu_n/\gamma)}{\nu_n^4 I_1^{'}(\nu_n/\gamma)} \left(\nu_n (-1)^n - 1\right)}{\displaystyle\sum_{n=0}^{\infty} \frac{I_1(\nu_n/\gamma)}{\nu_n^3 I_1^{'}(\nu_n/\gamma)}}, \tag{2}$$

This work was supported by the Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under Project VEGA 1/0477/15.

Kamila Kotrasova is with the Department of Structural Mechanics, Institute of Structural Engineering, The Technical University of Kosice, Faculty of Civil Engineering, Vysokoskolska 4, 042 00 Kosice, Slovak Republic (corresponding author to provide phone: +421 55 6024294; e-mail: kamila.kotrasova@tuke.sk).

Eva Kormanikova is with the Department of Structural Mechanics, Institute of Structural Engineering, The Technical University of Kosice, Faculty of Civil Engineering, Vysokoskolska 4, 042 00 Kosice, Slovak Republic (e-mail: eva.kormanikova@tuke.sk).

$$m_{cn} = m\frac{2\tanh(\lambda_n H/L)}{(\lambda_n H/L)(\lambda_n^2-1)}, \tag{3}$$

$$h_{cn} = H\left(1+\frac{1-\cosh(\lambda_n \cdot H/L)}{(\lambda_n \cdot H/L)\cdot\sinh(\lambda_n \cdot H/L)}\right), \tag{4}$$

$$k_{cn} = \omega_{cn}^2 m_{cn} \tag{5}$$

$$\omega_{cn}^2 = g\frac{\lambda_n}{L}\tanh(\lambda_n \gamma) \tag{6}$$

where $v_n = \pi\frac{2n+1}{2}$, $\gamma = H/R$, $I_1(\cdot)$ and $I_1^{'}(\cdot)$ denote the modified Bessel function of order 1 and its derivate. The derivate can be expressed in terms of the modified Bessel functions of order 0 and 1 as: $I_1^{'}(x) = \frac{dI_1(x)}{dx} = I_0(x) - \frac{I_1(x)}{x}$.

λ_n are soliution of Bessel function of first order, λ_1=1,8412; λ_2=5,3314; λ_3=8,5363, λ_4=11,71, λ_5=14.66 and $\lambda_{5+i}=\lambda_5+5\ i$ (i=1,2,...)).

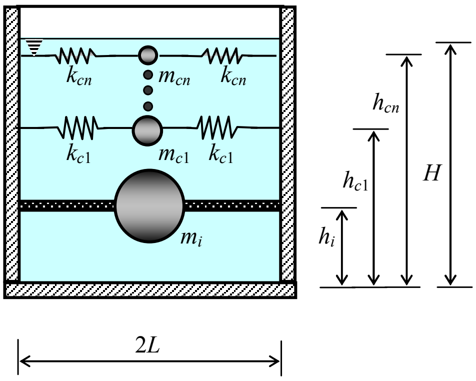


Fig. 1 liquid-filled tank modelled by generalised single degree of freedom systems

For a horizontal earthquake ground motion, the response of various SDOF systems may be calculated independently and then combined to give the base shear and overturning moment. The most tanks have slimness of tank γ, whereby 0,3 < γ < 3. Tank's slimness is given by relation $\gamma = H/L$, where H is the filling height of fluid in the tank and 2L is inside width of tank.

III. FEM - Fluid-structure interaction

For the fluid-structure interaction analysis, there are possible three different finite element approaches to represent fluid motion, Eulerian, Lagrangian and mixed methods. In the Eulerian approach, velocity potential (or pressure) is used to describe the behavior of the fluid, whereas the displacement field is used in the Lagrangian approach. In the mixed approaches, both the pressure and displacement fields are included in the element formulation, [1-2, 7].

In fluid-structure interaction analyses, fluid forces are applied into the solid and the solid deformation changes the fluid domain. For most interaction problems, the computational domain is divided into the fluid domain and solid domain, where a fluid model and a solid model are defined respectively, through their material data, boundary conditions, etc. The interaction occurs along the interface of the two domains. Having the two models coupled, we can perform simulations and predictions of many physical phenomena, [14, 18].

In many fluid flow calculations, the computational domain remains unchanged in time. Such the problems involve rigid boundaries and are suitable handled in Eulerian formulation of equilibrium equations [1, 11]. In the case where the shape of the fluid domain is expected to change significantly, modified formulation called Arbitrary Lagrangian-Eulerian (ALE) formulation was adopted to simulate the physical behavior of the domain of interest properly. The ALE description is designed to follow the boundary motions rather than the fluid particles. Thus, the fluid particles flow through a moving FE-mesh. Basically there are two different algorithms available for generation of possible moving mesh:

- remeshing of fluid domain, which is computationally expensive procedure,
- rezoning of FE-mesh of fluid domain. This procedure is quite fast while precise enough if no dramatic, changes of fluid domain is expected.

A. Governing Equations

Dynamic equilibrium of fluid domain involving effect of moving mesh describes modified Navier-Stokes equations. Let us to assume temperature independent problem. Then the balance of momentum by ALE formulation is

$$\rho[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}-\mathbf{v}_b).\nabla\mathbf{v}] = \nabla(-p\,\mathbf{I}+\boldsymbol{\tau}) + \rho\,\mathbf{g}\ , \tag{7}$$

where ρ is density of fluid, $\mathbf{v}$ velocity of fluid, $\mathbf{v}_b$ velocity of moving FE-mesh, p fluid pressure, $\mathbf{I}$ unit matrix, $\boldsymbol{\tau}$ stress tensor and $\mathbf{g}$ gravity acceleration.

Dynamic equilibrium of solid domain governs balance of momentum, e.g. in Cauchy form it is

$$div\,\boldsymbol{\tau} + \rho_0(\mathbf{b}-\ddot{\mathbf{u}}) = 0\,, \tag{8}$$

where ρ_0 is density of solid in initial configuration, $\mathbf{u}$ displacement, $\mathbf{b}$ body load, $\boldsymbol{\tau}$ stress tensor.

Together with traditional boundary conditions defined for fluid domain (pressure and velocity), additional special conditions are considered:

- free surface, the interface between fluid and gas,
- FSI boundary, common boundary between solid and fluid.

B. Fluid Domain, Free Surface, Boundary conditions

Dynamic boundary condition for free surface express balance of forces between interactive forces of liquid and gas

$$\begin{gathered}\mathbf{f}_l.\mathbf{n} + \sigma K = -\mathbf{f}_g\ \mathbf{n}\,,\\ \mathbf{f}_l.\mathbf{t} + \sigma K = -\mathbf{f}_g\ \mathbf{t}\,,\\ \mathbf{f}_l.\mathbf{s} + \sigma K = -\mathbf{f}_g\ \mathbf{s}\,.\end{gathered} \tag{9}$$

where $\mathbf{f}_l$ resp. $\mathbf{f}_g$ are forces exerted by liquid, resp. gas, $\mathbf{t}$ a $\mathbf{n}$ tangent and normal to FSI surface and $\mathbf{s}$ is surface tension (if present).

Kinematic boundary condition states the velocity at a point of free surface moves together with point of FE-mesh. Thus

$$(\mathbf{v}-\mathbf{v}_b).\mathbf{n}=0\,. \qquad (10)$$

C. FSI Boundary, conditions

Dynamic boundary condition defines stresses at the common FSI boundary, which is opposite and equal

$$\sigma_f=\sigma_s\,. \qquad (11)$$

Kinematic boundary condition assumes velocities and displacements of FSI boundary are the same

$$\mathbf{v}_f=\mathbf{v}_s\,,$$

$$\mathbf{u}_f=\mathbf{u}_s\,, \qquad (12)$$

where indexes f, resp. s mean fluid, resp. solid, Fig. 2.

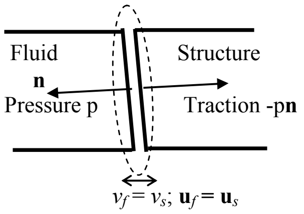


Fig. 2 common velocity and displacement of FSI boundary

D. Discretization by Finite Elements

Any of unknown physical variables in Finite element method is express in terms of nodal values instead of field value. That causes local discontinuity of the problem, but globally, with regards to whole FE model all governing equations are satisfied.

Unknown variables (displacement, velocity and pressure) are approximated using so called shape functions N.

$$\hat{\mathbf{u}}=\mathbf{N}\,\mathbf{U},\quad \hat{\mathbf{v}}=\mathbf{N}\,\mathbf{V},\quad \hat{\mathbf{p}}=\mathbf{N}\,\mathbf{P}\,, \qquad (13)$$

where **U**, **V**, resp. **P** are nodal values of initially unknown fields, **N** are shape functions.

Applying one of appropriate variation principle, governing equations are transformed into integral form, in which interpolations (13) are being easily incorporated and followingly proceeded in numerical calculation.

As the governing equations are basically nonlinear and time dependent, an appropriate linearization should be used together with a discretization in time domain. Plenty of methods by linearization and time discretization were published in the past. ADINA has implemented some of most popular of them [10,11,16-19].

IV. Numeric Example

In this study, the ground supported reinforced concrete rectangular tanks - endlessly long shipping channel is considered as shown in Fig. 3. The material characteristics of tank are: Young's modulus E = 37 GPa, Poisson ratio ν = 0.20, density ρ = 2550 kg/m^3. There is no roof slab structure covering the channel. The material characteristics of fluid filling (H_2O) are: bulk modulus $B = 2.1\cdot10^9$ N/m^2, density ρ_w = 1 000 kg/m^3. As the excitation input we consider horizontal earthquake load given by the accelerogram of the earthquake in Loma Prieta, California (18.10.1989), Fig. 4. In the analysis we use just the accelerogram for the seismic excitation in y - direction.

Dynamic time-history response of concrete open top rectangular liquid storage tanks - chipping channel was performed by application of Finite Element Method (FEM) utilizing software ADINA. Arbitrary-Lagrangian-Eulerian (ALE) formulation was used for the problem. Two way Fluid-Structure Interaction (FSI) techniques were used for simulation of the interaction between the structure and the fluid at the common boundary. The solid walls and base of the shipping channel was modeled by using 3D SOLID finite element under plain strain condition. The fluid inside the shipping channel was modeled by using 3D FLUID finite elements. As the excitation input was considered the load of input time dependent horizontal displacement measured during the earthquake Loma Prieta in California, in Fig. 6.

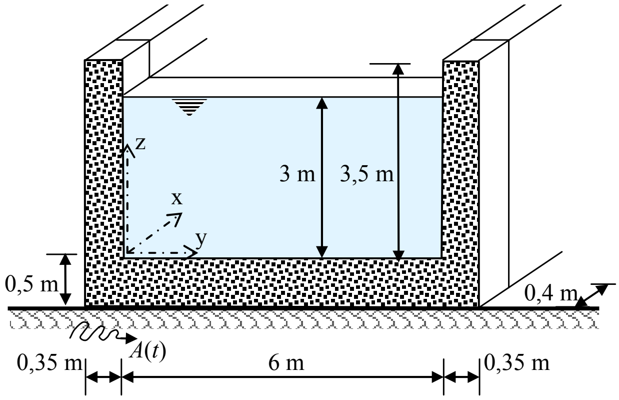


Fig. 3 details of tank geometry

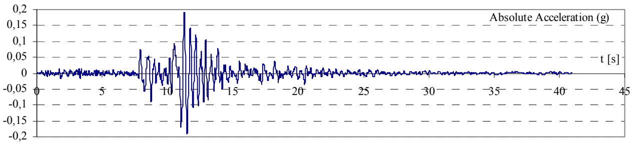


Fig. 4 considered ground motion Loma Prieta, California (18.10.1989)

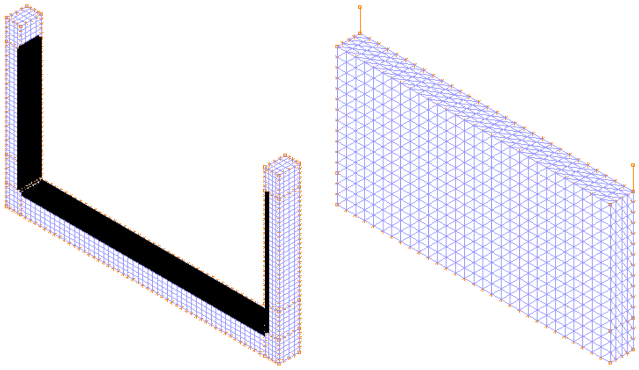


Fig. 5 FE-Model for 3D FSI analysis. FSI boundary on the solid domain at the left side, and fluid domain at the right side.

FE-Model for 3D FSI analysis was shown in Fig. 5. FSI boundary is on the solid domain (black color) at the left side, and fluid domain on the right side, Fig. 5.

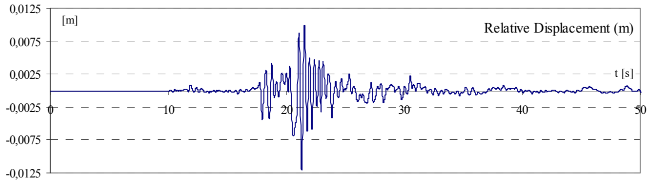


Fig. 6 input time dependent horizontal displacement measured during of earthquake Loma Prieta

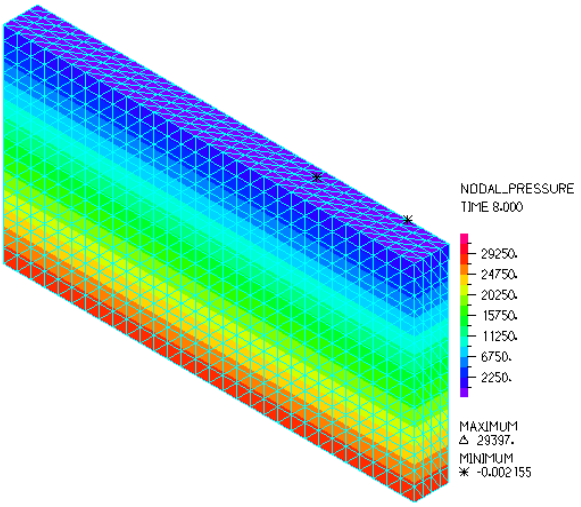


Fig. 7 pressure of fluid in time 8 s

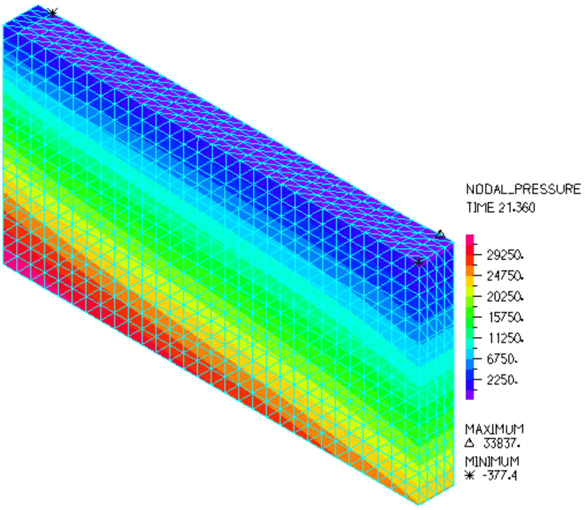


Fig. 8 pressure of fluid in time 21.36 s

The Fig. 7 presents distribution of pressure within fluid in time 8.0 s, whereas the same response of fluid shows Fig. 8 in time 21.36 s. In the time 8 s is influence hydrostatical pressure only and in time 21.36 s is time of obtained maximum pressure. The time dependent response of the pressure of fluid within time interval 17-27 s was described in Fig. 9 and Fig. 10, Fig. 9 in point ”DL” (Down Left edge of fluid region) and Fig. 10 in point ”DR” (Down Right edge of fluid region). It is seen that distribution of pressures in ”DL” and ”DR” points are asymmetric.

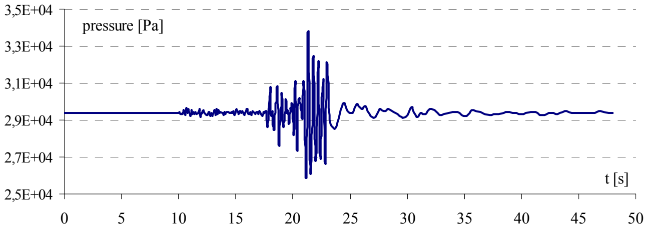


Fig. 9 time dependent response of the pressure of fluid in “DL” point

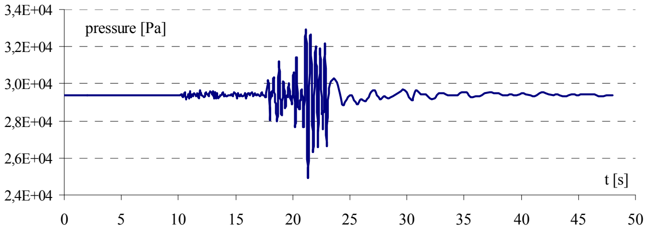


Fig. 10 time dependent response of the pressure of fluid in “DR” point

Fig. 11 shows more detail dependent responses of the fluid pressure within time interval 17-27 s together in points ”DL”, ”DR” and ”DM” (Down Middle of fluid region). In Fig. 12 is using smaller scale for dependent response of the fluid pressure within time interval 17-27 s in point ”DM”.

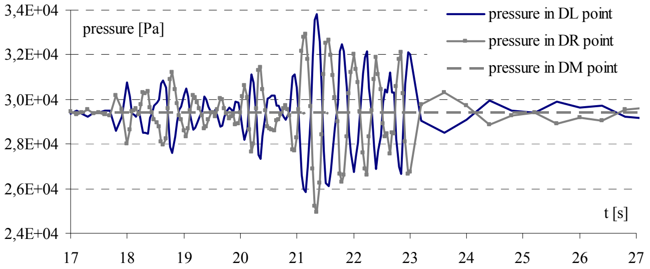


Fig. 11 time dependent response of the pressure of fluid in “DR”, “DL”, “DM” point

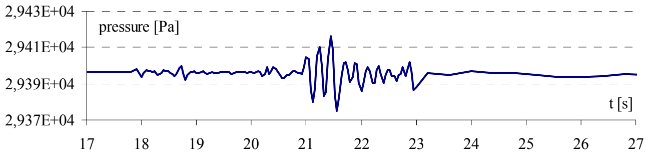


Fig. 12 time dependent response of the pressure of fluid in “DM” point

The Fig. 13 shows distribution of vertical displacement within fluid in time 21.56 s - the time of obtained maximum vertical displacement. The resulting time dependent vertical displacement of fluid in the point ”UL” (Up Left edge of fluid region on free surface) was presented in Fig. 14, whereas the same response of fluid in point ”UR” (Up Right edge of fluid region on free surface) was shown in Fig. 15. The timing of the peak response correlates well with peak excitation (Loma Prieta as in Fig. 4), which the numerical analysis makes realistic enough.

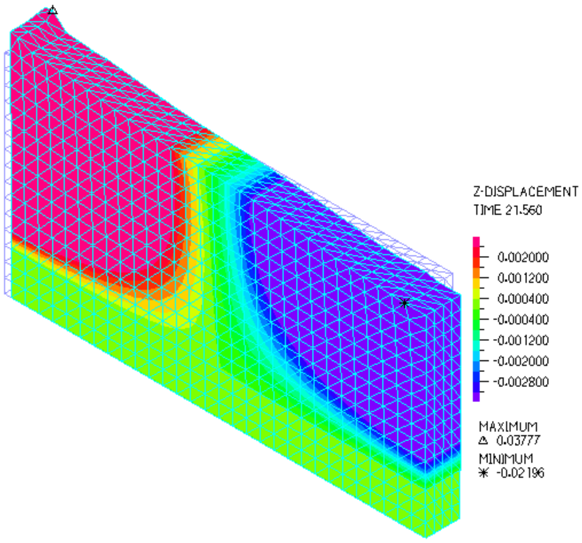


Fig. 13 shape of free surface and vertical displacement of fluid in time 21.56 s

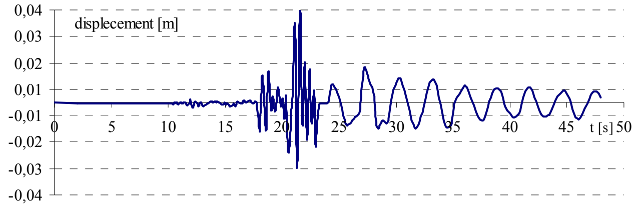


Fig. 14 time dependent response of the vertical displacement of fluid in "UL" point

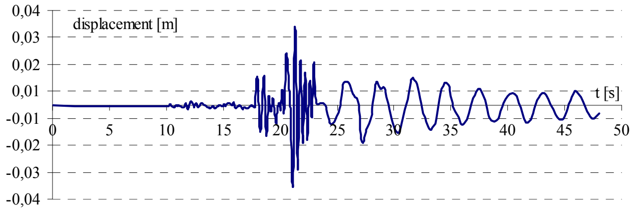


Fig. 15 time dependent response of the vertical displacement of fluid in "UR" point

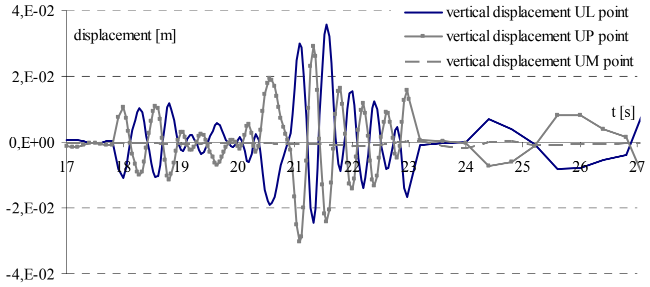


Fig. 16 time dependent response of the vertical displacement of fluid in "UR", "UL", "UM" point

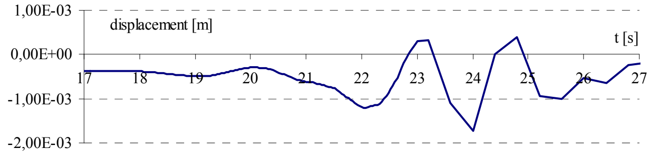


Fig. 17 time dependent response of the vertical displacement of fluid in "UM" point

Fig. 16 shows dependent responses of the fluid vertical displacement within time interval 17-27 s together in points "UL", "UR" and "UM" (Up Middle of fluid region on free surface). It is seen that distribution of vertical displacement in "DL" and "DR" points are asymmetric. In Fig. 17 is using smaller scale for dependent response of the fluid vertical displacement within time interval 17-27 s in point "DM".

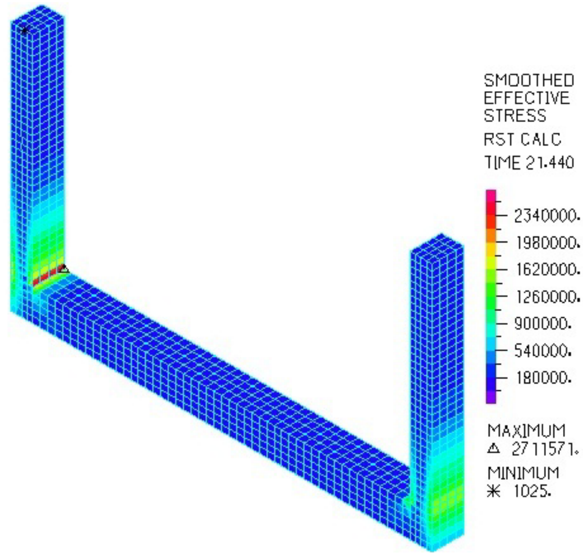


Fig. 18 shape and Von Mises stress of tank in time t = 21.44 s

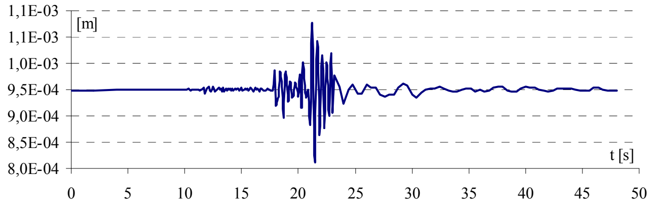


Fig. 19 time dependent response of tank relative horizontal displacement on the left corner

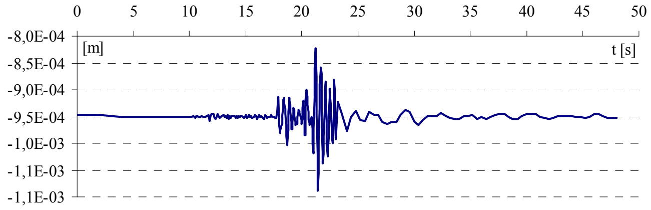


Fig. 20 time dependent response of tank relative horizontal displacement on the right corner

The Fig. 18 document distribution of Von Mises stress over the domain of interest in time t = 21.44 s, when peak responses were measured. The time dependent relative horizontal displacement of tank up corner to down corner was presented in Fig. 19 on the left wall, whereas the same response of fluid on the right wall in Fig. 20. Fig. 21 and Fig. 22 show dependent responses of relative horizontal of tank within time interval 17-27 s, Fig. 21 on left side and Fig. 22 on the right side.

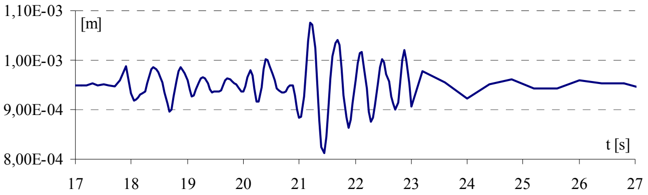


Fig. 21 time dependent response of tank relative horizontal displacement on the left corner within time interval 17-27 s

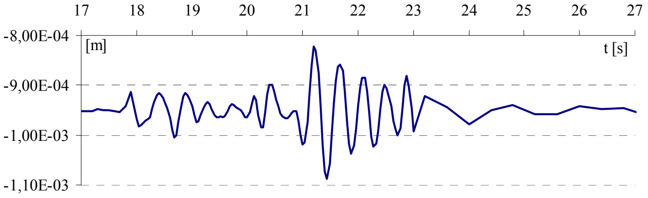


Fig. 22 time dependent response of tank relative horizontal displacement on the right corner within time interval 17-27 s

V. Conclusion

The ground supported rectangular endlessly long open top shipping channel was analyzed. The channel was excited by ground motion of Loma Prieta in California. Basic responses of the interest were: pressure in the fluid, displacement of the free fluid surface, structural deformation and stress distribution over the tank.

The peak hydrodynamic pressure and vertical displacement of fluid in the shipping channel along left side wall was similar, asymmetric and slightly higher to the peak values along the right wall.

Acknowledgment

This work was supported by the Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under Project VEGA 1/0477/15.

References

[1] G. K. Batchelor, An introduction to fluid dynamics. Cambridge: Cambridge University Press. 1967.

[2] A. Di Carluccio, G. Fabbrocino, E. Salzano, G. Manfredi, Analysis of pressurized horizontal vessels under seismic excitation In: ICSV18: 18th The World Conference on Earthquake Engineering: October 12 – 17, 2008, Beijing, China.

[3] N. Jendzelovsky, N., L. Balaz, Numerical Modeling of Cylindrical Tank and Compare with Experiment. In: Applied Mechanics and Materials. ISSN 1660-9336. Vol. 617: 6th International Scientific Conference on Dynamic of Civil Engineering and Transport Structures and Wind Engineering, DYN-WIND 2014, Donovaly, SR, 25.-29.5.2014 (2014), pp. 148-151.

[4] E. Kock, L. Olson, Fluid-structure interaction analysis by the finite element method a variational approach. International Journal for Numerical Methods in Engineering. Volume 31, Issue 3, pp. 463-491, March 1991, John Wiley & Sons, Ltd.

[5] K. Kralik, J. Kralik jr., Probability assessment of analysis of high-rise buildings seismic resistance, Advanced Materials Research, Volume 712-715, 2013, pp. 929-936.

[6] M. Krejsa, P. Janas, V. Krejsa, Software application of the DOProC method In: International Journal of Mathematics and Computers in Simulation Vol. 8, No. 1 (2014), pp. 121-126 ISSN: 1998-0159.

[7] K. Kotrasova, Sloshing of Liquid in Rectangular Tank. In: Advanced Materials Research. No. 969 (2014), p. 320-323. - ISBN 978-303835147-4, ISSN 1662-8985.

[8] K. Kotrasova, I. Grajciar, Dynamic Analysis of Liquid Storage Cylindrical Tanks Due to Earthquake. In: Advanced Materials Research. No. 969 (2014), pp. 119-124. - ISBN 978-303835147-4, ISSN 1662-8985.

[9] K. Kotrasova, I. Grajciar, E. Kormanikova, Dynamic Time-History Response of cylindrical tank considering fluid - structure interaction due to earthquake. Transport Structures and Wind Engineering. In: Applied Mechanics and Materials. No. 617 (2014), pp. 66-69, ISSN 1660-9336.

[10] H. Lamb, Hydrodynamics. 6th ed New York, Dover Publications; 1945.

[11] L. Meirovitch, Computational Methods in Structural Dynamics. Sijthoff & Noordhoff, 1980. Netherlands.

[12] J. Melcer, Experimental testing of a bridge. Applied Mechanics and Materials, Volume 486, 2014, pp. 333-340.

[13] J., Michel, M. Mihalikova, Degradation of pipes properties in creep conditions. In: Acta Metallurgica Slovaca. Roč. 6, č. 2, 2000, p. 108-115. ISSN 1335-1532

[14] E. Panulinova, S. Harabinova, Methods for Analyzing the Stability of an Earthen Dam Slope. In: Advanced Materials Research: SPACE 2013: 2nd International Conference on Structural and Physical Aspects of Civil Engineering: High Tatras, Slovakia, 27-29 November 2013. Vol. 969 (2014), p. 245-248. ISBN 978-303835147-4, ISSN 1022-6680.

[15] S. Rugonyi, K., J., Bathe,: On Finite Element Analysis of Fluid Coupled with Structural Interaction. In: CMES, vol. 2, no. 2, s. 195-212, 2001.

[16] O. Sucharda, J. Brozovsky, Bearing capacity analysis of reinforced concrete beams, International Journal of Mechanics, Volume 7, Issue 3, 192-200, 2013.

[17] B. Taraba, Z Michalec, V. Michalcova, T. Blejchar, M Bojko, M Kozubkova. CFD simulations of the effect of wind on the spontaneous heating of coal stockpiles. Fuel. 2014, vol. 118, pp. 107-112, ISSN 0016-2361, DOI: 10.1016/j.fuel.2013.10.064

[18] M. Zmindak, I. Grajciar, Simulation of the aquaplane problem. Computers and Structures. Vol. 64, Issue 5-6, September 1997, pp. 1155-1164.

[19] Manual ADINA. 71 Elton Ave, Watertown, MA 02472, USA, ADINA R&D, Inc., October 2005.

[20] Eurocode 8 – Design of structure for earthquake resistance – Part. 4: Silos, tanks and pipelines. Januar 2006.

K. Kotrasova graduated at the Technical University of Košice, Civil Engineering Faculty, study program - Building Construction. After finishing of the university she started to work at RCB in Spišská Nová Ves as designer and then at the Technical Technical University of Košice, Faculty of Mechanical Engineering, study program - Applied Mechanics. The research topics: seismic design of liquid storage ground-supported tanks, interaction problems of fluid, solid and subsoil.

E. Kormanikova graduated at the Technical University of Košice, Civil Engineering Faculty, study program - Building Construction. After finishing of the university she started to work at the Technical University of Košice, Civil Engineering Faculty, Department of Structural Mechanics as assistant. PhD. graduated at the Technical University of Košice, Faculty of Mechanical Engineering, study program - Applied Mechanics. Since 2009 she has worked at Civil Engineering Faculty TUKE, study program - Theory and Design of Engineering Structures, as associate professor. Her research topic is design and optimization of structural elements and structures made of composite materials.

The quick estimation of the stored heat in a cylindrical wall

M. I. Neacă, A. M. Neacă

Abstract**—**The problem of heat transfer through the walls with cylindrical symmetry occurs in many situations when analyzing electrical heating systems. It is all about heat transferred through multilayered cylindrical walls. Always heat transfer is accompanied by the stored heat of a part of the received heat energy into the wall. This paper aims to determine some relatively simple mathematical relationships which describe the heat storage processes in materials with cylindrical geometry. This allows the simulation of multilayer insulation systems under transient thermal regime, using Matlab-Simulink toolkit, in a manner similar to the one previously used by the authors, for walls with plane-parallel geometry.

Keywords**—**cylindrical symmetry, energy equivalence, heat transfer, stored energy.

I. Introduction

Modern times are characterized by the use of computer calculations in most various fields. This is due, mainly, to the increased speed and computing power, due to the improvement of hardware structures. But equally important is the quick development of some software packages that enable the development of software for numerical simulation, dedicated to different applications.

A simulation performed correctly has the advantage that one can find optimal solutions even before it starts the physical design of a plant. Also, through simulation can be determined the failure modes which, in many cases, can cause destructive effects on plants. So, during the design one can take the necessary steps to avoid their occurrence.

In the simulation of technical processes researchers are often faced with problems that are difficult to be mathematically modeled. In some other cases, mathematical models become extremely difficult. If for stady-state phenomena the mathematical models and simulations are quite well developed, in transient regime things get complicated. Often these regimes are carried out quickly, especially if one wants to achieve a plant driven in real-time by a computer system. Mathematical models that are based to the simulation should have a simple structure.

This paper aims to find a simple way to mathematical modeling of thermal systems with cylindrical structure, based on which to build future simulation software of multilayered cylindrical thermal insulation, commonly found in electro-thermal plants.

II. Preliminary

In the transient modeling and simulation, it should be pointed out that each layer of the composite wall will accumulate or give thermal energy from the heat flow that crosses it. This accumulated energy is proportional to the mass, specific heat of the layer and the temperature difference measured between two successive moments in time:

$$Q_s = \int_V m \cdot c_p \cdot \mathrm{d}\theta \quad [\mathrm{J}] \tag{1}$$

The problem in the modeling and simulation of this simple formula is to determine the change of the temperature field in the wall. This determination must be conducted with enough high speed, in order to allow its use in a real-time system for monitoring and control.

In order to increase the computation speed, for the homogenous layers, it is preferable to determine an equivalent increase of temperature ($\Delta\theta_{ech}$), then it will be used to determine the stored heat.

$$Q_s = m_{tot} \cdot c_p \cdot \Delta\theta_{ech} \quad [\mathrm{J}] \tag{2}$$

For the plane-parallel wall, crossed by a transverse heat flow, at which internal temperatures of each layer are distributed linearly between the temperatures of the two sides, the average temperature is calculated as the arithmetic average of the temperatures of the sides [1], [2].

The problem becomes more delicate when discussing about the heat transfer through a composite tubular wall, so based on a cylindrical model. In this case the thermal diffusion equation is written in cylindrical coordinates [4]:

$$\begin{aligned}&\frac{1}{r}\cdot\frac{\partial}{\partial r}\left(\lambda\cdot r\cdot\frac{\partial\theta}{\partial r}\right)+\frac{1}{r^2}\cdot\frac{\partial}{\partial\varphi}\left(\lambda\cdot\frac{\partial\theta}{\partial\varphi}\right)+\frac{\partial}{\partial z}\left(\lambda\cdot\frac{\partial\theta}{\partial z}\right)+w_q=\\&=\rho\cdot c_p\cdot\frac{\partial\theta}{\partial t}\end{aligned} \tag{3}$$

where:

M. I. Neacă is with the Department of Electrical, Energetic and Aerospace Engineering, Faculty of Electrical Engineering, University of Craiova, Craiova, Romania (e-mail: ineaca@elth.ucv.ro).

A. M. Neacă was with the Department of Electrical, Energetic and Aerospace Engineering, Faculty of Electrical Engineering, University of Craiova. She is now with HELLA Craiova (e-mail: neaca_andreea@yahoo.com).

λ = thermal conductivity [W/(m·K)];

w_q = the rate at which energy is generated per unit of volume of the environment [W/m^3];

$\rho \cdot c_p \cdot \dfrac{\partial \theta}{\partial t}$ = the change time rate of the thermal energy of the environment per unit of volume [W/m^3];

The paper presents a method of determining the representative layer (radius) within a tubular insulator. Its temperature can be used to calculate the thermal energy stored in the entire insulating layer.

It is envisaged a single insulating, cylindrical layer, of length L, crossed by a radial heat flow (Fig. 1).

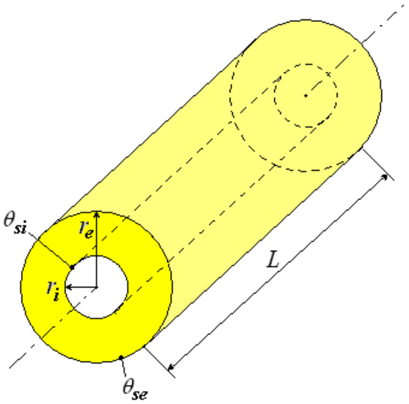


Fig.1. Cylindrical layer

In steady-state conditions, considering that within the insulating material there are no sources of heat, equation (3) becomes [3]:

$$\frac{1}{r} \cdot \frac{\partial}{\partial r}\left(\lambda \cdot r \cdot \frac{\partial \theta}{\partial r} \right) = 0 \tag{4}$$

Considering that $\theta_{si} > \theta_{se}$, the heat flow (constant in steady-state regime) that will cross the insulating layer, oriented from the inside to outside, will be:

$$\dot{Q}_r = \frac{\mathrm{d} Q_r}{\mathrm{d} t} \frac{\Delta\theta}{R_{t-cil}} = \frac{2\pi \cdot L \cdot \lambda \cdot (\theta_{si} - \theta_{se})}{\ln(r_i/r_e)} \tag{5}$$

For λ = const., considering the boundary conditions, equation (4) will have the solution for the temperature [5]:

$$\theta_r = \frac{\theta_{si} - \theta_{se}}{\ln(r_i/r_e)} \cdot \ln\left(\frac{r}{r_e} \right) + \theta_{se} \tag{6}$$

Relation (6) indicates that at each time the temperature in any point in the wall (at distance r from the axis) depends on the temperature difference between the two outer surfaces. During the transient process at least one parameter shall be modified.

III. Determining the Equivalent Computational Relationships

In determining the amount of energy stored in the cylindrical insulation wall, will be considered an elementary cylindrical volume, very thin, with the inner radius $(r - \mathrm{d}\,r/2)$, the outer radius $(r + \mathrm{d}\,r/2)$ and length L. The elementary volume considered for $L = 1\mathrm{m}$, will be:

$$V_{elm} = \pi(r + \mathrm{d}\,r/2)^2 - \pi(r - \mathrm{d}\,r/2)^2 = 2\pi \cdot r \cdot \mathrm{d}\,r \tag{7}$$

Since the elementary cylinder is very thin ($\mathrm{d}\,r \to 0$), it can be considered as limit that at a certain moment, the temperature is defined throughout its entire volume by equation (6). If we consider the initial temperature of the whole insulating coating as being equal to the ambient temperature (θ_0), one can define the warming of the outer surfaces also in the elementary volume:

$$\begin{cases} \Delta\theta_{si} = \theta_{si} - \theta_0 \\ \Delta\theta_{se} = \theta_{se} - \theta_0 \\ \Delta\theta_r = \theta_r - \theta_0 \end{cases} \tag{8}$$

If we modify equation (6) as:

$$\theta_r - \theta_0 = \frac{\theta_{si} - \theta_{se} + \theta_0 - \theta_0}{\ln(r_i/r_e)} \cdot \ln\left(\frac{r}{r_e} \right) + \theta_{se} - \theta_0$$

we obtain the formula for calculating the heating of the particles in the elementary volume:

$$\Delta\theta_r = \frac{\Delta\theta_{si} - \Delta\theta_{se}}{\ln(r_i/r_e)} \cdot \ln\left(\frac{r}{r_e} \right) + \Delta\theta_{se} \tag{9}$$

The energy stored in the elementary volume:

$$\begin{aligned} Q_{elm} &= m_{elm} \cdot c_p \cdot \Delta\theta_r = \rho \cdot c_p \cdot V_{elm} \cdot \Delta\theta_r = \\ &= 2(\pi \cdot \rho \cdot c_p) \cdot r \cdot \mathrm{d}\,r \cdot \left[\frac{\Delta\theta_{si} - \Delta\theta_{se}}{\ln(r_i/r_e)} \cdot \ln\left(\frac{r}{r_e} \right) + \Delta\theta_{se} \right] \end{aligned} \tag{10}$$

serves to calculate the part of the heat flow which is stored as heat in the insulating cylinder:

$$Q_s = \int_{r_i}^{r_e} 2(\pi \cdot \rho \cdot c_p) \cdot r \cdot \mathrm{d}\,r \cdot \left[\frac{\Delta\theta_{si} - \Delta\theta_{se}}{\ln(r_i/r_e)} \cdot \ln\left(\frac{r}{r_e} \right) + \Delta\theta_{se} \right] \tag{11}$$

Calculations finally lead to the value:

$$Q_s = 2\left(\pi \cdot \rho \cdot c_p\right)\cdot\left[\begin{array}{l}\dfrac{\Delta\theta_{si}-\Delta\theta_{se}}{\ln\left(r_i/r_e\right)}\left(\dfrac{r_i^2}{2}\ln\left(\dfrac{r_e}{r_i}\right)-\dfrac{1}{4}\left(r_e^2-r_i^2\right)\right)+\\ +\dfrac{\Delta\theta_{se}}{2}\left(r_e^2-r_i^2\right)\end{array}\right] \qquad (12)$$

Formula (11) can be implemented in a numerical simulation system, but involves making a large number of calculations for each simulation step.

In order to increase the simulation speed we will try to determine an equivalent heating (for the entire cylindrical insulating layer of unitary length). Considering the value of the equivalent heating as being equal to the heating of the particles at distance x from the axis ($\Delta\theta_{ech} = \Delta\theta_x$), according to equation (2) it is obtained:

$$Q_s = \pi \cdot \left(r_e^2 - r_i^2\right)\cdot \rho \cdot c_p \cdot \Delta\theta_x \qquad (13)$$

Equivalence in terms of energy implies that formulas (12) and (13) lead to the same result (after equivalence and calculations) and allows us to determine the equivalent heating:

$$\Delta\theta_x = \Delta\theta_{se} - \frac{\Delta\theta_{si}-\Delta\theta_{se}}{2\ln\left(r_i/r_e\right)} + \frac{r_i^2}{\left(r_e^2-r_i^2\right)}\cdot\left(\Delta\theta_{se}-\Delta\theta_{si}\right) \qquad (14)$$

Using the transformations defined by (8), will get the temperature at distance x from the axis:

$$\theta_x = \theta_{se} + K\cdot\left(\theta_{si}-\theta_{se}\right) \qquad (15)$$

where constant K can be determined based on the geometry of the cylindrical insulating layer:

$$K = \frac{1}{2\cdot\ln\left(r_e/r_i\right)} - \frac{1}{\left(r_e/r_i\right)^2 - 1} \qquad (16)$$

If in equation (6) we consider $r = x$ and θ_r is replaced by θ_x from equation (15), we can determine the distance measured from the axis of the cylinder, where the particles were heated with $\Delta\theta_{ech}$. This value is:

$$x = \frac{r_e}{\sqrt{e}}\cdot\left(\frac{r_e}{r_i}\right)^{\frac{r_i^2}{r_e^2-r_i^2}} \qquad (17)$$

IV. THE ANALYSIS OF RESULTS

Graphical representation of the results expressed by equations (15) and (17) allow verifying their correctness and highlighting some important conclusions. For this we plotted the sizes $\frac{x}{r_e}$ and K as functions of the ratio $\frac{r_e}{r_i}$. In the following figures are shown such representations for different areas of variation of ratio $\frac{r_e}{r_i}$.

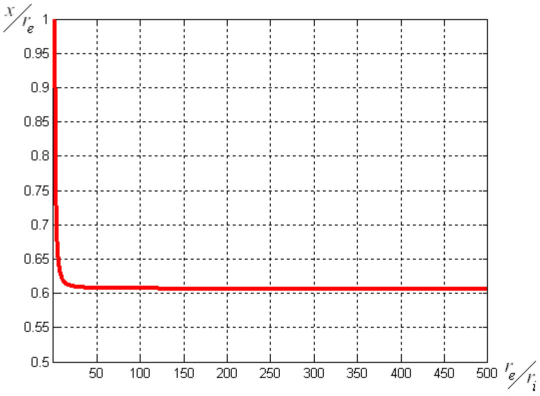


Fig.2. Equivalent radius x for [$0 < r_e/r_i < 500$]

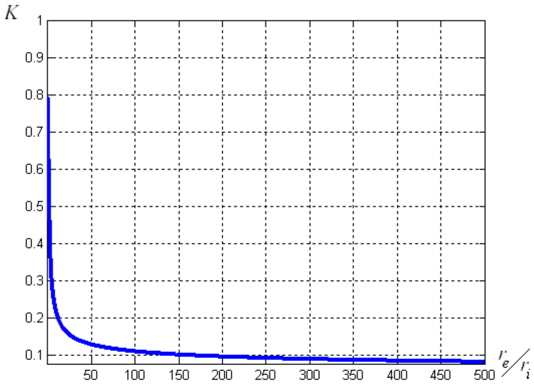


Fig.3. $K = K\left(r_e/r_i\right)$ for [$0 < r_e/r_i < 500$]

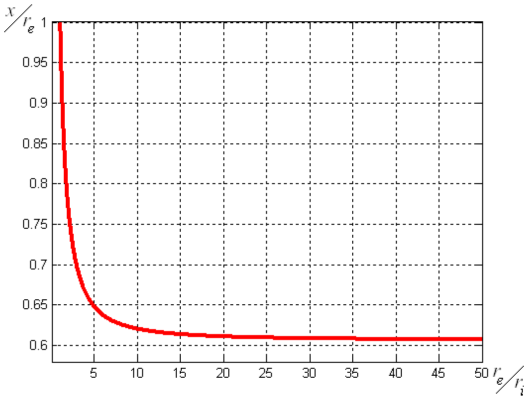


Fig.4. Equivalent radius x for [$0 < r_e/r_i < 50$]

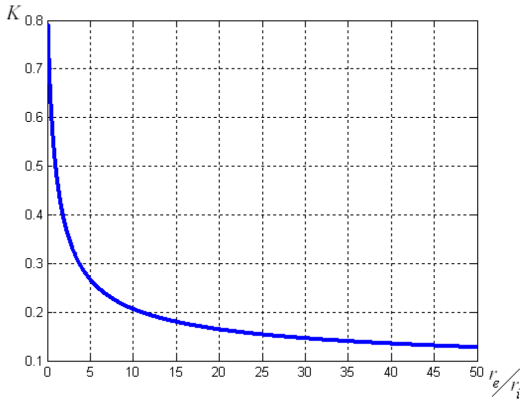


Fig.5. $K = K(r_e/r_i)$ for [$0 < r_e/r_i < 50$]

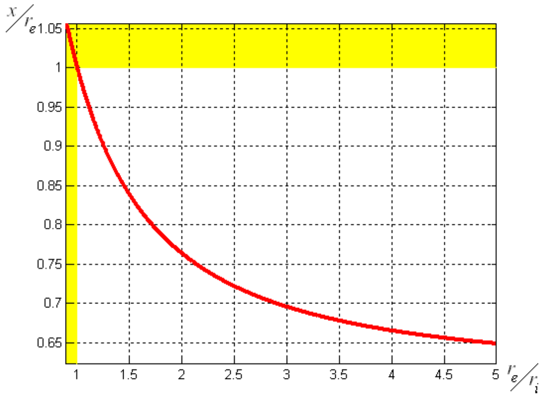


Fig.6. Equivalent radius x for [$0 < r_e/r_i < 5$]

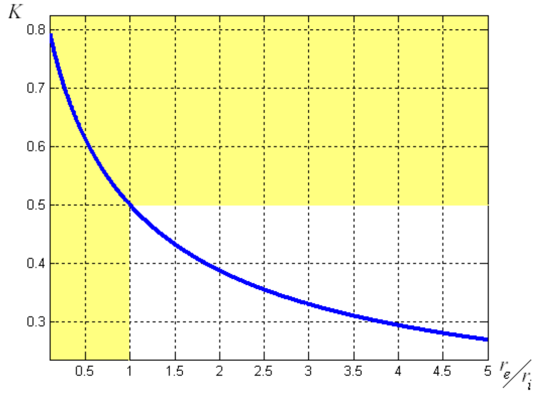


Fig.7. $K = K(r_e/r_i)$ for [$0 < r_e/r_i < 5$]

Solutions correctness is easily accomplished considering the limit case where the insulating cylinder is extremely thin, i.e. $r_e \to r_i$. In this limit case, their relationship becomes unitary and the value for $r_i \le x \le r_e$ will lead to $x/r_e \to 1$ (according to Fig. 6). For such very thin coating it can be considered that all the points have the same temperature (by default $\theta_i = \theta_e = \theta$). According to Fig. 7, we obtain $K = 0{,}5$ (the maximum value possible) and substituting it in equation (15) it results:

$$\theta_x = \theta_{se} + K \cdot (\theta_{si} - \theta_{se}) = \theta_{se} + 0{,}5 \cdot (\theta_{si} - \theta_{se}) = \frac{\theta_{si} + \theta_{se}}{2} = \theta$$

The analysis of the curves in Fig. 2 ÷ Fig. 7 shows a rapid decrease of the ratio x/r_e from the unitary value to about 0.65 for thin tubular insulating materials (with the ratio $r_e/r_i < 5$). Further increase of the thickness of the insulating material leads to a slow decrease of the ratio x/r_e , which tends to stabilize at a value less over 0.6.

Parameter K has a relatively similar evolution, except that at high values of r_e/r_i its speed of decrease is more pronounced.

A synthetic representation of the above conclusions can be followed in Table 1.

Table 1.

r_e/r_i	x/r_e	K
1	1	0.5
2	0.7642	0.3880
3	0.6958	0.3301
5	0.6486	0.2690
10	0.6208	0.2070
20	0.6111	0.1644
50	0.6075	0.1274
100	0.6068	0.1085
200	0.6066	0.0943
500	0.6065	0.0805

V. Conclusion

Generally, an equivalent mathematical relations system for describing a system can be achieved if first we establish a criterion which is the base of the equivalence. In this case the criterion was "the equivalence in terms of the energy stored" by the material. Relations obtained will be used for the simulation, using Matlab-Simulink toolkit, of some electro-thermal systems with cylindrical symmetry.

References

[1] NEACĂ, A.M., STOENESCU E., NEACĂ,M.I, *The simulink model of a multilayer wall for the study of the thermal transfer in transient regime,* Analele Universităţii din Craiova, seria Inginerie Electrică, no.34, 2010, ISSN 1842-4805, pag.200 – 205.

[2] NEACĂ, A.M., NEACĂ,M.I., *Thermal Transfer Trough the Walls of an Enclosed Area – Different Methods of Simulation,* Analele Universităţii din Craiova, seria Inginerie Electrică, no.36, 2012, ISSN 1842-4805, IEEE Catalog Number CFP1299S-PRT, ISBN 978-1-4673-1808-2, pag.470.

[3] BERGMAN,T.L., LAVINE,A.S., INCROPERA,F.P, DEWITT,D.P, *Fundamentals of Heat and Mass Transfer,* John Wiley & Sons, Inc.,U.S.A., 2011, ISBN 13 978-0470-50197-9

[4] LIENHARD IV,J.H., LIENHARD V,J.H., *A Heat Transfer Textbook,* Phlogiston Press, Cambridge, Massachusetts, U.S.A., 2008

[5] BEJAN,A., KRAUS,A.D., *Heat Transfer Handbook,* John Wiley & Sons, Inc., Hoboken, New Jersey, U.S.A., 2003, ISBN 0-471-39015-1

Determination Of Unknown Spacewise-Dependent Coefficient In a Parabolic Equation

Emine Can[a], Afet Golayoglu Fatullayev[b] and M.Aylin Bayrak[c]

[a]*Department of Physics, Kocaeli University, Kocaeli, Turkey*

[b]*Faculty of Commercial Science, Baskent University, Ankara, Turkey*

[c]*Department of Mathematics, Kocaeli University, Kocaeli,Turkey*

***Abstract*—**An inverse problem of determining an unknown spacewise-dependent coefficient of lower derivative in one-dimensional parabolic equation is considered. Using the additional condition the unknown coefficient is eliminated and problem is reformulated as a nonclassical parabolic equation along with initial and boundary conditions. Finite diffeerence method and appropriate iterative procedure are applied for discretization and for numerical solution. The effectiveness of the proposed numerical method and regularization technique is illustrated by some examples**.**

***Keywords*—**Finite Difference Method, Inverse Problems, Parabolic equation, Inverse problem, Unknown spacewise coefficient, Tikhonov regularization..

I. Introduction

Consider the inverse problem of finding $((u(x,t), b(x))$ in the parabolic equation

$$\rho(x,t)u_t - u_{xx} + b(x)u_x + d(x,t)u = f(x,t), \ (x,t) \in Q, \ (1)$$

with initial condition

$$u(x,0) = u_0(x), \quad x \in [0,l], \qquad (2)$$

the Dirichlet boundary conditions

$$u(0,t) = \beta_1(t), \quad u(l,t) = \beta_2(t), \quad t \in [0,T], \qquad (3)$$

and the additional final temperature condition

$$\int_0^T u(x,t)\chi(t)dt = \varphi(x), \quad x \in [0,l]. \qquad (4)$$

The inverse problem (1)-(4) is nonlinear and belongs to the class of so-called coefficient inverse problems in which, in addition to finding the function $u(x,t)$ one should find one or several unknown coefficients of the equation. The inverse problem has recently the attention of many mathematical stud-ed. Coefficient inverse problems were considered by numerous authors [1, 2, 3,4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. Recently, in [14, 15, 16, 17, 18] an additional assumption on the final measurements was used to guarantee uniqueness identification of spacewise source and heat conductivity and regularization approaches for parameter identification in parabolic differential equation.

We must point out only the paper [19], in which Solov'ev considered the problem of finding the coefficient multiplying the term $h(u,u_x)$ in a multidimensional quasilinear parabolic equations with smooth coefficients independent of t under the so-called terminal observation condition $u(x,T) = \varphi(x)$.

If (1)- (4) is used to describe the heat transfer system, the coefficient $b(x)$ is called radiative coefficient which is often dependent on the medium property [1].

For a given coefficient $b(x)$, the parabolic Eq. (1) which is defined as a direct problem consists of determination of the solution from the given initial condition.

It is well-known that in all cases, the inverse problem is ill-posed in the sense of Hadamard, while the direct problem is well-posed [2, 3, 4]. The ill-posedness, the numerical instability is the main difficulty for Eq.(1). Since data errors in the additional condition $\varphi(x)$ are inevitable, small changes in $\varphi(x)$ may lead to arbitrarily large changes in $b(x)$. Therefore, in order to obtain a stable solution, the output least-squares method with Tikhonov regularization is applied to inverse problem and the numerical solution is provided by the finite difference method.

Set $Q_T = [0,l] \times [0,\tau]$, $\tau \in [0,T]$, and Q_T=Q. The spaces

$L_2([0,l]),\ L_2(Q_T),\ L_\infty([0,l]), W_p^1([0,l]), W_p^2([0,l]),$

$W_2^{1,2}(Q_T), C^{0,\gamma}(Q_T),$

and

$C^{1,\gamma}(Q_T)$, $(1 \le p \le \infty)$, $0 < \gamma < 1$ used with the corresponding norms are treated in the usual sense [20,21]. Assumed that the functions appearing in the original data for the Eq.(1) and its initial and boundary conditions are measurable and satisfied the following conditions (A)-(D :) $(A)\ 0 < \rho_1 \le \rho(x,t) \le \rho_2$ and $|\rho_t| \le K_\rho, (x,t) \in Q.$

(B) $\left|d(x,t)\right| \le K_d$ and $\left|f(x,t)\right| \le K_f, \quad (x,t) \in Q.$

(C) $u_0(x) \in W_\infty^2([0,l])$ and $\beta_1(t) \in W_\infty^1([0,T]), \ i = 1, 2;$ $\|u_0\|_{w_\infty^2([0,l])} \le M_0$ and $\|\beta_i\| W_\infty^1([0,T]) \le K_\beta ; u_0(0) = \beta_1(0)$ and $u_0(l) = \beta_2(0).$

(D) $\chi(t) \in W_1^1([0,T]); \varphi(x) \in W_\infty^2([0,l]), \ \varphi'(x) \ge \varphi_1 > 0,$ and $\left|\varphi''(x)\right| \le K_\varphi, x \in [0,l]; \left|\chi(t)\right| \le \chi_0, t \in [0,T],$ and

$\|\chi'\|_{L_1([0,T])} \le K_\chi ; \int_0^T \beta_1(t)\chi(t)dt = \varphi(0)$

$\int_0^T \beta_1(t)\chi(t)dt = \varphi(0)$ and $\int_0^T \beta_2(t)\chi(t)dt = \varphi(l)$ where $\rho_1, \rho_2, M_0, K_\beta, \varphi = const > 0$ and $K_\rho, K_d, K_f, K_\varphi, K_\chi = const \ge 0.$

In $L_\infty([0,l]), B_R = \{b(x) \in L_\infty([0,l]) : \left\| \quad \right\|_{-L_\infty([0,l])} \le R\}, R = const > 0.$

Then the solution $(u(x,t), b(x))$ of the inverse problem (1)- (4) is exist and unique. These results were obtained in [18]. The paper is organized as follows. In Section 2, the numerical procedure for the solution of the inverse problem using implicit scheme combined with the iteration method is given. In Section 3, the description of used regularization method is given. Finally, in Section 4 numerical examples are presented.

II. NUMERICAL PROCEDURE

Let $\tau = \Delta t > 0$ and $h = \Delta x > 0$ be step length on time and space coordinate, $\{0 = t_0 < t_1 < .... < t_M = T\}$ and $\{0 = x_0 < x_1 < .... < x_N = l\}$ where $t_n = n\tau, n = 0,1,....M, \ x_i = ih, i = 0,1,....N,$ denote a partitions of the [0, T] and [0, l], respectively.

We multiply Eq.(1) by $\chi(t)$ _(t) and integrate the resulting relation over [0, T]. By taking into account conditions (4) and (A) - (D), we obtain the relation

$$b(x) = \frac{1}{\varphi'(x)}\{\int_0^T f(x,t)\chi(t)dt + \varphi''(x) - \rho(x,T)\chi(T)u(x,T)$$

$$+ \rho(x,0)\chi(0)u(x,0) + \int_0^T [(\rho\chi)_t - d\chi]u(x,t)dt\} \qquad (5)$$

The implicit finite difference approximation of the system (1)-(4) can be written in the form

$$\rho_i^{n+1}\frac{u_i^{n+1} - u_i^n}{\tau} - \frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}}{h^2} + b_i \frac{u_{i+1}^{n+1} - u_{i-1}^{n+1}}{2h} \qquad (6)$$

$$+ d_i^n u_i^{n+1} = f_i^{n+1}, \quad 1 \le i \le N-1, \quad 0 \le n \le M-1$$

$$u_i^0 = u_0(x_i), \quad 0 \le i \le N \qquad (7)$$

$$u_0^n = \beta_1(t_n), \quad u_N^n = \beta_2(t_n) \quad 1 \le n \le M \qquad (8)$$

The finite difference approximation of (5) is

$$b_{i+1} = \frac{2h}{\varphi_{i+2} - \varphi_i}\{F_{i+1} + \frac{\varphi_{i+2} - 2\varphi_{i+1} + \varphi_i}{h^2} - \rho_{i+1}^T \chi^T u_{i+1}^T$$

$$+ \rho_{i+1}^0 \chi^0 u_{i+1}^0 + G_{i+1}, \qquad (9)$$

where $\varphi_i = \varphi(x_i), F_i = \int_0^T f(x_i,t)\chi(t)dt$ and

$$G_i = \int_0^T [(\rho\chi)_t - d\chi]u(x_i,t)dt, \quad i = 1, 2\cdots$$

At the initial time step we take $u_i^{M(0)} = \beta_2(t_M), i = 1, 2\cdots, N-1.$ To find the next values $u_i^{M(s)}$ we solve the following system for s = 1, 2,….

$$\rho_i^{n+1}\frac{u_i^{n+1(s)} - u_i^{n(s)}}{\tau} - \frac{u_{i+1}^{n+1(s)} - 2u_i^{n+1(s)} + u_{i-1}^{n+1(s)}}{h^2} + b_{i+1}^{(s)}\frac{u_{i+1}^{n+1(s)} - u_{i-1}^{n+1(s)}}{2h}$$

$$+ d_i^n u_i^{n+1(s)} = f_i^{n+1}, \quad 1 \le i \le N-1, \quad 0 \le n \le M-1 \qquad (10)$$

$$u_i^{0(s)} = u_0(x_i), \quad 0 \le i \le N \qquad (11)$$

$$u_0^{n(s)} = \beta_1(t_n), \quad u_N^{n(s)} = \beta_2(t_n), \quad 1 \le n \le M \qquad (12)$$

where

$$b_{i+1}^{(s)} = \frac{2h}{\varphi_{i+2} - \varphi_i}\{F_{i+1} + \frac{\varphi_{i+2} - 2\varphi_{i+1} + \varphi_i}{h^2} - \rho_{i+1}^T \chi^T u_{i+1}^T$$

$$+ \rho_{i+1}^0 \chi^0 u_{i+1}^0 + G_{i+1}^{(s)}, \qquad (13)$$

$$G_i^s = \int_0^T [(\rho(x_i,t)\chi(t))_t - d(x_i,t)\chi(t)]u^s(x_i,t)dt, \qquad (14)$$

If for some s

$$\max_i \left|u_i^{M(s)} - u_i^{M(s-1)}\right| \le \varepsilon \qquad (15)$$

for a given ε then we stop the iteration and obtain the approximation $u_i^{M(s)}$

After that we find

$$b(x_i) = \frac{2h}{\varphi_{i+2} - \varphi_i}\{F_{i+1} + \frac{\varphi_{i+2} - 2\varphi_{i+1} + \varphi_i}{h^2} - \rho_{i+1}^T \chi^T u_{i+1}^T$$

$$+ \rho_{i+1}^0 \chi^0 u_{i+1}^0 + G_{i+1}^{(s)}, \qquad (16)$$

Solution of system (10)- (12) at each iteration step is obtained using TDMA (Three Diagonal Matrix Algorithm) method.

III. REGULARIZATION METHOD

Since the problem (1) - (4) is an ill-posed problem we apply the Tikhonov regularization method [22, 23, 24]. Let us shortly describe this method for the linear algebraic system in general form:

$$Az = w$$

Note that A is not necessarily a square matrix, in general case, when Tikhonov regularization is considered. Generally, the linear algebraic system may have no solution in classical sense. However, we can speak about the normal solution relative to some given vector z_0 (note that z_0 expresses guessed solution and is determined from physical considerations). The normal solution for any linear system exists and is unique. The problem of inding the normal solution is ill-posed, i.e. arbitrarily small changes in the input data can cause arbitrarily large changes in the solution. In the present work, finding normal solution of the Tikhonov regularization method is applied which is stable relative to small perturbations of the right-hand side of the system (experimental data). It is supposed that instead of true data we know their approximate values, i.e. instead of the vector w we have a vector $\tilde{w}$ such that $\|w - \tilde{w}\|. \le \delta$ where δ is the error of the measurements. The problem is to find for each value of δ an approximate solution $z_\delta \delta$, which converges to the exact normal solution z^* as $\delta \to 0$.

In Tikhonov regularization method z_δ is defined as a vector z_α, minimizing the Lagrange functional

$$L_\alpha[z, \tilde{w}] = \|Az - \tilde{w}\|^2 + \alpha\|z - z_0\|^2$$

with the regularization parameter $\alpha > 0$, which satisfies the condition $\|Az_\alpha - \tilde{w}\| = \hat{\delta}$, where $\hat{\delta} = \min_z \|Az_\alpha - \tilde{w}\| + 2\delta \cdot$

Minimization problem is equivalent to solving the Euler equation representing the noise level of the input data. Then one gets

$$(A^T A + \alpha I)z = A^T \tilde{w} + \alpha z_0$$

Note that as a vector z_0 we take the solution at the preceding time step when we apply Tikhonov regularization method for solving the linear algebraic system in the presented numerical algorithm.

IV. NUMERICAL RESULTS

Example. In the example we take,

$$\rho(x,t) = x + t, d(x,t) = xt, \ \beta_1(t) = t, \beta_2(t) = 1 + t$$

$$f(x,t) = x^3(t+2) + x(t^2+1) + t - 2, \chi(\mathrm{t}) = 1,$$

and additional condition

$$\int_0^T u(x,t)\ \chi(t)dt = \varphi_T(x) = x^2 + \frac{1}{2}, x \in [0,1], \ l = 1, T = 1$$

and the exact solution as

$$u(x,t) = x^2 + t$$

and the identifying coefficient as

$$b(x) = x^2$$

By solving the direct problem with these data by using implicit finite difference approximation (10) - (12), the solution values of $b(x)$ were recorded. In all numerical experiments, the noisy free output data to

$$u_T{}^\delta := u_T + \mu\delta \max_{[0,l]} |u_T|$$

where δ is the percentage of noise and μ are N normally distributed numbers mean 0 and variance 1 generated by the MATLAB function *randn(1,N).*

Then the inverse problem was solved with the additional condition to determine the unknown coefficient $b(x)$. Result of determination of $b(x)$ by the presented numerical procedure is illustrated in Fig. 1, corresponds to result with grids $N \times M = 50 \times 50, 50 \times 100, 50 \times 200, \varepsilon = 0.00005.$ It is seen that approximation of $b(x)$ is improved by increasing the number of nodes and that for sufficiently large number of nodes the agreement between numerical and exact solution becomes uniformly good.

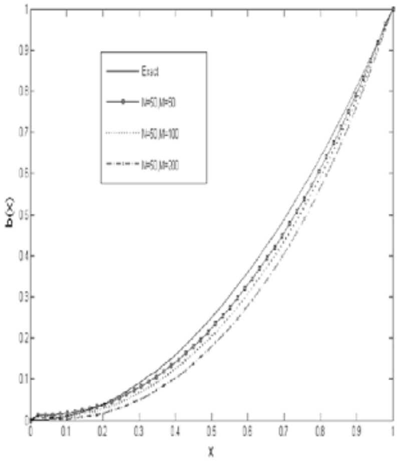


Fig1. Exact and numerical results of b(x).

In Figs. 2 - 3 are illustrated solution of (1), satisfying the initial and boundary condition (2) -(4) for various percentages of noise $\delta \in \{5, 15\}\%$ .Here,the integral condition are numerically calculated using Simpsons rule of integration. The best approach of value $n = 20$ and $m = 50$ for exact solution, these graphs are plotted for different noise value. The larger error value increases uctuation as a result of effect of regularization is clearly seen in the figures.

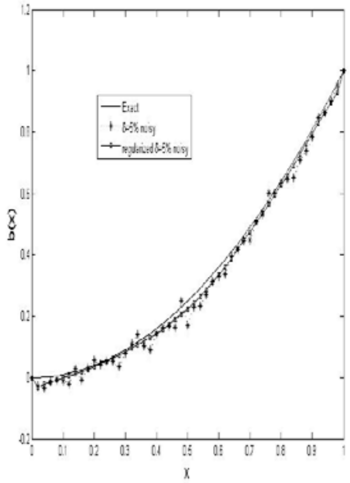


Fig2. Results of δ = 5% noisy with additional condition $u_T(x)$.

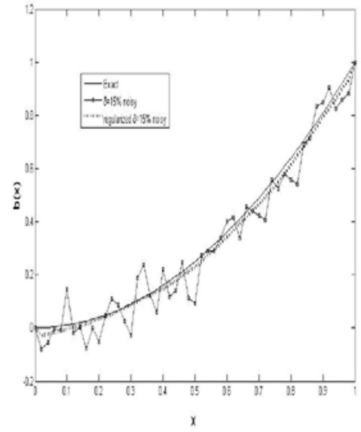


Fig3. Results of δ =1 5% noisy with additional condition $u_T(x)$.

We have also calculated the relative root mean square (rrmse) to analyze theerror between the exact and estimated coefficients, defined as,

$$rrmse = \sqrt{\frac{1}{N+1}\sum_{i==}^{N}\left(\frac{b_{app}(x_i) - b_{ex}(x_i)}{b_{ex}(x_i)}\right)^2} \ .$$

One of the main difficulties when we solve inverse and ill-posed problems is how to choose an appropriate regularization parameter α which must compromise between accuracy and stability. In our work, we have used trial and error.

In Table 1 , we present $rrmse$ values of the identified coefficient, respectively, for $\delta \in \{1,3,10\}\%$ noise. It can be observed that we obtain stable and reasonable accurate solution for $b(x)$ when we choose $\alpha = 10^{-1}$ which has minimum $rrmse$ value for b .

V. CONCLUSIONS

The inverse problem of finding the coefficient of a lower derivative in a parabolic equation on the plane has been considered. Implicit finite difference scheme combined with the iteration method and regularized minimization problem which were solved Tikhonov-type regularization approach with the additional final data are presented. The numerically obtained results are shown to be stable and accurate. The proposed method is readily extendable to solve some other ill-posed problems.

Table I. The *rrmse* values of estimated coefficients.

	$\alpha = 0$	$\alpha = 10^{-1}$	$\alpha = 10^{-2}$	$\alpha = 10^{-3}$
$\delta = 0$	5.4948E-4	5.4131E-4	5.4866E-4	5.4940E-4
$\delta = 1\%$	5.5195E-4	5.4259E-4	5.4571E-4	5.4918E-4
$\delta = 3\%$	5.7585E-4	5.3445E-4	5.4994E-4	5.7716E-4
$\delta = 0.1\%$	5.4978E-4	5.4085E-4	5.4763E-4	5.4888E-4

REFERENCES

[1] J. Cannon, “Determination of an unknown coefficient in a parabolic differential equation” , Duke Math. J., vol. 30, no. 2, pp. 313–323, 1963.

[2] N. Beznaschchenko, “On the determination of a coefficient in a parabolic equation” , Differ.Uravn..,vol.10, no. 1, pp.24–35, 1974.

[3] A. Prilepko and V. Solov'ev, “On the solvability of inverse boundary value problems for the determination of the coefficient preceding the lower derivative” , Differ.Uravn.,vol.23, no. 1, pp.136–143, 1987.

[4] A. Prilepko and A. Kostin, “Inverse problems of determining the coefficient in a parabolic equation” , I.Sibirsk.Mat.Zh.,vol.33, no. 3, pp.146–155, 1992.

[5] A. Prilepko, and A. Kostin, “ Inverse problems of determining the coeffcient in a parabolic equation”, II. Sibirsk. Mat. Zh. vol.34 , no. 5 pp. 147-162, 1993.

[6] A. Prilepko, and I. Tikhonov, “The principle of positiveness of a solution to a linear inverse problem and its application to the heat conduction coefficient problem”, Dokl.Akad.Nauk., vol.36, no.1,pp. 21-23,1999.

[7] A. Kozhanov, “ On the solvability of the inverse problem of determining the thermal conductivity coefficient”, Sibirsk. Mat.Zh.vol. 46, no.5, pp.1053-1071, 2005

[8] V. Kamynin, “On the inverse problem of determining the leading coefficient in a parabolic equation”, Mat. Zametki, vol.84, no. 1, pp.48-58, 2008.

[9] V. Kamynin, and A. Kostin, “ Two inverse problems of the determination of a coefficient in a parabolic equation”, Differ. Uravn., vol.4, no.3,,pp.372-383, 2010.

[10] A. Fatullayev, “Numerical method of identification of an unknown source
term in a heat equation”, Math. Prob. in Eng., vol. 82,pp.161-168,2002.

[11] A. Fatullayev, “Numerical procedure for the determination of an unknown coefficients in parabolic equations”, Comp. Physics Comm., vol.144, pp.29-33, 2002.

[12] A. Fatullayev, and S. Cula, “An iterative procedure for determining an unknown spacewise-dependent coefficient in a parabolic equation”, Appl. Math. Lett., vol.22, pp.1033-1037, 2009.

[13] L. Yan, F. Yang and C. Fu, “A meshless method for solving an inverse spacewise-dependent heat source problem”, J. of Comp. Phys., vol. 228, pp.123-136, 2009.

[14] D. Hao, P. Thanh, D. Lesnic and M. Ivanchov, “Determination of a source in the heat equation form integral observations”, Journal of Computational
and Applied Mathematics, vol. 264, pp. 82-98, 2014.

[15] F. Yang and C. Fu, “A simplified tikhonov regularization method for determining heat source”, Applied Mathematical Modelling, vol. 34, pp.3286-3299, 2010.

[16] A. Hazanee, D. Lesnic, M. Ismailov and N. Kerimov, “ An inverse time-
dependent source problem for the heat equation with a non-classical bound-
ary condition”, Applied Mathematical Modelling ,2015 1{15doi:http: //dx.doi.org/10.1016/j.apm.2015.01.058.

[17] A. D. Cezaro and B. Johansson, “A note on uniqueness in the identification of a spacewise dependent source and diffusion coefficient for the heat equation”, Springer,New York, vol.1996, pp.1-17, 2012.

[18] V. Kamynin and A. Kostin, “ Inverse problem of finding n coefficients of lower derivatives in a parabolic equation”, Differ. Eq., vol.50, no. 4,pp. 476-488,2014..

[19] V. Solov'ev, “On the control of a coefficient in a semilinear equation of the
parabolic type”, In Upravlenie nelineinymi sistemami (control for nonlinear systems), vol. 4, pp.36-40,1991.

[20] V. Kamynin, “Inverse problem of finding the coefficient of a lower derivative in a parabolic equation on the plane”, Diff. Eq. , vol. 448, no.2, pp. 214-223, 2012.

[21] M. Golberg, and C. Chen, “The method of fundamental solutions for elliptic boundary value problems”, Adv. Comput. Math., vol.9, no. (1-2), pp.69-95, 1998.

[22] P. Hansen, “Rank-deficient and discrete ill-posed problems”, SIAM,Philadelphia, 1998.

[23] A. Kirsch, “An introduction to the mathematical theory of inverse problems”,Springer,New York, 1996.

[24] H. Engl, M. Hanke, and A. Neubauer, “ Regularization of inverse problems”,Springer, The Netherlands, 1996.

Assc. Prof. Dr. Emine Can received her Ph.D. degree in Physics from Yildiz Technical University, Turkey, in 2000.
Her research interests include the areas of mathematical modelling, numerical methods for inverse problems, fuzzy set and systems and finitee difference methods for PDES. E. Can is an Associate Professor of Mathematical Physics at Kocaeli University, Kocaeli, Turkey.

Prof. Dr. Afet Golayoglu Fatullayev was born in Gecegozlu, Fizuli, Azerbaijan, in 1962. He got his B.Sc. and Ph.D. degrees in Applied Mathematics from Lomonosov Moscow State University. His research interests include computational methods and algorithms for inverse problems, low temperature plasma physics, finite difference methods for PDEs. A.G. Fatullayev is a Professor of Applied Mathematics at Baskent University, Ankara, Turkey.

Ass. Prof. Dr. Mine Aylin Bayrak r received B.Sc.and Ph.D. degrees in Applied Mathematics from Kocaeli University, in 2003. Her research interests include inverse problems, numerical methods, fuzzy differential equations

SFedU Software Package for Nucleotide Sequence Analysis

Boris J. Steinberg, Jumana M. Abu-Khalil, Mikhail G. Adigeyev, Sergey V. Avdyakov, Andrey A. Bout, Anton V. Kermanov, Evgeny A. Pshenichnyy, Galina V. Ramanchauskayte and Natalia Ponomareva

Abstract— We present a software package for nucleotide sequence analysis. Programs are being under development in the Southern Federal University (SFedU), Russia (available at mmcs.sfedu.ru/bio/). Currently, the package consists of modules for a number of bioinformatics problems including novel post-genomic challenges (like de-novo motif discovery). The most resource-consuming calculations are optimized by means of special data structures, parallel running, etc. For example, the package includes two variants of pairwise alignment algorithms (parallel block and parallel block with optimal memory usage) tailored for parallel run on multi-core processors and accelerators. The performance tests have confirmed that the former one is faster than Needleman-Wunsch algorithm by ~60% and EMBOSS tool by ~30% . The latter one aligns long sequences faster than EMBOSS Stretcher by 40% . The paper describes the current state of the project, some performance evaluation results and plans and approaches for future improvements.

Keywords— Bioinformatics, high performance computing, nucleotide sequence analysis.

This work was supported by project of the Ministry of Education and Science of Russia №6.703.2014/K.

Boris J. Steinberg is with the Southern Federal University, Dept of Math, Mechanics and Comp. Sci., Rostov-on-Don, Russia (e-mail: borsteinb@mail.ru).

Jumana M. Abu-Khalil is with the Southern Federal University, Dept of Math, Me-chanics and Comp. Sci., Rostov-on-Don, Russia (e-mail: jumana.abukhalil@gmail.com).

Mikhail G. Adigeyev is with the Southern Federal University, Dept of Math, Mechan-ics and Comp. Sci., Rostov-on-Don, Russia, (e-mail: madi@math.sfedu.ru).

Sergey V. Avdyakov is with the Southern Federal University, Dept of Math, Me-chanics and Comp. Sci., Rostov-on-Don, Russia (e-mail: sergeyavdya-kov@gmail.com).

Andrey A. Bout is with the Southern Federal University, Dept of Math, Mechanics and Comp. Sci., Rostov-on-Don, Russia (e-mail: a-bout@yandex.ru).

Anton V. Kermanov is with the Southern Federal University, Dept of Math, Mechan-ics and Comp. Sci., Rostov-on-Don, Russia and Research Institute for Plague Con-trol, Rostov-on-Don, Russia (e-mail: av-kermanov@mail.ru).

Evgeny A. Pshenichnyy is with the Southern Federal University, Research Institute of Biology, Rostov-on-Don, Russia, (corresponding author to provide phone: 007-863-2975070; address: 344090 pr. Stachki 104/1 Rostov-on-Don, Russia; e-mail: pshenichniy.eugene@gmail.com).

Galina V. Ramanchauskayte is with the Southern Federal University, Dept of Math, Mechanics and Comp. Sci., Rostov-on-Don, Russia (e-mail: galinka@lastbit.com).

Natalia Ponomareva is with the Southern Federal University, Research Institute of Biology, Rostov-on-Don, Russia (e-mail: nsponomareva@sfedu.ru).

I. Introduction

Bioinformatics is an interdisciplinary field where the close collaboration is required between mathematicians, computer scientists and biologists. The rapid growth of data, mostly genomic, due to appearance of high throughput sequence technologies (HST) leads to need of effective algorithms for nucleotide sequence data manipulations.

One of the most important bioinformatics problems is nucleotide sequence alignments. Global alignment is applied for the analysis of conservative parts of sequences, for pinpointing sequence relations and usually is the basic step in molecular phylogenetics inference [1]. The problem of pairwise global alignment can be stated as follows [2]: given a pair of sequences, build a two-row matrix such that the rows contain the characters of the sequences in order, interspersed with some spaces. Each alignment is assigned a numerical characteristic called 'score'. The score reflects the degree of similarity of the sequences. The problem is to build a maximum score alignment. The definition of the pairwise local alignment problem is essentially similar to the definition of global alignment, but the goal is to find a pair of substrings, one in each sequence, that maximizes the score. Whereas global alignment is used in bioinformatics for evaluation of similarity between two sequences, local alignment is used for detection of similar fragments within functionally related sequences. A number of dynamic programming algorithms have been designed to find global (Needleman-Wunsch algorithm [3]) or local alignments (Goad and Kanehisa [4], Sellers [5], Smith and Waterman [3], Waterman and Eggert [6], Hall and Myers [7]). There are fast local alignment algorithms, such as BLAST [8] reducing the amount of alignment time at the cost of exactness.

Search of regulatory sequences seems to be one of the most important task taking into account the data about non-coding genome regulatory roles obtained due to ENCODE project [9]. This problem could be formulated as a motif discovery [10]. The motif of choice even if it is known, however, could be subjected to changes and variation leading to other task of pattern search (inexact) for relatively huge DNA datasets in order to find genes joined together into regulatory loop.

Practical research in bioinformatics involves not only solving computationally difficult problems but also some additional tasks which are worth of automation. SFedU package provides several utilities for such tasks. Lots of bioinformatical software packages are available for biologists. However, most of them are either primarily user interface wrappers integrating third party tools, or specialized heuristic procedures [8],[11],[12]. The main focus of our package is the enhanced analyses of DNA sequences. The package is oriented both for long strings (applicable for alignments, as an example) and for substring search (motifs and patterns).

II. Global Alignments

Our package includes several alignment procedures. The main features of these procedures are running alignments in parallel blocks and memory usage optimization. The basic data structure of Needleman-Wunsch and other dynamic programming algorithms is similarity matrix. A similarity matrix is an (m+1) by (n+1) matrix where m and n are the lengths of the sequences to be aligned. Such model presents serious challenges for efficient parallel execution on present computers.

The basic idea of our algorithm is to divide the similarity matrix into blocks and then apply anti-diagonal approach. This algorithm uses less amount of memory than classic Needleman-Wunsch algorithm, as it does not save the similarity matrix as a whole. The size of each block is a flexible parameter, whose variation can lead to reduction of memory usage. Another our procedure for global alignment is based on the parallel global alignment algorithm with optimal memory usage. It performs alignment faster than Hirschberg's algorithm [3] and uses less memory than Needleman-Wunsch algorithm. This algorithm is based on [13] and the hyperplanes method [14]. We have performed a quantitative comparison of the two designed algorithms and other dynamic programming algorithms for optimal pairwise global alignment: Needleman-Wunsch, Hirshberg's algorithms and Myers and Miller algorithm [15], as well as other alignment tools: EMBOSS Stretcher [16], based on rapid modification of Myers and Miller algorithm, and Ngila [17], implementing a classic Miller and Myers algorithm. We used an Intel(R) Core(TM) i7 CPU @ 1.6 GHz computer with 4 GB RAM and 4 cores.

Tables 1 and 2 summarize the results of the tests. The columns correspond to algorithms: 'H' for Hirshberg's algorithm, 'NW' for Needleman-Wunsch algorithm, 'ES' for Emboss Strecher and 'N' for Ngila algorithm. The columns 'PAOM', 'PBAOM' and 'PB' correspond to procedures within our package: parallel algorithm with optimal memory usage, parallel block algorithm with optimal memory usage and parallel block algorithm respectively. The results in Table 1 show that the block algorithm is faster than Needleman-Wunsch algorithm by ~60% and Hirshberg's algorithm by ~80% on tested sets. The parallel block algorithm with optimal memory usage is faster than its base algorithm [13] by 35%. The table shows that the block algorithm is faster than EMBOSS Stretcher by ~30%. The parallel block algorithm with optimal memory usage aligns long sequences faster than EMBOSS Stretcher by 40% . Ngila tool is considerably slower comparing with designed algorithms.

Table 1. Calculation times for global alignment procedures

Seq. length (bp)	Time (sec)						
	H	NW	ES	N	PAOM	PBAOM	PB
2753 2517	0.4	0.2	0.082	0.27	0.3	0.2	0.1
8376 7488	3.3	1.3	0.53	2.41	1.2	0.8	0.5
268032 239616	2015.7	out of memory	584.9	14109	639.0	362.7	171.1

The memory usage comparison is presented in Table 2. The table shows that the parallel block algorithm with optimal memory usage requires less amount of memory than Needleman-Wunsch algorithm. The parallel block algorithm has been based on Needleman-Wunsch one, so it requires a lot of memory. However, it is more memory effective than Ngila tool. The designed algorithms are more memory intensive than EMBOSS Stretcher.

Table 2. Memory usage for global alignment procedures

Seq. length (bp)	Memory usage (MB)						
	H	NW	ES	N	PAOM	PBAOM	PB
2753 2517	0.7	27.0	3.95	13.9	1.3	1.4	1.4
8376 7488	0.9	240.0	4.2	121.3	2.9	3.0	5.4
268032 239616	11.2	out of memory	11.6	240.4	66.0	67.0	493.2

III Local Alignments

The parallel block procedure for local alignment is based on the Smith-Waterman dynamic programming algorithm combined with block division and anti-diagonal approach. The Parallel Block Local Alignment Algorithm with optimal memory usage is a combination of parallel block local alignment algorithm and the parallel block global alignment algorithm with optimal memory usage, discussed above.

We have compared the running time and memory usage of two presented algorithms and other algorithms for optimal local pairwise alignment, such as Smith-Waterman and Waterman-Eggert algorithms, as well as other alignment tools: EMBOSS [16] Water, which uses the Smith-Waterman algorithm (modified for speed enhancements), and EMBOSS Matcher, based on Bill Pearson's LALIGN application, version 2.0u4 (Feb. 1996). The experiments were performed on Intel(R) Core(TM) i7 CPU @ 1.6 GHz computer with 4 GB RAM and 4 cores. The results of running time testing are presented in Table 3. The columns correspond to the algorithms being compared: SW – Smith-Waterman, PBOM - Parallel Block algorithm with optimal memory usage, PB – Parallel Block algorithm, EW – EMBOSS Water and EM – EMBOSS Matcher. The table shows that both designed algorithms are faster than classic Smith-Waterman algorithm, the parallel block algorithm is faster than EMBOSS Water by ~95% and Matcher by ~90% on testing sets. The parallel block algorithm with optimal memory usage decrease the time of local alignment as compared to EMBOSS tools by 80%.

Table 3. Calculation times for local alignment procedures

Seq. length (bp)	Time (sec)				
	SW	PBOM	PB	EW	EM
2753 2517	0.28	0.14	0.06	0.42	0.80
8376 7488	0.61	0.60	0.27	5.05	3.19
268032 239616	out of memory	586.29	231.18	out of memory	2205.91

Table 4 reflects the memory usage of local alignment algorithms under study. The results of experiments show that EMBOSS Water, as Smith-Waterman algorithm, is memory intensive, and therefore could not perform alignment of long sequences. The EMBOSS Matcher uses less amount of memory than Smith-Waterman and parallel block algorithms. However, Matcher tool is more memory intensive than parallel block local alignment algorithm with optimal memory usage on short sequences; in case of long sequences, Matcher requires less space, but this difference is small.

Table 4. Memory usage for local alignment procedures

Seq. length (bp)	Memory Usage (MB)				
	SW	PBOM	PB	EQ	EM
2753 2517	53.63	2.09	1.51	56.86	4.19
8376 7488	479.59	3.78	3.92	482.29	4.71
268032 239616	Out of memory	108.15	155.91	Out of memory	21.84

IV Suffix Tree-Based Procedures

Pattern Search. The suffix tree module of presented bioinformatics package is responsible for solving several problems. Exact and inexact pattern search, palindrome search and motif discovery are among them. All of these problems deal with long sequences of characters of a fixed alphabet ({A,C,G,T} for DNA sequences). The huge amounts of data in bioinformatics require special data structures to be used for providing admissible performance. Among such data structures, suffix trees are considered as most appropriate. We have implemented several variants of suffix tree structure. The experiments show that truncated generalized suffix tree provides the best performance improvement for pattern search (both exact and inexact) and motif discovery problem, whereas for the palindrome search it does not give any speedup [18]. The truncation of tree may be useful for searching specific subsequences in a set of longer sequences [19]. There are different approaches for generalized suffix tree implementation [20], but we have designed a new modification that improves search time in a few cases. Our approach is based on branch and bounds method and could considerably reduce space and time requirements for specific problems. We have proposed a modification for Ukkonen suffix tree construction algorithm [21]. Two additional rules were suggested for decreasing both space and build time. These rules handle cases when maximum depth is reached for nodes. Our procedure uses special rules for updating information in leaf nodes as well.

As far as there are a few publicly available implementations, we have compared the results of our implementation with several software tools (MUMMER [22], SUDS [23], ERa [24]). The comparison demonstrates that our procedures either have a speedup or provide more functionality at the same speed. As an example, Table 5 shows that for exact pattern search our procedure has time complexity approximately equal to Mummer suffix tree procedure which has been declared to be the fastest tool for this problem [22]. Our suffix tree procedure has the additional ability of searching occurrences with mismatches. Results in Table 6 demonstrate some time parameters for our inexact search procedure.

Table 5. The times for exact pattern search

Sequence count/ length	Patterns count/ length	Mummer (sec)	Our package (sec)
1000/5000	100000/10	3.61	4.48
1000/5000	100000/30	6.82	4.21
1000/100000	100000/10	19.57	25.70
1000/100000	100000/30	23.11	20.21

Table 6. Time parameters for inexact pattern search

Sequence count/ length	Patterns count/ length	Number mismatches	Running time (sec)
1000/5000	100/10	2	531.5
1000/5000	100/10	5	981.6
1000/100000	10/30	2	255.1
1000/100000	10/30	12	935.1

Motif Discovery. Motifs are defined as sequence fragments (patterns) whose occurrences in a given set of sequences are statistically significant. The difference between motif discovery and pattern search problems is that in motif discovery problem the motif (pattern) is not given but should be found ('discovered') as the solution. There are several mathematical models for motif discovery problem, and different algorithms and software tools tend to use different models.

Table 7. The times for motif discovery (motif lengths: 10 - 30)

Q_P	Q_N	MERCI (sec)	Our package (sec)
100	1	130.91	5.92
500	1	36.14	5.85
1000	1	8.46	5.66
100	5	165.21	5.79
500	5	37.57	6.61
1000	5	11.49	5.78
100	10	164.11	6.70
500	10	34.41	5.91
1000	10	14.12	5.64

The motif discovery tool implemented within our package looks for exact (not heuristic) solutions of the following problem: given a positive set of sequences S = {S1, …, Sn} and a negative (or control) set C = {C1, …, Cm}, find all strings (motifs) of length between Lmin and Lmax such that each of the motifs occurs at least in QP positive sequences and in no more than QN negative sequences. The motif discovery module of our package uses truncated suffix tree for speeding up calculations.

We have compared the performance of our module and the software tool MERCI [25] which uses similar motif discovery model. The results are shown in Table 7. All experiments were performed on dataset with 1000 positive and 10 negative sequences of length 1000, the variated parameters were quorums (QP and QN).

V Other modules

In addition to the above-mentioned procedures the package includes several modules that are useful for practical bioinformaticians. These modules allow: 1) automatically download sequences from online genomic databases; 2) calculate character and short words statictics; 3) search long sequences for palindromes.

The problem of loading large amounts of nucleotide sequences arises frequently when analyzing bioinformatical data. Most genomic databases can be accessed using public APIs. We have implemented a separate module for batch loading information from such databases. The module allows fetching nucleotide sequences from Ensembl and NCBI databases though the Ensembl API and the Entrez Programming Utilities (eUtils). Obtained sequences and additional meta-information can be stored within any SQL relational database. Database storage allows the data to be accessed simultaneously from several processes, including processes running on different computers.

Character and short words (oligos, oligomers) statistics are important sources for exploring patterns of DNA organization. The SfedU package includes a module for fast calculation of DNA statictics.

A 'palindrome' in bioinformatics is a sequence which is equal to its own reverse complement and flanks it. The palindromes are significant for genetic research because such sites formed in exact parts of nucleic acids could be the reason for stalling polymerase during RNA synthesis in a cell or DNA replication. It occurs due to hairpin-like structure of these elements interferring the enzyme to act appropriately [26]. The SFedU bioinformatics package includes a module for fast searching palindromes in given nucleotide sequences.

VI Plans for Future Development

Degenerate (universal) primers are widely used for PCR (polymerase chain reaction) of variable sequences. Such approach is valid to identify bacteria in metagenomic studies, in particular for pathogen research [27]. Theoretical basis and limitations for in silico primer construction have been discussed in [28]. We have implemented heuristic search algorithm. Other widely used algorithms [28, 29] will be included with parallel programming options and user-friendly interface with lots of parameters which could help in fine-tuning the expected result.

Sequences of one species could be searched for SNPs (single nucleotide polymorphisms) or other type of mutations allowing creating mutational profile distinguishing species from another one. It could help in solving problems of biological classification, for instance, of pathogenic bacteria [30]. We are in progress with development of a program making a mutational profile for species or other taxonomic groups using a database with specific sequence marker defined computationally or by user. The other program application is the convenient graphical interface with sequence specific mutations pointed out against reference.

Suffix tree based procedures suffer from excessive memory requirements. Thus, reducing memory complexity while preserving the option of inexact pattern search is an urgent issue in our plans. There a several approaches addressing this problem, and ‘compressed suffix tree’ [31] looks promising enough. In addition, in the next version of the package we plan to implement switching between truncated tree and sparse suffix tree according to patterns size. It seems to improve suffix tree pattern search, both for exact and inexact cases.

References

1. Thompson, C. C., Chimetto, L., Edwards, R. A., Swings, J., Stackebrandt, E., Thompson, F. L.: Microbial Genomic Taxonomy. BMC genomics, 14(1), 913 (2013)
2. Pevzner, P.: Computational Molecular Biology: an algorithmic approach. Cambridge, Mass. MIT Press (2000).
3. Gusfield, D.: Algorithms on Strings, Trees and Sequences. Cambrige University Press, 556 p. (1997)
4. Goad, W. B., Kanehisa, M. I.: Pattern Recognition in Nucleic Acid Sequences I: A General Method for Finding Local Homologies and Symmetries. Nucl. Acids Res. 10, 247-263 (1982)
5. Sellers, P. H.: Pattern Recognition in Genetic Sequences by Mismatch Density. Bull. Math.Biol. 46, 501-514 (1984)
6. Waterman, M. S., Eggert, M.: A New Algorithm for Best Subsequence Alignments with Application to tRNA-rRNA Comparisons. J. Mol. Biol. 197, 723-728 (1987)
7. Hall, J. D., Myers, E. W.: A Software Tool for Finding Locally Optimal Alignments in Protein and Nucleic Acid Sequence. CABIOS 4, 35-40 (1988)
8. Basic Local Alignment Search Tool, http://blast.ncbi.nlm.nih.gov/
9. ENCODE Project, http://www.genome.gov/10005107
10. McGuire, A. M., Hughes, J. D., Church, G. M.: Conservation of DNA Regulatory Motifs and Discovery of New Motifs in Microbial Genomes. Genome Research, 10(6), 744-757.1 (2000)
11. Okonechnikov, K.; Golosova, O.; Fursov, M.; the UGENE team: Unipro UGENE: A Unified Bioinformatics Toolkit. Bioinformatics, 28, 1166-1167 (2012)
12. Molecular Evolutionary Genetics Analysis, http://www.megasoftware.net/
13. Abu-Khalil, Z. M., Morylev, R. I., Steinberg, B. Y.: Parallel Global Alignment Algorithm with the Optimal Use of Memory. Modern Problems of Science and Education, 1 (2013) http://www.science-education.ru/en/1 07-8139 (in Russian
14. Lamport, L.: The Parallel Execution of DO Loops. Commun. ACM, 83–93 (1974)
15. Myers, E.W., Miller, W.: Optimal Alignments in Linear Space. Computer Applications in the Biosciences, 4, 11-17 (1988)
16. Rice, P., Longden. I., Bleasby A.: EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics, 16, (6), 276-277 (2000).

17. Cartwright, R.A.: Ngila: Global Pairwise Alignments with Logarithmic and Affine Gap Costs. Bioinformatics. 23(11),1427-1428 (2007)
18. Adigeyev, M.G., Bout, A.A.: Efficiency Analysis of Applying Suffix Trees for Solving Some Bioinformatics Problems. Modern Problems of Science and Education. 6, http://www.science-education.ru/en/106-7418 (2012) (in Russian)
19. Schulz, M.H., Bauer, S., Robinson, P.N.: The Generalised k-Truncated Suffix Tree for Time- and Space-Efficient Searches in Multiple DNA or Protein Sequences. In: Int. J. Bioinformatics Research and Applications, 81-95. Inderscience Publishers (2008)
20. Bieganski, P., Ned1, J., Cadis, J.V., Retzel, E.E.: Generalized Suffix Trees for Biological Sequence Data: Applications and Implementation. In: System Sciences. Proc. of the Twenty-Seventh Hawaii International Conference, vol. 5 (1994).
21. Ukkonen, E.: On-line Construction of Suffix Trees. Algorithmica, 14, 249–260 (1995)
22. MUMmer, http://mummer.sourceforge.net/
23. SuDS project, http://www.cs.helsinki.fi/group/suds/cst/
24. Mansour, E., Allam, A., Skiadopoulos, S., Kalnis, P.: ERA: Efficient Serial and Parallel Suffix Tree Construction for Very Long Strings. PVLDB, 5(1), 49-60 (2011)
25. Vens, C., Rosso, M.N., Danchin, E.G.: Identifying Discriminative Classification-Based Motifs in Biological Sequences. Bioinformatics, 27, 1231–1238 (2011)
26. Lewis, S. M., Coté, A.G.: Palindromes and Genomic Stress Fractures: Bracing and Repairing the Damage. DNA repair 5.9, 1146-1160 (2006)
27. Miller, R. R., Montoya, V., Gardy, J. L., Patrick, D. M., Tang, P.: Metagenomics for Pathogen Detection in Public Health. Genome medicine, 5(9), 81 (2013)
28. Linhart, C., Shamir, R.: The Degenerate Primer Design Problem. Bioinformatics 18.suppl 1, S172-S181 (2002)
29. Rose, T. M., Henikoff, J.G., Henikoff, S.: CODEHOP (COnsensus-DEgenerate hybrid oligonucleotide primer) PCR Primer Design. Nucleic Acids Research 31.13, 3763-3766 (2003)
30. Nhung, P. H., Shah, M. M., Ohkusu, K., Noda, M., Hata, H., Sun, X. S.,Ezaki, T.: The dnaJ Gene as a Novel Phylogenetic Marker for Identification of Vibrio Species. Systematic and applied microbiology, 30(4), 309-315 (2007)
31. Sadakane, K.. Compressed Suffix Trees with Full Functionality. Theo. Comp. Sys., 41(4), 589-607 (2007).

Iterative solution methods for parabolic optimal control problem with constraints on time derivative of state function

E. Laitinen, A. Lapin

Abstract—An iterative solution method is proposed and investigated for the finite difference approximation of a parabolic optimal control problem with constraints on time derivative of the state function. Convergence analysis of the iterative methods is made. It is based on the general results on the convergence of iterative methods for constrained saddle point problem ([1], [2], [3]).The main feature of the constructed iterative solution methods is their easy implementation. Computational experiments confirm the theoretical results.

Index Terms—terative methods, saddel point problem, constraints in time derivativeterative methods, saddel point problem, constraints in time derivativei

I. Problem formulation

Let $\Omega = [0,1]^n, n \geqslant 1$, $\partial\Omega$ be its boundary, $Q_T = \Omega \times (0,T]$ and $\Sigma_T = \partial\Omega \times (0,T]$. Define a state problem with distributed control:

$$\frac{\partial y}{\partial t} - \Delta y = f + u \text{ in } Q_T; \quad \begin{array}{l} y = 0 \text{ on } \Sigma_T; \\ y = 0 \text{ for } t = 0,\, x \in \Omega, \end{array} \tag{1}$$

where function $f(x,t) \in L_2(Q_T)$ is given, while $y(x,t)$ and $u(x,t)$ are unknown state and control functions. This problem has a unique weak solution $y \in L_2(0,T;H_0^1(\Omega))$ such that $\dfrac{\partial y}{\partial t} \in L_2(Q_T)$.

Let objective function be defined by the equality

$$J(y,u) = \frac{1}{2}\int\limits_{Q_T}(y(x,t) - y_d(x,t))^2 dxdt + \frac{\alpha}{2}\int\limits_{Q_T} u^2 dxdt, \quad \alpha > 0, \tag{2}$$

with given observation function $y_d(x,t) \in L_2(Q_T)$.

Finally, define the sets of the constraints:

$$\begin{array}{l} U_{ad} = \{u \in L_2(Q_T) :\ |u| \leqslant \bar{u} \text{ a.e. } Q_T\}; \\ Y_{ad} = \{y :\ \dfrac{\partial y}{\partial t} \in L_2(Q_T) \text{ and } y_{\min} \leqslant \dfrac{\partial y}{\partial t} \leqslant y_{\max} \text{ a.e. } Q_T\}, \end{array} \tag{3}$$

with given constants $\bar{u} > 0, y_{\min}$ and $y_{\max}$.

We will solve the following optimal control problem:

$$\begin{array}{c} \min\limits_{(y,u)\in K} J(y,u), \\ K = \{(y,u) \in Y_{ad} \times U_{ad} :\ \text{equation (1) holds}\}. \end{array} \tag{4}$$

Lemma 1. Problem (4) has a unique solution (y,u) if $K \neq \emptyset$.

E. Laitinen is with the Department of Mathematical Sciences, University of Oulu, Oulu, Finland. E-mail: erkki.laitinen@oulu.fi

A. Lapin is with the Department of Computational Mathematics and Cybernetics, Kazan Federal University, ul. Kremlevskaya, 18, Kazan 420008, Russia. E-mail: avlapine@mail.ru

II. Finite difference approximation of the optimal control problem

We suppose for the simplicity that $f(x,t)$ is a continuous function in $\bar{\Omega} \times [0,T]$. Let ω_x be the uniform mesh of the meshsize h on $\bar{\Omega}$, $\operatorname{card}\omega_x = N_x$. By A we denote the mesh approximation of Laplace operator with homogeneous Dirichlet boundary conditions. Then the spectrum of symmetric and positive definite matrix A belongs to the segment $[\nu_{\min}(A), \nu_{\max}(A)]$, where $\nu_{\max}(A)$ has an order h^{-2}, while $\nu_{\min}(A) > 0$ is limited from below by a constant which doesn't depend on h. For the mesh functions defined on the mesh ω_x and the vectors from $\mathbb{R}^{N_x}$ of their nodal values we will use the same notations. By $(.,.)_x$ and $\|.\|_x$ we denote the inner product and euclidian norm in $\mathbb{R}^{N_x}$. Further, let $\omega_t = \{t_j = j\tau,\ j = 0, 1, \dots M;\ M\tau = T\}$ be a uniform mesh on the segment $[0,T]$. Denote by $y_j = y(x,t_j)$ a mesh function on a time level $t_j \in \omega_t$, or equivalently the vector $y_j \in \mathbb{R}^{N_x}$ of its nodal values.

Let us approximate state equation (1) by weighted finite difference:

$$\frac{1}{\tau}(y_j - y_{j-1}) + A(\delta y_j + (1-\delta)y_{j-1}) = f_j + u_j,\ j = 1,\dots,M,\ y_0 = 0 \tag{5}$$

with $\delta \in [0,1]$. We suppose that the stability condition $\tau < 2(\nu_{\max}(A)(1-2\delta))^{-1}$ in the case $\delta < 1/2$ is satisfied. In the case $\delta \geqslant 1/2$ this finite difference problem is unconditionally stable.

Matrix $L \in \mathbb{R}^{MN_x \times MN_x}$:

$$\begin{array}{c} (Ly)_j = \{\dfrac{1}{\tau}(y_j - y_{j-1}) + A(\delta y_j + (1-\delta)y_{j-1}) \text{ for} \\ j = 2,\dots,M;\ \dfrac{1}{\tau}y_1 + \delta A y_1 \text{ for } j = 1\} \end{array}$$

is positive definite (the stability condition condition is supposed to be satisfied in the case $\delta < 1/2$).

The objective function (2) is approximated by the mesh objective function

$$I(y,u) = \frac{1}{2}\sum_{j=1}^{M}\|y_j - y_{dj}\|_x^2 + \frac{\alpha}{2}\sum_{j=1}^{M}\|u_j\|_x^2, \tag{6}$$

while the mesh approximations of the constraints sets (3) are

$$\begin{array}{l} U_{ad}^h = \{u : |u(x,t)| \leqslant \bar{u}\, \forall x \in \omega_x, \forall t \in \omega_t\}, \\ Y_{ad}^h = \{y : \tau y_{\min} \leqslant y_j - y_{j-1} \leqslant \tau y_{\max}\ (y_0 = 0)\, \forall x \in \omega_x, \forall t \in \omega_t\}. \end{array}$$

Now mesh optimal control problem reads as follows:

$$\begin{array}{c} \min\limits_{(y,u)\in K_h} I(y,u), \\ K_h = \{(y,u) \in Y_{ad}^h \times U_{ad}^h :\ \text{equation (5) holds}\}. \end{array} \tag{7}$$

Lemma 2. Problem (7) has a unique solution if $K_h \neq \emptyset$.

III. Saddle point problem

Let us define matrix $R \in \mathbb{R}^{MN_x \times MN_x}$, $(Ry)_j = \{y_j - y_{j-1}$ for $j = 2, \ldots, M;$ y_1 for $j = 1\}$, and vector $p = Ry$. Then we can replace the constraint $y \in Y_{ad}^h$ in the optimal control problem by the following constraint: $p \in P_{ad}^h = \{\tau y_{\min} \leqslant p_j \leqslant \tau y_{\max},\ j = 1, 2, \ldots M\}$. Let further θ and φ be indicator functions of the sets P_{ad}^h and U_{ad}^h: $\theta(p) = \{0$ if $p \in P_{ad}^h;$ $+\infty$ otherwise$\}$, $\varphi(u) = \{0$ if $u \in U_{ad}^h;$ $+\infty$ otherwise$\}$. Then mesh optimal control problem (6) can be written as

$$\min_{Ly=f+u,\ p=Ry} \{I(y,u) + \theta(p) + \varphi(u)\}.$$

Define Lagrange function

$$\mathcal{L}(y,u,\lambda) = I(y,u) + \theta(p) + \varphi(u) + (\lambda, Ly - u - f) + (\mu, Ry - p),$$

where $(.,.)$ is the inner product in $\mathbb{R}^{MN_x}$. Its saddle point satisfies the following system (cf. e.g. [4]):

$$\begin{pmatrix} E & 0 & 0 & L^T & R^T \\ 0 & \alpha E & 0 & -E & 0 \\ 0 & 0 & 0 & 0 & -E \\ L & -E & 0 & 0 & 0 \\ R & 0 & -E & 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ u \\ p \\ \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} 0 \\ \partial\varphi(u) \\ \partial\theta(p) \\ 0 \\ 0 \end{pmatrix} \ni \begin{pmatrix} y_d \\ 0 \\ 0 \\ f \\ 0 \end{pmatrix}, \tag{8}$$

where $E \in \mathbb{R}^{MN_x \times MN_x}$ is unit matrix, $\partial\varphi(u)$ and $\partial\theta(p)$ are the subdifferentials of φ and θ respectively.

Lemma 3. Let the strengthened variant of the assumption $K_h \neq \emptyset$ be satisfied:

> There exists a pair $(y^*, u^*) \in \operatorname{int} Y_{ad}^h \times \operatorname{int} U_{ad}^h$ such that $Ly^* = f + u^*$.
>
> Then saddle point problem (8) has a nonempty solution set $X = \{(w, \eta)\}$ and w is unique.

IV. Iterative methods

Using the notations $w = (y,u,p)^T$, $\eta = (\lambda, \mu)^T$, $g_1 = (y_d, 0, 0)^T$, $g_2 = (f, 0)^T$, $\partial\psi(w) = (0, \partial\varphi(u), \partial\theta(p))^T$ and

$$\mathcal{A} = \begin{pmatrix} E & 0 & 0 \\ 0 & \alpha E & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathcal{B} = \begin{pmatrix} L & -E & 0 \\ R & 0 & -E \end{pmatrix}$$

problem (8) can be written in the following compact form:

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^T \\ \mathcal{B} & 0 \end{pmatrix} \begin{pmatrix} w \\ \eta \end{pmatrix} + \begin{pmatrix} \partial\psi(w) \\ 0 \end{pmatrix} \ni \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}. \tag{9}$$

The degenerate matrix $\mathcal{A}$ is an obstacle to the application of Uzawa-type iterative methods for solving (9). To overcome this deficiency we use two equivalent transformations of (8) and obtain the saddle point problems with positive definite matrices instead of $\mathcal{A}$. In both transformations we use the last equation of system (8), and obtain the variants of saddle point problem (9) with the matrices

$$\mathcal{A}_{1r} = \begin{pmatrix} E & 0 & 0 \\ 0 & \alpha E & 0 \\ -rR & 0 & rE \end{pmatrix} \text{ or}$$

$$\mathcal{A}_{2r} = \begin{pmatrix} E + rR^T R & 0 & -rR^T \\ 0 & \alpha E & 0 \\ -rR & 0 & rE \end{pmatrix}, \ r > 0,$$

instead of $\mathcal{A}$ and with the same matrix $\mathcal{B}$, function ψ and vectors g_1, g_2.

Lemma 4. Matrix $\mathcal{A}_{1r}$ is positive definite for $0 < r < 1$ and matrix $\mathcal{A}_{2r}$ is positive definite for any $r > 0$. Moreover, for these parameters r they are energy equivalent to block-diagonal matrix $\mathcal{A}_0 = \operatorname{diag}\begin{pmatrix} E & \alpha E & rE \end{pmatrix}$ with constants of the equivalence, which depend only on r:

$$(1 - \sqrt{r})(\mathcal{A}_0 z, z) \leqslant (\mathcal{A}_{1r} z, z) \leqslant (1 + \sqrt{r})(\mathcal{A}_0 z, z),$$

$$\sigma_0(r)(\mathcal{A}_0 z, z) \leqslant (\mathcal{A}_{2r} z, z) \leqslant \sigma_2(r)(\mathcal{A}_0 z, z), \ \forall z = (y, u, p),$$

where $\sigma_0(r) = (1 + 2r + 2\sqrt{r + r^2})^{-1}$, $\sigma_1(r) = 2r(1 + 5r + \sqrt{1 + 6r + 25r^2})^{-1}$.

A preconditioned Uzawa-type iterative method for solving saddle point problem (9) reads as

$$\begin{aligned} &\mathcal{A}w^{k+1} + \partial\psi(w^{k+1}) \ni \mathcal{B}^T \eta^k + g_1, \\ &\frac{1}{\rho} D(\eta^{k+1} - \eta^k) + \mathcal{B}w^{k+1} = g_2, \end{aligned} \tag{10}$$

where D is a symmetric and positive definite matrix (preconditioner), $\rho > 0$ is an iterative parameter.

Due to [1] iterative method (10) converges for any initial guess η^0 (convergence means $(w^k, \eta^k) \to (w^*, \eta^*) \in X$ for $k \to \infty$) if the pair "preconditioner D - parameter ρ" satisfies one of the following (equivalent) assumptions:

$$\mathcal{A}_s \geqslant \frac{(1+\varepsilon)\rho}{2} \mathcal{B}^T D^{-1} \mathcal{B} \text{ or } D \geqslant \frac{(1+\varepsilon)\rho}{2} \mathcal{B} \mathcal{A}_s^{-1} \mathcal{B}^T, \ \varepsilon > 0,$$

where $\mathcal{A}_s = 0.5(\mathcal{A} + \mathcal{A}^T)$ is the symmetric part of $\mathcal{A}$. The optimal preconditioner D is a matrix which is spectrally equivalent to $\mathcal{B}\mathcal{A}_s^{-1}\mathcal{B}^T$: $c_0 \mathcal{B}\mathcal{A}_s^{-1}\mathcal{B}^T \leqslant D \leqslant c_1 \mathcal{B}\mathcal{A}_s^{-1}\mathcal{B}^T$, with smallest ratio $\dfrac{c_1}{c_0}$.

Our goal is to construct a preconditioner D such that the constants c_0, c_1 don't depend on meshsizes h and τ and on the parameter α, while D is "easily invertible".

Due to Lemma 4 the matrix $B\mathcal{A}_s^{-1}B^T$ is spectrally equivalent to $\mathcal{B}\mathcal{A}_0^{-1}\mathcal{B}^T = \begin{pmatrix} LL^T + \alpha^{-1}E & LR^T \\ RL^T & RR^T + r^{-1}E \end{pmatrix}$ for any choice $\mathcal{A} = \mathcal{A}_{1r}$ or $\mathcal{A} = \mathcal{A}_{2r}$. In turn, this matrix is spectrally equivalent to a block-diagonal one. More precisely, the following statement takes place:

Lemma 5. Matrix

$$D = \begin{pmatrix} (L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E) & 0 \\ 0 & r^{-1}E \end{pmatrix}$$

is spectrally equivalent to $\mathcal{B}\mathcal{A}_0^{-1}\mathcal{B}^T$ with constants, which depend only on r.

Method (10) for problem (8) with $\mathcal{A} = \mathcal{A}_{1r}$ and with preconditioner D reads as follows:

$$\begin{aligned} &y^{k+1} = y_d - L^T\lambda^k - R^T\mu^k, \\ &\alpha u^{k+1} + \partial\varphi(u^{k+1}) \ni \lambda^k, \\ &rp^{k+1} + \partial\theta(p^{k+1}) \ni rRy^{k+1} + \mu^k, \\ &(L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E)\frac{\lambda^{k+1} - \lambda^k}{\rho} = Ly^{k+1} - u^{k+1} - f, \\ &\frac{\mu^{k+1} - \mu^k}{r\rho} = Ry^{k+1} - p^{k+1}. \end{aligned} \tag{11}$$

Theorem 1. Method (11) converges if $r \in (0,1)$ and $0 < \rho < 2(1 - \sqrt{r})(\sqrt{1 + r} - r)^2$.

Implementation. On every step of method (11) we have to solve two inclusions, for u^{k+1} and for p^{k+1}, and the system of equations with the matrix $(L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E)$. Solving the inclusions reduces to pointwise projections on the corresponding sets of the constraints. On the other hand, solving a system of linear equations with the matrix $(L + \alpha^{-1/2}E)(L^T + \alpha^{-1/2}E)$ consists of sequential solving the systems with the matrices $L + \alpha^{-1/2}E$ and $L^T + \alpha^{-1/2}E$. In the case of explicit finite difference scheme ($\sigma = 0$) these matrices are triangle ones and the solutions can be found by explicit calculations.

Let now

$$\mathcal{A}_{2r} = \mathcal{A}_1 + \mathcal{A}_2, \quad \mathcal{A}_1 = \begin{pmatrix} E + rR^T R & 0 & 0 \\ 0 & \alpha E & 0 \\ -rR & 0 & rE \end{pmatrix}, \qquad \mathcal{A}_2 = \begin{pmatrix} 0 & 0 & -rR^T \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{12}$$

Block relaxation-Uzawa iterative method for solving saddle point problem (9) reads as follows:

$$\begin{aligned} &\mathcal{A}_1 w^{k+1} + \mathcal{A}_2 w^k - \mathcal{B}^T \eta^k + \partial\psi(w^{k+1}) \ni g_1, \\ &\frac{1}{\rho} D(\eta^{k+1} - \eta^k) + \mathcal{B} w^{k+1} = g_2. \end{aligned} \tag{13}$$

Due to [2] this method converges for any initial guess (w^0, η^0) if there exist constants $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ and a continuous and non-negative function ρ, $\rho(0) = 0$, such that

$$\begin{aligned} (\mathcal{A}_1 w, w) + (\mathcal{A}_2 v, w) \geqslant \;& \varepsilon_1 \|w\|^2 + \frac{(1+\varepsilon_2)\rho}{2} (D^{-1}\mathcal{B}w, \mathcal{B}w) \\ & + \rho(w) - \rho(v) \; \forall w, v. \end{aligned} \tag{14}$$

Method (13) for problem (9) with the matrix $\mathcal{A} = \mathcal{A}_{2r}$ splitted into the sum as mentioned in (12) with the same preconditioner D as above takes the form:

$$\begin{aligned} & y^{k+1} + rR^T R y^{k+1} = y_d - L^T \lambda^k - R^T \mu^k + rR^T p^k, \\ & \alpha u^{k+1} + \partial\varphi(u^{k+1}) \ni \lambda^k, \\ & rp^{k+1} + \partial\theta(p^{k+1}) \ni rRy^{k+1} + \mu^k, \\ & (L + \alpha^{-1/2} E)(L^T + \alpha^{-1/2} E) \frac{\lambda^{k+1} - \lambda^k}{\rho} = Ly^{k+1} - u^{k+1} - f, \\ & \frac{\mu^{k+1} - \mu^k}{r\rho} = Ry^{k+1} - p^{k+1}. \end{aligned} \tag{15}$$

Theorem 2. Method (15) converges if $r > 0$, $0 < \rho < 1$.

Implementation. The implementation of method (15) differs from the implementation of method (11) only in the equation for y^{k+1}. Namely, now we have to solve a system of linear equations with the matrix $E + rR^T R$ for finding y^{k+1}. The corresponding calculations reduce to solving for every fixed node of ω_x a system with tridiagonal matrix (with respect to time variable), so, can be implemented by Thomas algorithm.

Acknowledgment

In this work, the first and second authors were supported by grants n:o 278488 and n:o 278029 from Academy of Finland.

References

[1] A. Lapin: Preconditioned Uzawa type methods for finite-dimensional constrained saddle point problems, Lobachevskii J. Math. - V.31, 4 - P.309-322 (2010).

[2] N.S. Kashtanov, A.V. Lapin: Efficiently implementable iterative methods for linear elliptic variational inequalities with constraints on the gradient of solution, Matematica. - N.7 -P.10-24 (2015) (in Russian).

[3] E. Laitinen, A. Lapin and S. Lapin: Easily implementable iterative methods for variational inequalities with nonlinear diffusion-convection operator and constraints to gradient of solution, Russian J. Numer. Analysis Math. Modeling - V.30,1 - P.43 - 54 (2015).

[4] I. Ekeland and R. Temam: Convex analysis and variational problems –Amsterdam: North- Holland. – 1976.

Ph.D Erkki Laitinen is a university lecturer at the Department of Mathematical Sciences of University of Oulu, and an adjunct professor of Computer Science at the University of Jyvāskylā, Finland. His research interests include numerical analysis, optimization and optimal control. He is active in promoting these techniques in practical problem solving in engineering, manufacturing, and industrial process optimization. He has published more than hundred peer reviewed scientific papers in international journals and conferences. He has participated in several applied projects dealing with optimization and control of production processes or wireless telecommunication systems.

D. Sc. Alexander Lapin is a professor of Kazan Federal University (Russia) at the Institute of Computational Mathematics and Information Technology. His research interests include numerical analysis of free boundary problems and optimal control. He participated in the implementation of numerus applications including dam problem, nonlinear filtration problem, Stefan problem etc. He has published more than hundred peer reviewed scientific papers in journals and conferences proceedings.

Genetic networks inferring of the activating-repressive type under the Boolean network model

Evgeny Pshenichnyy, Dmitry Romanov, and Natalia Ponomareva

Abstract— This paper considers the problem of identification of synchronous Boolean gene regulatory networks. Under this model, a gene network is a directed graph with each vertices corresponding to a gene. Each gene is associated with a Boolean variable enoting gene activity state, and with a Boolean function that determines the dependence of the state of a gene in the next moment of time on the state of genes connected with this one in the current moment of time. In this paper we propose an algorithm for inferring this specific type of Boolean networks, and show its efficiency in comparison with generic Boolean network identification algorithms, like REVEAL.

Keywords— Bioinformatics, Boolean networks, inferring of gene network.

I. Introduction

Genes function in an ensemble and form gene networks, the coordinated work of which regulates all the processes in the organism and stipulates his phenotypical characteristics. Boolean networks are the simplest models of networks, which, nevertheless, possess system properties similar to those, which are possessed by the real biological networks [1]. Firstly these networks were adopted by Kauffman S. as a model of genetic regulation of the biological processes, proceeding in a cell. Boolean networks as a model of genetic regulation of biological processes is based on hypothesis, such that every gene through the proceeded product under its control may influence the expression of any other genes. Thus, the expression of every other gene depends on the pattern of other genes expression at any time.

To study the behavior of the biological gene networks using the model of boolean networks the Boolean network is created, which is in line with visible data. This task is also known as a task of reconstruction (Inferring) of gene network. The structure recovery of the graph and the Boolean functions is included in the task of inferring of the Boolean networks, related to each of the nodes, conformed to the experimental data about the genes expression of genes. The data about genes expression is given in the form of pair sets of the successive in times states of a boolean network. Unlike these researches[2-4], in the paper it is suggested that the nodes connection properties in the network regards the biological nature of the modeling objects.

The algorithm of reconstruction of the networks is presented in the paper, considering the fact that genes in the gene network can have an activating or suppress impact on each other. The suggested algorithm appeared to be more effective than the universal one for the Boolean gene networks of the activating-repressive type.

II. BOOLEAN NETWORKS AND INFERENCE PROBLEM

The Boolean network G is an oriented graph (X,E), $|X| = n$, where each top node consists of a Boolean variable. A Boolean function $f_i(x_1, x_2, ..., x_l), l \le n$ is connected to each top node $x □ X$, where $x_i □ X$ are the parameters of those nodes, which have outgoing arcs, ended at x [5].

Values of Boolean variables in the nodes of the graph generate the sequences $x_i(t), t \in Z$, t – discrete time, i – index of the top node in the network. Consider a state of the Boolean network at time t as a set of values of all Boolean variables at the time t $S(t) = (x_1(t), ..., x_n(t))$.

The Boolean network passes from the condition S(t) to S(t+1) synchronously, i.e. all the parameter values in the nodes regenerate consistently, regards to the Boolean functions in the nodes:

$$x_i(t+1) = f_i\left(x_{i_1}(t), x_{i_2}(t), ..., x_{i_l}(t)\right)$$

The table of states T consists of the pair sets of the input and output states $\{I_i / O_i\}_{i=1}^{m}$. Such set of the input-output states is considered to be a table of visualized states, or briefly – a table o states 1.

This work was supported by project of the Ministry of Education and Science of Russia №6.703.2014/K.

Evgeny A. Pshenichnyy is with the Southern Federal University, Research Institute of Biology, Rostov-on-Don, Russia, (corresponding author to provide phone: 007-863-2975070; address: 344090 pr. Stachki 104/1 Rostov-on-Don, Russia; e-mail: pshenichniy.eugene@gmail.com).

Dmitry Romanov is with the Southern Federal University, Research Institute of Biology, Rostov-on-Don, Russia(e-mail: rdme@yandex.ru).

Natalia Ponomareva is with the Southern Federal University, Research Institute of Biology, Rostov-on-Don, Russia(e-mail: nsponomareva@sfedu.ru).

The table of states T can be divided into two parts – inputs and outputs, fig. 1. Each part consists of n columns. The input columns are to be denoted as i^k , while the output columns - o^k .

Table 1. The structure of the table of states.

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
[9] (	[10]:	[11](	[12]:	[13]:	[14](	[15](	[16]:
[17]:	[18](	[19](	[20]:	[21](	[22](	[23](	[24]:
[25]:	[26](	[27]:	[28](	[29]:	[30](	[31]:	[32](
[33]:	[34]:	[35](	[36](	[37]:	[38](	[39]:	[40]:
[41](	[42]:	[43]:	[44](	[45]:	[46]:	[47](	[48](
[49]:	[50](	[51]:	[52](	[53]:	[54](	[55]:	[56]:
[57].	[58].	[59].	[60].	[61].	[62].	[63].	[64].

The task of the Boolean gene network reconstruction is a task of searching of the Boolean network G, coordinated with the given table of states $T = \{I / O\}_{i=1}^{m}$.

III. AR-BOOLEANNETWORKS

Consider construction of the Boolean network for the cooperation of genes in the process of regulation the transcription. The mechanism of regulation the transcription was widely studied in prokaryotic and eukaryotic organisms. In many cases initiation of the gene transcription happens under control of promoter and many other regulator elements. DNA – sites of coupling proteins recognize these regulatory sequences and have an activating or repressive impact on the expression through the interaction with the promoter and RNA polymerase [6,7].

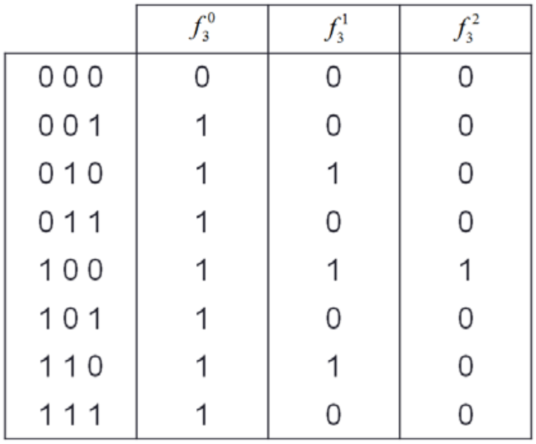
Consider the gene networks with two types of interactions between genes: activating interaction and repressive interaction. Taking into account this constraint, consider the task of reconstructing for the Boolean networks, where in the nodes of which are only activating – repressive Boolean functions of the type $\boldsymbol{f(a_1,...,a_s,r_1,...,r_t) = (a_1 \square ... \square a_s) \square \overline{(r_1 \square ... \square r)}}$. Here, parameters-activators are defined as a, and parameters-repressors are deduced as r. In this case, $s + t = d,\ 0 < s \le d,\ t \ge 0$.

It should be mentioned that the function of the type activator-repressor should have at least one activator and, at the same time it could not have no parameters-repressors. In brief, the function of type activator-repressor is defined as AR-function, and the network with respective functions in the nodes – AR-networks.

In spite of this, activating-repressive Boolean functions satisfy the constraint of the “number of channels” according to Kaufman [1]. This constraint requires, that the definition of the Boolean function can be defined by any of the variables, despite the values of other Boolean functions. Parameters-repressors possess this property. Namely, if at least one repressor is active, it means that the gene would be repressive and the value of the Boolean function would be equal to zero.

The set of parameters-activators of the AR-function f define as A(f), and the set of parameters-repressors - R(f).
In the tbl. 2 is presented the truth table for three Boolean AR functions of three variables with different number of repressors. The first function has no repressors, while the second has one repressor (the first two parameters are active, the third is a repressor), and the third function has only one activator as an argument (the first one) and two repressors (the second and third arguments).

Table 2. The truth table for three AR-functions of three variables



	f_3^0	f_3^1	f_3^2
0 0 0	0	0	0
0 0 1	1	0	0
0 1 0	1	1	0
0 1 1	1	0	0
1 0 0	1	1	1
1 0 1	1	0	0
1 1 0	1	1	0
1 1 1	1	0	0

IV. A-REVERSE ALGORITHM

Provide an algorithm for solving the task of exact reconstruction of the Boolean gene activating-repressor network. Assume that it is given the complete table of states T generated by the unknown activating-repressor gene network of the size n.
A-Reverse:

1. For each input o^j do the step 2.
2. Find the set of activators and repressors:

   1.1 Construct the down-sized table Ψ_j from the rows of the table T, for which the value in o^j column is equal to one. The down-sized table for the inputs is created by crossing out all the rows, the output of which has the zero value.

   1.2 For each column of Ψ_j compute the number of zeroes and ones.

   1.3 Partition the column sets of Ψ_j into three sets:

      1) Indexes of the columns with respect to the inputs, from which does not depend the output. The input column in the down-sized table would have equal number of zeroes and ones, so the frequency of occurrence of 1 in these columns would equal to ½.
      $$U_j = \left\{ i \in \{1, n\},\ fr(i, \Psi_j) = \frac{1}{2} \right\}$$

2) Indexes of the columns of repressors inputs. According to the definition, the inputs of the repressors cannot have 1 in the rows, where the input is equal to 1. In the other words, the columns of the down-sized table with zeroes would hit to the set of parameters-repressors with the sequence of occurrence of 1 in such columns equal to zero. $R_j = \left\{ i \in \{1,n\},\ fr(i,\Psi_j) = 0 \right\}$

3) The columns, which have not hit in the previous sets, are the columns equal to the inputs of the activators. $A_j = \{1,n\} \setminus R_j \cup U_j$

Thus, for each output o^j there would be corresponded two disjoint sets A_j, R_j. Such range of sets uniquely assigns the activating-repressor gene network.

Evaluate the complexity of the algorithm of the gene A-Reverse network reconstruction. The algorithm consists of the cycle, on the outputs of which the sets of activators and repressors are determined. The cycle on outputs consists of n steps.

At the second stage of the A-Reverse algorithm a table Ψ_j is created and the frequency setting $fr(i^1,\Psi_j)...fr(i^n,\Psi_j)$ is computed. Deduce the number of ones in the column o^j as q. For one iteration of the cycle the time proportional to qn would be required. In the worst case, the number of ones in the column o^j is equal to $2^n - 1$, that is why in the worst case it is required to calculate the number of ones for n-columns with the length of $2^n - 1$. Hence, for the algorithm execution the time proportional to $n \cdot qn \le n^2 \cdot (2^n - 1)$ would be required. Write down the computational cost of the algorithm:

$$O(n \cdot qn) \approx O(n^2 \cdot (2^n - 1)) = O(n^2 \cdot 2^n)$$

V. Numerical test and comparison with REVEAL

The number of numerical test has been carried out to evaluate the difference in productivity of the REVEAL [7] and A-Reverse algorithms. Accomplish this, the reconstruction of the gene networks with different number of genes and various levels of the tope nodes has been obtained. In all the tests the results appeared to be equal and the networks have been reconstructed correctly.

The time results for the corresponding algorithms REVEAL and A-Reverse are presented in the table 3.

Table 3. Time in seconds, taken for reconstruction of the networks using the algorithms A-Reverse and REVEAL

[65] N\D	[66] 2		[67] 3		[68] 4	
	[69] A-Reverse	[70] REVEAL	[71] A-Reverse	[72] REVEAL	[73] A-Reverse	[74] REVEAL
[75] 10	[76] 0,02	[77] 0,33	[78] 0,02	[79] 1,26	[80] 0,01	[81] 5,95
[82] 12	[83] 0,14	[84] 2,76	[85] 0,12	[86] 18,21	[87] 0,10	[88] 23,69
[89] 14	[90] 0,48	[91] 19,41	[92] 0,54	[93] 32,22	[94] 0,63	[95] 149,13
[96] 16	[97] 2,57	[98] 107,98	[99] 2,79	[100] 420,33	[101] 2,98	[102] 1352,33
[103] 18	[104] 14,36	[105] 678,38	[106] 14,10	[107] 2989,08	[108] 14,57	[109] 9773,00

Both of the algorithms have exponential complexity. However, it is observed that the A-Reverse algorithm runs significantly faster.

VI. Conclusion

In the paper there have been suggested the method for the Boolean gene networks identification, optimized for reconstruction of the networks with properties appropriate to the biological gene networks. That is why, unlike the generic algorithms, as REVEAL, the A-Reverse one reconstructs the network, satisfying to the above-described biological properties. Limitation of the Boolean function in the nodes was based on the works [1,8,9], in which there have been presented properties of the biological networks – such that, small maximum level of inputs in the network and activating-repressive relationships between the nodes. That is why, unlike the universal algorithms, as REVEAL, the A-Reverse one reconstructs the network, satisfying to the above-described biological properties. It appears that the suggested methodology allows not only obtaining more accurate networks, in the context of biology, but also accelerating the solution of the reconstruction problem.

References

1. Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. Journal of Theoretical Biology, 22:437-467.,1969
2. T. Akutsu, S. Kuhara, O. Maruyama, and S. Miyano. Identification of gene regulatory networks by strategic gene disruptions and gene overexpressions under a Boolean model. Theor. Comput. Sci., 298:235–51, 2003

3. Martin S. et al. Boolean dynamics of genetic regulatory networks inferred from microarray time series data //Bioinformatics. – 2007. – T. 23. – №. 7. – C. 866-874.
4. S. Liang, S. Fuhrman, and R. Somogyi. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. In Pac. Symp. Biocomputing, volume 3, pages 18–29, 1998
5. Aluru S., Handbook of Computational Molecular Biology Handbook Taylor & Francis Group, LLC, 2006
6. Ptashne M. A. Genetic Switch. Oxford: Cell Press, 1986
7. Ratner VA Genetics, molecular cybernetics // Personalities and Problems, Moscow, Nauka, 2002
8. Shmulevich I., Lahdesmaki H.The role of certain Post classes in Boolean network models of genetic networks, PNAS September 16, 2003 vol. 100 no. 19 10734-10739
9. Milo R. et al. Network motifs: simple building blocks of complex networks //Science. – 2002. – T. 298. – №. 5594. – C. 824-827.

Study the existence and uniqueness of the weak solution of antiplane electro-elastic problem with the power-law friction

DALAH Mohamed
Department of Mathematics, Faculty of Sciences,
University Constantine 1, UMC
Constantine, Algeria
e-mail (Corresponding Author) dalah.mohamed@yahoo.fr

DERBAZI Ammar
Department of Mathematics, Faculty of MI,
University of El-Bordj: BBA
Bordj Bou Arreridj, Algeria

Abstract— In this paper the material used is electro-elastic and the friction and it is modeled with Tresca's law and the foundation is assumed to be electrically conductive. First we derive the well posedness mathematical model. In the second step, we give the classical variational formulation of the model which is given by a system coupling an evolutionary variational equality for the displacement field and a time-dependent variational equation for the potential field. Then we prove the existence of a unique weak solution to the model by using the Banach fixed-point Theorem.

Keywords— Tresca's friction, electro-elastic material, variational inequality, weak solution, fixed point, antiplane shears deformation.

Mathematics Subject Classification— 74G25, 49J40, 74F15, 74M10

1. INTRODUCTION

We consider the antiplane contact problem for electro-elastic materials with Tresca friction law. In this new work, we assume that the dispalcement is parallel to the generators of the cylinder and is dependent of the axial coordinate. Our interest is to describe a physical process (for more details see [1, 4, 5, 6, 7, 8]) in which both antiplane shear, contact, state of material with Trescafriction law and piezoelectric effect are involved, leading to a well posedness mathematical problem. In the variational formulation, this kind of problem leads to an integro-differential inequality. The main result we provide concerns the existence of a unique weak solution to the model, see for instance [2, 3, 6] for details.

The rest of the paper is structured as follows. In Section s:2 we describe the well posedness mathematical model of the frictional contact process between electro-elastic body and a conductive deformable foundation. In Section s:3 we derive the variational formulation. It consists of a variational inequality for the displacement field coupled with a time-dependent variational equation for the electric potential. We state our main result, the existence of a unique weak solution to the model in Theorem 3.1. The Proof of the Theorem is provided in the end of Section s: 4, where it is based on arguments of evolutionary inequalities, and a fixed point Theorem.

2. THE MODEL

In this section, we consider a piezoelectric body $\mathbf{B}$ identified with a region in IR^3 it occupies in a fixed and undistorted reference configuration. We assume that $\mathbf{B}$ is a cylinder with generators parallel to the x_3 -axes with a cross-section which is a regular region Ω in the x_1 , x_2 -plane, $Ox_1x_2x_3$ being a Cartesian coordinate system. The cylinder is assumed to be sufficiently long so that the end effects in the axial direction are negligible. Thus, $\mathbf{B} = \Omega \times (-\infty, +\infty)$. The cylinder is acted upon by body forces of density $\mathbf{f}_0$ and has volume free electric charges of density q_0 . It is also constrained mechanically and electrically on the boundary. To describe the boundary conditions, we denote by $\partial\Omega = \Gamma$ the boundary of Ω and we assume a partition of Γ into three open disjoint parts Γ_1 , Γ_2 and Γ_3 , on the one hand, and a partition of $\Gamma_1 \cup \Gamma_2$ into two open parts Γ_a and Γ_b , on the other hand. We assume that the one-dimensional measure of Γ_1 and Γ_a , denoted meas Γ_1 and meas Γ_a , are positive.

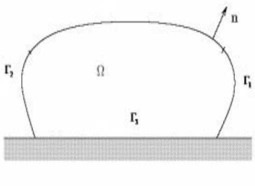


FIGURE 1. Deformable solid Ω on contact with a rigid foundation

The cylinder is clamped on $\Gamma_1 \times (-\infty, +\infty)$ and therefore the displacement field vanishes there. Surface tractions of density $\mathbf{f}_2$ act on $\Gamma_2 \times (-\infty,+\infty)$. We also assume that the electrical potential vanishes on $\Gamma_a \times (-\infty,+\infty)$ and a surface electrical charge of density q_2 is prescribed on $\Gamma_b \times (-\infty,+\infty)$. The cylinder is in contact over $\Gamma_3 \times (-\infty,+\infty)$ with a conductive obstacle, the so called foundation. The contact is frictional and is modeled with Tresca's law. We are interested in the deformation of the cylinder on the time interval $[0,T]$. We assume that

$$\mathbf{f}_0 = (0,0,\ f_0)\ with\ f_0 = f_0(x_1,\ x_2,t):\ \Omega \times [0,T] \to \mathrm{IR}, \quad (1)$$

$$\mathbf{f}_2 = (0,0,\ f_2)\ with\ f_2 = f_2(x_1,\ x_2,t):\ \Gamma \times [0,T]\ \to \mathrm{IR}, \quad (2)$$

$$q_0 = q_0(x_1,\ x_2,t)\ :\ \Omega \times [0,T] \to \mathrm{IR}, \quad (3)$$

$$q_2 = q_2(x_1,\ x_2,t):\ \Gamma_b \times [0,T] \to \mathrm{IR}. \quad (4)$$

The forces (1), (2) and the electric charges (3), (4) would be expected to give rise to deformations and to electric charges of the piezoelectric cylinder corresponding to a displacement $\mathbf{u}$ and to an electric potential field φ which are independent on x_3 and have the the form

$$\mathbf{u} = (0,0,\ u) \qquad with \qquad u = u(x_1,\ x_2,t):\ \Omega \times [0,T] \to \mathrm{IR}, \quad (5)$$

$$\varphi = \varphi(x_1,\ x_2,t)\ :\ \Omega \times [0,T] \to \mathrm{IR}. \quad (6)$$

Such kind of deformation, associated to a displacement field of the form (3), is called an antiplane shear.

The infinitesimal strain tensor is denoted $\boldsymbol{\varepsilon}(\mathbf{u}) = (\varepsilon_{ij}(\mathbf{u}))$ and the stress field by $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_{ij})$. We also denote by $\mathbf{E}(\varphi) = (E_i(\varphi))$ the electric field and by $\mathbf{D} = (D_i)$ the electric displacement field. Here and below, in order to simplify the notation, we do not indicate the dependence of various functions on x_1, x_2, x_3 or t and we recall that

$$\varepsilon_{ij}(\mathbf{u}) = \frac{1}{2}(u_{i,j} + u_{j,i}),\ \ E_i(\varphi) = -\varphi_{,i}.$$

The material's is modeled by the following electro-elastic constitutive law with Tresca friction law

$$\sigma = \lambda(\mathbf{tr}\ \in(\mathbf{u}))\mathbf{I} + 2\mu \in(\mathbf{u}) - \mathbf{E}^{*}\mathbf{E}(\varphi), \quad (7)$$

$$\mathbf{D} = \mathbf{E}\boldsymbol{\varepsilon}(\mathbf{u}) + \beta\mathbf{E}(\varphi), \quad (8)$$

where λ and μ are the Lame coefficients $\varepsilon(\mathbf{u}) = (\in ij(\mathbf{u}))$, $\mathbf{I}$ is the unit tensor in IR^3 , β is the electric permittivity constant, $\mathbf{E}$ represents the third-order piezoelectric tensor and $\mathbf{E}^*$ is its transpose. In the antiplane context (5), (6), using the constitutive equations (7), (8) it follows that the stress field and the electric displacement field are given by

$$\boldsymbol{\sigma} = \begin{pmatrix} 0 & 0 & \boldsymbol{\sigma}_{13} \\ 0 & 0 & \boldsymbol{\sigma}_{23} \\ \boldsymbol{\sigma}_{31} & \boldsymbol{\sigma}_{32} & 0 \end{pmatrix}, \quad (9)$$

$$\mathbf{D} = \begin{pmatrix} eu_{,1} - \beta\varphi_{,1} \\ eu_{,2} - \beta\varphi_{,2} \\ 0 \end{pmatrix} \quad (10)$$

where

$$\boldsymbol{\sigma}_{13} = \boldsymbol{\sigma}_{31} = \mu\partial_{x_1} u$$

and

$$\boldsymbol{\sigma}_{23} = \boldsymbol{\sigma}_{32} = \mu\partial_{x_2} u.$$

We assume that

$$\mathbf{E}\boldsymbol{\varepsilon} = \begin{pmatrix} e(\varepsilon_{13} + \varepsilon_{31}) \\ e(\varepsilon_{23} + \varepsilon_{32}) \\ e\varepsilon_{33} \end{pmatrix}\ \forall \boldsymbol{\varepsilon} = (\varepsilon_{ij}) \in \mathbf{S}^3, \quad (11)$$

where e is a piezoelectric coefficient. We also assume that the coefficients μ, β and e depend on the spatial variables x_1, x_2 , but are independent on the spatial variable x_3 . Since $\mathbf{E}\boldsymbol{\varepsilon} \cdot \mathbf{v} = \boldsymbol{\varepsilon} \cdot \mathbf{E}^*\mathbf{v}$ for all $\boldsymbol{\varepsilon} \in \mathbf{S}^3$, $\mathbf{v} \in IR^3$, it follows from (e) that

$$\mathbf{E}^*\mathbf{v} = \begin{pmatrix} 0 & 0 & ev_1 \\ 0 & 0 & ev_2 \\ ev_1 & ev_2 & ev_3 \end{pmatrix}\ \forall \mathbf{v} = (v_i) \in \mathrm{IR}^3. \quad (12)$$

We assume that the process is mechanically quasistatic and electrically static and therefore is governed by the equilibrium equations

$$\mathbf{Div}\sigma + \mathbf{f}_0 = 0,\ D_{i,i} - q_0 = 0 \quad \text{in} \quad B \times [0,T] \ ,$$

where $\mathbf{Div}\sigma = (\sigma_{ij,\,j})$ represents the divergence of the tensor field $\boldsymbol{\sigma}$. Taking into account (1), (3), (5), (6), (9) and (10), the equilibrium equations above reduce to the following scalar equations

$$\mathbf{div}(\mu\nabla u + e\nabla\varphi) + f_0 = 0,\ in\ \Omega \times [0,T], \ (13)$$

$$\mathbf{div}(e\nabla u - \beta\nabla\varphi) = q_0,\ in\ \Omega \times [0,T]. \ (14)$$

Here and below we use the notation

$$\mathbf{div}\tau = \tau_{1,1} + \tau_{1,2} \quad \text{in} \quad \tau = (\tau_1(x_1,\ x_2),\ \tau_2(x_1,\ x_2))$$

and

$$\nabla v = (v_1,\ v_2) \quad , \quad \partial_{\iota /} v = v_1\nu_1 + v_2\nu_2 \quad \text{for} \quad v = v(x_1,\ x_2)$$

We now describe the boundary conditions. During the process the cylinder is clamped on $\Gamma_1 \times (-\infty,+\infty)$ and the electric potential vanish on $\Gamma_1 \times (-\infty,+\infty)$; thus, (5) and (6) imply that

$$u = 0 \ \text{ on } \Gamma_1 \times [0,T], \quad (15)$$

$$\varphi = 0 \text{ on } \Gamma_a \times [0,T]. \quad (16)$$

Let $\boldsymbol{\nu}$ denote the unit normal on $\Gamma \times (-\infty,+\infty)$. We have

$$\nu = (\nu_1,\ \nu_2, 0) \ with\ \nu_i = \nu_i(x_1,\ x_2):\ \Gamma \to \mathrm{IR},\ i = 1,2. \quad (17)$$

For a vector $\mathbf{v}$ we denote by v_ν and $\mathbf{v}_\tau$ its normal and tangential components on the boundary, given by

$$v_\nu = \mathbf{v} \cdot \boldsymbol{\nu},\ \ \mathbf{v}_\tau = \mathbf{v} - v_\nu \boldsymbol{\nu}. \quad (18)$$

For a given stress field $\boldsymbol{\sigma}$ we denote by σ_ν and $\boldsymbol{\sigma}_\tau$ the normal and the tangential components on the boundary, that is

$$\sigma_\nu = (\boldsymbol{\sigma\nu}) \cdot \boldsymbol{\nu},\ \ \boldsymbol{\sigma}_\tau = \boldsymbol{\sigma\nu} - \sigma_\nu \boldsymbol{\nu}. \quad (19)$$

From (9), (10) and (17) we deduce that the Cauchy stress vector and the normal component of the electric displacement field are given by

$$\boldsymbol{\sigma\nu} = (0,0,\mu\partial_\nu u + e\partial_\nu \varphi),\ \ \mathbf{D}\cdot\boldsymbol{\nu} = e\partial_\nu u - \beta\partial_\nu \varphi. \quad (20)$$

Taking into account (2), (4) and (20), the traction condition on $\Gamma_2 \times (-\infty,\infty)$ and the electric conditions conditions on $\Gamma_b \times (-\infty,\infty)$ are given by

$$\mu\partial_\nu u + e\partial_\nu \varphi = f_2 \text{ on } \Gamma_2 \times 0,T],[(21)$$
$$e\partial_\nu u - \beta\partial_\nu \varphi = q_2 \text{ on } \Gamma_b \times [0,T).(22)$$

We now describe the frictional contact condition and the electric conditions on $\Gamma_3 \times (-\infty,+\infty)$. First, from (5) and (17) we infer that the normal displacement vanishes, $u_\nu = 0$, which shows that the contact is bilateral, that is, the contact is kept during all the process. Using now (5) and (17)-(19) we conclude that

$$\mathbf{u}_\tau = (0,0,u),\ \ \boldsymbol{\sigma}_\tau = (0,0,\sigma_\tau) \quad (23)$$

where

$$\sigma_\tau = (0,0,\mu\partial_\nu u + e\partial_\nu \varphi).$$

We assume that the friction is invariant with respect to the x_3 axis and is modeled with Tresca's friction law, that is

$$\boldsymbol{\sigma}_\tau(t) = \begin{cases} 0, \text{if u} = 0, \\ -g|u|^{s-1}, & if\ \ \mathrm{u} \neq 0 \ \text{ on } \Gamma_3 \times (0,T).(24) \end{cases}$$

Here $g : \Gamma_3 \to \mathrm{IR}_+$ is a given function, the friction bound, and $\mathbf{u}_\tau$ represents the tangential velocity on the contact boundary. Using now (23) it is straightforward to see that the friction law (24) implies

$$\mu\partial_\nu u + e\partial_\nu \varphi = \begin{cases} 0, \text{if u} = 0, \\ -g|u|^{s-1}, & if\ \ \mathrm{u} \neq 0 \ \text{ on } \Gamma_3 \times (0,T).(25) \end{cases}$$

Next, since the foundation is electrically conductive and the contact is bilateral, we assume that the normal component of the electric displacement field or the free charge is proportional to the difference between the potential on the foundation and the body's surface. Thus,

$$\mathbf{D}\cdot\boldsymbol{\nu} = q_2 \text{ on } \Gamma_b \times (0,T),$$

Then, we get

$$\begin{pmatrix} eu,_1 - \beta\varphi,_1 \\ eu,_2 - \beta\varphi,_2 \\ 0 \end{pmatrix} .\boldsymbol{\nu} = q_2 \text{ on } \Gamma_b \times (0,T). \ (26)$$

Finally, we use (20) and the previous equality to obtain

$$e\partial_\nu u - \beta\partial_\nu \varphi = q_2 \text{ on } \Gamma_b \times (0,T). \quad (27)$$

We collect the above equations and conditions to obtain the following mathematical model which describes the antiplane shear of an electro-viscoelastic cylinder in frictional contact with a conductive foundation.

Problem P. Find the displacement field $u : \Omega \to \mathrm{IR}$ and the electric potential $\varphi : \Omega \to \mathrm{IR}$ such that

$$\mathbf{div}(\mu\nabla u) + \mathbf{div}(e\nabla\varphi) + f_0 = 0,\ in\,\Omega,\ (28)$$

$$\mathbf{div}(e\nabla u) - \mathbf{div}(\alpha\nabla\varphi) = q_0\ \mathbf{in}\,\Omega,\ \ (29)$$

$$u = 0 \text{ on } \Gamma_1, (30)$$

$$\mu\partial_\nu u + e\partial_\nu \varphi = f_2 \text{ on } \Gamma_2, (31)$$

$$\mu\partial_\nu u + e\partial_\nu \varphi = \begin{cases} 0, \text{if u} = 0, \\ -g|u|^{s-1}, & if\ \ \mathrm{u} \neq 0 \ \text{ on } \Gamma_3,(25) \end{cases}$$

$$\varphi = 0 \text{ on } \Gamma_a,\ (33)$$

$$e\partial_\nu u - \alpha\partial_\nu \varphi = q_2 \text{ on } \Gamma_b.\ (34)$$

Note that once the displacement field u and the electric potential φ which solve **Problem P** are known, then the stress tensor $\boldsymbol{\sigma}$ and the electric displacement field $\mathbf{D}$ can be obtained by using the constitutive laws (9) and (10), respectively.

3. Variational Formulation

For a real Banach space $(X, \|.\|_X)$ we use the usual notation for the spaces $L^p(0,T;X)$ and $W^{k,p}(0,T;X\ )$ where $1 \le p \le \infty,\ k = 1,2,...$; we also denote by $C([0,T];X)$ the space of continuous and continuously differentiable functions on $[0,T]$ with values in X, with the norm

$$\|x\|_{C([0,T];X)} = \max_{t\in[0,T]} \|x(t)\|_X$$

and we use the standard notations for the Lebesgue space $L^2(0,T;X)$ as well as the Sobolev space $W^{1,2}(0,T;X)$. In particular, recall that the norm on the space $L^2(0,T;X)$ is given by the formula

$$\|u\|^2_{L^2(0,T;X)} = \int_0^T \|u(t)\|_X^2 \, dt$$

and the norm on the space $W^2(0,T;X)$ is given by the formula

$$\|u\|^2_{W^{1,2}(0,T;X)} = \int_0^T \|u(t)\|_X^2 \, dt + \int_0^T \|\dot{u}(t)\|_X^2 \, dt. \quad (38)$$

Finally, we suppose the argument X when $X = IR$; thus, for example, we use the notation $W^2(0,T)$ for the space $W^2(0,T;IR)$ and the notation $\|\cdot\|_{W^2(0,T)}$ for the norm $\|\cdot\|_{W^2(0,T;\mathrm{R})}$.

In the study of the **Problem P** we assume that the viscosity coefficient satisfy:

and the electric permittivity coefficient satisfy

$$\beta \in L^\infty(\Omega) \text{ and there exists } \beta^* > 0 \text{ such that } \beta(\mathbf{x}) \geq \beta^* \text{ a.e. } \mathbf{x} \in \Omega. \quad (39)$$

We also assume that the Lame coefficient and the piezoelectric coefficient satisfy

$$\mu \in L^\infty(\Omega) \;\; (40)$$

$$\text{and}$$

$$\mu(\mathbf{x}) > 0 \;\; \text{a.e. } \mathbf{x} \in \Omega, (41)$$

$$e \in L^\infty(\Omega).(42)$$

The forces, tractions, volume and surface free charge densities have the regularity

$$f_0 \in L^2(\Omega),(43)$$

$$f_2 \in L^2(\Gamma_2),(44)$$

$$q_0 \in L^2(\Omega),(45)$$

$$q_2 \in L^2(\Gamma_b).(46)$$

The friction bound function g satisfies the following properties

$$g \in L^\infty(\Gamma_3) \;\text{ and }\; g(\mathbf{x}) \geq 0 \;\text{ a.e. } \mathbf{x} \in \Gamma_3.(47)$$

and, moreover,

$$a_\mu(u_0,v)_V + j(v) \geq (f(0),v)_V \;\; \forall v \in V.(48)$$

We define now the functional $j : V \to \mathrm{IR}_+$ given by the formula

$$j(v) = \frac{1}{s+1}\int_{\Gamma_3} g\,|\,v\,|^{s+1} da \;\; \forall v \in V.(49)$$

We also define the mappings $f \in V$ and $q \in W$, respectively, by

$$(f,v)_V = \int_\Omega f_0 v\,dx + \int_{\Gamma_2} f_2 v\,da,(50)$$

and

$$(q,\psi)_W = \int_\Omega q_0 \psi\,dx - \int_{\Gamma_b} q_2 \psi\,da,(51)$$

for all $v \in V$, $\psi \in W$ and $t \in [0,T]$. The definition of f and q are based on Riesz's representation theorem; moreover, it follows from assumptions by (42)-(43), that the integrals above are well-defined and

$$f \in L^2(\Omega),(52)$$

$$q \in L^2(\Omega).(53)$$

Next, we define the bilinear forms a_μ : $V \times V \to \mathrm{IR}$, a_e: $V \times W \to \mathrm{IR}$, , and a_α : $W \times W \to \mathrm{IR}$, by equalities

$$a_\mu(u,v) = \int_\Omega \mu \nabla u \cdot \nabla v\,dx,(54)$$

$$a_e(u,\varphi) = \int_\Omega e \nabla u \cdot \nabla \varphi\,dx,(55)$$

$$a_\alpha(\varphi,\psi) = \int_\Omega \beta \nabla \varphi \cdot \nabla \psi\,dx,(56)$$

for all $u, v \in V$, $\varphi,\psi \in W$. Assumptions (49)-(51) imply that the integrals above are well defined and, using (37) and (18), it follows that the forms a_μ and a_e are continuous; moreover, the forms a_μ and a_α are symmetric and, in addition, the form a_α is W -elliptic, since

$$a_\alpha(\psi,\psi) \geq \alpha^* \|\psi\|_W^2 \;\; \forall \psi \in W. \quad (57)$$

4. Main Results

The variational formulation of **Problem P** is based on the

Lemma 1 For all (u,φ) in space $X = V \times W$, then we get

$$\int_\Omega \mu \nabla u.\nabla(v-u)dx + \int_\Omega e \nabla \varphi.\nabla(v-u)dx + \frac{1}{s+1}\int_{\Gamma_3}(|v|^{s+1} - |u|^{s+1})da \geq$$

$$\int_\Omega f_0(v-u)dx + \int_{\Gamma_2} f_2(v-u)da, \forall(v,\psi) \in X = V \times W, \forall(u,\varphi) \in X = V \times W.(58)$$

Proof. We introduce relation (50) in the previous relation, then, we have

$$a_\mu(u,v-u) + a_e(\varphi,v-u) + j(v) - j(u) \geq (f,v-u)_V, \forall v \in V, \forall \varphi \in W,(71)$$

which conclude the proof of lemma 1.

Lemma 2. For all element $\psi \in W$ and for all $(u,\varphi) \in V \times W$, then, we have

$$a_e(\varphi,\psi) - a_\beta(u,\psi) = (q,\psi)_W, \forall \psi \in W.(72)$$

Proof. It is immediately by using (29), (33) and (34).

We collect the above equations and conditions to obtain the following variational formulation which describes the antiplane shear of an electro-viscoelastic cylinder in frictional contact with a conductive foundation.

Problem PV 1. Find a displacement field $u : \Omega \to V$ and an electric potential field $\varphi : \Omega \to W$ such that

$$a_{\mu}(u,v-u)+a_{e}(\varphi,v-u)+j(v)-j(u)\ge(f,v-u)_{V},\forall v\in V,\forall\varphi\in W,(77)$$

$$a_{e}(\varphi,\psi)-a_{\beta}(u,\psi)=(q,\psi)_{W},\forall\,\psi\in W.(78)$$

Let now using the bilinear form:

$$\begin{aligned}&a(.,.):X\times X\to R\\&(x,y)\mapsto a(x,y)=a_{\mu}(u,v)+a_{e}(\varphi,v)+a_{\beta}(\varphi,\psi)-a_{e}(u,\psi),\forall x=(u,\varphi)\in X,\forall y=(v,\psi)\in X,(79)\end{aligned}$$

the functional

$$\begin{aligned}&J(.):X\to R\\&x\mapsto J(x)=j(u),\forall x=(u,\varphi)\in X,(80)\end{aligned}$$

and the function

$$F=(f,q)\in X.(81)$$

Now, using notations (79)-(81), the Problem (77)-(78) take the final form:

Problem PV 2. Find a couple $x=(u,\varphi)\in X$ such that

$$a\ \ (x,y-x)+J(x)-J(y)\ge(F,y-x)_{X},\forall y\in X.(82)$$

Theorem 3. The **Problem PV 1** and **Problem PV 2** are equivalent.

Proof. We have two step to proof our Theorem.

Step 1: Problem PV 1⇒ Problem PV 2

In the first step we ill suppose that $x=(u,\varphi)\in X$ is solution of **Problem PV 1.** We change in (78) the element

$\psi\in W$ by $(\psi-\varphi)\in W$ and we add the resulting equation to the two sides of the inequality (77), hence, we obtain:

$$\begin{aligned}&a_{\mu}(u,v-u)+a_{e}(\varphi,v-u)+a_{\beta}(\varphi,\psi-\varphi)-a_{\beta}(u,\psi-\varphi)+\\&+j(v)-j(u)\ge(f,v-u)_{V}+(q,\psi-\varphi)_{W},\forall v\in V,\forall\varphi\in W.(83)\end{aligned}$$

Using now notations (79), (80) and (81) then for all $\psi\in W$ and for all $y\in X$, we get

$$a\ \ (x,y-x)+J(x)-J(y)\ge(F,y-x)_{X},\forall y\in X.(84)$$

which conclude the proof of the first step.

Step 2: Problem PV 2⇒ Problem PV 1

In ths second step we will suppose that $x=(u,\varphi)\in X$ is solution of **Problem PV 2.** We change the bilinear form $a(.,.)$ by (79), $(F,y-x)_{X}$ by (81) and the functional $J(.)$ by (80); then, for all $(v,\psi)\in X$, we obtain

$$\begin{aligned}&a_{\mu}(u,v-u)+a_{e}(\varphi,v-u)+a_{\beta}(\varphi,\psi-\varphi)+\\&+j(v)-j(u)\ge(f,v-u)_{V}+(q,\psi-\varphi)_{W},\forall v\in V,\forall\varphi\in W.(85)\end{aligned}$$

We test in the last inequality (85) with $\psi=\varphi$, then we obtain (77). Next, we take $v=u$ and $\psi-\varphi=\varphi\pm\psi-\varphi$ in (84), it follows that for all $\psi\in W$:

$$a_{\beta}(\varphi,\pm\psi)-a_{e}(\varphi,\pm\psi)\ge(q,\pm\psi)_{W},\forall v\in V,\forall\varphi\in W,(86)$$

which conclude the proof of the second. Then, the **Problem PV 1** and **Problem PV 2** are equivalent.

Our main existence and uniqueness result, which we state now and prove in the next section, is the following:

Theorem 4. Assume that (39)-(57) hold. Then the variational **Problem PV 2** possesses a unique solution $x=(u,\varphi)\in X$ satisfies

$$a\ \ (x,y-x)+J(x)-J(y)\ge(F,y-x)_{X},\forall y\in X.(87)$$

We note that an element $x=(u,\varphi)$ which solves **Problem PV 1** is scalled a weak solution of the antiplane contact **Problem PV 1.** We conclude by Theorem 3 that the element $x=(u,\varphi)$ also solves **Problem PV 2**, then the element x is called a weak solution of the antiplane contact **Problem PV 2.** Hence, the antiplane contact Problem P has a unique weak solution, provided that (39)-(57).

Proof of Theorem 4.

The Proof of Theorem 4 which will be carried out in several steps and it is immediately to obtain our result of existence and uniqueness of the weak solution.

Acknowledgment

The authors would like to thank our team for his/her thorough verify and highly appreciate the comments, remarks and suggestions, which significantly contributed to improving the quality of the publication. Finally, to Dr. K. Fernane, for making so many things possible.

References

[1] W. Han and M. Sofonea, Quasistatic Contact Problems in Viscoelasticity and Viscoplasticity **30**, Studies in Advanced Mathematics, Americal Mathematical Society, Providence, RI-International Press, Somerville, MA, 2002.

[2] C. O. Horgan, Antiplane shear deformation in linear and nonlinear solid mechanics, *SIAM Rev.* **37** (1995), 53-81.

[3] M. Dalah, Analyse of a Electro-Viscoelastic Antiplane Contact Problem With Slip-Dependent Friction, *Electronic Journal of Differential Equations,* Vol. 2009(2009), No. 118, pp. 1-15.

[4] M. Sofonea, M. Dalah, Antiplane Frictional Contact of Electro-Viscoelastic Cylinders, *Electronic Journal of Differential Equations.* **161** (2007), 1-14.

[5] M. Sofonea, M. Dalah and A. Ayadi, Analysis of an antiplane electro-elastic contact problem, *Adv. Math. Sci. Appl.* **17** (2007), 385-400.

Grid Computing for Multi-Objective Optimization Problems

Aouaouche El-Maouhab
Networks Division
Research Center on Scientific and
Technic Information, CERIST
Algiers, Algeria
Email: elmaouhab@wissal.dz

Hassina Beggar
Networks Division
Research Center on Scientific and
Technic Information, CERIST
Algiers, Algeria
Email: h.beggar@grid.arn.dz

***Abstract*—Solving multiobjective discrete optimization applications has always been limited by the resources of one machine: by computing power or by memory, most often both. To speed up the calculations, the grid computing represents a primary solution for the treatment of these applications through the parallelization of these resolution methods. In this work, we are interested in the study of some methods for solving multiple objective integer linear programming problem based on Branch-and-Bound and the study of grid computing technology. This study allowed us to propose an implementation of the method of Abbas and Al on the grid by reducing the execution time. To enhance our contribution, the main results are presented.**

Keywords–Multi-objective optimization, integer linear programming, grid computing.

I. Introduction

In many practical situations, it is necessary to use discrete variables in the modelisation of the problem, for instance, to represent an investment choice, a production level, a problem of manufacture planning or the number of units produced must be integer. This is called discret or integer linear programming. The purpose of this introductory section is to present some basics about the problems of linear integer programming.

II. Multi-objective Integer Linear Programming

A. Formulation of the problem

A linear program (LP) is a problem in which the variables are real and must satisfy a set of linear equations and / or inequalities (called constraints) and the value of a linear function of these variables (called objective function) must be made maximum (or minimum).
Without loss of generality, we assume thereafter that we consider maximization problems.
If in the linear program, the variables are constrained to be integer then we obtain an integer linear program (ILP), which is defined as:

$$\begin{cases} Max\ z = cx & \\ Ax = b & x \geq 0 \\ x \in Z^n & \end{cases}$$

where $x = (x_j)_{j=1,...,n}$ is a vector of Z^n, z is the objective fonction of the problem, $A = (a_{ij})$ is a $m \times n$-matrix of $Z^{m\times n}$, b is an m-vector of Z^m and c is the n-vector cost of R^n.
The two main families of currently known methods for solving these programs are tree methods and methods of truncation (cuts) such as the cutting method propsed in 1958 by R.Gomory [1].

A multi-objective optimization problem (MOP) consists in optimizing simultaniously k objective fonctions ($k \geq 2$) often conflicting (two objectives are conflicting when the decrease of one results in an increase of the other). Therefore, the notion of optimality is meaningless, we try to propose compromise solutions.
Associating an integer linear program (ILP) with a MOP gives a **M**ultiple **O**bjective **I**nteger **L**inear **P**rogram (MOILP), which can be formulated as follows:

$$\begin{cases} Max & z_1 = c^1 x \\ & \vdots \\ Max & z_k = c^k x \\ & x \in S = \{x \in Z^n | Ax = b,\, x \geq 0\} \end{cases}$$

where $A \in Z^{m\times n}$, $b \in Z^m$ and $c^i \in R^{1\times n}$, $i = \overline{1,k}$.

B. Concepts and Definitions

Let (P) a multi-objective integer linear program. S the set of feasible solutions of (P) in the decision space (the space Z^n where S is).

1) Feasible solutions in the criteria space: The set Z given by:

$$Z = \left\{ z \in R^k | z = Cx,\, x \in S \right\}$$

represents all feasible points in the criteria space (the space R^k where Z is situated). In other words, Z is the set of images of all points of S.
A partial order relation is imposed (a solution may be better than another on specific objectives and worse on others) on this set of points, called *d*ominance relation.

2) Efficient solution: A solution x is known as efficient, if there is not another solution y such that $Z^q(y) \geq Z^q(x)$ for all $q \in 1, ..., r$ and $Z^q(y) > Z^q(x)$ for at least one $q \in 1, ..., r$. Otherwise, x is not efficient and the vector $Z(y)$ dominates the vector $Z(x)$, where $Z(x) = (Z^q(x))_{q=(1\bar{,}r)}$.

III. ABBAS AND AL METHOD [2]

The method generates the set of all integer non dominated solutions for MOILP. It is a method based on the concept of branching in integer linear programming and efficient cuts in the criteria space which means cuts using criteria.
Let (P) be the multi-objective Integer Linear Program (MOILP) :

$$(P)\begin{cases} Max & Cx \\ & x \in X \\ & x\, integer\, vector \end{cases}$$

The set of feasible criteria vectors Y of MOILP is defined as follows:

$$Y = \left\{ z \in R^k | z^i = c^i x, x \in X \right\} = f(X) \qquad (1)$$

Let be $x_l \in X$ an integer solution of the problem (P), we define the following sets on x_l relatively to each criterion i; $i = 1, \ldots, k$,:

$$H_l^i = \left\{ j \in N_l | \hat{c}_j^i > 0 \right\} \qquad (2)$$

where N_j is the indices set of non-basic variables and $\hat{c}_j^i$ the component j of the reduced cost vector of the criterion i.

$$K_l = \left\{ i \in \{1, \ldots, k\} \, | H_l^i \neq \emptyset \right\} \qquad (3)$$

indicates the set of criteria that can be improved from x_l.
A constraint is an efficient cut for the problem (P) in the criteria space, if its addition does not eliminate non-dominated integer solutions from the set Y. From the simplex table at the point x_l; we have the following relationship:

$$c^i x = c^i x^l + \sum_{j \in N_l} \hat{c}_j^i x_j, \forall i \in \{1, \ldots, k\} \qquad (4)$$

and under the assumption that the matrix C is with integer coefficients, we deduce the following constraint for each critrion $i \in K_l$:

$$c^i x \geq c^i x^l + \sum_{j \in N_l \setminus H_l^i} \left\lfloor \hat{c}_j^i \right\rfloor x_j + \max\left\{1, \left\lfloor \hat{c}_{j0}^i \right\rfloor \right\} \qquad (5)$$

where $\hat{c}_{j0}^i = \min_{j \in H_i} \left\{ \hat{c}_j^i \right\}$ for $i \in K_l$. This constraint defines an efficient cut in the criteria space for (P).
The algorithm describing the set of all integer non dominated solutions of problem (P) is presented in the following steps:

- Step 0 (Initialization). Denote by NDS The set of non dominated solutions of (P). Set NDS to the empty set. Solve the linear program (P_0). Set x the optimal solution found.
- Step r $(r \geq 1)$. Solve the linear program (P_l), $0 \leq l < r$. Let x^l be an optimal solution of (P_l).
  If x^l is non integer, choose a coordinate j of x^l whose value is not integer and separate the node l on this coordinate into two new nodes and return to step r.
  If x^l is integer then:
  - ○ If the corresponding criterion vector Cx^l is not dominated by any criterion vector Z from NDS then update (NDS) by adding Cx^l.
  - ○ If there exists a solution $Z \in NDS$ that is dominated by the criterion vector Cx^l then replace Z by Cx^l in NDS.

  Determine the set K_l of criteria that can be improved from x_l, for any $i \in K_l$. add the constraint (5) for K_l new linear programs $(P_k), k > l$ and go back to step r.
- stopping condition. The procedure stops when $H_l = \emptyset$ which means no criterion can be improved or (P_l) admits no feasible solutions for any stage l such that $0 \leq l \leq r$.

Numerical Example: Consider the following MOILP problem:

$$(P)\begin{cases} Max\ z_1 = x_1 + 3x_2 \\ Max\ z_2 = \qquad -x_2 \\ \qquad\qquad (x_1, x_2) \in S \end{cases}$$

Where:

$$S = \left\{ (x_1, x_2) \in Z^2 | 2x_1 + 3x_2 \leq 5, 2x_1 + x_2 \leq 4, x_1 \geq 0, x_2 \geq 0 \right\}$$

The set of feasible solutions of the relaxed problem is:

$$X = \left\{ (x_1, x_2) \in R^2 | 2x_1 + 3x_2 \leq 5, 2x_1 + x_2 \leq 4, x_1 \geq 0, x_2 \geq 0 \right\}$$

The linear program (P_0) is:

$$(P_0)\begin{cases} Max\ z(x) = z_1(x) + z_2(x) = x_1 + 2x_2 \\ \qquad (x_1, x_2) \in X \end{cases}$$

The resolution of (P_0) gives:

	b	x_1	x_2	x_3	x_4
x_2	$\frac{5}{3}$	$\frac{2}{3}$	1	$\frac{1}{3}$	0
x_4	$-\frac{7}{3}$	$\frac{4}{3}$	0	$-\frac{1}{3}$	0
$-z$	$-\frac{10}{3}$	$-\frac{1}{3}$	0	$-\frac{2}{3}$	0
z_1	-5	-1	0	-1	0
z_2	$\frac{5}{3}$	$\frac{2}{3}$	0	$\frac{1}{3}$	0

The optimal solution is not integer $(0, \frac{5}{3})$. Then we separate on the variable x_2 and we obtain two sub-problems:

$$(P_1)\begin{cases} (P_0) \\ x_2 \leq 1 \end{cases} \qquad (P_2)\begin{cases} (P_0) \\ x_2 \geq 2 \end{cases}$$

The problem (P_2) being not feasible, the corresponding node is fathomed.
The resolution of (P_1) gives:

	b	x_1	x_2	x_3	x_4	x_5
x_2	1	0	1	0	0	1
x_4	1	0	0	-1	1	2
x_1	1	1	0	$\frac{1}{2}$	0	$-\frac{3}{2}$
$-z$	-3	0	0	$-\frac{1}{2}$	0	$-\frac{1}{2}$
$-z_1$	-4	0	0	$-\frac{1}{2}$	0	$-\frac{3}{2}$
$-z_2$	1	0	0	0	0	1

The integer optimal solution is $x^1 = (1,1)$ and the corresponding criterion vector is:
$(f_1(x^1), f_2(x^1)) = (4,-1)$.
Thus $NDS := \{(4,-1)\}, N_1 = 3,5, H_1^1 = \emptyset, H_1^2 = 5, K_1 = 2, N_1 \setminus H_1^2 = 3$ and $\min_{j \in H_1^2} \{\hat{c}_j^2\} = \hat{c}_5^2 = 1$.
As $|K| = 1$, one problem is created, (P_3), by adding the efficient cut: $f_2(x \geq f_2(x^1) + \lfloor \hat{c}_3^2 \rfloor x_3 + \max\{1, \lfloor \hat{c}_5^2 \rfloor\}$, which expression is given by: $-x_2 \geq -1 + \lfloor 0 \rfloor x_3 + 1$.

The resolution of (P_3) gives :

	b	x_1	x_2	x_3	x_4	x_5	x_6
x_2	0	0	1	0	0	0	1
x_3	1	0	0	1	1	0	-2
x_1	2	1	0	0	$\frac{1}{2}$	0	$-\frac{1}{2}$
x_5	1	0	0	0	0	1	-1
$-z$	-2	0	0	0	$-\frac{1}{2}$	0	$-\frac{3}{2}$
$-z_1$	-2	0	0	0	$-\frac{1}{2}$	0	$-\frac{5}{2}$
$-z_2$	0	0	0	0	0	0	1

The integer optimal solution is $(2,0)$ and the corrseponding criterion vector is $(2,0)$.
$NDS := \{(4,-1),(2,0)\}$ since $(2,0)$ is not dominated by $(4,-1)$, $N_3 = 4,6, H_3^1 = \emptyset, H_3^2 = 6, K_3 = 2, N_3 \setminus H_3^2 = 4et \min_{j \in H_3^2} \{\hat{c}_j^2\} = \hat{c}_6^2 = 1$.
A new problem (P_4) is created from (P_3) after adding the efficient cut: $-x_2 \geq 0 + 0.x_4 + 1$.

	b	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_2	-1	0	1	0	0	0	0	1
x_3	3	0	0	1	1	0	0	-2
x_1	$\frac{5}{2}$	1	0	0	$\frac{1}{2}$	0	0	$-\frac{1}{2}$
x_5	2	0	0	0	0	1	0	-1
x_6	1	0	0	0	0	0	1	-1
$-z$	$-\frac{1}{2}$	0	0	0	$-\frac{1}{2}$	0	0	$-\frac{3}{2}$
$-z_1$	$\frac{1}{2}$	0	0	0	$-\frac{1}{2}$	0	0	$-\frac{5}{2}$
$-z_2$	-1	0	0	0	0	0	0	1

The problem (P_4) is infeasible, as a result the corresponding node is fathomed.
All the nodes of the tree being fathomed, the algorithm terminates and the set of non dominated solutions of the problem (P) is $NDS := \{(4,-1),(2,0)\}$.
It is well known that for large size problems, the enumeration of all efficient solutions and / or non-dominated is no longer possible, it is in this context that we were motivated in this work to study the grid computing technology in order to know what it offers as opportunities for improvement in terms of execution time of such methods, this study is the subject of the next section.

IV. Grid Computing

Grid computing is an architecture that allows multiple communities to share computer resources. The central idea introduced by the grid is the concept of virtual organization (VO) which means a number of people, institutions or organizations that have a common goal in their use of the grid. May be mentioned the best known virtual organizations : LHC (Large Hadron Collider) particle collider at CERN (European Center for Nuclear Research) [3] produces extraordinary amounts of data per year which will be viewed and analyzed by 6000 physicists in the world via computational grids spread over four levels. The value of VO is to provide a simplified access to shared by the organization, which prevents users from bringing increasing amounts of data in their own PC data.

A. *Some definitions of Grid Computing:*

The Grid represents the mutualization of a set of computing resources geographically distributed in different sites. Nevertheless, there are no very precise definition on grid computing.

- **Buyya** defines the grid as an infrastructure that involves the use and integrates collaborative computers, networks, databases of data and scientific instruments owned and managed by multiple organizations [4].
- **CERN**, one of the largest consumers of computing power through the technology of grid computing, defines the Grid in his website Grid cafe [5] as a service that links together computing resources such as PCs, workstations, servers, storage elements, and provides the mechanism needed to access them.
- But the most complete definition for us is that given by **Ian Foster** in his article ”What is the Grid? A Three Point Checklist” [6], where he made a synthesis of several definitions and obtained a list of three items that a grid is a system that:
  1) Coordinates computing resources whose administration is not centralized,
  2) using standardized methods and standards, open for general purposes (a grid is constructed from the multi-purpose protocols and interfaces that enable authentication, resource discovery and access to resources, ...)
  3) provides a significant quality of service (with respect to response time, for example, availability, safety, and /or co-allocation of multiple types of resources to meet the complex needs of the users.

B. *Topology of a grid computing*

The grid is physically composed of nodes, which are processors with their disks, the set being interconnected via a network, these nodes are more or less powerful machines or PCs, or groups of computers (called clusters[1]).
A software driver is installed on each node. It ensures the good running of the local activity of the node (depending on the role of that one) and allows all the grid component elements (people

[1] **Cluster:** Groupped computer units that cooperate for a common end goal and form a single virtual computer unit functionally and administratively.

and resources) to work. All software ensuring the management of the grid is referred to as middleware of the grid. This is sort of the grid ”engine”, it handles all of the grid resources (computing and storage nodes), and is constantly informed of their status, must not only find the data scientists need but also use the programs and computing resources to execute them. The tasks should be distributed anywhere in the world whenever there is sufficient available resource, and then return the result to the scientists.
The middleware provides users and applications the following services:

- reservation and allocation of resources
- scheduling and launching of jobs[2]
- activity monitoring
- system administration

The middleware used by a large number of grid projects is GT (Globus Toolkit). The American scientific grid ”OGS” (Open Grid Science) [7] and the European grid ”EGI” (European Grid Infrastructure) [8] use this middleware since their creation even if EGI used after some years an evolution of the middleware called gLite (Lightweight Middleware for Grid Computing) which was upgraded now into EMI (European Middleware Initiative) [9]. The latter is also the middleware used by the Euro-Mediterranean grid EUMEDGRID [10] and the Algerian grid (DZ e-Science Grid) [11] .
EMI :European Middleware Initiative (gLite before) It is a middleware developed by the EGI (European Grid Infrastructure). The architecture of a grid with EMI middleware consists of a set of services, we present the most important of them [12] (see Figure 1):

- **Security:** To access to the grid resources user must be authenticated to these resources, and must have a ”digital certificate” X.509 (e-passport) which is an encrypted public key associated with a private key. The certificate is issued by a certificate authority ”CA” (Certification Authority).
  The user must then be registered in a ”VO” (Virtual Organization)[3]. With his certificate, he may register in one or more VOs. His request will be validated by the ”VO Manager”.
  The user certificate is used to generate and sign a temporary certificate, called a ”proxy certificate” (or simply a proxy) which allows the identification of the user and its jobs during their execution.
- **The user interface (UI):** The access point to the resources of the grid is the user interface, it is a machine on which the user must have an account where he installs his certificate.
  A UI allows through CLI (Command Line Interface) provided by EMI middleware to perform the basic operations of the grid:
  - ○ list all the appropriate resources to complete a task;
  - ○ submit jobs;
  - ○ cancel jobs;
  - ○ query the status of a job and retrieve its results;
  - ○ copy, reproduce, and delete files in the grid;
  - ○ retrieve the status of the various resources from the information system.
- **The computing element (CE):** It is a service that represents a computing resource, it is the entry point on the computing farms. The CE is composed of:
  - ○ information System;
  - ○ local batch system [4](example: Torque / MAUI or lsf);
  - ○ set of Worker Nodes (homogeneous machines in general).

  The CE submits jobs to worker nodes via the local batch system.
  Applications and software needed by users are installed on this node on a shared directory with the Worker nodes.
- **Worker nodes (WNs):** These are the nodes that provide the computing and execution of jobs sent by the CE, each site must have several WNs.
- **Workload Management System (WMS):** is the conductor of the grid, it allows to accept the job submitted by the user, assigns it to the appropriate CE, records their status and retrieves its results.
  Submitted jobs are described using the JDL (Job Description Language) as a file that specifies what executables and which parameters are to be launched on the grid. Using a process called ”match-maker” the job is submitted to the CE on the basis of the best information in the JDL file.
  The purpose of WMS is to schedule and manager grid resources by optimizing their use. It allows users to submit jobs and run them as soon as possible.
- **The storage element (SE):** These nodes provide access to mass storage drives. disk space is managed by a lightweight solution with the DPM (Disk Pool Manager).
- **Information Service (IS):** provides information on the Grid resources and their status as well as jobs. The information published on the grid by the IS used for supervision and the study of resource performance. The IS used by EMI is BDII (Berkeley Database Information Index), there are two types of BDII machines:
  - ○ the Site-BDII: installed at each site, it reports the status across the site itself.
  - ○ the Top-BDII: represents the top node of the grid, collects information from all sites belonging to the grid through the sites-BDII.

After presenting the grid computing technology based on EMI middleware, we detail in the next section the proposed parallel model for the method of Abbas and Al [2] and its implementation on the grid and finally, experimental results are presented and discussed.

[2] **The job** is the task (work, scientific application) executed by a user on the grid.

[3] VO or Virtual Organization means a number of people, institutions or organizations that have a common purpose in their use of the grid.

[4] Batch denotes that operations are done by a queue system: the jobs are not made upon arrival, but first all grouped in the queue and executed.

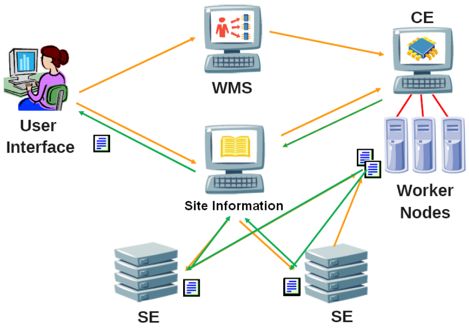


Figure 1. Schema of the grid with EMI (gLite before) [13].

V. Contribution and Implementation

To propose a parallel model for this method we studied the different possibilities to parallelize Branch and Bound algorithms.

A. Parallelism Sources

The use of parallelism has been developed to improve the search performance during the Branch and Bound search tree. Three basic approaches now classic have been classified by researchers in the field (such as Crianic, Le Cun, Gendron and Roucairol [14] and [15]) for the design of parallel branch and bound algorithms:

1) **Type 1 parallelism (based on nodes)**, which introduces parallelism on a particular operation, mainly at a node or a sub problem generated: Parallel analysis of the lower or upper bounds, inversion a matrix, the parallel evaluation of the son, cuts generation, and so on.
   This class of strategies is not intended to alter the path of exploration in the tree, nor its size. The speed is the only objective.
2) **Type 2 parallelism (based on the tree)**, which aims to build and explore the branch and bound tree in parallel by simultaneously performing operations on multiple sub-problems, which means offering to all the processors the ability to select in parallel vertices to explore. Each processor executes the decomposition cycle of the same manner as sequential.
3) **Type 3 parallelism**, which aims to define multiple branch and bound to explore the same area in parallel with or without communication between processors (”Multisearch” approach). Such parallelism can also involve decomposing the domain of feasible solutions and using Branch and Bound to solve the problem on each partition.

Unlike the type of parallelism 1, Type 2 and 3 can lead to modification of the initial structure of the tree.
These strategies are not mutually incompatible. Indeed, when the problems of instances are particularly large and difficult, several strategies can be combined into an overall algorithmic design. For example, a strategy based on the nodes could start the search and quickly generate interesting sub-problems, followed by a parallel exploration of the tree. In other cases, several Branch and Bound can be considered simultaneously to explore a same area, each tree being constructed and explore in parallel.
Parallelization strategies based on the tree give irregular algorithms and corresponding challenges were clearly identified in the project STRATAGEMME [16]:

- tasks are created dynamically in the algorithm.
- The structure of the tree to explore is not known in advance.
- The dependency graph between tasks is unpredictable.
- The assignment of tasks to processors must be made dynamically.

In addition, the parallelism can create tasks that are redundant or unnecessary to be performed, or reduce the research performance in the branch and bound. In the eighties, several researchers have shown that, under certain conditions, parallel implementations may sometimes not provide the hoped acceleration.
It is therefore important to consider the algorithmic aspects at the time of execution, such as sharing, balancing workload or transmition of information between processors.
We explain in the following, what the Matlab environment offers in terms of Parallel programming and integration technology of grid computing. To eventually provide a complete design of the parallel version of the method in question.

B. Parallel Computing in Matlab

Parallel computing consists on the simultaneous execution of the same task, partitioned and adapted in order to be distributed among multiple processors for faster processing of larger problems.
Matlab is a high-level technical language and interactive environment for algorithm development, it is widely used to solve problems in many application areas.
There are toolkits (Matlab toolboxes) used to support parallel computing in Matlab. Two products were released in 2004: ”Parallel Computing Toolbox”, which is the toolkit that integrates parallelism in Matlab code and the ”MATLAB Distributed Computing Server” that distributes the computation over several processors [17], [18].

C. How does the parallel mode work under MATLAB?

Matlab parallel mode consists of two main parts as shown in Figure 2:

- **The user part:** On his workstation, the user must have Matlab Parallel Computing Toolbox (PCT) license. It has the ability to parallelize the code matlab with functions and predefined structures offered by the PCT, the code is divided into sub-tasks that can run in parallel, and it belongs to the user to decide how to divide it.
  They can then test and debug the program locally before launching it remotely on the grid.
- **The distributed part:** The program once arrived at the distributed system (cluster or a grid for example), is managed by a component called ”MATLAB Distributed Computing Server (MDCS)”. It consists

of a scheduler that receives jobs (tasks, calculations) sent by the user and which is at the head of a set of computing nodes called ”Matlab Workers” who will perform the calculations.

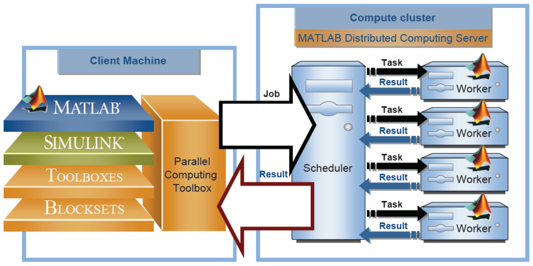


Figure 2. MDCS-Matlab Distributed Computing Server

D. How did we integrate Matlab on the DZ e-Science Grid?

The integration of Matlab on the DZ e-Science Grid came so naturally in order to benefit from Matlab parallel mode. We installed the same components as those illustrated in the previous figure on the elements that make up a grid based on EMI middleware (this have been the subject of a study in the previous section) as shown in Figure 3:
The Matlab Distributed Computing Server is installed on the Computing Element (CE) and its Worker Nodes (WNs); the scheduler is installed on the Computing Element, and Matlab Workers will be launched on the Worker Nodes of the grid.
On his workstation, the user selects the configuration of DZ e-Science Grid and enables access to the grid through the User Interface (UI). When he launches the execution of the program, the MDCS launches calculations in parallel on Matlab Workes of the grid. Once the execution is completed the output files are copied to the Storage Element (SE) and Matlab will transfer them to the user’s workstation.

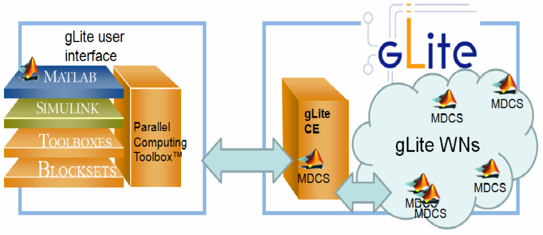


Figure 3. Matlab Distributed Computing Server on EMI (gLite before)

E. The parallel language

It exists in matlab a group of buildings, functions and data structures dedicated to parallel programming, these buildings are divided into two types [19]:

- high-level parallel programming: as parallel loops and distributed arrays;
- low-level or advanced parallel programming: using the protocol of the message passing programming, better known under the name of ”Message Passing Interface” (MPI) [5].

The parallel for loop ”parfor”: This loop can replace a loop in a program, provided that the iterations of the loop are completely independent of each other. Thus, during execution, the parfor loop iterations will be distributed through the available Matlab workers .
For example, suppose that a code comprises a loop to create a sinusoidal wave to draw the waveform:

```
for i = 1: 1024
A(i) = sin(i * 2 * pi / 1024);
end
plot(A)
```

Each iteration being completely independent of each other, it is possible to replace the loop by parfor, as follows:

```
parfor i = 1: 1024
A(i) = sin(i * 2 * pi / 1024);
end
plot(A)
```

At run time each iteration will be calculated by Matlab worker as shown in figure 4 [17].

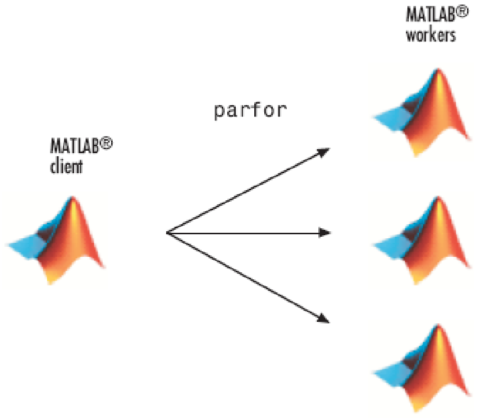


Figure 4. Example of execution parallel with the loop parfor

The distributed Matlab tables: It is possible to distribute a matrix through Matlab workers and working with large data sets. Matlab operations like multiplication, matrix decomposition can use distributed arrays of Matlab. Each worker treats a portion of the matrix and all the workers communicate and are informed of the part that each worker treating. Figure 5 illustrates this [20] principle.

Message passing programming (The Message Passing Interface): Parallel computing experts have the option to use the features of advanced parallel programming to exercise

[5] **Message Passing Interface (MPI):** This protocol was designed in 1993-94, it is a standard that defines a library of functions for use with C, C ++ and Fortran. It allows to use message passing on nodes running parallel programs on distributed memory systems such as computational grids.

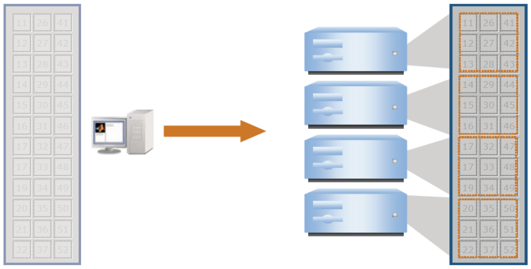


Figure 5. Matlab distributed arrays

more control over their Matlab applications. When the program requires communication (sending and reception of data, synchronization) between nodes (Matlab workers) ”Message Passing Interface” protocol would be the solution.

F. Parallel Model of Abbas and Al Algorithm

The parallel model for the method of Abbas and Al [2] is based on a combination of two parallel strategies that are type 1 and 2.
In step r $(r \geq 1)$; after the linear program Resolution (P_l), $0 \leq l < r$ and determining an optimal solution x^l of (P_l).
If the optimal solution found x^l is integer, which corresponds to a node of type 2 of the method and in order to find a new integer solution, the set K_l of criteria that can be improved from x_l is determined.
For every $i \in K_l$, a constraint is added to (P_l) to obtain $|K_l|$ new linear programs. The $|K_l|$ linear programs will be in turn resolved and so on.
Since every linear program is solved independently of the other, then we proposed to solve these $|K_l|$ linear programs in parallel using the Matlab parfor loop, which means that each program will be resolved on a worker of the grid.
For this model we have chosen to implement the parallelism in the type of node 2 of the algorithm (parallelism based on nodes or type1) and treat in parallel way the tree that is generated by this node (parallelism based on the tree or type2).
The parallel version of the method of Abbas and Al can be described as follows:

- Step 0 (Initialization). Denote by NDS The set of non dominated solutions of (P). Set NDS to the empty set. Solve the linear program (P_0). Set x the optimal solution found.
- Step r $(r \geq 1)$. Solve the linear program (P_l), $0 \leq l < r$. Let x^l be an optimal solution of (P_l).
  If x^l is non integer, choose a coordinate j of x^l whose value is not integer and separate the node l on this coordinate into two new nodes and return to step r.
  If x^l is integer then:
  ◦ If the corresponding criterion vector Cx^l is not dominated by any criterion vector Z from NDS then update (NDS) by adding Cx^l.
  ◦ If there exists a solution $Z \in NDS$ that is dominated by the criterion vector Cx^l then replace Z by Cx^l in NDS.
  Determine the set K_l of criteria that can be improved from x_l, for any $i \in K_l$.
  parfor $i \in K_l$: add the constraint (5) for K_l new linear programs $(P_k), k > l$ and go back to step r.
- stopping condition. The procedure stops when $H_l = \emptyset$ which means no criterion can be improved or (P_l) admits no feasible solutions for any stage l such that $0 \leq l \leq r$.

G. Implemantation and Results

The parallel algorithm were implemented in a Matlab 7.0 environment equipped with Parallel Computing Toolbox Licence and tested on randomly generated MOILP problems. For each instance n represents the number of variables, $m$the number of constraints, and $k$the number of objectives. These experiments were performed on instances of MOILP problems with two, three and four objectives. The data are integers uniformly distributed in the interval [-10,99] for constraints coefficients.
The weights for the criteria are all taken equal to 1.
Note that the machine where is installed Matlab and PCT is a quad-core server with 4 GB of memory configured with Scientific Linux 6.8.
DZ-01 cluster part of DZ e-Science Grid on which is installed the MDCS is a cluster equipped with 5 workers, each worker is a server with two processors Intel Pentium 4 running at 3.6 GHz 4 GB of RAM and 160 GB disk, also configured with Scientific Linux 6.8.
All procedures have been programmed, including methods simplex and dual simplex.
For each instance we carried out the execution locally on the UI machine first and then on the grid. And this in order to compare the performance between one machine and parallel execution on 5 machines.
The following table shows for each instance and for both types of performance, the average of five independent runs, computation time in seconds and the number of non-dominated solutions obtained:

Instances			Computing Time		Number of
n	m	k	Execution on one machine	Execution on DZ-01 Cluster	non dominated solutions
15	5	2	17.32	18.9	12.40
20	5	2	92.49	93.53	12.80
25	5	2	400.6	414.3	13.60
15	5	3	20.2	21.45	18.50
20	5	3	99.87	101.85	25.50
25	5	3	356.67	340.3	21.20
20	10	3	64.34	50.45	24
25	10	3	200.23	170.65	20.60
15	5	4	18.12	21.45	19.10
20	5	4	128.1	119.36	130
25	5	4	378.7	350.99	128.90
30	10	4	632.82	500.34	89
50	10	4	x	10800	75.30
60	10	4	x	18034	64
70	10	4	x	x	x

From the table, it appears that for small instances no improvement is obtained when running parallel through the Matlab Workers. There may even be a light slowing down compared to the execution on a single machine as in the case of instances (15,5,2), (20,5,3) and (15,5,4). This slowdown is due to the fact that the execution time is negligible comparing to the data transfer time and communic ation between the scheduler and its Matlab and workers as well as workers together. The running time in parallelis then slowed.

We are starting to see a light improvement in the time to parallel execution of medium-sized instances, such as for instances (20,10,3), (25,10,3) and (30,10,4). Here the runtime instances becomes considerable comparing to communication data transfer time and this is why we get improvements.

In large bodies such as (50,10,4) and (50,10,4) it is clear that the performance on a single machine does not release results after 24 hours of computation as the execution time and the memory required for the calculation becomes very large, however, although with a running time which is very large compared to the preceding instances , parallel execution gives the set of all non-dominated solutions and manages to solve instances in question.

Beyond Instance (70,10,4) no results are obtained and after a wait time that exceeds 24 hours.

Given these results, it appears that distributing computing through several workers is beneficial for instances with execution times larger than the communication time across the network.

We can also deduce that through the parallelization of the method, we were able to solve up to a limit of large instances which is not negligible.

VI. Conclusion

In this paper, we explained our approach to parallelize Abbas and Al method [2]. We concluded that parallel computing techniques can help reduce the time required to reach a solution. To take advantage of parallelization, it is important to choose a correct approach to the optimization problem.

After testing the parallelization method by parfor loop and to solve larger instances, it would be interesting in the near future to test other methods of parallelization as message passing programming "MPI" allowing better manage of communication between Matlab workers or use distributed arrays Matlab and maybe even use a combination of several parallel methods.

References

[1] R. Gomory, "Outline of an algorithm for integer solutions to linear programs," Bulletin of the American Mathematical Society, vol. 64, 1958, pp. 275–278.

[2] M. Abbas, M. Chergui, and M. A. Mehdi, "Efficient cuts for generating the non-dominated vectors for multiple objective integer linear programming," International Journal of Mathematics in Operational Research (IJMOR), Vol. 4, No.3 pp. 302-316, vol. 4, no. 3, 2012.

[3] "CERN Accelerating science," URL: http://home.web.cern.ch/.

[4] R. Buyya and S. Venugopal, "A gentle introduction to grid computing and technologies," Computer Society of India, 2005.

[5] "What is Grid Computing," URL: http://www.gridcafe.org/EN/what-is-the-grid.html [accessed: 2014-05].

[6] I. Foster, "What is grid? a three point checklist," July 2002, URL: http://www.mcs.anl.gov/ itf/Articles/WhatIsTheGrid.pdf.

[7] "Open Science Grid," URL: http://www.opensciencegrid.org/.

[8] "EGI site," URL: www.egi.eu.

[9] "EMI-European Middleware Initiative," URL: http://www.eu-emi.eu/.

[10] "EUMEDGRID Support," URL: www.eumedgrid.eu.

[11] "DZ e-Science GRID," URL: www.grid.arn.dz,.

[12] S. Burke, S. Campana, E. Lanciotti, P. M. Lorenzo, V. Miccio, C. Nater, R. Santinelli, and A. Sciab, GLITE 3.1 USER GUIDE, 18 December 2009.

[13] F. Galeazzi, Architecture of the gLite Data Management System. Presentation in EUMEDGRID School for Application porting, Algiers, July 2010.

[14] T. Crainic, B. L. Cun, and C. Roucairol, Parallel Branch-and-Bound Algorithms, Chapitre 1. John Wiley Sons, Inc, 2009.

[15] B. Gendron and T. Crainic, Parallel Branch-and-Bound Algorithms:Survey and Synthesis. Operations Research, 1994, vol. 42, no 6.

[16] C. Roucairol, A Parallel Branch and Bound Algorithm for the Quadratic Assignment Problem. Discrete Applied Mathematics, No 18, pages 211-225, 1987.

[17] Parallel Computing Toolbox User's Guide. The MathWorks, Natick, MA, March 2008.

[18] MathWorks, Matlab Distributed Computing Server System Administrator's Guide. The MathWorks, Natick, MA, March 2008.

[19] A. Chakravarti, S. Grad-Freilich, E. Laure, M. Jouvin, G. Philippon, C. Loomis, and E. Floros, Enhancing e-Infrastructures with Advanced Technical Computing:Parallel MATLAB on the Grid. The MathWorks, Inc, 2008.

[20] A. Chakravarti and E. Chan, Hands-On Session for Parallel Computing with MATLAB and gLite. The MathWorks, Inc, 2008.

Evaluation of melt pool geometry during pulsed laser welding of Ti6Al4V alloy

Mohammad Akbari, Seyfolah Saedodin, Afshin Panjehpour, Samaneh N aghieh and Masoud Afrand

***Abstract*—*In this paper, laser welding of titanium alloy (Ti6Al4V) is investigated as regarded a numerical and experimental study. Modeling for the temperature distribution is performed through a transient three-dimensional problem to predict the heat affected zone (HAZ), depth and width of the molten pool. The experiments were performed at different welding conditions to estimate the thermal model results with the depth, width and microstructure of the welded samples. It was observed that the thermal model was in good agreement with the experimental data. The model prediction error was found to be in the 2–17% range with most numerical values falling within 7% of the experimental values*.

***Keywords*—** *Laser welding, Titanium alloy, numerical study, Temperature distribution*.

I. Introduction

THE increased interest by industry in laser welding is because this technique has shown high efficiency and low production cost compared to other welding methods. Laser welding can provide a significant benefit for the welding because of its precision and rapid processing capability. Titanium and its alloys have been widely used due to low density, good corrosion resistance, high operating temperature, etc. Some applications of titanium alloys in aerospace, biomedical, nuclear and automotive industries are reported by Wang et al. [1]. During the laser welding of Ti6Al4V alloy some sequential types of the phase transitions such as alpha–beta titanium trans-formation in the solid phase, melting, evaporation, ionization occur and then in the reverse order during cooling stage. Joining Ti6Al4V titanium alloys using pulsed Nd:YAG laser welding method was done by Akman et al. [2]. Their results showed that it was possible to control the penetration depth and geometry of the laser weld bead by precisely controlling the laser output parameters. Yang et al. [3] provided a finite element model to predict the depth and width of HAZ in laser heating of Ti6Al4V alloy plate and found that the depth and width of HAZ were decreased with an increase of laser scan speed. Frewin and Scott [4] produced a time-dependent 3D model of heat flow during pulsed Nd:YAG laser welding. By ignoring the convective flows in the melt pool and assuming a Gaussian energy distribution, they calculated transient temperature cross-sections along with the dimensions of the fusion and HAZ. They found that the fusion and HAZ produced numerically were extremely close to those produced experimentally. Review of aforementioned articles showed that various parameters were investigated in different studies. From each work, different aspects of laser welding were studied and therefore, different results were obtained, each of which could be useful in its position. However, no comprehensive study in these research fields was found to predict the width and depth of molten pool by using temperature history. We have carried out a numerical and experimental study of laser welding for modeling of temperature distribution and molten pool shape to predict the depth and width of the molten pool and HAZ dimensions. The objective of this paper was to examine the effect of welding speed on the temperature distribution, weld depth and weld width.

II. Experimental setup

Experiments were performed to characterize the temperature measurements and HAZ dimensions in laser welding. The sample was Ti6Al4V alloy plate (50 mm × 20 mm with the thickness of 3 mm). A model IQL-10 pulsed Nd:YAG laser with a maximum mean laser power of 350 W and wavelength 1.06 μm was used as the laser source. The laser parameter ranges were 0.2–25 ms for pulse duration, 1–1000 Hz for pulse frequency and 0–40 J for pulse energy. The spot diameter on the surface of the plate was set at about 0.7 mm. For the purpose of shielding, the pure argon gas from a coaxial nozzle was used with the flow rate at 15 l/min. Fig. 1 shows a schematic illustration of the experimental setup. K-type thermocouples with an operative range between -40° C and +1260° C and the accuracy between ±1% were used for temperature measurement. Because the temperature of the molten pool was very high, the thermocouples were attached

This work was supported in part by the Najafabad Branch, Islamic Azad University, Iran.

Mohammad Akbari is with the Department of Mechanical Engineering, Najafabad Branch, Islamic Azad University, Isfahan, Iran (corresponding author to provide phone: 00989133154773; fax: 00983142291111; e-mail: m.akbari.g80@gmail.com).

Seyfolah Saedodin is with the Department of Mechanical Engineering, Semnan University, Semnan, Iran (e-mail: s_sadodin@semnan.ac.ir).

Afshin Panjehpour is with the Department of Mechanical Engineering, Najafabad Branch, Islamic Azad University, Isfahan, Iran e-mail: (afshin_p2010@yahoo.com)

Samaneh Naghieh is with the Department of Mechanical Engineering, Najafabad Branch, Islamic Azad University, Isfahan, Iran e-mail: (s_naghieh@gmail.com).

Masoud Afrand is with the Department of Mechanical Engineering, Najafabad Branch, Islamic Azad University, Isfahan, Iran e-mail: (masoud_afrand@yahoo.com).

on the top surface at 2 mm lateral distance from the center of the molten pool. The locations of thermocouples are specified in Fig. 1 (points A and B). The data were recorded using the data acquisition card (model: Advantech USB 4718). For the metallographic preparation, all the samples were mounted, polished using the standard metallographic techniques and etched using Kroll's reagent (Distilled water-92 ml, nitric acid-6 ml and hydrofluoric acid-2 ml). The width and depth of the molten pool were measured using an Olympus SZ-X16 stereoscopic microscope.

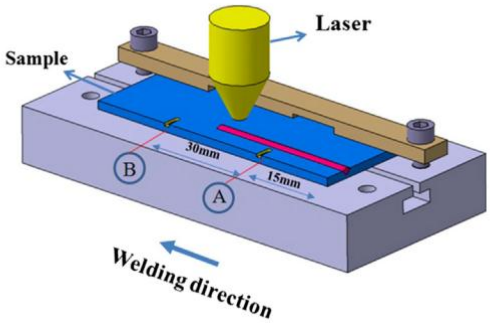


Fig1. Schematic of the laser welding experiments.

III. Numerical simulation

The purpose of this study was the numerical modeling and experimental investigation of temperature distribution and molten pool dimensions to predict the depth and width of the molten pool and HAZ dimensions. Due to the fact that the workpiece is moving in x direction, the temperature distribution and the melting pool shape are not asymmetric. As a result, the problem is a transient three-dimensional problem. For simplicity, the weld pool surface was considered to be flat and the complex physical phenomenon causing the formation of keyhole was not considered. The governing equations and their boundary conditions were discretized by control volume schemes. The SIMPLE algorithm and first upwind discretization method [5] were used to calculate the fluid flow and heat transfer phenomena.

A. Governing equations

The mathematical formulation of the model is based on the following assumptions:

The initial temperature for workpiece is at 293 K. The laser beam and coordinate system are fixed and workpiece moves in the x direction with a constant velocity. The thermophysical properties of the material are temperature dependent. The equations of continuity, energy and momentum in the Cartesian coordinate system can be written as follows [6]:

Continuity equation

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho u)}{\partial x}+\frac{\partial(\rho v)}{\partial y}+\frac{\partial(\rho w)}{\partial z}=0 \qquad (1)$$

X -momentum equation

$$\frac{\partial(\rho u)}{\partial t}+\frac{\partial(\rho uu)}{\partial x}+\frac{\partial(\rho uv)}{\partial y}+\frac{\partial(\rho uw)}{\partial z}=-\frac{\partial P}{\partial x}+ \\ \frac{\partial}{\partial x}(\mu\frac{\partial u}{\partial x})+\frac{\partial}{\partial y}(\mu\frac{\partial u}{\partial y})+\frac{\partial}{\partial z}(\mu\frac{\partial u}{\partial z})-\frac{\mu}{K}(u-v_w) \qquad (2)$$

Y -momentum equation

$$\frac{\partial(\rho v)}{\partial t}+\frac{\partial(\rho vu)}{\partial x}+\frac{\partial(\rho vv)}{\partial y}+\frac{\partial(\rho vw)}{\partial z}=-\frac{\partial P}{\partial y}+ \\ \frac{\partial}{\partial x}(\mu\frac{\partial v}{\partial x})+\frac{\partial}{\partial y}(\mu\frac{\partial v}{\partial y})+\frac{\partial}{\partial z}(\mu\frac{\partial v}{\partial z})-\frac{\mu}{K}v \qquad (3)$$

Z -momentum equation

$$\frac{\partial(\rho w)}{\partial t}+\frac{\partial(\rho wu)}{\partial x}+\frac{\partial(\rho wv)}{\partial y}+\frac{\partial(\rho ww)}{\partial z}=-\frac{\partial P}{\partial z}+\frac{\partial}{\partial x}(\mu\frac{\partial w}{\partial x}) \\ +\frac{\partial}{\partial y}(\mu\frac{\partial w}{\partial y})+\frac{\partial}{\partial z}(\mu\frac{\partial w}{\partial z})-\frac{\mu}{K}w+\rho g\beta(T-T_{ref}) \qquad (4)$$

Energy equation

$$\rho C_p(\frac{\partial T}{\partial t}+(u-v_w)\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}+w\frac{\partial T}{\partial z})=\frac{\partial}{\partial x}(k\frac{\partial T}{\partial x})+\frac{\partial}{\partial y}(k\frac{\partial T}{\partial y}) \\ +\frac{\partial}{\partial z}(k\frac{\partial T}{\partial z})+S-\frac{\partial}{\partial x}(\rho u\Delta H)-\frac{\partial}{\partial y}(\rho v\Delta H)-\frac{\partial}{\partial z}(\rho w\Delta H) \qquad (5)$$

In momentum equations, K is the permeability coefficient, which is related to the liquid volume fraction with the Koreny–Carman equation [7]. It enables us to have a smooth transition of velocity from zero in the solid region to a large value in the fully liquid region for the fixed-grid numerical method [8]. In energy equation, according to a suitable latent updating form during each interaction within a time step updated, a source-based method is used to deal with the latent heat of fusion, ΔH , as an additional heat source. By introducing the permeability factor and source-based method, the Eqs. (1)–(5) are unique for both liquid and solid phases. Therefore, it is not necessary to track the melt-solid interface and specify a boundary at that location. In the weld pool, the surface tension force, the Lorentz force and the buoyancy force interact with each other. The Marangoni convection which is the main driving force of the fluid flow, acts because of the temperature dependence of the surface tension [9, 10].

B. Boundary and initial conditions

The boundary conditions at the upper surface are as follows

For the weld pool (in the liquid region)

$$\mu\frac{\partial u}{\partial z}=-\frac{\partial\gamma}{\partial T}\frac{\partial T}{\partial x}; \qquad \mu\frac{\partial v}{\partial z}=-\frac{\partial\gamma}{\partial T}\frac{\partial T}{\partial y}; \qquad w=0 \qquad (6)$$

where $\frac{\partial\gamma}{\partial T}$ is the temperature coefficient of surface tension

For the solid region

$$u=v_{welding}; \qquad v=0; \qquad w=0 \qquad (7)$$

The initial condition at time $t=0$ is given as

$$T(x,y,z,0) = 293 \text{ K} \tag{8}$$

The convection and radiation boundary conditions on all surfaces are considered. In addition, on the top surface, a transient heat flux (which is produced by the beam laser) is considered [11]

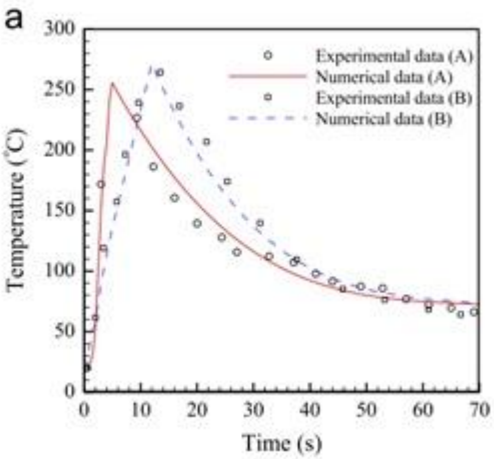
$$\text{for} \quad z=0 \qquad -k\frac{\partial T}{\partial z} = q - \varepsilon(T)\sigma(T^4 - T_\infty^4) - h(T - T_\infty) \tag{9}$$

where h is the convective heat transfer coefficient, σ is Stefan–Boltzmann constant $=5.67\times10^{8}\, W\!/_{m^2K^4}$ and ε is emissivity. For radiation and convection problems, the following lumped convection coefficient was used as suggested by Frewin and Scott [4]

$$h = 2.4\times10^{-3}\varepsilon T^{1.61} \tag{10}$$

IV. Results and discussion

The effect of process parameters on temperature history is investigated in Fig. 2. The temperature histories for specific points (A and B) representing the positions of thermo- couples have been plotted. In the experiment corresponding to this figure, the temperature histories of welding process were studied with the variation in the welding speed by keeping the remaining parameters equal to each other. As shown in this figure, the finite volume (FV) thermal model (numerical simulation) was in good agreement with the experimental data. Also, we observed that the temperature histories for different welding speeds had similar trends in the case of identical welding speed. This figure shows that by decreasing the welding speed, the peak value of temperature diagram is increased and its maximum value occurs at a longer time. It can be observed that the temperature histories of both the numerical and experimental data had similar shapes when they were compared.



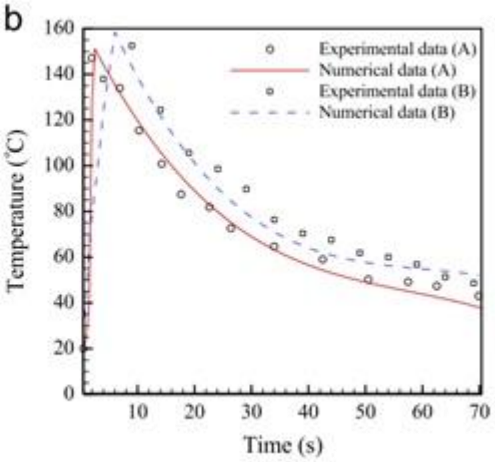


Fig.2 Simulated and experimental results of temperature distribution for points A and B as a function of welding speed (a) v =3 mm/s, (b) v =6 mm/s

Fig3. reports the welding width versus welding speed achieved by performing an experimental setup with 25 Hz pulse frequency, 4.2 ms pulse duration and average power of 240 W. The welding speed was varied between 3 and 9 mm/s. According to the results, the width was decreased with increasing laser welding speed. This increase in the welding speed has an inverse effect on the welding width. In laser welding, a good weld is not only a weld with sufficient penetration but also, it is one with an acceptable weld surface (width). Hence from this figure, we can conclude that at sufficient low welding speed, we have a larger width and the weld surface has an acceptable quality.

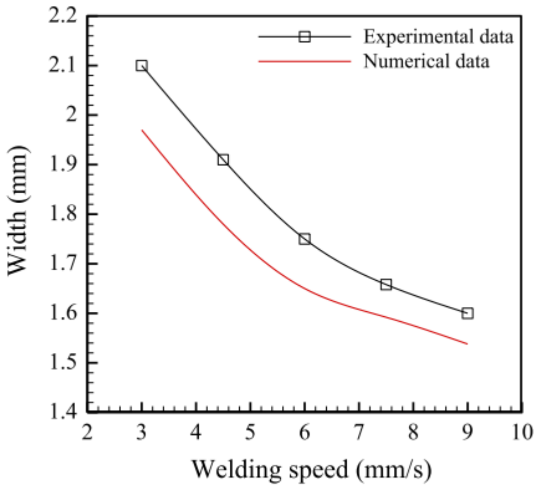


Fig3. Experimental and numerical results: width versus welding speed variation

Fig. 4 shows the melt pool depth as a function of welding speed. It is clearly seen that the melt pool depth was decreased by increasing the laser welding speed. This means that at a given laser power, a larger welding speed did produce lower penetration depth. It should be noted that the numerical results agreed with the experimental data.

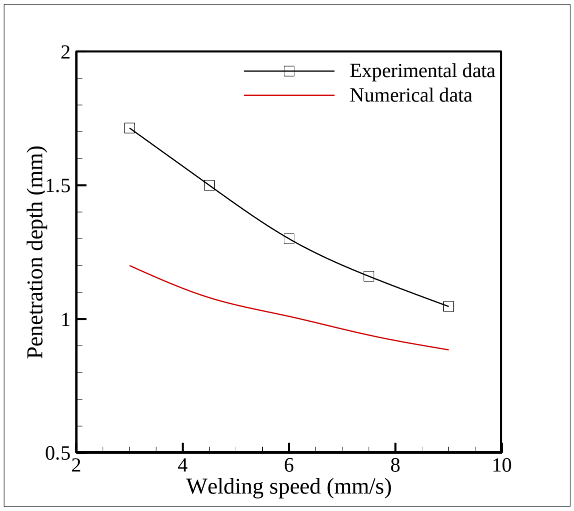


Fig4. Experimental and numerical results: penetration depth versus welding speed variation

A numerically predicted temperature contour versus the experimental micrograph for welding speed of 3 mm/s is plotted in Fig. 5 for comparison between numerical and experimental results. It can be concluded that the predicted numerical temperature contours give a good insight related to phase transformation in the molten pool and HAZ as observed in Fig. 5.

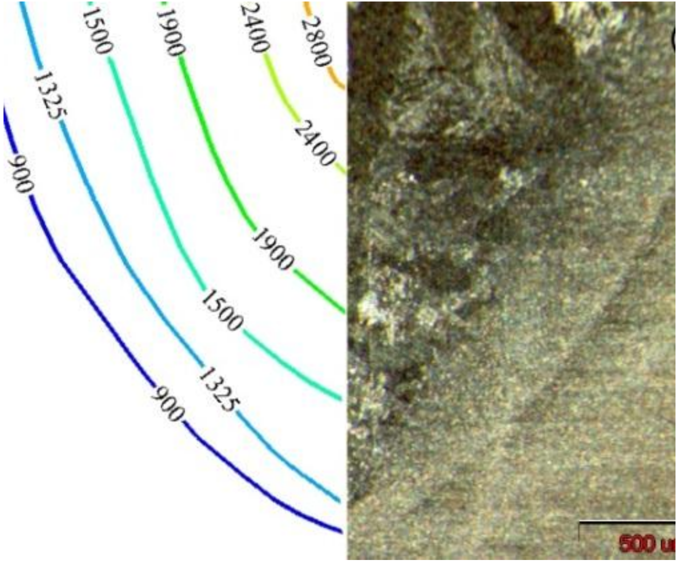


Fig5. The cross sectional area of the sample numerical and experimental results v =3 mm/s

In laser welding different zones are always identified such as the fusion zone (FZ), the heat-affected zone (HAZ), and the base metal (BM). According to the Fig 6, the HAZ microstructure consists of a mixture of martensitic α' and primary α. The FZ microstructures are identified as α' martensite. Martensitic phase could be obtained because the welding pool temperature reached the β transus (980° C for Ti6Al4V) and the cooling rate was high.

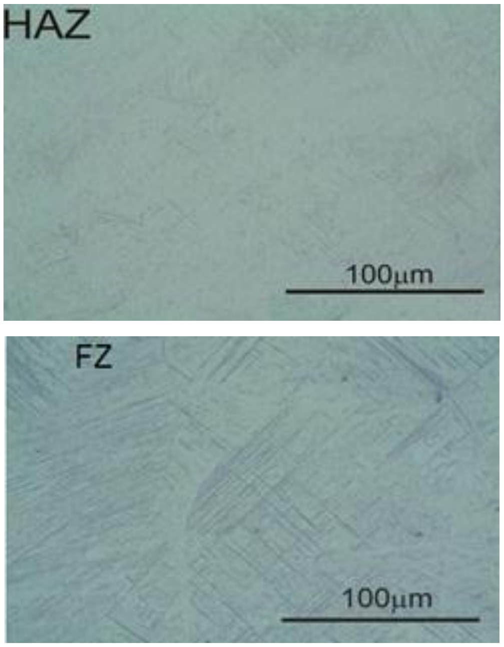


Fig6. Microstructure of the heat-affected zone and fusion zone (FZ)

In this paper a numerical and experimental study of laser welding was conducted for prediction of temperature distribution and molten pool geometry. The variations in the weld geometry (width and depth) that affected by laser welding parameters indirectly estimated with considering the temperature variations around the molten pool which obtained from the numerical model. The finite volume thermal model is in good agreement with the experimental data. The model can predicts the influences of laser welding speed on the weld pool shape and size related to temperature variations. The predicted numerical temperature contours give a good insight related to phase transformation in the molten pool and HAZ.

References

[1] SH. Wang, MD. wei and LW. Tsay, "Tensile properties of LBW welds in Ti–6Al–4V alloy at evaluated temperatures below 450 °C", Materials Letters, vol. 57, pp. 1815–1823, 2003.

[2] E. Akman, A. Demir, T.Canel and T.Sınmazcelik, "Laser welding of Ti6Al4V titanium alloys", Journal of Materials Processing Technology, vol. 209, pp. 3705–3713, 2009.

[3] J. Yang, S. sun, M. Brandt and W. Yan, "Experimental investigation and 3D finite element prediction of the heat affected zone during laser assisted machining of Ti6Al4V alloy", Journal of Materials Processing Technology, vol. 210, pp. 2215–2222, 2010.

[4] MR. Frewin and DA. Scott, "Finite element model of pulsed laser welding", Welding journal, vol. 78, pp. 15–22, 1999.

[5] S.V. Patankar and D.B. Spalding, "A calculation procedure for heat, mass and momentum transfer in three-dimensional parabolic flows", International Journal of Heat and Mass Transfer, vol. 15, pp. 1787–1806, 1972

[6] K. Abderrazak, S. Bannour, H. Mhiri, G. Lepalec and M. Autric, "Numerical and experimental study of molten pool formation during continuous laser welding of AZ91 magnesium alloy", Computational Material Science, vol. 44, pp. 858–866, 2009.

[7] W.D. Bennon and F.P. Incropera, "A continuum model for momentum, heat and species transport in binary solid–liquid phase change systems-

1model formulation", International Journal of Heat and Mass Transfer, vol. 30, pp. 2161–2170, 1987.

[8] V.R. Voler and C. Prakash, "A fixed grid numerical modeling methodology for convection diffusion mushy region phase-change problems", International Journal of Heat and Mass Transfer, vol. 30, pp. 1709–1719, 1987.

[9] EJ. Ha and WS. Kim, "A study of low-power density laser welding process with evolution of free surface", International Journal of Heat and Fluid Flow, vol. 26, pp. 613–21, 2005.

[10] HG. Fan, HL. Tsai and SJ. Na, "Heat transfer and fluid flow in a partially or fully penetrated weld pool in gas tungsten arc welding", International Journal of Heat and Mass Transfer, vol. 44 pp. 417–28, 2001.

[11] M. Akbari, S. Saedodin, D. Toghraie, R. Shoja-Razavi and F. Kowsari., "Experimental and numerical investigation of temperature distribution and melt pool geometry during pulsed laser welding of Ti6Al4V alloy", Optics & Laser Technology, vol. 59, pp. 52–59, 2014

Investigation on the film cooling effectiveness from cylindrical and row trenched cooling holes adjacent the combustor endwall surface

Ehsan Kianpour, Arezou Sayyedana, Arash Karimipour, Alireza Shirneshan, Iman Golshokouh

***Abstract*—** The current research was accomplished in order to analyze the effects of cylindrical and row trenched cooling holes with alignment angle of +60 degrees on the film cooling effectiveness adjacent the combustor end wall surface at blowing ratio of 3.18. This research included a linear representation simulation of a true Pratt and Whitney gas turbine engine. The present investigation has been done with Reynolds-averaged Navier-Stokes turbulence model (RANS) on internal cooling passages. The combustor simulator contains a combination of two rows of dilution jets interaction. These jets were staggered in the stream wise direction and aligned in the span wise direction. The entire findings of the study indicated that when, the row trenched holes used the, near the enwall surface, film cooling effectiveness is almost two times more than film cooling performance of baseline case.

***Keywords*—** gas turbine, film-cooling, trenched holes, cylindrical holes, brayton cycle.

I. Introduction

USING a Brayton cycle is a key to get higher engine efficiency and power to weight ratio. According to this cycle it is needed to increase outlet combustion temperature. This temperature increment creates harsh environment for the downstream components. So, it is needed to design a cooling technique in this area. Film cooling is the common way that is used. In this method, a thin thermal boundary layer such as buffer zone is created by cooling holes and attached on the protected surface. Cylindrical and trenched cooling holes are two forms of the holes. A broad literature survey was conducted to collect the information. Azzi and Jurban [1] tested some methods to investigate the film cooling thermal field. They used standard k-ε turbulence model to solve the Reynolds averaged Navier-Stokes equation. In concurred with Rozati and Danesh Tafti [2], findings declare that the film cooling effectiveness was improved at low blowing ratios. Kianpour et. al [3,4], simulated the combustor end wall cooling holes with two different layouts. The results declared that while, the central part of the jets stayed nominally at the same temperature level for both configurations. Yiping et. al [5] tested the effects of depth and width of trenches on the film cooling under overall cooling effectiveness of φ=0.6 as determined by Maikell [6]. Also it is found that the third (w=2.0D and d=0.75D) and fourth (w=3.0D and d=0.75D) cases were more effective than others. According to the importance of this issue, more studies are required. There are many questions should be answered: How can trenched cooling holes improve the film cooling performance at the combustor end wall compared to the cylindrical holes? Also in order to measure the validity of the results, a comparison between the data gained from this study, Vakil and Thole [7] and Stitzel and Thole [8] was made.

II. Research Methodology

The combustor was a container with a width, height and length of 111.8 cm, 99.1 cm and 156.9 cm respectively. The contraction angle was 15.8 degrees and began at x=79.8cm. Also the inlet cross-sectional area and the exit cross sectional area was 1.11 m2 and 0.62 m2. The combustor simulator contained four stream wise series of film cooling panels. The starting point of these panels was approximately at 1.6 m upstream of the turbine vanes. The length of these panels was 39.2 cm, 40.6 cm, 36.8 cm and 43.2 cm respectively (Fig. 1). The low thermal conductivity (k= 0.037 w/mk) of combustor panels were 1.27 cm in thickness, which allowed for adiabatic surface temperature measurements. There were two rows of dilution holes within the second and third cooling panels. These dilution holes were located at 0.67 m and 0.90 m downstream of the beginning of the combustor liner panels. The diameter of the first and second row of dilution holes was 8.5 cm and 11.9 cm. The centerline of the second row was staggered with respect to the first row of dilution holes. The model was a three dimensional representation of a true Pratt

This work was supported by Universiti Teknologi Malaysia and Ministry of Higher Education of Malaysia. This research was supported by Research University Grant 03H58.

Ehsan Kianpour is with the Mechanical Engineering Department, Najafabad Branch, Islamic Azad University, Isfahan, Iran (corresponding author phone: +98-31-42292881; e-mail: eianpour@pmc.iaun.ac.ir).

Arezou Sayyedana, (e-mail: arezousayyedana@gmail.com).

Arash Karimipour is with the Mechanical Engineering Department, Najafabad Branch, Islamic Azad University, Isfahan, Iran (e-mail: arashkarimipour@gmail.com).

Alireza Shirneshan is with the Mechanical Engineering Department, Najafabad Branch, Islamic Azad University, Isfahan, Iran (e-mail: arshirneshan@yahoo.com).

Iman Golshokouh is with the the Mechanical Engineering Faculty, Izeh Branch, Islamic Azad University, Izeh, Iran (e-mail: golshokooh@yahoo.com).

and Whitney gas turbine aero-engine, however, the current simulator included two schemes of cooling holes.

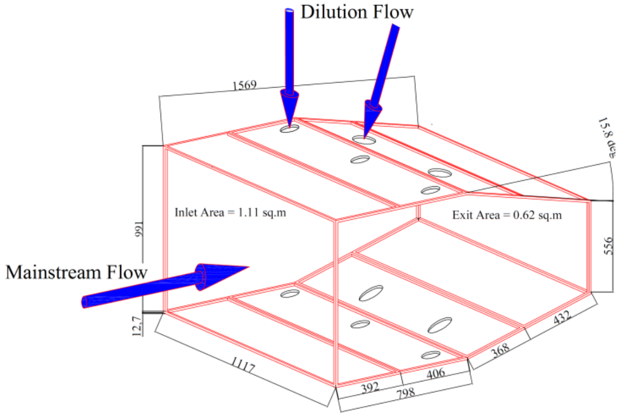


Fig. 1 The 3-D view of the combustor simulator.

For the second case, the cooling holes placed within a row trench with alignment angle of +60 degrees. The film-cooling holes were placed in equilateral triangles. The film-cooling holes diameter of both models was 0.76 cm. They drilled at an angle of 30 degrees from the horizontal surface. For the baseline case, the length of each cooling hole was 2.5 cm. Furthermore, the trench depth and width was 0.75D and 1.0D respectively. Also a global coordinate system (x, y, and z) was selected. Also dimensionless variables were defined for both the coolant and the dilution flow (Table I).

Table I. Typical operating conditions for main flow and coolant

Parameter	Location	Combustor Simulator
Density $\left(\frac{Kg}{m^3}\right)$	Main Flow	1
	Dilution Row 1	(air) 1.17
	Dilution Row 2	
	Cooling Panels	
Velocity $\left(\frac{m}{s}\right)$	Main Flow	1.62
	Dilution Row 1	17.36
	Dilution Row 2	8.68
	Cooling Panels	(First Velocity) 4.6
		(Second Velocity) 1.8

The temperature of the coolant, dilution jets and main flow was considered equal to 295.5 K and 332 K respectively.

The film cooling effectiveness (η) distribution in the combustor simulator was collected along the specific observation planes. These observation planes are illustrated in Fig. 2. To solve the combustor simulator and getting more accurate data and reasonable time consumption, about 8×106 tetrahedral meshes were used and this is in concurred with Stitzel and Thole study [8]. The meshes were denser around the cooling and dilution holes and even wall surfaces. According to the specific flow ratio at the inlet of volume control, mass flow inlet and the condition was defined at the inlet of combustor simulator.

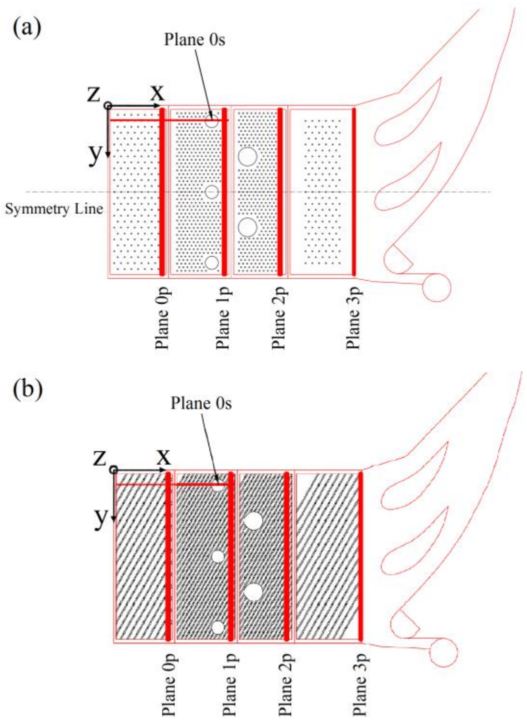


Fig 2. Location of the observation planes (a) baseline (b) case 2.

Wall boundary and slip less boundary conditions were applied to limit the interaction zone between fluid and solid layer. Also at the end of volume control the pressure outlet boundary condition was used. In addition, both cases were completely symmetric along the x-y and x-z planes. According to this issue, symmetry boundary condition was applied. In addition these equations ere used as well.

continuity equation:

$$\frac{\partial}{\partial t}(\rho u_i)+\frac{\partial}{\partial x_j}=(\rho u_i u_j)=-\frac{\partial P}{\partial x_i}+\frac{\partial \tau_{ij}}{\partial x_i}+\rho g_i+\vec{F}_1 \qquad (1)$$

momentum equation

$$\frac{\partial \rho}{\partial t}+\frac{\partial}{\partial x}\frac{dx}{dt}+\frac{\partial \rho}{\partial y}\frac{dy}{dt}+\frac{\partial \rho}{\partial z}\frac{dz}{dt}=-\rho(\nabla.V) \qquad (2)$$

energy and RNG *k*-*ε* equations

$$\begin{gathered}\frac{\partial}{\partial t}(\rho E)+\frac{\partial}{\partial x_i}\left(u_i(\rho E+P)\right)= \\ \frac{\partial}{\partial x_i}\left(K_{eff}\frac{\partial T}{\partial x_i}-\sum_j h_j J_j+u_i\left(\tau_{ij}\right)_{erf}\right)+S_h\end{gathered} \qquad (3)$$

$$\frac{\partial}{\partial t}(\rho k)+\frac{\partial}{\partial x_i}(\rho k u_i)=\frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right]+P_k-\rho\varepsilon \tag{4}$$

$$\frac{\partial}{\partial t}(\rho\varepsilon)+\frac{\partial}{\partial x_i}(\rho\varepsilon u_i)= \\ \frac{\partial}{\partial x_j}\left[\left(\mu+\frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right]+C_{1\varepsilon}\frac{\varepsilon}{k}P_k-C^*_{2\varepsilon}\rho\frac{\varepsilon^2}{k} \tag{5}$$

To understand the thermal field results, the quantities should be defined. Film cooling effectiveness is defined as below:

$$\eta=\frac{T-T_\infty}{T_c-T_\infty} \tag{6}$$

In the above equation, T is the local temperature, T_∞ is the main stream temperature, and T_c is the temperature of coolant.

III. Findings and Discussion

Fig. 3 illustrates the fluctuations of film cooling observed in the current study, the experimental findings and the CFD prediction by Vakil and Thole [7] and Stitzel and Thole [8].

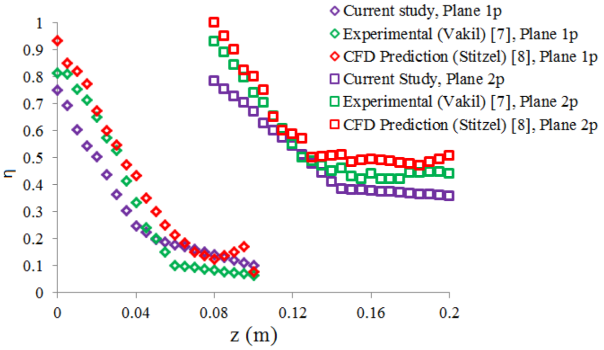


Fig 3. The film cooling effectiveness comparison of planes 1p and 2p along y/W=0.4.

The plot clearly shows that for plane 1p, different solutions are satisfactorily consistent. However, it appears that a good prediction can be made based on the present study because the value gained in this analysis is closer to the CFD prediction data for z=4.5cm. Within observation plane of 2p, Vakil's model and Stitzel's CFD prediction possess reasonable consistency, the lowest performance could be seen in the present study. The investigation was under-predicted and this was intensified adjacent the combustor end wall surface (z=1cm). Also, the variation between the current study computation and benchmarks were calculated as follows:

$$\%\text{Diff}=\frac{\sum_{i=1}^{n}\frac{x_i-x_{i,\text{benchmark}}}{x_{i,\text{benchmark}}}}{n}\times 100 \tag{7}$$

The deviation of film cooling effectiveness from experimental data gathered by Vakil and Thole [7] and CFD prediction by Stitzel and Thole [8] was equal to 11.14% and 15.79% and 6.35% and 11.14% for observation planes of 1p and 2p, respectively.

The changes of film cooling effectiveness of different configurations for the observation planes can be seen in Fig. 4. For the observation plane of 0p, the trenched holes with alignment angle of +60 degrees performed the most effectively. This was 1.4 times as much the film cooling effectiveness of the baseline case. For both cases the film cooling effectiveness reduces continuously from the combustor end wall surface to further distance. Also For the next observation plane, the film exiting the trenches created a new boundary and enhanced heat transfer coefficients immediately downstream of the trench. The baseline was the lowest among all cases. Even in observation plane of 2p, the trenched holes with alignment angle of +60 degrees remained at the highest level. The baseline had a 13% reduction in cooling performance with an increase in its blowing ratio. And for the last plane, the results showed that the highest film cooling effectiveness enhancement was seen by trenched holes with alignment angle of +60 degrees (38.8%).

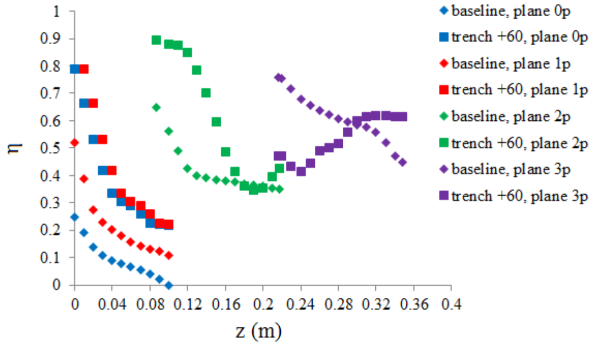


Fig 4. The variation of film cooling effectiveness for different measurement planes.

Findings related to film cooling effectiveness in plane 2p at BR=3.18 can be observed in Fig. 5. The trenched cases showed higher cooling performance. Film cooling effectiveness reduced remarkably in baseline cases. Hence, a warmer area ($0.15<\eta<0.20$) expanded at the right (0cm<y<4cm) and left (48cm<y<52cm) sides of the thermal field contour. At this blowing ratio, the layer reached to z=18cm for the trenched holes with alignment angles of +60 degrees, while the thickness of the baseline case hardly reached to z=9 cm. On the left of Fig. 5, the thermal contours of the second row of dilution (depicted by the solid lines and arrow) are visible. Also Fig. 5 shows the *v* and *w* velocity ectors superimposed on the top of the thermal field contours in plane 2p. It is visible for all arrangements that the coolant sweeps toward the second row of dilution jet and accelerates near the combustor wall.

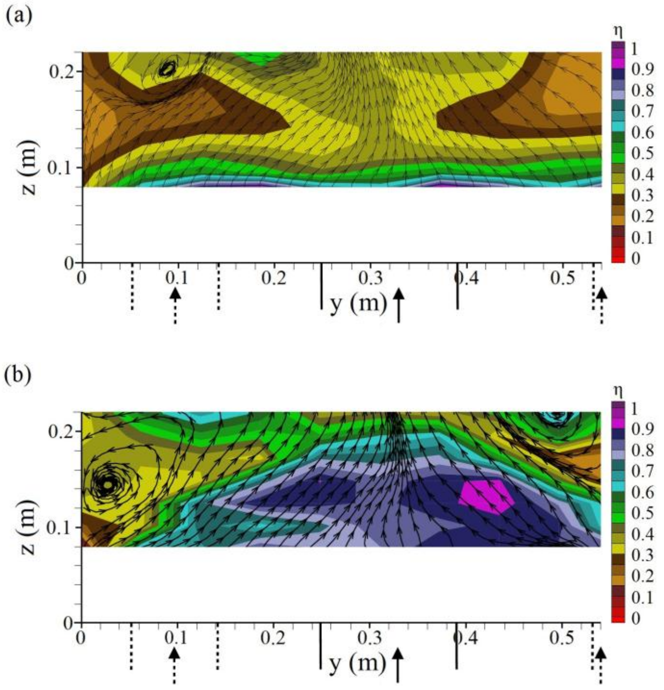


Fig 5. Vectors of *v* and *w* on the film-cooling effectiveness contours of plane 2p (a) baseline (b) case 2.

IV. CONCLUSION AND RECOMMENDATION

The objective of the current research was to analyze the effects of different cooling holes configurations of cylindrical and row trenched holes with alignment angle of +60 degrees at BR=3.18 on the film cooling effectiveness at the end of combustor simulator. In this study a 3-D representation of a Pratt and Whitney aero-engine was simulated and analyzed. To sum up, the usage of trenched cooling holes significantly to development of the film cooling layer. Also, the central part of the plane 2p showed the intense penetration of the coolant and a thick film cooling layer creation in the trenched case. Initially, the results declared that for the observation planes of 0p, 1p and 2p trenching cooling holes has intense effect on film cooling effectiveness especially for plane 1p and 2p. comparison between experimental and computational results show that the prediction of the film cooling for the different observation planes exhibited a thinner film cooling layer for the current study. The new configurations of cooling holes could be used in gas turbine, aerospace and refinery industries. Based on the results and conclusions of the study, in future research within this area, different configurations of trenched cooling holes and baseline case should be considered for different cooling panels.

REFERENCES

[1] Azzi, and B. A. Jubran, “Influence of leading edge lateral injection angles on the film cooling effectiveness of a gas turbine blade,” J. of Heat and Mass Transfer, vol. 40, May 2004, pp. 501–508.

[2] A. Rozati, and K. Danesh Tafti, “Effect of coolant–mainstream blowing ratio on leading edge film cooling flow and heat transfer – LES investigation,” J. of Heat and Fluid Flow, vol. 29, August 2008, pp. 857–873.

[3] E. Kianpour, N. A. C. Sidik, and I. Golshokouh, “Measurement of film effectiveness for cylindrical and row trenched cooling holes at different blowing ratios,” J. Numerical Heat Transfer, Part A, vol. 66, November 2014, pp. 1154–1171.

[4] E. Kianpour, N. A. C. Sidik, and A. S. M. Bozorg, “Dynamic Analysis of Flow Field at the End of Combustor Simulator,” Jurnal Teknologi, vol. 58, 2012, pp. 5-12.

[5] L. Yiping, A. Dhungel, S. V. Ekkad, and R. S. Bunker, “Effect of Trench Width and Depth on Film Cooling From Cylindrical Holes Embedded in Trenches,” J. Turbomachinery, vol. 131, September 2008, pp. 011003-1-011003-13.

[6] J. Maikell, D. Bogard, J. Piggush, and A. Kohli, “Experimental Simulation of a Film Cooled Turbine Blade Leading Edge Including Thermal Barrier Coating Effects,” J. Turbomachinery, vol. 133, September 2010, pp. 011014-1-011014-7.

[7] S. S. Vakil, and K. A. Thole, “Flow and Thermal Field Measurements in a Combustor Simulator Relevant to a Gas Turbine Aeroengine,” J. Engineering Gas Turbines Power, vol. 127, April 2005, pp. 257-267.

[8] S. Stitzel, and K. A. Thole, “Flow field computations of combustor-turbine relevant to a gas turbine engine,” J. Turbomachinery. Vol. 126, March 2004, pp. 122-129.

Hamming distance between partitions, clustering comparison and information

Giovanni Rossi

***Abstract*—Measuring the distance between partitions is useful for clustering comparison in different fields. For example, in bioinformatics the measuring mostly obtains through a *maximum matching* distance MMD, although this is algorithmically demanding and hardly fits certain instances. In fact, another distance measure is being tested, namely one based on information theory and termed *variation of information* VI. Alternatively, this paper proposes the *Hamming* distance HD, displaying large range and great measurement sensitivity, while also relying on a neat binary string representation of partitions. Novel distance HD is computationally handy and shares with VI important characterizing axioms. Developing from the combinatorial concern to translate the traditional Hamming distance from subset to partition lattices, HD constitutes a valuable computational tool for clustering and information processing where a distance between partitions is to be measured.**

***Index Terms*—Combinatorial problem, partition lattice, Hamming distance, clustering, information theory, bioinformatics.**

I. Introduction

A Partition or clustering of a (finite) set N divides this latter into disjoint subsets whose union yields the whole (partitioned) set N. There exists a massive literature on cluster analysis and clustering algorithms, which is important in a variety of fields at the interface of computer science, artificial intelligence and engineering, including pattern recognition/learning, web mining and bioinformatics. In particular, measuring the distance between partitions is useful for information processing in terms of comparisons between clustering results. For example, a local-search clustering algorithm generally provides different outputs for varying initial candidate solutions (or inputs), and it may be relevant to measure the distance between any two of them, or between each of them and a given one (see [22, section 1]). Specifically, in bioinformatics a great deal of attention is being paid to measuring the distance between clusterings of populations, either natural or experimental, for sibling relationship reconstruction. In practice, thus far the focus has been placed almost exclusively on a unique distance measure (see [17], [6], [31], [19], [18], [30], [2], [8], [5], [10], [3]), namely one relying on maximum matching and denoted MMD in the sequel. After its first appearance [4], this measure was subsequently shown [15] to be computable via the assignment problem (see [20, p. 236] and [12, subsection 4.1]). In this view, a notable exception relies on using an alternative distance measure [9], called *variation of information* VI and derived from information theory [22].

The present paper provides a further solution to the problem of measuring a distance between partitions. This proposed measure is of the Hamming type (see section 2), indeed obtained by translating the traditional Hamming distance between subsets [7, p. 3] in terms of partitions. As long as lattice theory is concerned, the result is thus combinatorially consistent. Furthermore, it shares with VI important characterizing axioms. On the other hand, it also displays two features that appear interesting for biologists and practitioners in general. Firstly, as the Hamming distance HD between partitions provided here obtains simply by means of scalar products between (input) vectors, no proper computational issue arises, while determining MMD and VI is algorithmically demanding (see [15], [19], [18] and [12, chapter 4]). Secondly, HD has a large range and hence great measurement sensitivity.

II. Hamming and partition distances

FOR a finite set $N = \{1, \ldots, n\}$, let $(2^N, \cap, \cup)$ and $(\mathcal{P}^N, \wedge, \vee)$ denote the corresponding subset and partition lattices, with inclusion $\supseteq$ and coarsening $\geqslant$ as order relations, respectively. Both are atomic (and atomistic), but the fomer is distributive while the latter is geometric [1], [32]. For populations or data sets, n is the number of individuals or data points being partitioned.

The distance between elements of a ordered set is to be measured in terms of the order relation. Also, measures of the distance between elements of any set are called *Hamming distances* when these elements are represented as arrays and the distance between two of them is the number of entries where their array representations differ. The Hamming distance $d(A, B)$ between two subsets $A, B \in 2^N$ is $d(A, B) =$

$$= |A\Delta B| = |A\backslash B| + |B\backslash A| = r(A \cup B) - r(A \cap B), \quad (1)$$

$r : 2^N \rightarrow \mathbb{Z}_+$ being the rank function: $r(A') = |A'|$ for all $A' \in 2^N$. In words, it counts how many $i \in N$ are included in either A or else B, but not in both. Elements $i \in N$, when regarded as 1-cardinal subsets $\{i\} \in 2^N$, are *atoms* in subset lattice $(2^N, \cap, \cup)$. Expression (1) defines indeed a Hamming distance in that subsets $A \in 2^N$ are firstly represented as binary n-vectors $\chi_A \in \{0, 1\}^n$ (or vertexes of the n-dimensional unit hypercube $[0, 1]^n$) through their *characteristic function* $\chi_A : N \rightarrow \{0, 1\}$, defined by $\chi_A(i) = 1$ if $i \in A$ and $\chi_A(i) = 0$ if $i \in N\backslash A$. Next, the distance between any two of them $A, B \in 2^N$ is the number of entries where χ_A and χ_B differ. That is, the cardinality of their symmetric difference $A\Delta B$ [7], [1].

A partition $P = \{A_1, \ldots, A_{|P|}\} \subset 2^N$ of N is a collection of pairwise disjoint subsets, called blocks (or clusters), whose union yields N. Any subset $A \in 2^N$ has a unique complement $A^c = N\backslash A$. For all partitions $P \in \mathcal{P}^N$ and all non-empty

Giovanni Rossi is with the Department of Informatica Scienza Ingegneria (DISI), University of Bologna, via Mura Anteo Zamboni 7, Bologna 40127, Italy; e-mail: giovanni.rossi6@unibo.it

subsets $\emptyset \subset A \subseteq N$, denote the partition of A induced by P by $P^A = \{B \cap A : B \in P, \emptyset \neq B \cap A\}$. The maximum matching distance $MMD : \mathcal{P}^N \times \mathcal{P}^N \rightarrow \{0, 1, \ldots, n-1\}$ between any two partitions P, Q is

$$MMD(P, Q) = \min\{|A^c| : \emptyset \subset A \subseteq N, P^A = Q^A\}. \quad (2)$$

This is the minimum number of elements $i \in N$ that must be deleted in order for the two residual induced partitions to coincide. Also, $MMD(P, Q)$ *is the minimum number of elements that must be moved between clusters of P so that the resulting partition equals* Q ([15, p. 160]). It is computable as a maximum matching (or assignment) problem. Firstly recall that in a (simple) graph a matching is a set of pairwise disjoint edges (i.e. the endpoints are all different vertexes). Now consider the bipartite graph $G = (P \dot{\cup} Q, E)$ with $|P| + |Q|$ vertexes ($\dot{\cup}$ is the union of disjoint sets), one for each block of each partition, and join any two of them $A \in P$ and $B \in Q$ with an edge $(A, B) \in E$ if $A \cap B \neq \emptyset$. In addition, let $|A \cap B|$ be the weight of the edge. Then, determining $MMD(P, Q)$ amounts to find a maximum weighted matching E^* in G, that is one where the sum $\sum_{(A,B) \in E^*} |A \cap B|$ of edge weights is maximal. In fact, the (minimal) number $MMD(P, Q)$ of elements that must be removed for the two residual partitions to coincide is the sum $\sum_{(A,B) \in E^*} |A \Delta B|$ over all selected edges of the cardinality of the symmetric difference between the associated endpoints.

Another important measure of the distance between two partitions P, Q is the variation of information $VI(P, Q)$. Firstly obtained axiomatically from information theory [22], this distance is now useful in bioinformatics as well [9]. Entropy $H(P) = -\sum_{A \in P} \frac{|A|}{n} \log\left(\frac{|A|}{n}\right)$ of a partition P and *mutual information* $I(P, Q) = \sum_{A \in P, B \in Q} \frac{|A \cap B|}{n} \log\left(\frac{n|A \cap B|}{|A||B|}\right)$ between P and Q (binary logarithm) enable to measure the distance between these latter as variation of information

$$VI(P, Q) = H(P) + H(Q) - 2I(P, Q). \quad (3)$$

While MMD ranges over all integer values $\{0, 1, \ldots, n-1\}$ in $[0, n-1] \subset \mathbb{R}_+$, VI ranges in $[0, \log n]$.

Apart from MMD and VI, there exist several other partition distance measures (see [13, sections 10.2 and 10.3, pp. 191-193] and [11], [16], [33], [23]). One was proposed as the *Hamming distance between (matrices representing) partitions* [22], [24], [25], and thus shall be briefly distinguished from the object of this paper. A binary relation R on N is a subset $R \subseteq N \times N$ of ordered pairs (i, j) of elements $i, j \in N$. The collection of all such binary relations is subset lattice $(2^{N \times N}, \cap, \cup)$. If symmetry $(i, j) \in R \Rightarrow (j, i) \in R$ and *transitivity* $(i, j), (j, h) \in R \Rightarrow (i, h) \in R$ hold, then R is an *equivalence* relation, or a partition of N into equivalence classes: maximal subsets $A \in 2^N$ such that $(i, j), (j, i) \in R$ for all $i, j \in A$ are precisely its blocks. A binary relation R may be represented as a $n \times n$ matrix $M^R \in \{0, 1\}^{n \times n}$ with entries $M^R_{ij} = 1$ if $(i, j) \in R$ and $M^R_{ij} = 0$ if $(i, j) \notin R$. Now let two equivalence relations R, R' have associated partitions P, P' and representing matrices $M^R, M^{R'}$. The distance $d(R, R')$ between subsets $R, R' \in 2^{N \times N}$can be computed through expression (1) above: $d(R, R') = |R \Delta R'| = |R \cup R'| - |R \cap R'|$. While providing a distance $\delta(P, P')$ between partitions, this factually is a Hamming distance between subsets. In particular, $|R \Delta R'|$ is the number of 1s in matrix $M^{R \Delta R'} = M^R + M^{R'}$ modulo 2 (see [1, p. 338]). The point is that $2^{N \times N}$ contains many lattice elements (or binary relations) that do not correspond to partitions (or equivalence relations).

Roughly speaking, partition lattice $(\mathcal{P}^N, \wedge, \vee)$ is compressed into a larger subset lattice, with which some elements are shared while some others are not. Apart from binary relations just described, another example comes from noticing that partitions P are collections of subsets, i.e. $P \in 2^{2^N}$, and thus the distance between P and Q may be computed as the Hamming distance $|P \Delta Q|$ between elements of subset lattice $(2^{2^N}, \cap, \cup)$, i.e. the number of subsets $A \in 2^N$ that are blocks of either one but not both. Again, there are many set systems (or collections $S \in 2^{2^N}$ of subsets) that do not correspond to partitions. This feature is maintained even when P and Q are represented as joins of atoms, for they generally admit several such representations. The Hamming distance HD between partitions proposed below relies on representing partitions as $\{0, 1\}$-valued and $\binom{n}{2}$-dimensional arrays, because $\binom{n}{2}$ is the number of atoms. But not all $2^{\binom{n}{2}}$ such arrays correspond to partitions.

III. Atoms, rank and size

The rank function $r : \mathcal{P}^N \rightarrow \mathbb{Z}_+$ for the partition lattice is $r(P) = n - |P|$, that is, the maximum conceivable number of blocks for a partition of a n-set (which is n, of course) minus the actual number of blocks of any partition P. The unique partition $P_\perp = \{\{1\}, \ldots, \{n\}\}$ with rank $r(P_\perp) = 0$ is the bottom element. By definition, atoms are immediately above, with rank 1. That is, atoms populate level 1 (above level 0) of the associated *Hasse diagram* (see [1], [32], [22]), with coarsening $\geqslant$ as the order relation (i.e. coarser partitions in upper levels). This means that an atom, in the partition lattice, is any P with $n - 1$ blocks, out of which $n - 2$ must be singletons, while the remaining one is a pair, and there are $\binom{n}{2}$ unordered pairs.

For notational convenience, let $[ij] \in \mathcal{P}^N$ denote the atom where the unique 2-cardinal block is (unordered) pair $\{i, j\} \in [ij]$, noting that $2^{\binom{n}{2}}$ is the number of simple graphs with n vertexes. In fact, partitions are the *transitive closure* of such graphs: the former obtain from the latter by adding all edges needed to render complete (or fully connected) each component [1, Point 2.31, p. 54]. The order relation among partitions is coarsening $\geqslant$, where $P \geqslant Q$ means that every block of Q is included in some block of P. Let $\mathbf{1} \in \{0, 1\}^n$ denote the n-vector with all entries equal to 1 and $\langle x, y \rangle$ be the scalar product of any two vectors x, y. For any subset $A \in 2^N$, clearly $|A| = r(A) = \langle \chi_A, \mathbf{1} \rangle$. Also, given any $A, B \in 2^N$,

$$d(A, B) = |A \Delta B| = |A| + |B| - 2|A \cap B| = \quad (4)$$

$$= \langle \chi_A, \mathbf{1} \rangle + \langle \chi_B, \mathbf{1} \rangle - 2\langle \chi_A, \chi_B \rangle. \quad (5)$$

Let $\mathcal{P}^N_{(1)} = \{[ij] : 1 \leq i < j \leq n\}$ denote the $\binom{n}{2}$-set of atoms of the partition lattice. The analog of the characteristic function χ_A of subsets $A \in 2^N$ for partitions $P \in \mathcal{P}^N$ is the *indicator function* $I_P : \mathcal{P}^N_{(1)} \rightarrow \{0,1\}$ defined by

$$I_P([ij]) = \begin{cases} 1 \text{ if } P \geqslant [ij] \\ 0 \text{ if } P \ngeqslant [ij] \end{cases} \quad \text{for all } P \in \mathcal{P}^N, [ij] \in \mathcal{P}^N_{(1)}.$$

In words, if pair $\{i,j\}$ is included in some block A of P (i.e. $\{i,j\} \subseteq A \in P$), then partition P is coarser than atom $[ij]$, and the corresponding position $I_P([ij])$ of indicator array I_P has entry 1. Otherwise, that position is 0.

By reversing the order relation, $P \geqslant [ij]$ turns into $[ij] \leqslant P$, i.e. atom $[ij]$ is finer than P. The number of atoms finer than $P = \{A_1, \ldots, A_{|P|}\}$ is the *size* $s : \mathcal{P}^N \rightarrow \mathbb{Z}_+$ (see [27]), i.e.

$$s^P = \sum_{1 \leq k \leq |P|} \binom{|A_k|}{2} = \langle I_P, \mathbf{1} \rangle, \tag{6}$$

with $s^P = s(P)$ and $\mathbf{1} \in \{0,1\}^{\binom{n}{2}}$ now denoting the unitary $\binom{n}{2}$-vector. That is to say, $s^P = |\{[ij] \in \mathcal{P}^N_{(1)} : [ij] \leqslant P\}|$.

TABLE I

Available sizes of partitions of n-sets, $1 \leq n \leq 7$.

$\lvert N \rvert = n$	$\{s^P : P \in \mathcal{P}^N\}$ Available sizes
1	$\{0\}$
2	$\{0,1\}$
3	$\{0,1,3\}$
4	$\{0,1,2,3,6\}$
5	$\{0,1,2,3,4,6,10\}$
6	$\{0,1,2,3,4,6,7,10,15\}$
7	$\{0,1,2,3,4,5,6,7,9,10,11,15,21\}$

While the cardinality $|A|$ of subsets $A \in 2^N$ takes every integer value between 0 and n, the size $s^P = \langle I_P, \mathbf{1} \rangle$ of partitions $P \in \mathcal{P}^N$ does not the same between 0 and $\binom{n}{2}$. In fact, a main difference between characteristic function χ_A of subsets $A \in 2^N$ and indicator function I_P of partitions $P \in \mathcal{P}^N$, is that the former spans the whole vertex set of the unit n-dimensional hypercube (that is $\{\chi_A : A \in 2^N\} = \{0,1\}^n$), while the latter only spans a (rather small) proper subset of the vertex set of the unit $\binom{n}{2}$-dimensional hypercube (that is $\{I_P : P \in \mathcal{P}^N\} \subset \{0,1\}^{\binom{n}{2}}$, where $\subset$ denotes strict inclusion $\subsetneq$; again, see [1, Point 2.31, p. 54]). This is due to *linear dependence* [34], characterizing geometric lattices in general. The number of non-spanned vertexes is $2^{\binom{n}{2}} - \mathcal{B}_n$, where $\mathcal{B}_n = |\mathcal{P}^N|$ is the n-th *Bell number* or number of partitions of a n-set [29], [14] [18, Section 2.2]. The smallest value of n where linear dependence may be appreciated is $n = 3$, in that there are $\mathcal{B}_3 = 5$ partitions, namely the finest $\{\{1\},\{2\},\{3\}\}$ and coarsest $\{\{1,2,3\}\}$ ones, and $\binom{3}{2} = 3$ atoms: $[12] = \{\{1,2\},\{3\}\}$, $[13] = \{\{1,3\},\{2\}\}$ and $[23] = \{\{2,3\},\{1\}\}$. Thus, there is no partition with size equal to 2, as $[12] \vee [23] = [12] \vee [13] = [13] \vee [23] = \{\{1,2,3\}\} = = [12] \vee [13] \vee [23]$, where for any two partitions $P, Q \in \mathcal{P}^N$, the join $P \vee Q$ is the finest partition coarser than both P and Q, while the meet $P \wedge Q$ is the coarsest partition finer than both P and Q (see [1], [32]). Available sizes for $1 \leq n \leq 7$ are in Table I above.

A. Representations as joins of atoms: rank and size

In atomic (or atomistic, see [1], [32]) lattices every element admits some representation as a join of atoms. Yet, while subsets $A \in 2^N$ admit a unique such a representation, namely $A = \underset{i \in A}{\cup} \{i\}$, partitions generally admit several such representations. Again, this results from linear dependence and is observable already when $n = 3$ (see above), in that the coarsest partition $\{\{1,2,3\}\}$ may be represented either as the join of any two atoms, or even as the join of all the three available atoms at once. In particular, the rank $r(P) = n - |P|$ of any partition P is the minimum number of atoms needed for representing P as a join of atoms, while the size $s^P = \sum_{A \in P} \binom{|A|}{2}$ is the maximum number of atoms available for representing P as a join of atoms. Hence, keeping the coarsest partition of a 3-cardinal set as the simplest example, the rank is $r(\{\{1,2,3\}\}) = 3 - 1 = 2$, while the size is $s^{\{\{1,2,3\}\}} = 3 = \binom{3}{2}$.

The rank or cardinality of subsets is a peculiar lattice function, namely a *valuation* of subset lattice $(2^N, \cap, \cup)$ [1, p. 187]), meaning that $|A \cup B| + |A \cap B| = |A| + |B|$ for all $A, B \in 2^N$. On the other hand, neither the rank nor the size of partitions are valuations of partition lattice $(\mathcal{P}^N, \wedge, \vee)$. In fact, valuations of the partition lattice are constant (partition functions), taking the same value on each and every partition [1, Exercise 12 (ii) p. 195]. The following definitions enable to appreciate the features displayed by the rank and the size of partitions, regarded as lattice functions $r, s : \mathcal{P}^N \rightarrow \mathbb{Z}_+$.

For any two partitions $P, Q \in \mathcal{P}^N$, the strict coarsening relation $>$ is $P > Q$, meaning $P \geqslant Q, P \neq Q$. The rank is easily seen to be a strictly monotone partition function, that is to say $P > Q$ entails $r(P) > r(Q)$. The same is established hereafter for the size.

Proposition 1: The size is a strictly monotone partition function:

$$s^P > s^Q \text{ for all } P, Q \in \mathcal{P}^N \text{ such that } P > Q.$$

Proof: If $P > Q$, then every block $A \in P$ is the union of some blocks $B_1, \ldots, B_{|Q^A|} \in Q$, with[1] $|Q^A| > 1$ for at least one block $A \in P$. The union of any two such blocks $B, B' \in Q$ increases the size (toward s^P, i.e. while reaching an higher level of the Hasse diagram toward P) by

$$\binom{|B| + |B'|}{2} - \left(\binom{|B|}{2} + \binom{|B'|}{2} \right) = |B||B'|,$$

which is strictly positive as blocks are non-empty. ■

If $f(P) + f(Q) \geq f(P \vee Q) + f(P \wedge Q)$ for all $P, Q \in \mathcal{P}^N$, then partition function $f : \mathcal{P}^N \rightarrow \mathbb{R}$ is sub-modular. Super-modularity obtains by reversing the inequality. As partitions are the transitive closure of graphs, partition lattice $\mathcal{P}^N$ is the *polygon matroid* (see again [1, pp. 54 and 274-5]), and matroids have sub-modular rank [1, Rank axioms 6.14 p. 265]. Conversely, the size is a super-modular partition function. To see this, firstly consider the *covering* relation $>^*$, with $P >^* Q$ meaning: $P > Q$ and there is no P' such that $P > P' > Q$.

[1] Recall that Q^A is the partition of $A \in 2^N$ induced by $Q \in \mathcal{P}^N$, see section 2 above.

Proposition 2: The size is a super-modular partition function:

$$s^{P\vee Q}+s^{P\wedge Q}\geq s^P+s^Q \text{ for all } P,Q\in\mathcal{P}^N.$$

Proof: If the two partitions are comparable, say $P\geqslant Q$, then $P=P\vee Q$ and $Q=P\wedge Q$, which makes the statement satisfied with equality. Otherwise, if $P\ngeqslant Q\ngeqslant P$, then there are two maximal chains of partitions (see [1], [32]), one meets $P\wedge Q$ and P as well as $P\vee Q$, while the other meets $P\wedge Q$ and Q as well as $P\vee Q$. Focusing on the relevant part or segment of the former chain, there are $\hat{P}_{\hat{r}}>^*\cdots>^*\hat{P}_1>^*\hat{P}_0$, with $\hat{r}=r(P\vee Q)-r(P\wedge Q)$, such that $\hat{P}_0=P\wedge Q$ and $\hat{P}_{\hat{r}}=P\vee Q$ as well as $\hat{P}_{k_P}=P$ for some $k_P, 0<k_P<\hat{r}$. Similarly, focusing on the relevant segment of the latter chain[2], there are $\hat{Q}_{\hat{r}}>^*\cdots>^*\hat{Q}_1>^*\hat{Q}_0$ such that $\hat{Q}_0=P\wedge Q$ and $\hat{Q}_{\hat{r}}=P\vee Q$ as well as $\hat{Q}_{k_Q}=Q$ for some $k_Q, 0<k_Q<\hat{r}$. Note that $r(P)=r(Q)\Leftrightarrow k_P=k_Q$.

Difference $s^{P\vee Q}+s^{P\wedge Q}-(s^P+s^Q)$ may be counted by summing across (Hasse diagram) levels for the two segments. The fact is that frequently many atoms are both *(a)* finer than $P\vee Q$, and *(b)* $\geqslant$-incomparable with respect to both P and Q. Atoms $[ij]\leqslant P\wedge Q$ may be disregarded because they are counted in the size of all the four involved partitions, namely $P,Q,P\wedge Q,P\vee Q$. As for the remaining ones, firstly observe that sets

$$S_P=\{[ij]\in\mathcal{P}_1^N:P\geqslant[ij]\nleqslant P\wedge Q\} \text{ and}$$

$$S_Q=\{[ij]\in\mathcal{P}_1^N:Q\geqslant[ij]\nleqslant P\wedge Q\}$$

are disjoint, i.e. have empty intersection $S_P\cap S_Q=\emptyset$. In fact, if an atom $[ij]\nleqslant P\wedge Q$ satisfied $P\geqslant[ij]\leqslant Q$, then $(P\wedge Q)\vee[ij]$ (and not $P\wedge Q$) would be the coarsest partition finer than both P,Q, i.e. a contradiction.

Now consider going upwards through the Hasse diagram from $P\wedge Q$ to $P\vee Q$ *twice*, starting with all atoms finer than $P\vee Q$ apart from those also finer than $P\wedge Q$. Specifically, a first route is through segment $\hat{P}_0,\ldots,\hat{P}_{\hat{r}}$ of the former maximal chain, and at each partition reached up to $\hat{P}_{k_P}=P$ all atoms finer than the current partition but not also finer than the preceding one are removed. A second, subsequent route starts with only the *residual* atoms and is through segment $\hat{Q}_0,\ldots,\hat{Q}_{\hat{r}}$ of the latter maximal chain. Again, up to $\hat{Q}_{k_Q}=Q$ at each reached level all atoms finer than the current partition but not also finer than the preceding are removed. The above intersection being empty, it is not possible that the same atom is found to be removed twice, and at the end of the second route there still remains a non-empty (generally large) collection of atoms, namely all those for reaching $P\vee Q$ from either P or Q. ■

IV. Hamming distance between partitions

In order to parallel the traditional Hamming distance between subsets given by (1) above, HD has to count the number of atoms finer than either one of any two partitions but not finer than both. Thus, in terms of cardinalities of subsets of atoms, distance $HD:\mathcal{P}^N\times\mathcal{P}^N\rightarrow\mathbb{Z}_+$ is given, for all $P,Q\in\mathcal{P}^N$, by

$$\begin{aligned}HD(P,Q)&=|\{[ij]\in\mathcal{P}_{(1)}^N:P\geqslant[ij]\nleqslant Q\}|+\\&+|\{[ij]\in\mathcal{P}_{(1)}^N:P\ngeqslant[ij]\leqslant Q\}| \quad (7)\\&=|\{[ij]\in\mathcal{P}_{(1)}^N:P\geqslant[ij]\text{ OR }Q\geqslant[ij]\}|+\\&-|\{[ij]\in\mathcal{P}_{(1)}^N:P\geqslant[ij]\leqslant Q\}|, \quad (8)\end{aligned}$$

where OR in second expression (8) means precisely that the corresponding subset contains all atoms finer than *at least one* of the two partitions, while the other (second) subset clearly contains all atoms finer than both.

The size and the indicator functions allow to reproduce these expressions in a computationally efficient manner. Considering the size first, note that

$$P\wedge Q=\bigvee_{\substack{[ij]\in\mathcal{P}_{(1)}^N\\P\geqslant[ij]\leqslant Q}}[ij], \quad (9)$$

and this is the maximal representation of $P\wedge Q$ as a join of atoms, namely that utilizing $s^{P\wedge Q}$ atoms. Accordingly, a further expression of HD obtains immediately as follows

$$HD(P,Q)=s^P+s^Q-2s^{P\wedge Q} \text{ for all } P,Q\in\mathcal{P}^N. \quad (10)$$

In words, this subtracts twice the number $s^{P\wedge Q}$ of atoms finer than both P and Q from the sum of the number s^P of atoms finer than P and the number s^Q of atoms finer than Q.

Remark 3: In proposition 2 above, subtracting $2s^{P\wedge Q}$ from both sides yields

$$s^{P\vee Q}-s^{P\wedge Q}\geq s^P+s^Q-2s^{P\wedge Q}.$$

If the two partitions are comparable (say $P\geqslant Q$), then equality holds, but the converse is not true: there exist incomparable partitions P,Q, that is $P\ngeqslant Q\ngeqslant P$, where equality holds as well. For example, $N=\{1,2,3,4\}$ and $P=[12],Q=[34]$. Also, the left hand side is the *size-based* distance [28], classifiable as a *modular* partition distance measure.

Turning to the indicator function, there are two distinct ways of using it for the current computational needs. Let $\mathbf{1}$ denote the $\binom{n}{2}$-vector each of whose entries equals 1, as before. Then,

$$\begin{aligned}HD(P,Q)&=\sum_{[ij]\in\mathcal{P}_{(1)}^N}\big(I_P([ij])-I_Q([ij])\big)^2=\\&=\langle(I_P-I_Q)^2,\mathbf{1}\rangle \quad (11)\\&=\langle I_P,\mathbf{1}\rangle+\langle I_Q,\mathbf{1}\rangle-2\langle I_P,I_Q\rangle, \quad (12)\end{aligned}$$

where $\langle I_P,I_Q\rangle=\langle I_{P\wedge Q},\mathbf{1}\rangle$ (see expression (9) above), hence this is the analogue of expression (5) above.

From another perspective, for all $[ij]\in\mathcal{P}_{(1)}^N$ define

$$\begin{aligned}I_{P,Q}^{\max}([ij])&=\max\{I_P([ij]),I_Q([ij])\},\\I_{P,Q}^{\min}([ij])&=\min\{I_P([ij]),I_Q([ij])\}.\end{aligned}$$

Then, in terms of the indicator function, expression (8) yields

$$HD(P,Q)=\langle I_{P,Q}^{\max},\mathbf{1}\rangle-\langle I_P,I_Q\rangle=\langle I_{P,Q}^{\max}-I_{P,Q}^{\min},\mathbf{1}\rangle=$$
$$=\langle I_P+I_Q \text{ modulo } 2,\mathbf{1}\rangle.$$

[2] The length $\hat{r}$ is the same for both the two segments.

Remark 4: While $I^{\min}_{P,Q} = I_{P \wedge Q}$, in general both $\langle I^{\max}_{P,Q}, \mathbf{1}\rangle$ and $\langle I_P + I_Q \text{ modulo } 2\ , \mathbf{1}\rangle$ may well fail to be available sizes (see section 3 above), as

$$s^{P \vee Q} = \langle I_{P \vee Q}, \mathbf{1}\rangle \geq \langle I^{\max}_{P,Q}, \mathbf{1}\rangle =$$

$$= |\{[ij] \in \mathcal{P}^N_{(1)} : P \geqslant [ij] \text{ OR } Q \geqslant [ij]\}|.$$

Attention is now placed on comparing HD with both MMD and VI described above. While VI displays a precise axiomatic characterization [22], its applicative possibilities and computational complexity are not yet extensively investigated [12]. Conversely, MMD is widely used in bioinformatics applications and tightly bounded in terms of required computations [15], [19], [18].

V. HD and VI: axioms

PARTITION distance measures HD and VI display very similar axiomatic behaviors [22], as detailed hereafter.

Proposition 5: HD is a *metric*: for all $P, P', Q \in \mathcal{P}^N$,

1) $HD(P, Q) \geq 0$, with equality if and only if $P = Q$,
2) $HD(P, Q) = HD(Q, P)$,
3) $HD(P, P') + HD(P', Q) \geq HD(P, Q)$.

Proof: The second condition is obvious. The first one is also immediate as $\min\{s^P, s^Q\} \geq s^{P \wedge Q}$. Concerning the third one, known as triangle inequality,

$$HD(P, P') + HD(P', Q) - HD(P, Q) =$$

$= 2(s^{P'} - s^{P \wedge P'} - s^{P' \wedge Q} + s^{P \wedge Q})$. It must be shown that this is positive for all $P, P', Q \in \mathcal{P}^N$. For any such a triplet, the quantity is minimized when $s^{P \wedge P'} + s^{P' \wedge Q}$ is maximized, which in turn occurs for $P' \leqslant P, Q$ or $P \wedge P' = P' = P' \wedge Q$. This entails $s^{P'} = s^{P \wedge P'} = s^{P' \wedge Q} \leq s^{P \wedge Q}$. ■

Triangle inequality is satisfied with equality by both HD and VI when $P' = P \wedge Q$.

Proposition 6: HD satisfies *horizontal collinearity*:

$$HD(P, P \wedge Q) + HD(P \wedge Q, Q) = HD(P, Q)$$

for all $P, Q \in \mathcal{P}^N$.

Proof: Observe that summing $HD(P, P \wedge Q) = s^P - s^{P \wedge Q}$ and $HD(P \wedge Q, Q) = s^Q - s^{P \wedge Q}$ yields $HD(P, Q)$. ■

A further axiom characterizing both HD and VI is expressed in terms of those two partitions $P_\perp = \{\{1\}, \ldots, \{n\}\}$ and $P^\top = \{\{N\}\}$ occupying the bottom and top levels of the Hasse diagram, with $r(P^\top) = n - 1$.

Proposition 7: HD satisfies *vertical collinearity*:

$$HD(P_\perp, P) + HD(P, P^\top) = HD(P_\perp, P^\top) \text{ for all } P \in \mathcal{P}^N.$$

Proof: Sum $HD(P_\perp, P) + HD(P, P^\top) = s^P + s^{P^\top} - s^P$ yields $s^{P^\top} = \binom{n}{2} = HD(P_\perp, P^\top)$. ■

Two partitions P, Q are each a complement of the other if $P \wedge Q = P_\perp$ and $P \vee Q = P^\top$. While every subset $A \in 2^N$ admits a unique complement $A^c = N \backslash A$, a partition generally admits several complements. In particular, the Hamming distance between subsets given by expression (1) above yields $d(A, A^c) = n = d(\emptyset, N)$. That is, the distance between any two complements is the same as the distance between the bottom and top elements. Analogously, a partition distance measure $\delta : \mathcal{P}^N \times \mathcal{P}^N \to \mathbb{R}_+$ satisfies such a *complement maximality* condition [28] if for any two complements P', Q', it holds $\delta(P', Q') = \max\limits_{(P,Q) \in \mathcal{P}^N \times \mathcal{P}^N} \delta(P, Q) = \delta(P_\perp, P^\top)$.

For generic $P = \{A_1, \ldots, A_{|P|}\}$, consider P_* with size $s^{P_*} = |P| - 1$ and maximal representation as a join of atoms (see above) $P_* = [ij]_1 \vee \cdots \vee [ij]_{|P|-1}$. Also, for $1 \leq k < |P|$, let $P \vee [ij]_k = \{A_1, \ldots, A_{k-1}, A_k \cup A_{k+1}, A_{k+2}, \ldots, A_{|P|}\}$ (assume $|A| \geq 2$ for all $A \in P$). In words, P_* is the join of $|P| - 1$ pair wise disjoint atoms each merging two distinct blocks of P. Such a P_* is a complement of P. A further complement of P is $P^* = B \cup P_\perp^{B^c}$ with $|B \cap A| = 1$ for every block $A \in P$. This partition P^* has $n - |P| + 1$ blocks: one is B containing $|P|$ elements, while all remaining ones are singletons. A sensible partition distance measure should distinguish between such complements. That is, it should *not* satisfy complement maximality. Indeed, this is the behavior displayed by all distances MMD, VI and HD, and is easily checked by means of an example: the partitioned set is $N = \{1, \ldots, 7\}$ while partitions are $P = 123|456|7$ and $P_* = 1|2|34|5|67$ as well as $P^* = 147|2|3|5|6$, with $|$ separating blocks. Then, $VI(P, P_*) = \frac{6}{7}\log 6 - \frac{2}{7} < \frac{4}{7}\log 9 + \frac{2}{7}\log 3 - \frac{1}{7} = VI(P, P^*)$ as well as $HD(P, P_*) = 8 < 9 = HD(P, P^*)$. On the other hand, $MMD(P, P_*) = 4 > 2 = MMD(P, P^*)$.

VI. HD and MMD: computations and examples

A Computational procedure determining HD accepts as inputs two $\binom{n}{2}$-dimensional and $\{0, 1\}$-valued arrays as inputs. These are the indicator functions I_P, I_Q of the two partitions P, Q between which the distance is to be computed. In fact, the indicator function alone constitutes a novel, neat and mostly natural way of coding partitions (while how to code partitions as inputs when computing MMD is not straightforward [15], [19], [18]). Most importantly, HD does not rise any algorithmic issue. In fact, consider expression (12) above. The number of elementary operations (or running time [20]) needed to compute HD is easily counted. Each of the three scalar products requires $\binom{n}{2}$ multiplications (between simple numbers, namely 0s or 1s) and $\binom{n}{2} - 1$ additions (between positive integers never exceeding $\binom{n}{2}$ itself). Next, another multiplication and two further additions provide total $3\left(2\binom{n}{2} - 1\right) + 3 = 3n(n-1)$. This is evidently better than the best running time $O(n^3)$ when computing MMD in bioinformatics [18].

Beside computations, HD and MMD can also be compared through examples showing the different sensitivity of their normalized versions (see below). This seems achievable even for the small value $n = 7$ applying to Table 2 below. Consider that the partition lattice has n levels. Maintaining the notation used thus far for level 1 (above level 0), populated by $\binom{n}{2}$ atoms, let $\mathcal{P}^N_{(n-k)}$ denote level $n - k$, for $k = 1, \ldots, n$. Each level has a number of elements given by the *Stirling numbers of the second kind* $\mathcal{S}_{n,k}$ [1], [14]. For $k = 1, \ldots, n$,

$$\mathcal{S}_{n,k} = \frac{1}{k!} \sum_{0 \leq h \leq n} (-1)^{k-h} \binom{k}{h} h^n = |\mathcal{P}^N_{(n-k)}|,$$

TABLE II
Comparing HD and MMD, normalized values in parentheses.

#	P	Q	$HD(P,Q)$	$MMD(P,Q)$
1	12|3|4|5|6|7	1|2|3|4|5|67	2 (0.095)	2 (0.333)
2	12|3|4|5|67	1|2|345|6|7	5 (0.238)	4 (0.667)
3	123|4|5|6|7	1|2|3|4|567	6 (0.286)	4 (0.667)
4	1234|5|6|7	1|2|3|4|567	9 (0.429)	5 (0.833)
5	1234|5|6|7	1235|4|6|7	6 (0.286)	2 (0.333)
6	123|45|67	17|25|36|4	8 (0.381)	4 (0.667)
7	123|4567	1456|237	10 (0.472)	5 (0.833)
8	123|45|67	1|2|34567	11 (0.524)	4 (0.667)
9	123|45673	1234567	8 (0.381)	5 (0.833)
10	123|45|67	123|4567	4 (0.190)	3 (0.5)
11	123|4567	167|2345	12 (0.571)	4 (0.667)
12	12|34|56|7	1|23|45|67	6 (0.286)	3 (0.5)
13	123|45|6|7	167|2|34|5	8 (0.381)	5 (0.833)
14	1|2|3|4|5|6|7	1234567	21 (1)	6 (1)
15	12|34|56|7	17|23|45|6	6 (0.286)	4 (0.667)
16	123|4|56|7	1|72|345|6	8 (0.381)	3 (0.5)
17	1|234567	12|3|4|5|6|7	16 (0.762)	5 (0.833)
18	1234567	12|3|4|5|67	19 (0.905)	5 (0.833)
19	1234567	12|3|4|5|6|7	20 (0.952)	5 (0.833)
20	1234567	1|234567	6 (0.286)	1 (0.167)

where $h > k \Rightarrow \binom{k}{h} = 0$. These numbers are known to be such that most populated levels are central ones [1, Proposition 3.30 p. 91]. Also, upper levels are comparatively more populated than lower ones. For $n = 7$,

$$\begin{aligned}
\mathcal{S}_{7,7} &= 1 = \mathcal{S}_{7,1}, \\
\mathcal{S}_{7,6} &= 21 = \binom{7}{2}, \\
\mathcal{S}_{7,5} &= 140, \\
\mathcal{S}_{7,4} &= 350, \\
\mathcal{S}_{7,3} &= 301, \\
\mathcal{S}_{7,2} &= 63.
\end{aligned}$$

Examples 1-20 in Table II above compare measurement sensitivity between MMD and HD with partitions picked initially from less populated levels, and subsequently from more populated ones.

It seems worth noting again that while MMD takes all n integer values $0, 1, \ldots, n-1$ in its interval $[0, n-1]$, neither VI nor HD have a range displaying the same even distribution over associated intervals $[0, \log n]$ and $[0, \binom{n}{2}]$, respectively. Concerning HD, the size of partitions does not take all values $0, 1, \ldots, \binom{n}{2}$, as explained in section 1.3. In particular, some parts of the interval are more densely populated by range elements than others. This is another consequence of linear dependence, characterizing all (non-corrected) partition distance measures based on counting pairs; see [22, section 2.1]. Most importantly, the number of range elements is greater with HD (and VI) than with MMD, i.e. $|rg_n(HD)| > |rg_n(MMD)|$ as soon as $n > 3$, where $rg_n(MMD) = \{0, 1, \ldots, n-1\}$ and $rg_n(HD) \subseteq \{0, 1, \ldots, \binom{n}{2}\}$ denote the range of MMD and HD, respectively, for any given n. These are sets of integer numbers, and a larger range provides greater measurement sensitivity. Finally, any distance measure may be normalized so to range in the unit interval $[0, 1]$. That is, $\frac{MMD(P,Q)}{n-1}, \frac{VI(P,Q)}{\log n}, \frac{HD(P,Q)}{\binom{n}{2}} \in [0, 1]$ for all $P, Q \in \mathcal{P}^N$. These normalized values are within parentheses in Table II.

VII. Conclusion

In statistical classification, measuring the distance between partitions is an important combinatorial problem, attracting attention since the '60s [26], [21], [11]. Solution HD provided here parallels the (traditional) Hamming distance between subsets by counting atoms in the corresponding lattices. From an applicative viewpoint, HD is added to MMD and VI as tools for clustering comparison in bioinformatics. Easily computed, HD has fine measurement sensitivity and displays suitable axiomatic features.

References

[1] M. Aigner. *Combinatorial Theory*. Springer, 1997. Reprint of the 1979 edition.

[2] A. Almudevar. A commentary on some recent methods in sibling group reconstruction based on set coverings. *Optimization Methods and Software*, 26(6):993–1003, 2011.

[3] A. Almudevar and E. C. Anderson. A new version of PRT software for sibling groups reconstruction with comments regarding several issues in the sibling reconstruction problem. *Molecular Ecology Resources*, 12(1):164–178, 2012.

[4] A. Almudevar and C. Field. Estimation of single-generation sibling relationships based on DNA markers. *Journal of Agricultural, Biological and Environmental Statistics*, 4(2):136–165, 1999.

[5] M. V. Ashley, I. C. Caballero, W. A. Chaovalitwongse, B. DasGupta, P. Govindan, S. I. Sheikh, and T. Y. Berger-Wolf. KINALYZER, a computer program for reconstructing sibling groups. *Molecular Ecology Resources*, 9(4):1127–1131, 2009.

[6] T. Y. Berger-Wolf, S. I. Sheikh, B. DasGupta, M. V. Ashley, I. C. Caballero, W. Chaovalitwongse, and S. L. Putrevu. Reconstructing sibling relationship in wild populations. *Bioinformatics*, 23(13):i49–i56, 2007.

[7] B. Bollobas. *Combinatorics. Set Systems, Hypergraphs, Families of Vectors, and Combinatorial Probability*. Cambridge University Press, 1986.

[8] D. G. Brown and T. Y. Berger-Wolf. Discovering kinship through small subsets. *Algorithms in Bioinformatics*, LNCS 6293:111–123, 2010.

[9] D. G. Brown and D. Dexter. Sibjoin: a fast heuristic for half-sibling reconstruction. *Algorithms in Bioinformatics*, LNCS 7534:44–56, 2012.

[10] W. A. Chaovalitwongse, T. Y. Berger-Wolf, B. DasGupta, and M. V. Ashley. Set covering approach for reconstruction of sibling relationships. *Optimization Methods and Software*, 22(1):11–24, 2007.

[11] W. H. E. Day. The complexity of computing metric distances between partitions. *Mathematical Social Sciences*, 1(3):269–287, 1981.

[12] D. Dexter. *Reconstruction of half-sibling population structures*. Master thesis, Computer Science, Waterloo, 2012.

[13] M. M. Deza and E. Deza. *Encyclopedia of Distances - Second Edition*. Springer, 2013.

[14] R. Graham, D. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley, 1994.

[15] D. Gusfield. Partition-distance: A problem and class of perfect graphs arising in clustering. *Information Processing Letters*, 82:159–164, 2002.

[16] L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.

[17] D. A. Konovalov. Accuracy of four heuristics for the full sibship reconstruction problem in the presence of genotype errors. *Series on Advances in Bioinformatics and Computational Biology*, 3:7–16, 2006.

[18] D. A. Konovalov, N. Bajema, and B. Litow. Modified Simpson $O(n^3)$ algorithm for the full sibship reconstruction problem. *Bioinformatics*, 21(20):3912–3917, 2005.

[19] D. A. Konovalov, B. Litow, and N. Bajema. Partition-distance via the assignment problem. *Bioinformatics*, 21(10):2463–2468, 2005.

[20] B. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms (2nd edition)*. Springer, 2002.

[21] I. C. Lerman. *Classification et Analyse Ordinale des Données*. Dunod, 1981.

[22] M. Meila. Comparing clusterings - an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, 2007.

[23] B. G. Mirkin. *Mathematical Classification and Clustering*. Kluwer Academic Press, 1996.

[24] B. G. Mirkin and L. B. Cherny. Measurement of the distance between distinct partitions of a finite set of objects. *Automation and Remote Control*, 31(5):786–792, 1970.

[25] B. G. Mirkin and I. Muchnik. Some topics of current interest in clustering: Russian approaches 1960-1985. *Electronic Journal for History of Probability and Statistics*, 4(2):1–12, 2008.

[26] S. Rénier. Sur quelques aspects mathématiques des problémes de classification automatique. *ICC Bulletin*, 4:175–191, 1965. Reprinted in Mathématiques et Sciences Humaines 82:13-29, 1983.

[27] G. Rossi. Information functions and expectation. In *RUD Proceedings*, 2004. www.kellogg.northwestern.edu/research/risk/rud/risk_papers.htm.

[28] G. Rossi. *Partition distances*. arXiv:1106.4579v1, 2011.

[29] G.-C. Rota. The number of partitions of a set. *American Mathematical Monthly*, 71:499–504, 1964.

[30] S. I. Sheikh, T. Y. Berger-Wolf, M. V. Ashley, I. C. Caballero, W. Chaovalitwongse, and B. DasGupta. Error-tolerant sibship reconstruction in wild populations. *Computational Systems Bioinformatics*, 7:273–284, 2008.

[31] S. I. Sheikh, T. Y. Berger-Wolf, A. A. Khokhar, I. C. Caballero, M. V. Ashley, W. Chaovalitwongse, C.-A. Chou, and B. DasGupta. Combinatorial reconstruction of half-sibling groups from microsatellite data. *Journal of Bioinformatics and Computational Biology*, 8(2):337–356, 2010.

[32] M. Stern. *Semimodular Lattices. Theory and Applications. Encyclopedia of Mathematics and its Applications 73*. Cambridge University Press, 1999.

[33] M. J. Warrens. On the equivalence of Chen's Kappa and the Hubert-Arabie adjusted Rand index. *Journal of Classification*, 25(1):177–183, 2008.

[34] H. Whitney. On the abstract properties of linear dependence. *American Journal of Mathematics*, 57:509–533, 1935.

Comparison of ACO and GA Techniques to Generate Neural Network Based Bezier-PARSEC Parameterized Airfoil

Waqas Saleem[1], Riaz Ahmad[2], Athar Kharal[3], Ayman Saleem[4]

***Abstract*—** This research uses Neural Networks to determine two dimensional airfoil geometry using Bezier-PARSEC parameterization. Earlier, Ant Colony Optimization (ACO) techniques have been used to solve combinatorial optimization problems like TSP. This work extends ACO method from TSP problem to design parameters for estimating unknown Bezier-PARSEC parameters that define upper and lower curves of the airfoil. The efficiency and the performance of ACO technique was compared to that of GA. The work established that ACO exhibited improved performance than the GA in terms of optimization time and level of precision achieved. In the next phase, Neural Network is implemented using Cp as input in terms of C_l, C_d and C_m for learning and targeting the corresponding Bezier-PARSEC parameters. Neural Networks including Feed-forward back propagation, Generalized Regression and Radial Basis were implemented and were compared to evaluate their performance. Similar to earlier work with GA and Neural Nets, this work also established Feed-forward back propagation Neural Network as a preferred method for determining the design of airfoil since the technique presented better approximation results than other neural nets.

***Keywords*—** Airfoil Optimization, Ant Colony Optimization, Bezier-PARSEC, C_p, Neural Network

I. INTRODUCTION

Airfoil design is one of the most challenging processes [1] in development of aircraft aerodynamic surfaces as it affects various aircraft performance parameters like lift, drag, spin-stall, cruise and turning radius [2]. Studies indicate that selecting the right design of airfoil with required characteristics reduces overall cost and improves the performance of air vehicle. Airfoil design largely depends on desire for high lift to drag ratio that is in conflict with the performance requirements [3].

There are two major techniques for designing an airfoil; direct and inverse [4]. First method involves designing a new or modifying an existing airfoil (UIUC Airfoil Database [5] and computing pressure distribution across the surface to achieve desired set of parameters. This approach may limit the approximation for desired specifications due to inherent limitations in airfoil's aerodynamics. For faster approximations, reduced degrees of freedoms are required but such reduction results in computational errors like round off, truncation and discretization error. In fact, determining the airfoil geometry should be based on requirements for aircraft's performance. Thus later method involves using desired operational characteristics and performance parameters unless the airfoil geometry so generated meets the desired criteria. To reduce the computational time and meet the required design criteria various techniques including CFD, fuzzy logic, neural networks [6] and heuristics based algorithms like PSO [7] and GA [8] have been implemented to advantage the aerodynamic design process.

This research, largely inspired by Saleem and Kharal [9], uses neural network based approach for airfoil generation exploiting Bezier-PARSEC 3434 parameterization rather than full coordinates for a given Cp. However, this research implements ACO to optimize Bezier-PARSEC unknown parameters instead of GA as in earlier work.

II. ARITIFICIAL NEURAL NETWORK

In machine learning and data mining, Artificial Neural Network is a set of learning algorithms modeled after neural network structure of the cerebral cortex and is used to approximate functions involving a larger number of the unknown input variables [10] Each neuron receives input from external sources or neighbors in the network, computes output and propagates to other neurons. Another function is the weight adjustments in the connections between neurons. Incremental learning is the technique by gathering information on cumulative error and consequently adjusting weight coefficients, w_{ij}. Mathematically, a Neural Network can be defined as a triple (N, C, w) where N is the set of neurons, C $\{(i, j)|i, j \in N\}$ is a set of connections, and function w((i, j)), shortened as w_{ij} is called weights between neurons i and j. For every neuron, there is an external input ϑ_j and an activation function F_j to establish the new activation level based on effective input of a neuron S_j and is determined by following propagation rule in "(1)".

$$S_j = \sum_i w_{ij}(t)\, y_i(t) + \theta_j(t) \tag{1}$$

W. Saleem is with School of Mechanical and Manufacturing Engineering National University of Sciences, Pakistan (Phone: +923224362442, e-mail: waqas_jeral@hotmail.com

R. Ahmad is currently director research at CIE Building, Research Directorate National University of Sciences and Technology, Pakistan (e-mail: dresearch@nust.edu.pk)

A. Kharal is currently Associate Professor at College of Aeronautical Engineering National University of Sciences and Technology, Pakistan (atharkharal@gmail.com)

A. Saleem is with College of Aeronautical Engineering National University of Sciences and Technology (aimen1173@hotmail.com)

Besides, a threshold is also introduced as linear, non-linear or sigmoidal function [11] that helps avoid the situation when training is not successful at ||σ||>0. A threshold function for each neuron is given by "(2)"

$$F_{S_j} = \frac{1}{1 + e^{-S_j}} \tag{2}$$

A. Feed-forward Back propagation Neural Network

A feed forward Back Propagation Neural Network (FFBP) contains a multi layered interconnected feed forward structure where every layer gets input from below and gives output to layer above it. Back propagation is a learning technique where output values are compared to a desired value to calculate the error using a pre-determined error function. This value of error is then fed back through the network repeatedly for minimizing through neural network algorithm by adjusting weights for each network connection until the network converges to a bare minimum acceptable level of error [12] Generally, a non-linear optimization method called gradient descent is implemented where derivative of the error function is determined w.r.t. weights, that are adjusted till the reduction of error.

B. Radial Basis Function Neural Network

A Radial Basis Function Neural Network (RBF) consists of an input layer, a hidden layer with non-linear Radial Basis activation function and an output layer. For Radial Basis Neural Network, the input is modeled as vector of real numbers (R^n) while output is a scalar function φ, given in "(3)" by [13]

$$\varphi(x) = \sum_{i=1}^{n} a_i p(||x - c_i||) \tag{3}$$

where n is number of neurons, a_i is weight of neuron and c_i is center vector.

In Radial Basis Neural Networks, neurons respond to inputs close to their center in contrast with other neural networks. Although Radial Basis Neural Network requires more neurons for high dimensional input spaces, it can be trained faster than standard multi layered neural networks and have proven efficiency in regression and classification problems.

C. Generalized Regression Neural Network

A Generalized Regression Neural Network (GRNN) consists of one each input layer, pattern layer, summation layer and output layer. Training patterns are presented by neurons in pattern layer. In GRNN, pattern layer is connected to summation layer. Sum of weighted responses and un-weighted responses of pattern neurons are computed by two neurons in summation layers [9] The summation layer consists of both summation and single division units. Normalization of output is performed together both by summation and output layers. GRNN exhibit single pass learning algorithm with high parallel structure for estimating continuous variables and do not require iterative process as in multi-layered networks. GRNN converges to optimal regression even in noisy environments given a large number of sample data is available. Generalized Regression Neural Network is particularly advantageous with sparse data but as the training data increase, the error converges to zero.

III. PARSEC PARAMETERIZATION & BEZIER CURVES

PARSEC parameterization has the capability to describe the airfoil shape and its flow using engineering parameters [10] On the other hand, a Bezier curve is a parametric curve of degree n defined by polygon of n+1 vertex points called control points of nth order Bezier curve and is given by "(4)"

$$P(t) = \sum_{k=0}^{n} P_k \binom{n}{k} t^k (1-t)^{n-k} \tag{4}$$

where P_k is the kth control point while parameter t ranges from 0 to 1 with 0 at the zeroth control point and 1 at the nth control point. Eq. (5) gives Third order Bezier Curve

$$\begin{cases} x = x_a (1-t)^3 + 3x_b(1-t)^2 t + 3x_c(1-t)t^2 + x_d t^3 \\ y = y_a (1-t)^3 + 3y_b(1-t)^2 t + 3y(1-t)t^2 + y_d t^3 \end{cases} \tag{5}$$

Eq. (6) present fourth order Bezier Curve

$$\begin{cases} x = x_a (1-t)^4 + 4x_b(1-t)^3 t + 6x_c(1-t)^2 t^2 + 4x_d(1-t)t^3 \\ y = y_a (1-t)^4 + 4y_b(1-t)^3 t + 6y_c(1-t)^2 t^2 + 4y_d(1-t)t^3 \end{cases} \tag{6}$$

IV. BEZIER-PARSEC PARAMETERIZATION

Bezier-PARSEC parameterization is a technique in which Bezier Curves are described using PARSEC parameterizations [14] and is further subdivided into BP3333 and BP3434.

A. BP3333 Parameterization

BP3333 Parameterization employs third order Bezier Curves for camber shape and thickness of airfoil [15] Twelve PARSEC parameters represent Bezier control points as shown in Fig 1.

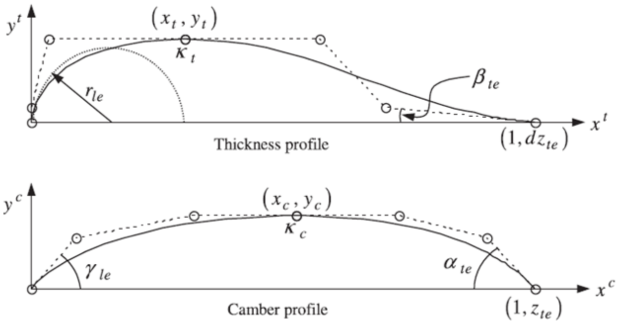


Fig 1 : BP3333 Bezier PARSEC Control Points and Respective Airfoil Geometry

Main advantages of BP3333 include close relevance to airfoil aerodynamics parameters, faster optimization, continuity characteristics, reduced deviation of design process and avoidance of sharp leading edges. Disadvantage of this technique is reduced degree of freedom resulting in failure to parameterize airfoils having radical camber trailing edge

B. BP3434 Parameterization

BP3434 Parameterization depends on 10 PARSEC parameters and 5 Bezier parameters for airfoil shape representation. Here, camber and thickness leading edge of airfoil is defined by third order Bezier Curves while fourth order Bezier Curves are used to define camber and thickness trailing edge of airfoil shape [15] This allows increased degree of freedom for trailing edge parameterization of airfoil as shown in Fig 2.

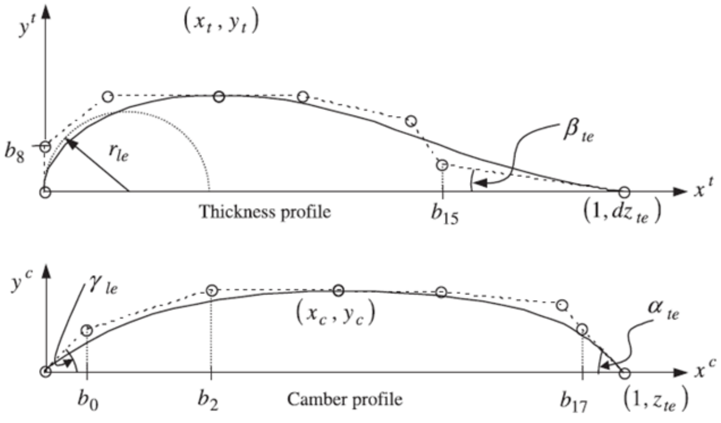


Fig 2 : BP3434 Bezier PARSEC Control Points and Respective Airfoil Geometry

BP3434 proves to be efficient than BP3333 when airfoil camber becomes negative along any part of the chord length. However, the convergence speed for this method reduces due to greater number of variables as compared to BP3333. In presence of high computing numerical computers, the convergence speed of BP3434 can be compensated for effective application of the method.

V. ANT COLONY OPTIMIZATION

Ant Colony Optimization (ACO) is the meta-heuristic motivated from the working of natural ants that suggests that ants follow different paths to reach food source initially. Thus the ants with shortest path would reach the source in least time than the longer paths [16] Ant Colony Optimization is an algorithm where artificial ants are used to probabilistically construct solutions guided by higher pheromone trails and promising heuristic information [17] In actual, ants implement a randomized construction heuristics that differ from greedy heuristics by adding a probabilistic component to partial solution than a deterministic one. Generally, ACO algorithm consists of two phases. In first phase, artificial ants construct a solution where in second phase, pheromone trail is updated by first reducing by an evaporation factor to avoid unlimited accumulation followed by adding pheromone proportionate to quality of their solutions [18]. Thus most important is to update pheromone for generating quality solutions in future iterations of algorithm. ACO algorithms can be considered as competitive solution technique where previous solutions known to be part of good solutions are used to generate even better solutions in future cycles [19].

VI. METHODOLOGY

In this research work, our methodology was quite similar to earlier work; however, ACO was preferred as a choice for optimization technique instead of GA to determine unknown Bezier-PARSEC parameters.

A. *Airfoil Representation*

A vector of 71 points is used to represent x-y coordinates of an airfoil where x_i ranges from 1 to 0 for upper airfoil curve and lower airfoil curve, thus only values for y change which determine the shape of both curves.

Mean Camber Line is a line at equal distance from both upper and lower surfaces of airfoil. Therefore, camber curve y points were obtained by taking average of upper and lower coordinates corresponding to the same x coordinate. These upper and lower coordinates were divided by chord length for non-dimensionalizing. The camber profile of an airfoil is calculated by "(7)", "(8)", "(9)" and "(10)"

c-|x$_1$-x$_{36}$| for i=1 to 36 and j=36 to 71

$$y_i^u = \frac{y_i}{c} \text{ and } \qquad y_j^l = \frac{y_j}{c} \qquad (7)\ \&\ (8)$$

$$x_i^c = \frac{x_i}{c} \qquad \text{and} \qquad y_i^c = \frac{y_i^u + y_j^l}{2} \qquad (9)\ \&\ (10)$$

Thickness curve used to define the airfoil thickness is the difference between the camber curve and upper curve of the airfoil i.e.

$$y_i^t = y_i^u - y_i^c \qquad (11)$$

Next a two dimensional analysis of airfoil was carried out using Panel Method to obtain values for lift coefficient C_l quarter-chord pitching moment coefficient C_m and drag coefficient C_d at ten angles of attack α. Thus the airfoil would be represented by x_i^c, y_i^c, C_l, C_d, C_m and α.

B. *Calculating Bezier-PARSEC Parameterization*

Table 1 presents the required parameters for Bezier-PARSEC

Table I : Known Bezier-PARSEC Parameters

Parameters	**Caculations**
Maximum Thickness Point	$y_t = C^t\left(\left\{x_t \left\lvert \frac{dC^t}{dx_i}\right\rvert_{x_i=x_t} = 0\right\}\right)$
Maximum Camber Point	$y_c = C^c\left(\left\{x_c \left\lvert \frac{dC^c}{dx_i}\right\rvert_{x_i=x_c} = 0\right\}\right)$
Trailing Edge Vertical Displacement	$Z_{te} = C^c(x)\lvert_{x=1}$ and $dZ_{te} = C^t(x)\lvert_{x=1}$
Trailing Camber Line Angle	$\alpha_{te} = -tan^{-1}(\frac{dC^c}{dx}\lvert_{x=1})$
Trailing Wedge Angle	$\beta_{te} = -tan^{-1}(\frac{dC^t}{dx}\lvert_{x=1})$
Leading Edge Direction	$\beta_{te} = -tan^{-1}(\frac{dC^t}{dx}\lvert_{x=1})$
Leading Edge Radius	r_{le}

While ten parameters are calculated using Bezier-PARSEC equations, there is no specific mathematical expression for finding remaining five parameters i.e., b_0, b_2, b_8, b_{15} and b_{17} and therefore are calculated by curve fitting. Since actual airfoil is known, Bezier Curves with correct five control points would suffice given a smallest Sum-of-Least-Square Error.

Table II shows the four curves and corresponding unknown Bezier points.

Table II : Unknown Bezier-PARSEC Parameters

Curve	Bezier Curve	Order	Unknown Bezier Control Points
Camber	Leading Edge	3rd	b_0, b_2
	Trailing Edge	4th	b_{17}
Thickness	Leading Edge	3rd	b_8
	Trailing Edge	4th	b_8, b_{15}

C. *Optimization of Unknown Bezier Control Points Using ACO*

To determine optimal value of these unknown parameters, Ant Colony Optimization was implemented requiring fitness functions for each Bezier Curve that was equal to the difference between Bezier generated and actual airfoil. For this a Simple ACO code was written to determine each of these parameters i.e., b_0, b_2, b_8, b_{15} and $b_{17.}$ In ACO, 6 ants were used to determine the optimal path to the destination and since the destination point was unknown; therefore, SSE for each curve was calculated for each generated point. Thus, a decrease in SSE over the path indicates that the ant is close to the destination point and vice versa. The pheromone is inversely proportional to the distance so the path with least distance or least SSE would have maximum pheromone. For each value of b_0, a corresponding value of b_2 is calculated through ACO. Thus a number of combinations (pair of b_0 and b_2 values) are made where pair with the least SSE is finally chosen. Same approach was used for b_8 and b_{15} while value of b_{17} was calculated separately. Fig 3 present flow charts for the method used.

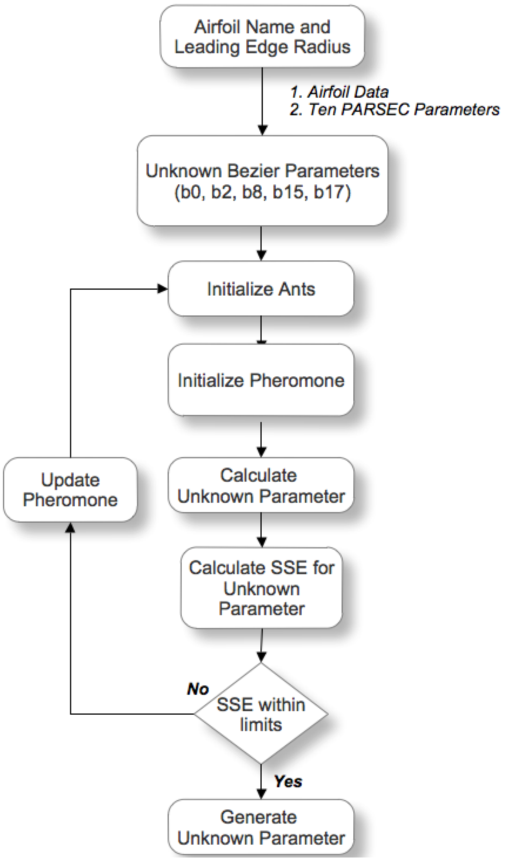

Airfoil Name and Leading Edge Radius
1. Airfoil Data
2. Ten PARSEC Parameters
Unknown Bezier Parameters (b0, b2, b8, b15, b17)
Initialize Ants
Initialize Pheromone
Calculate Unknown Parameter
Calculate SSE for Unknown Parameter
SSE within limits
No
Yes
Update Pheromone
Generate Unknown Parameter

Fig 3 : Ant Colony Optimization Methodology

D. *Error Calculation*

All 15 BP3434 parameters determined are used for generation of airfoil geometry. The error is calculated by comparing Bezier generated airfoil with actual airfoil. To calculate this error, at a certain x-value, y-value from parameterized and actual airfoils should relate to this x-value. The main challenge was to determine y-values of Bezier parameterized airfoil corresponding to these x-values. After generating x and y values of trailing and leading edge of thickness curve, these are arranged into a single set of x-y array in which first element corresponds to leading edge followed by trailing edge. Then cubic spline interpolation is used to fit a curve in the vector of x and y values which is then evaluated for 36 x-values of actual airfoil. Same procedure was followed for camber curve. These thickness and camber curves can be used to determine the shape of airfoil. The airfoil geometries of parameterized and actual airfoils are then plotted against same axis for comparison.

Fig 4 shows flow chart for SSE calculations while Fig 5 presents results for Eppler 433 sailplane parameterized airfoil.

E. *Neural Networks Estimations*

Neural Networks of three types as discussed in Section 2 were implemented in this research work. A 10X4 matrix of Cl, Cd, Cm at ten angles of attack for 500 heterogeneous airfoils was input to neural network while target was 15 Bezier-PARSEC parameters for airfoil generation.

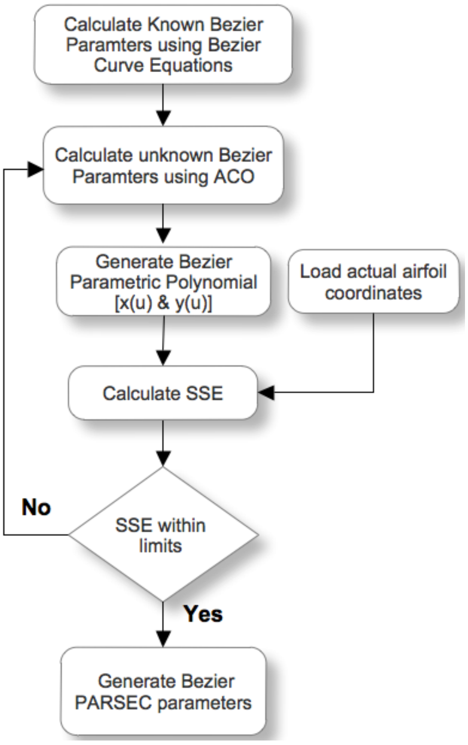

Calculate Known Bezier Paramters using Bezier Curve Equations
Calculate unknown Bezier Paramters using ACO
Generate Bezier Parametric Polynomial [x(u) & y(u)]
Load actual airfoil coordinates
Calculate SSE
SSE within limits
No
Yes
Generate Bezier PARSEC parameters

Fig 4 : SSE Calculations for Bezier-PARSEC Parameters

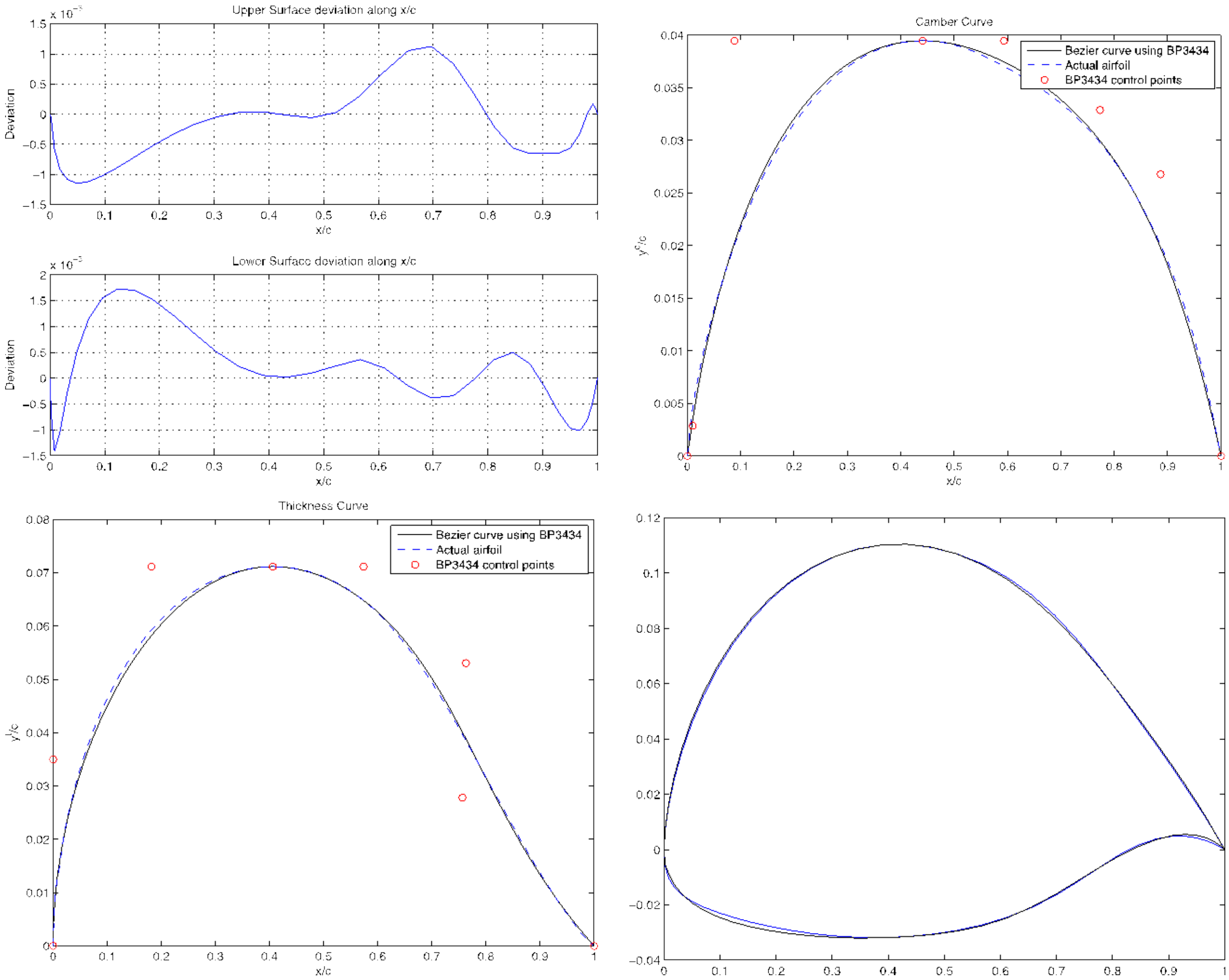


Fig 5 : Bezier Parameterization Results for Eppler 433 Sailplane Airfoil

VII. RESULTS AND DISCUSSIONS

A. Comparison of ACO Results with GA

Implementation of ACO for finding unknown Bezier Curve parameters proved to be more efficient than Genetic Algorithm. We were able to achieve a precision level of $\leq 1X10^{-5}$ as compared to GA based version of the program. Also time to optimize the missing BP3434 parameters was greatly reduced. For example, Eppler 433 Sailplane airfoil took 30.905144 seconds to optimize BP3434 missing parameters using ACO as compared to GA that took 87.869966 seconds for optimization of said airfoil using 2.7GHz Processor and 4GB RAM. Table III gives a comparison of ACO and GA optimizations for few airfoils for reference.

Table III : Comparison of ACO and GA Optimization Results

Airfoil	Ant Colony Optimization		Genetic Algorithm	
	Time (Seconds)	**Level**	**Time (Seconds)**	**Level**
Eppler E433	30.905144	$\leq 1X10^{-5}$	87.869966	$\leq 1X10^{-4}$
NACA 65(4)-421	55.187357	$\leq 1X10^{-5}$	90.952194	$>1X10^{-4}$
Eppler E335	65.389595	$\leq 1X10^{-5}$	109.694796	$>1X10^{-4}$
Gottingen GOE426	44.489090	$\leq 1X10^{-5}$	82.259980	$\leq 1X10^{-4}$
Eppler E399	55.089536	$\leq 1X10^{-5}$	94.445729	$\leq 1X10^{-4}$

From Table III, we see that optimization time has remarkably been reduced to almost half for above airfoils.

B. Results of Neural Networks

As discussed above, three types of neural networks were implemented and tested against 500 airfoils for training and 200 airfoils unknown to the neural nets. Consolidated results for these airfoils is shown in Table IV.

The results from Table IV show that Feed Forward and Back Propagation has proved to be more promising in terms of better performance as indicated by increased fraction of both known and unknown airfoils within acceptable MSE values. On the other hand, GRNN and RBF showed improved efficiency with known airfoils than for the unknown airfoils. Comparison of Results for a known to network airfoil (Eppler 399 airfoil) and an unknown to network airfoil (Gottingen 426 airfoil) to the three types of neural networks is shown in Fig 6

The plots for Gottingen 426 airfoil and Eppler 399 airfoil support application of Feed Forward Back Propagation Neural Network for solving this problem. However, results from RBF and GRNN largely favour known to network airfoils than unknown airfoils as is evident from RBF and GRNN plots for Gottingen 426 airfoil. Results for 200 airfoils unknown to network also support similar findings. MSE for GRNN and RBF is higher than FFBP with RBF performing the worst with a high MSE.

Table IV : Comparison of Test Results for Three Neural Nets

Artificial Neural Network	$\leq 1X10^{-5}$		$\geq 1X10^{-5}$	
	Count	**%age**	**Count**	**%age**
Feed Forward and Back Propagation				
Known Airfoils (500)	273	54.6	227	45.4
Unknown Airfoils (200)	113	56.5	87	43.4
Radial Basis Neural Network				
Known Airfoils (500)	394	78.8	106	21.2
Unknown Airfoils (200)	47	23.5	153	76.5
Generalized Regression Neural Network				
Known Airfoils (500)	363	72.6	137	27.4
Unknown Airfoils (200)	78	39.0	122	61.0

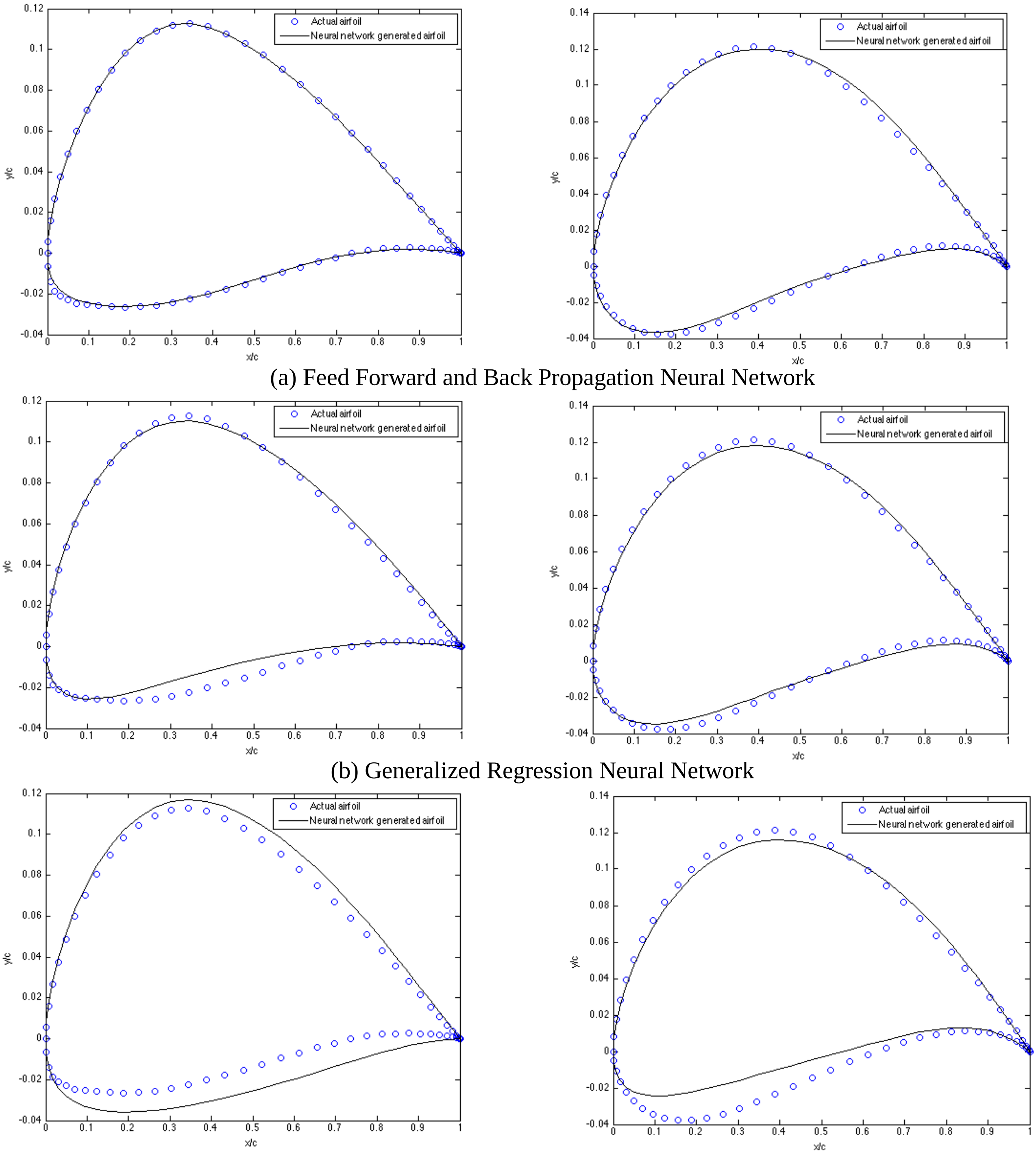
C. Regression Analysis

A post training regression analysis was performed to analyze the neural networks. In this analysis, the output of neural networks for known targets was compared. Thus neural network output would match the target values and would ideally be a straight line with 45° slope passing through the origin as shown in Fig 7.

Fig 7 shows that performance of Feed Forward and Back Propagation is better than other two types of neural nets as indicated by the high regression values and low training R-values. On the other hand, both Generalized Regression Neural Network and Radial Basis Neural Network have higher R-values but shown poor results with test and validation data. The main reason is their architecture as both determine distance between input and weight vectors, which are incrementally multiplied by biased vectors. This would lead an input close to weight vector, produce an output close to unity while output would be close to zero if input is different from weight vector.

VIII. CONCLUSION

This work determines airfoil geometry for a given C_p using Neural Network and Bezier-PARSEC parameters. The main consideration of this paper is to use Ant Colony Optimization technique to optimize missing BP3434 parameters instead of complete set of airfoil coordinates. Further, three types of Neural Networks; Feed Forward and Back Propagation, Radial Basis and Generalized Regression were employed. Similar to earlier findings with GA based code, we proved that Feed-forward and Back Propagation exhibited greater efficiency than the other two types of Neural Networks. However, we were able to achieve higher precision with reduced time for optimization using ACO to determine missing BP3434 parameters. Besides, percentage of known and unknown airfoils with precision $\leq 1X10^{-5}$ has shown a slight increase.



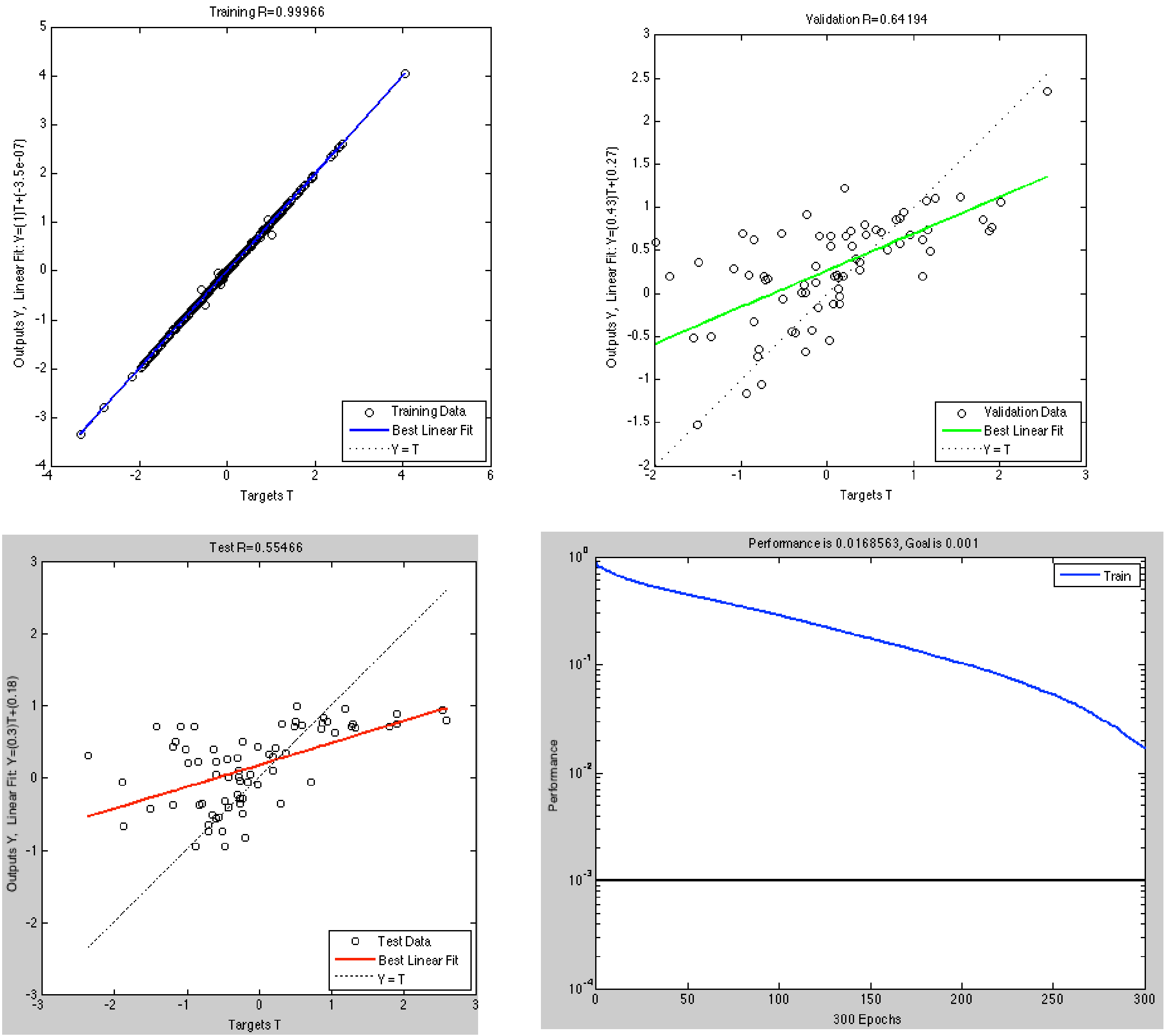
(a) Feed Forward and Back Propagation Neural Network

(b) Generalized Regression Neural Network

(c) Radial Basis Neural Network

Gottingen GOE426 Airfoil Eppler E399 Airfoil

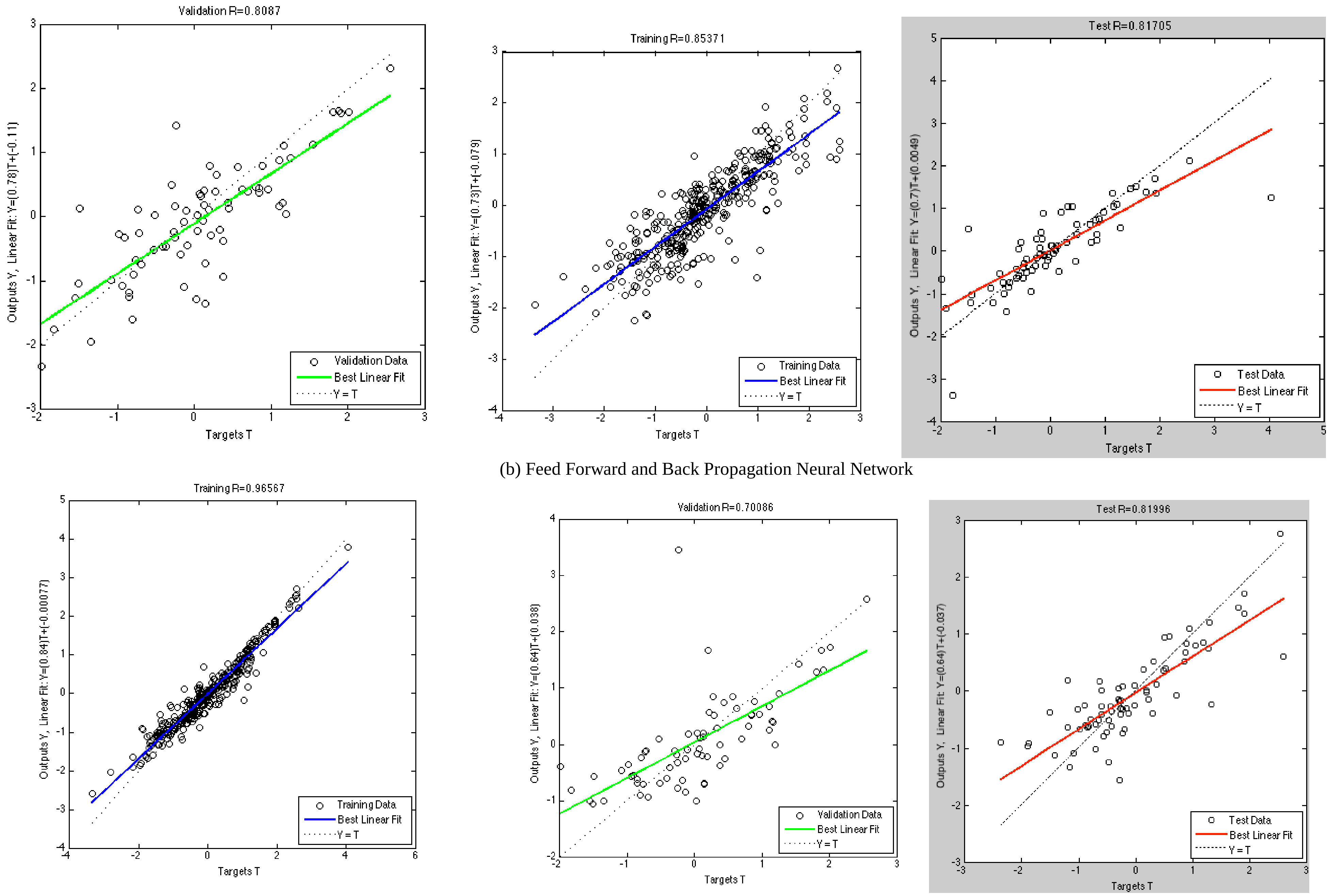
Fig 6 : Comparison of Results of Known to Unknown Airfoil to Neural Network



(a) Radial Basis Neural Network

IX. FUTURE WORK

We have implemented Simple ACO in this research work. Future works may consider implementation of other extensions of ACO techniques like Elitist AS, Ant-Q, Max-Min As, Hyper-cube AS and etc to achieve high performance in order to further reduce the optimization level and attain higher level of precision.



(b) Feed Forward and Back Propagation Neural Network

(c) Generalized Regression Neural Network

Fig 7 : Regression Analysis

REFERENCES

[1] G. Trapani, T. Kipouros and M. Savill, 'Computational Aerodynamic Design for 2D High-Lift Airfoil Configurations', 2010.

[2] K. Lane, 'Novel Inverse Airfoil Design Utilising Parametric Equations', Graduate, California Polytechnic State University, 2010.

[3] Z. Zhu, F. Hongyan, Y. Rixin and L. Jie, 'Multi-objective Optimization Design of Airfoil and Wing', *Science in China Series E Technological Sciences*, vol. 47, no. 1, pp. 15-25, 2004.

[4] P. Hewitt and S. Marques, 'Aerofoil Optimisation Using CST Parameterisation in SU2', in *Royal Aeronautical Society Applied Aerodynamics Group Conference*, Bristol, 2014.

[5] M-selig.ae.illinois.edu, 'UIUC Airfoil Data Site', 2015. [Online]. Available: http://m-selig.ae.illinois.edu/ads/coord_database.html. [Accessed: 11- Jan- 2014].

[6] A. Hacioglu, 'Fast Evolutionary Algorithm for Airfoil Design via Neural Network', *AIAA Journal*, vol. 45, no. 9, pp. 2196-2203, 2007.

[7] U. Wickramasinghe, R. Carrese and X. Li, 'Designing Airfoils Using a Reference Point Based Evolutionary Many-objective Particle Swarm Optimization Algorithm', *IEEE Congress on Evolutionary Computation*, 2010.

[8] M. Ebrahimi and A. Jahangirian, 'Aerodynamic Optimization of Airfoils Using Adaptive Parameterization and Genetic Algorithm', *Journal of Optimization Theory and Applications*, vol. 162, no. 1, pp. 257-271, 2013.

[9] A. Kharal and A. Saleem, 'Neural Networks Based Airfoil Generation for a Given Cp Using Bezierâ€“PARSEC Parameterization', *Aerospace Science and Technology*, vol. 23, no. 1, pp. 330-344, 2012.

[10] G. Sun, Y. Sun and S. Wang, 'Artificial neural network based inverse design: Airfoils and wings', *Aerospace Science and Technology*, vol. 42, pp. 415-428, 2015.

[11] K. Thinakaran and R. Rajasekar, 'Design of Airfoil Using Inverse Procedure and Neural Network', *An International Journal of Advanced Computer Technology*, vol. 2, no. 9, pp. 285-290, 2013.

[12] L. Prabhu and J. Srinivas, 'Flutter Prediction Using Neural Networks-based Hybrid Optimization Scheme', *International Journal of Research in Aeronautical and Mechanical Engineering*, vol. 2, no. 4, 2014.

[13] M. Ahmed and N. Qin, 'Surrogate-Based Aerodynamic Design Optimization: Use of Surrogates in Aerodynamic Design Optimization', in *13th Conference on Aerospace Sciences & Aviation Technology*, Cairo, 2009.

[14] R. Derksen and T. Rogalsky, 'Bezier-PARSEC: An Optimized Aerofoil Parameterization for Design', *Advances in Engineering Software*, vol. 41, no. 7-8, pp. 923-930, 2010.

[15] N. Salunke, R. Juned and S. Channiwala, 'Airfoil Parameterization Techniques: A Review', *American Journal of Mechanical Engineering*, vol. 2, no. 4, pp. 99-102, 2014.

[16] M. Darigo and L. Gambardella, 'Ant Colonies for the Travelling Salesman Problem', *BioSystems*, vol. 43, pp. 73-81, 1997.

[17] M. Darigo and T. Stutzle, *Ant Colony Optimization*. Cambridge: The MIT Press, 2004.

[18] J. Bell and P. McMullen, 'Ant Colony Optimization techniques for The Vehicle Routing Problem', *Advanced Engineering Informatics*, vol. 18, no. 1, pp. 41-48, 2004.

[19] G. Fainekos and K. Giannakoglou, 'Inverse Design of Airfoils Based on a Novel Formulation of the Ant Colony Optimization Method', *Inverse Problems in Engineering*, vol. 11, no. 1, pp. 21-38, 2003.

Analytical solution of a problem on MHD flow in a rectangular duct

Elena Ligere, Ilona Dzenite and Aleksandrs Matvejevs

***Abstract* —** This paper presents an analytical solution of the MHD problem on a fully developed flow of a conducting fluid in a duct with the rectangular cross-section, located in a uniform external magnetic field, and under a slip boundary condition on side walls of the duct. The flow is driven by a constant pressure gradient. The case of perfectly conducting Hartmann walls and insulating side walls is considered. The solution is derived by using integral transforms.

***Keywords* —** Integral transforms, magnetohydrodynamic duct flow, slip boundary condition.

I. Introduction

A FLOW of a conducting fluid in the presence of external magnetic field produces a variety of new effects, studied by magnetohydrodynamics (MHD), the discipline combining the classical fluid mechanics and electrodynamics. The MHD effects are widely exploited both in technical devices (e.g., in pumps, flow meters, generators) and industrial processes in metallurgy, material processing, chemical industry, industrial power engineering and nuclear engineering. Channels, in particular rectangular and circular channels, are common parts of many MHD devices. Therefore, investigation of MHD phenomena in channels with conducting fluids is quite important.

The motion of conducting fluid in external magnetic field is described by the system of MHD equations, containing Navier-Stokes equation for the motion of incompressible viscous fluid with the additional term corresponding to the Lorentz force and Maxwell's equations (see [1]). In MHD the number of exact solutions, obtained analytically, is limited due to the nonlinearity of the Navier-Stokes equation. The exact solutions have been obtained only for very specific problems; however, numerical methods are widely used for solving MHD
problems.

The fully developed flows in rectangular ducts are well studied for different electric conductivities of the walls, but under "no slip" condition on the duct walls (for example, see [1]). Recently, in [2] three classic MHD problems are revisited on assuming a hydrodynamic slip condition at the interface between the electrically conducting fluid and the insulating walls. One of the problems studied analytically in [2] is the problem on a fully developed flow in the rectangular duct with insulating walls and a slip condition on the Hartmann walls (the walls perpendicular to the magnetic field).

This paper presents an analytical solution of the MHD problem on a fully developed flow of a conducting fluid in the duct with the rectangular cross-section, located in a uniform external magnetic field, and under a slip boundary condition on side walls of the duct. The obtained solution seems absent in literature.

The use of integral transforms or series expansion (see [3]) is one of the powerful method for obtaining analytical solutions of problems in mathematical physics. Also in MHD some problems with specific geometry of the flow and boundary conditions are well-solved by integral transforms (for example, see the author's works [4] - [7]).

The MHD problem of this paper is also solved by using integral transforms, but at first, the kernels of integral transforms has been derived and then used for solving the problem.

II. Problem Formulation

Consider the MHD problem on a fully developed flow of a conducting fluid in the rectangular duct with the perfectly conducting Hartmann walls at $z = \pm 1$ and non-conducting side walls at $y = \pm d$ (the walls parallel to the external magnetic field) with the slip boundary condition on the side walls (see Fig.1).

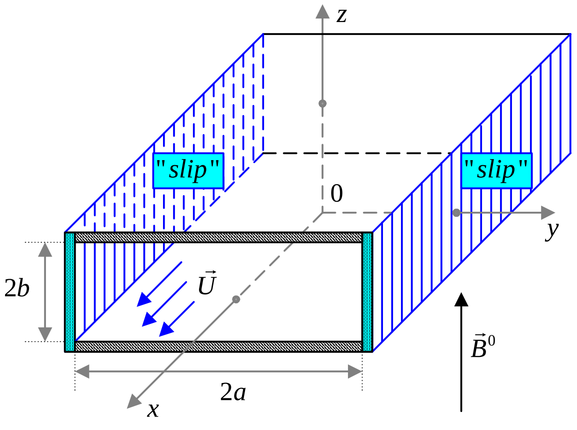


Fig. 1 MHD duct flow with a slip boundary condition

Elena Ligere is with the Department of Engineering Mathematics, Riga Technical University (RTU), Riga, Latvia (e-mail: jelena.ligere@rtu.lv).

Ilona Dzenite with the Department of Engineering Mathematics, RTU, Riga, Latvia (e-mail: ilona.dzenite@rtu.lv).

Aleksandrs Matvejevs with the Department of Engineering Mathematics, RTU, Riga, Latvia (e-mail: aleksandrs.matvejevs@rtu.lv).

The dimensionless MHD equations, describing the problem, have the form ([1], [2]):

$$\frac{\partial^2 U}{\partial z^2} + \frac{\partial^2 U}{\partial y^2} + 1 + Ha\frac{\partial b_x}{\partial z} = 0\,, \tag{1}$$

$$\frac{\partial^2 b_x}{\partial z^2} + \frac{\partial^2 b_x}{\partial y^2} + Ha\frac{\partial U}{\partial z} = 0\,, \tag{2}$$

where $\vec{V} = U(y,z)\cdot\vec{e}_x$ is the velocity of the fluid,

$\vec{b} = b_x(y,z)\cdot\vec{e}_x$ is the induced magnetic field,

$Ha = B_0 h\sqrt{\sigma/\rho\nu}$ is the Hartmann number, which characterizes the ratio of electromagnetic force to viscous force; σ, ρ, ν are the conductivity, the density and the viscosity of the fluid, respectively.

The boundary conditions are

$$z = \pm 1:\quad U = 0\,,\qquad \frac{\partial b_x}{\partial z} = 0\,, \tag{3}$$

$$y = \pm d:\; U \pm \alpha\frac{\partial U}{\partial y} = 0\,,\quad b_x = 0\,, \tag{4}$$

where α is the slip length. The slip condition is given by the 3rd kind boundary condition ([2]).

III. Problem Solving

The problem is solved by using the integral transforms

$$\bar{u}(\lambda, y) = \int_{-1}^{1} U(y,z)\,K_1(\lambda,z)\,dz\,, \tag{5}$$

$$\bar{b}(\lambda, y) = \int_{-1}^{1} b_x(y,z)\,K_2(\lambda,z)\,dz\,, \tag{6}$$

where $K_1(\lambda,z)$ and $K_2(\lambda,z)$ are unknown kernels.

In order to find the unknown kernels, (1) is multiplied by $K_1(\lambda,z)$, (2) by $K_2(\lambda,z)$, and integrated with respect to z.

Thus, it yields

$$\frac{\partial U}{\partial z}K_1\bigg|_{z=1} - \frac{\partial U}{\partial z}K_1\bigg|_{z=-1} - UK_1'\big|_{z=1} + UK_1'\big|_{z=-1} + \int_{-1}^{1} UK_1''dz + \frac{d^2\bar{u}}{dy^2} + \int_{-1}^{1} K_1\,dz + Ha\left(b_xK_1\big|_{z=1} - b_xK_1\big|_{z=-1} - \int_{-1}^{1} b_xK_1'\,dz\right) = 0\,, \tag{7}$$

$$\frac{\partial b_x}{\partial z}K_2\bigg|_{z=1} - \frac{\partial b_x}{\partial z}K_2\bigg|_{z=-1} - b_xK_2'\big|_{z=1} + b_xK_2'\big|_{z=-1} + \int_{-1}^{1} b_xK_2''\,dz + \frac{d^2\bar{b}}{dy^2} + Ha\left(UK_2\big|_{z=1} - UK_2\big|_{z=-1} - \int_{-1}^{1} UK_2'\,dz\right) = 0 \tag{8}$$

Due to the boundary conditions (3), the following terms are equal to zero:

$$UK_1'\big|_{z=1} = UK_1'\big|_{z=-1} = 0\,,\qquad \frac{\partial b_x}{\partial z}K_2\bigg|_{z=1} = \frac{\partial b_x}{\partial z}K_2\bigg|_{z=-1} = 0\,,$$

$$UK_2\big|_{z=1} = UK_2\big|_{z=-1} = 0\,. \tag{9}$$

The following additional conditions for the kernels are to be applied [3]:

$$K_1''(\lambda,z) = -\lambda^2 K_1(\lambda,z)\,,\quad K_1'(\lambda,z) = \beta K_2(\lambda,z)\,, \tag{10}$$

$$K_1\big|_{z=1} = K_1\big|_{z=-1} = 0\,. \tag{11}$$

The solution of (10) with the boundary conditions (11) has the form

$$K_1(\lambda,z) = \cos(\lambda_n z)\,,\quad K_2(\lambda,z) = \sin(\lambda_n z)\,, \tag{12}$$

where

$$\lambda_n = \frac{\pi}{2} + \pi n,\quad n = 0,1,2,\ldots \tag{13}$$

Hence, the inverse integral transform for (5)-(6) has the form

$$U(y,z) = \sum_{n=0}^{\infty} \bar{u}(\lambda_n, y)\cdot\cos(\lambda_n z)\,, \tag{14}$$

$$b_x(y,z) = \sum_{n=0}^{\infty} \bar{b}(\lambda_n, y)\cdot\sin(\lambda_n z)\,. \tag{15}$$

Then system (7)-(8), describing the problem, takes the form:

$$\frac{d^2\bar{u}}{dy^2} - \lambda_n^2\bar{u} + \frac{2}{\lambda_n}(-1)^n + Ha\lambda_n\bar{b} = 0\,, \tag{16}$$

$$\frac{d^2\bar{b}}{dy^2} - \lambda_n^2\bar{b} - Ha\lambda_n\bar{u} = 0 \tag{17}$$

with the following boundary conditions, obtained from (4) by using the integral transform (5):

$$y = \pm d:\quad \bar{u} \pm \alpha\frac{d\bar{u}}{dy} = 0\,,\quad \bar{b} = 0\,. \tag{18}$$

The following ordinary differential equations for the unknown functions $\bar{u}(\lambda_n, y)$ and $\bar{b}(\lambda_n, y)$ can be obtained from (16)-(17):

$$\bar{u} = \frac{1}{\lambda_n Ha}\left(\bar{b}'' - \lambda_n^2 \bar{b}\right), \tag{19}$$

$$\bar{b}^{(4)} - 2\lambda_n^2 \bar{b}'' + \lambda_n^2\left(\lambda_n^2 + Ha^2\right)\bar{b} = 2(-1)^{n+1} Ha \;. \tag{20}$$

The characteristic equation of the corresponding homogeneous equation of (20) is

$$k^4 - 2\lambda_n^2 k^2 + \lambda_n^2\left(\lambda_n^2 + Ha^2\right) = 0\,, \tag{21}$$

with the roots

$$k_{1,3} = \pm\sqrt{\lambda_n^2 + i\cdot\lambda_n Ha}\,, \qquad k_{2,4} = \pm\sqrt{\lambda_n^2 - i\cdot\lambda_n Ha}\,. \tag{22}$$

Taking into account that the function $\bar{b}(\lambda_n, y)$ is even with respect to y, the solution of (20) takes the form

$$\bar{b}(\lambda_n, y) = A\cosh\left(k_1 y\right) + B\cosh\left(k_2 y\right) + \frac{2(-1)^{n+1} Ha}{\lambda_n^2\left(\lambda_n^2 + Ha^2\right)}\,. \tag{23}$$

Then it follows from (19) and (23) that

$$\bar{u}(\lambda_n, y) = i\left(A\cosh\left(k_1 y\right) - B\cosh\left(k_2 y\right)\right) - \frac{2(-1)^{n+1} Ha}{\lambda_n\left(\lambda_n^2 + Ha^2\right)}\,, \tag{24}$$

where the coefficients *A* and *B* are determined from the boundary conditions (18) and are equal to

$$A = \frac{2\,(-1)^n}{\left(\lambda_n{}^2 + Ha^2\right)\cdot\lambda_n{}^2}\times$$

$$\times\frac{\left(Ha + i\,\lambda_n\right)\cosh\left(k_2 d\right) + \alpha\,Ha\,k_2\sinh\left(k_2 d\right)}{\Delta}\,, \tag{25}$$

$$B = \frac{2\,(-1)^n}{\left(\lambda_n{}^2 + Ha^2\right)\cdot\lambda_n{}^2}\times$$

$$\times\frac{\left(Ha - i\,\lambda_n\right)\cosh\left(k_1 d\right) + \alpha\,Ha\,k_1\sinh\left(k_1 d\right)}{\Delta}\,. \tag{26}$$

Applying the inverse integral transforms (14)-(15) to the (23)-(24), the solution of the problem (1)-(4) has the form:

$$U = \sum_{n=0}^{\infty}\frac{2(-1)^n i}{\left(\lambda_n^2 + Ha^2\right)\lambda_n^2}\left(\frac{\tilde{A}\cosh\left(k_1 y\right) - \tilde{B}\cosh\left(k_2 y\right)}{\Delta} - i\lambda_n\right)\cos(\lambda_n z) \tag{27}$$

$$b_x = \sum_{n=0}^{\infty}\frac{2(-1)^n}{\left(\lambda_n^2 + Ha^2\right)\lambda_n^2}\left(\frac{\tilde{A}\cosh\left(k_1 y\right) + \tilde{B}\cosh\left(k_2 y\right)}{\Delta} - Ha\right)\sin(\lambda_n z) \tag{28}$$

where

$$\tilde{A} = \left(Ha + i\cdot\lambda_n\right)\cosh\left(k_2 d\right) + \alpha\cdot Ha\cdot k_2\sinh\left(k_2 d\right), \tag{29}$$

$$\tilde{B} = \left(Ha - i\cdot\lambda_n\right)\cosh\left(k_1 d\right) + \alpha\cdot Ha\cdot k_1\sinh\left(k_1 d\right), \tag{30}$$

$$\Delta = 2\cosh\left(k_1 d\right)\cosh\left(k_2 d\right) + \alpha\cdot k_2\cosh\left(k_1 d\right)\sinh\left(k_2 d\right) +$$

$$+\alpha\cdot k_1\cosh\left(k_2 d\right)\sinh\left(k_1 d\right), \tag{31}$$

$$k_1 = \sqrt{\lambda_n^2 + i\cdot\lambda_n Ha}\,, \qquad k_2 = \sqrt{\lambda_n^2 - i\cdot\lambda_n Ha}\,.$$

References

[1] Muller U., Buhler L., *Magnetofluiddynamics in Channels and Containers*. Berlin, Heidelberg, New York, Barcelona, Hong Kong; London, Milan, Paris, Singapore, Tokyo: Springer, 2001.

[2] Smolentsev S., “MHD duct flow under hydrodynamic “slip” condition”, *J. Theor. Comput. Fluid Dyn.*, 23(6), 2009, pp. 557-570.

[3] Antimirov M.Ya., Kolyshkin A.A., Vaillancourt R. *Applied Integral Transforms*.- Rhole Island USA: American Mathematical Society, 1993.

[4] M.Ya.Antimirov, E.S Ligere. “Analytical solutions for the problems of the flowing into of the conducting fluid through the lateral side of the plane channel in a strong magnetic field”. *Magnetohydrodynamics.*, vol. 36 (1), 2000, pp. 47-60.

[5] Ligere E. “Remarks to the Solution of MHD Problem on an Inflow of Conducting Fluid into a Plane Channel through the Channel’s Lateral Side”. *Scientific Journal of Riga Technical University. Computer Science. - Boundary Field Problems and Computer Simulation,* vol. 50, 2011, pp. 30-39.

[6] Ligere E., Dzenite I., “Application of Integral Transforms for Solving Some MHD Problems” in *Proc 14th WSEAS Int. Conf. on Mathematical and Computational Methods in Science and Engineering. - Advances in Mathematical and Computational Methods,* Sliema (Malta), September 7-9, 2012, pp. 286-291.

[7] Ligere E., Dzenite I., “Analytical Solution for Some MHD Problems on a Flow of Conducting Liquid in the Initial Part of a Channel in the Case of Rotational Symmetry” in *Proc. 2014 Int. Conf. on Pure Mathematics, Applied Mathematics, Computational Methods (PMAMCM 2014). - Advances in Applied and Pure Mathematics,* Santorini Island, Greece, 2014, pp. 61-67.

Critical Exponents for the Multidimensional Heat Conduction Equation with a Nonlinear Boundary Condition and Variable Density

Mersaid Aripov, Zafar Rakhmonov

Abstract – In this paper considered the problem of nonlinear multidimensional non-Newtonian polytrophic filtration with nonlocal boundary condition in the fast diffusive case. It is established the conditions of global solvability and nosolvability of nonlinear filtration problem in an inhomogeneous medium by the method of standard equations, self-similar analysis and comparison principle. Obtained the critical Fujita exponent and the critical global existence exponent, that plays an important role in the study of qualitative properties of nonlinear models of reaction-diffusion, heat conduction, filtering, and other physical, chemical, and biological processes. In the case of the global solvability the leading term of the asymptotes of self-similar solutions were established. The asymptotic of solutions for the critical value of the parameters are proved.

Keywords—Critical global existence curve, Critical Fujita curve, Blow-up, Asymptotic, Numerical Solution, Self-Similar Analysis.

I. Introduction

Consider in $(x,t)\in R_+^N\times(0,+\infty)$ the following nonlocal problem of non-Newtonian filtration

$$\rho(x)u_t=\nabla\left(\left|\nabla u^m\right|^{p-2}\nabla u^m\right), \qquad (1)$$

$$-\left|\nabla u^m\right|^{p-2}\frac{\partial u^m}{\partial x_1}(0,t)=u^q(0,t), \quad t>0, \qquad (2)$$

$$u(x,0)=u_0(x)\geq 0, \quad x\in R_+^N, \qquad (3)$$

where $\rho(x)=(1+|x|)^n$, $n>0$, $m>0$, $q>0$, $1<p<1+1/m$, $u_0(x)$ are continuous nonnegative bounded functions.

Department of Informatics and Applied Programming
National University of Uzbekistan, 100174, University Street, 4. Tashkent, Uzbekistan
mirsaidaripov@mail.ru, zraxmonov@inbox.ru

The problem (1)-(3) appear in different models of a heat transfer process, population dynamics, chemical reactions, non-Newtonian fluids or certain diffusion [1, 2].

At first Fujita [3] to the problem Cauchy for semilinear equation

$$u_t=\Delta u+u^\beta, \quad x\in R^N, t>0$$

proved a condition of a global solvability $\beta>1+2/N$. Value of parameters $\beta_c=1+2/N$ is called the Fujita type a critical exponent in literature [1-3]. From then on, many similar results were established for different nonlinear evolution equations (see the survey papers [9, 10] and the reference therein). Among those, [3-13] are concerned with the equations with nonlinear boundary flux.

Huang, Yin and Wang [12] studied the porous media equation into multi-dimensional case

$$\begin{cases} u_t=\Delta u^m,\ x\in R_+^N,\ 0<t<T, \\ -\dfrac{\partial u^m}{\partial x_1}=u^q(x,t),\ x_1=0,\ 0<t<T, \\ u(x,0)=u_0(x), \quad x\in R_+^N. \end{cases}$$

They obtain the following value for a global existing of the considered problem $q_0=(m+1)/2$ and Fujita type critical exponent $q_c=m+1/N$.

For equation (1) with slow diffusion $(p>2)$, Wanjuan Du and Zhongping Li [9] considered the case $m=1$, $n=0$ of the equation (1) and obtained the critical global existence exponent $q_0=2(p-1)/p$ and the critical Fujita exponent $q_c=(1+1/N)(p-1)$.

The authors of [14] have studied the problem (1) - (3) in the fast diffusive case, when $N=1$. They obtained the critical exponent of the global existence of solutions $q_0 = \frac{\left(m(n+1)+1\right)(p-1)}{p+n}$ and the critical Fujita exponent $q_c = m(p-1)+\frac{p-1}{n+1}$ by constructing sub and supper solutions.

The Cauchy problem for the equation with double nonlinearity and variable density

$$\rho(x)u_t = \nabla\left(u^{m-1}|\nabla u|^{p-2}\nabla u\right),$$

where $m>1,\ p>2$ studied by the authors [15] and proved the nonlinear effect of finite speed of perturbation.

In this paper the conditions of a global solvability and nosolvability by reduction of equation (1) to the so called radially symmetrical form are studied. Developing results authors [9, 12] an asymptotical behavior of solution of the problem (1)-(3) including a critical value of the parameters are proved. Based on qualitative properties of a self-similar solution the numerical experiments carried out.

2. MAIN RESULTS

Introduce notations

$$q_0 = \frac{\left(m(n+1)+1\right)(p-1)}{p+n},\ q_c = m(p-1)+\frac{p-1}{N+n}.$$

Theorem 1. *If $0\le q\le q_0$, then each solution of problem (1)-(3) exists globally.*

Proof. Let

$$u_+(x,t) = e^{Lt}g(\xi),\ g(\xi) = \left(K+e^{-M\xi}\right)^{1/m},$$

$$\xi = (1+x_1)e^{Jt},\ x_i = 0,\ i=\overline{2,N},$$

where $L = J(p+n)/\left[1-m(p-1)\right]$, $J=(K+1)^2$, $M = (K+1)^{q/\left[m(p-1)\right]}$. A direct calculation yields

$$-\left|\frac{\partial u_+^m}{\partial x}\right|^{p-2}\frac{\partial u_+^m}{\partial x}\Bigg|_{x_1=0} = ,$$

$$= -e^{(p-1)(Lm+J)t}\left|\left(g^m\right)'\right|^{p-2}\left(g^m\right)'(0)$$

$$\frac{\partial}{\partial x}\left(\left|\frac{\partial u_+^m}{\partial x}\right|^{p-2}\frac{\partial u_+^m}{\partial x}\right)(x,t) = ,$$

$$= e^{\left(Lm(p-1)+J(p+n)\right)t}\left|\left(g^m\right)'\right|^{p-2}\left(g^m\right)'(\xi)$$

$$\rho(x)\frac{\partial u_+}{\partial t}(x,t) = e^{(L-Jn)t}\xi^n\left(Lg(\xi)+J\xi g'(\xi)\right).$$

Note that $(p-1)(Lm+J) = L-Jn$, $(Lm+J)(p-1)\ge q$, and hence, if

$$\xi^{1-N}\left(\xi^{N-1}\left|\left(g^m\right)'\right|^{p-2}\left(g^m\right)'\right)'(\xi) - , \qquad (4)$$
$$-J\xi^{n+1}g'(\xi) - L\xi^n g(\xi)\le 0$$

$$-\left|\left(g^m\right)'\right|^{p-2}\left(g^m\right)'(0)\ge g^q(0), \qquad (5)$$

then

$$\rho(x)\frac{\partial u_+}{\partial t}\ge\frac{\partial}{\partial x}\left(\left|\frac{\partial u_+^m}{\partial x}\right|^{p-2}\frac{\partial u_+^m}{\partial x}\right),$$

$$-\left|\frac{\partial u_+^m}{\partial x}\right|^{p-2}\frac{\partial u_+^m}{\partial x}(0,t)\ge u_+^q(0,t).$$

Not difficult to see that if $K\ge\|u_0\|_\infty^m$ is large enough, then (4) and (5) are satisfied. Also, we have $u_+(x,0)\ge u_0(x)$ and $u_+(0,0)>u_0(0)$, that by comparison, we conclude $u(x,t)\le u_+(x,t)$, which implies that u is global.

Theorem 2. *If $q>q_c$, then the problem (1)–(3) admits nontrivial global solutions with small initial data.*

Proof. The theorem is proved on the basis of a comparison principle. Equation (1) admits in $Q_T = \left\{(x,t):\ x\in R_+^N, 0<t<+\infty\right\}$ to self-similar solution of the following form

$$u_+(t,x) = (T+t)^{-\gamma} f(\xi), \qquad (6)$$

where $\xi = |\zeta|$, $\zeta_i = (1+x_i)(T+t)^{-\sigma}$, $i=\overline{1,N}$,

$$\gamma = \frac{p-1}{q(p+n)-(p-1)\left(m(n+1)+1\right)},$$

$$\sigma = \frac{q-m(p-1)}{q(p+n)-(p-1)\left(m(n+1)+1\right)}.$$

Construct a suppersolution of (1) - (3). In order for $u_{+}(t,x)$ was an suppersolution of problem (1) - (3) function $f(\xi)$ should be satisfy the following inequalities [5, 6]

$$\xi^{1-N}\frac{d}{d\xi}\left(\xi^{N-1}\left|\frac{df^{m}}{d\xi}\right|^{p-2}\frac{df^{m}}{d\xi}\right)+\sigma\xi^{n+1}\frac{df}{d\xi}+\gamma\xi^{n}f\le 0, \quad (7)$$

$$-\left|\left(f^{m}\right)'\right|^{p-2}\left(f^{m}\right)'(1)\ge f^{q}(1). \quad (8)$$

Consider the following function

$$\overline{f}(\xi)=\left(a+b\xi^{\frac{p+n}{p-1}}\right)_{+}^{-\frac{p-1}{1-m(p-1)}}, \quad (9)$$

where $b=\dfrac{1-m(p-1)}{m(p+n)}\sigma^{1/(p-1)}>0$, $a>0$.

Therefore, as

$$\overline{f}'(\xi)=-\frac{\sigma^{1/(p-1)}}{m}\xi^{(n+1)/(p-1)}\left(a+b\xi^{\frac{p+n}{p-1}}\right)^{-\left(\frac{p-1}{1-m(p-1)}+1\right)},$$

$$\left(\overline{f}^{m}\right)'(\xi)=-\sigma^{1/(p-1)}\xi^{\frac{n+1}{p-1}}\left(a+b\xi^{\frac{p+n}{p-1}}\right)^{-\left(\frac{m(p-1)}{1-m(p-1)}+1\right)},$$

$$\xi^{1-N}\left(\xi^{N-1}\left|\left(\overline{f}^{m}\right)'\right|^{p-2}\left(\overline{f}^{m}\right)'\right)'(\xi)=-(N+n)\sigma\xi^{n}\overline{f}+$$

$$+\sigma\xi^{\frac{p+n}{p-1}}\left(a+b\xi^{\frac{p+n}{p-1}}\right)^{-\left(\frac{p-1}{1-m(p-1)}+1\right)},$$

problem (8), (9) take the following form

$$-\left(\frac{(q-m(p-1))(N+n)-p+1}{q(p+n)-(p-1)(m(n+1)+1)}\right)\xi^{n}\overline{f}\le 0 \quad (10)$$

$$\sigma(a+b)^{-\frac{p-1}{1-m(p-1)}}\Big|_{\xi=1}\ge(a+b)^{-\frac{q(p-1)}{1-m(p-1)}}\Big|_{\xi=1} \quad (11)$$

From (10) follows directly that $q>q_c$ for all $\xi\in(0,+\infty)$. Thereby from (11) we have the following restrictions on the constants a:

$$a+b\le\sigma^{-\frac{1-m(p-1)}{(q-1)(p-1)}}.$$

In conclusion, we note that the obtained self-similar solution $u_{+}(t,x)$ is a suppersolution of (1)-(3). Owing to the comparison principle, the solution $u(t,x)$ of the problem (1)-(3) is global if the initial datum $u_0(x)$ is small enough.

Theorem 3. *If $q>q_0$, then the solution of the problem (1)-(3) with appropriately large initial data blows up in a finite time.*

Theorem 4. *If $q_0<q<q_c$, then each nontrivial solution of the problem (1)-(3) blows up in a finite time.*

Theorem 5. *Let $\dfrac{(N+n)(m+1)-n}{(N+n)m+1}<p<1+\dfrac{1}{m}$, then the solution of problem (7), (8) has the asymptotic*

$$f(\xi)\sim C\overline{f}(\xi),\ \xi\to+\infty,$$

where $C=\left(\sigma\left((N+n)(m(p-1)-1)+p+n\right)\right)^{1/\left[1-m(p-1)\right]}$

3. THE CRITICAL VALUE OF NUMERICAL PARAMETERS

The case $1-m(p-1)=0$ we will call a critical value of the numerical parameters. Below it is shown that in this case an asymptotic behavior of solution is different. In this case, the self-similar solution of problem (1) - (3) has the form

$$u_{+}(t,x)=(T+t)^{-\gamma_c}z(\xi), \quad (12)$$

where $\sigma_c=\dfrac{1}{p+n}$, $\gamma_c=\dfrac{p-1}{(q-1)(p+n)}$, $\xi=|\zeta|$,

$z(\xi)=e^{-d\xi^{\frac{p+n}{p-1}}}$, $\zeta_i=(1+x_i)(T+t)^{-\sigma_c}$ $i=1,\ldots,N$,

$d=\dfrac{p-1}{m(p+n)}\left(\dfrac{1}{p+n}\right)^{1/(p-1)}$. Using the well-known comparison principle can be shown that $u_{+}(t,x)=(T+t)^{-\gamma_c}z(\xi)$ is an suppersolution of problem (1)-(3) in $(x,t)\in R_{+}^{N}\times(0,+\infty)$

Theorem 6. *Let $1-m(p-1)=0$, then the solution of (7), (8) has the asymptotic*

$$f(\xi)\sim M\overline{f}(\xi),\ \xi\to+\infty,$$

where M is any positive number.

4. RESULTS OF NUMERICAL CALCULATION

It is known, in solving the problem by iterative method, much depends on the choice of the initial approximation. Therefore because of the nonexistence of solutions self-similar problems arises the problem of choosing an appropriate initial approximation preserving qualitative properties of nonlinear processes. Depending on the values of the numerical parameters of the equation, this difficulty can be overcome by the right choice of initial approximations, for which the calculations were taken above the established asymptotic formulas. On the basis of these qualitative results were numerically calculated. Below some results of numerical experiments in one dimensional case are illustrated. As showed the pictures the numerical calculation of an evolution of the studied problem has a property of a finite speed of propagation. Suggested an iteration process is convergent to a solution of the problem (1)-(3) quickly.

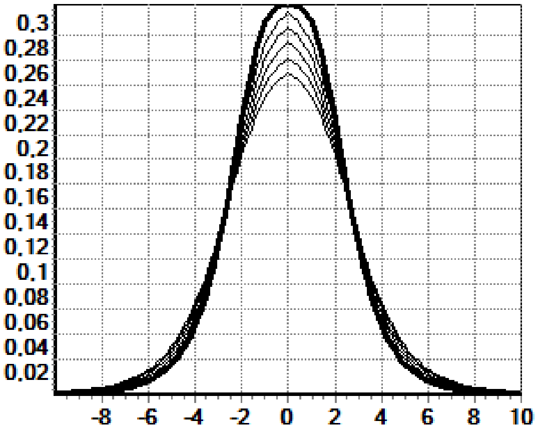


Fig 1. m=0.75, p=1.77, q=1.9, n=1

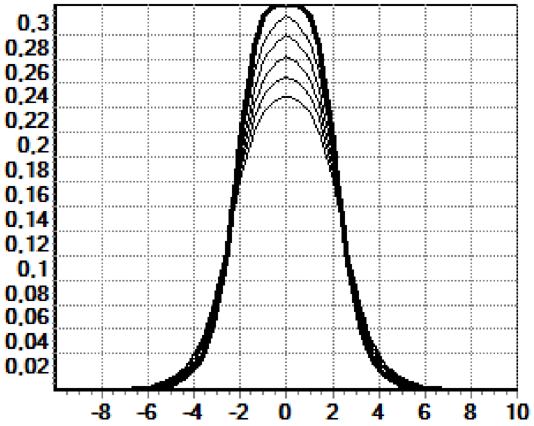


Fig 2. m=0.75, p=1.77, q=1.9, n=0.2

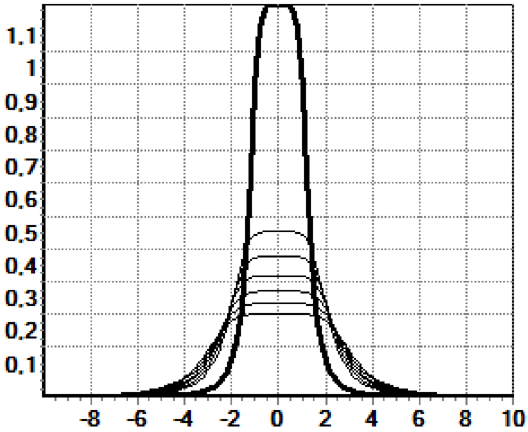


Fig 3. m=1. 5, p=1.32, q=2.9, n=0.8

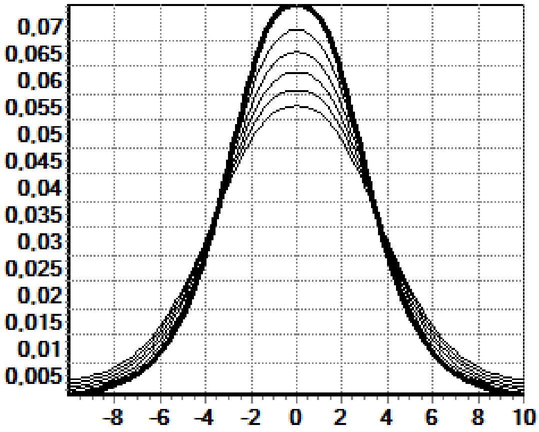


Fig 4. m=0.75, p=1.78, q=1.9, n=0.3

References:

[1]. M.Aripov. Standard Equation's Methods for Solutions to Nonlinear problems (Monograph), Tashkent, FAN, 1988.

[2]. A.S. Kalashnikov, Some problems of the qualitative theory of nonlinear degenerate second-order parabolic equations, Russian. Math. Surveys **42(2),** 1987, 169-222.

[3]. Victor A. Galaktionov and Juan L. Vazquez. The problem of blow-up in nonlinear parabolic equations. Discrete and continuous dynamical systems, vol. 8, №2, April 2002, 399-433.

[4]. Wang C.P. and Zheng S.N. Critical Fujita exponents of degenerate and singular parabolic equations. Proc. Roy. Soc. Edinburgh Sect. A 136, №2, 2006, 415-430.

[5]. Zejia W., Jingxue Y., Chunpeng W. Critical exponents of the non-Newtonian polytropic filtration equation with nonlinear boundary condition. Appl. Math. Lett. 20, 2007, 142-147.

[6]. Li Z., Mu Ch. Critical exponents for a fast diffusive polytrophic filtration equation with nonlinear boundary flux. J. Math. Anal. Appl. 346, 2008, 55-64.

[7]. Jiang Z. X. and Zheng S. N. Doubly degenerate parabolic equation with nonlinear inner sources or boundary flux, Doctor Thesis, Dalian University of Technology, China, 2009.

[8]. Zheng P., Mu Ch., Liu D., Yao X. and Zhou Sh. Blow-up analysis for a quasilinear degenerate parabolic equation with strongly nonlinear source. Abstract and Appl. Anal. vol. 2012, Article ID 109546, 19 p.

[9]. Wanjuan Du and Zhongping Li. Critical exponents for heat conduction equation with a nonlinear Boundary condition. Int. Jour. of Math. Anal. vol. 7, 11, 2013, 517-524.

[10]. Li Z., Mu Ch. and Du W. Critical Fujita exponent for a fast diffusive equation with variable coefficients. Bull. Korean Math. Soc. 50. №1, 2013, 105-116.

[11]. Galaktionov V. A., Levine H. A. On critical Fujita exponents for heat equations with nonlinear flux boundary condition on the boundary. Israel J. Math. 94, 1996, 125-146.

[12]. W. Huang, J. Yin, and Y. Wang, On critical Fujita exponents for the porous equation with nonlinear boundary condition, J. Math. Anal. Appl. 286, 2003, 369-377.

[13]. Aripov M., Rakhmonov Z. Asymptotic behavior of self-similar solutions of a nonlinear problem of polytrophic filtration with nonlinear boundary conditions. Jour. Comp. Tech., vol.18, no. 4, 2013, 50-55.

[14]. Aripov M., Rakhmonov Z. Numerical simulation of a nonlinear problem of a fast diffusive filtration with a variable density and nonlocal boundary conditions. Mathematical Models and Simulation in Science and Engineering, Series 23, 2014, 72-77.

[15]. Martynenko A. V. and. Tedeev A. F. On the behavior of solutions to the Cauchy problem for a degenerate parabolic equation with inhomogeneous density and a source, Comput. Math. Math. Phys. 48 2008, no. 7, 1145–1160.

Traffic signal control in congested road network

Alexander Krylatov*†, Victor Zakharov*†, Ovanes Petrosian*
*Saint-Peterburg State University, Saint-Petersburg, Russia,
198504, Universitetskii prosp., 35, E-mail: v.zaharov@spbu.ru
† Solomenko Institute of Transport Problems of the Russian Academy of Sciences, Saint-Petersburg, Russia,
199178, 12-th Line VO, 13, E-mail: aykrylatov@yandex.ru

***Abstract*—The most travel delays in daily trips in modern large urban areas take place primarily at signal-controlled junctions due to regular interruption by alternating traffic lights. Therefore in urban traffic control road networks alleviation increasing of traffic congestion at signalized junctions becomes one of the most significant issues facing decision makers at various levels of management. This paper is devoted to development of methodological tools to cope with problem of settings traffic signals for congested transportation network. The global traffic control system is assumed to define timing parameters of signals for the whole transportation network. Users of network are believed to react on any fixed signal setting assigning according to user-equilibrium of Wardrop. Thus the global optimal signal settings problem under user network equilibrium conditions is formulated as bi-level optimization program. When travel time is modeled by modified linear BPR cost function user-equilibrium flow patterns for two intersecting networks of parallel routes could be obtained as explicit condition of the lower programming level.**

I. Introduction

One of the main challenges for decision makers at different levels of management in modern worldwide large cities is coping with enormous traffic jams on their road networks. Authorities faced with such complicated problem are forced to implement various quite expensive arrangements. On the one hand they tend to put into operation new infrastructure facilities and on the other hand – try to reorganize road traffic. Wherein continuously growing travel demand encourages the development of advanced methodological tools and technological innovations to meet newly emerging requirements. Especially the need for innovations is felt in the area of traffic signal control as soon as numerous signalized junctions of the road networks contribute in congestion most significantly by alternating traffic lights. Extra complexity are appended by intricacy of large scale transportation networks and their inner nontrivial coherence. Therefore many researchers focus on optimization of signal control settings with certain and uncertain travel demand [3].

The present paper is devoted to the development of methodological tools for area traffic signal control under user-equilibrium flow pattern with certain demand. Mathematical programming approach is suggested to apply for achieving this purpose. Note that there exist researches where mathematical programming was successfully employed to solve constrained optimization problems of signal control settings [1], [6], [8]. Moreover bi-level programming technique was also implemented for tackling the problem of optimal signal control setting [9], [4]. However due to huge sizes of road networks in corresponding bi-level mathematical programs a solution of lower level cannot be expressed explicitly [5]. Thereby computationally tractable algorithms for a bi-level network design problems were appeared [2]. In this paper we offer mathematically explicit form of user-equilibrium assignment on the lower programming level for one type of networks. Explicit strategies of lower level allow manager of upper level to influence network performance directly solely by signal control setting.

The paper is organized as follows. Section II introduces a bi-level optimal signal setting program on the network of general topology. In Section III the problem of Wardrop user-equilibrium assignment on the network of parallel routes is considered. Equilibrium assignment is obtained in explicit mathematical form. Section IV is devoted to investigation of the network consisting of two intersecting subnetworks of parallel routes. Bi-level program is formulated for such a network. Due to explicit decisions obtained for the network of parallel routes formulated bi-level program is transformed in a way of direct signal control influence. Conclusions for this paper and topics of interest for further investigations are briefly summarized in Section V.

II. Optimal signal setting on the network of general topology

Consider the network presented by directed graph G that includes a set of consecutively numbered nodes N and a set of consecutively numbered arcs A. Let R denote the set of origin nodes and S – the set of destination nodes ($R \cap S = \emptyset$). We also use following notation: W – the set of origin-destination pairs (OD-pair) between R and S, $w \in W$; J – the set of signalized intersections in the network; A_j – the set of approaching links for intersection $j \in J$; K^w – the set of possible routes w OD-pair; F^w – demand between w OD-pair; f_k^w – traffic flow through route k, $k \in K^w$; x_a – traffic flow on the arc $a \in A$, $x = (\ldots, x_a, \ldots)$; λ_a – green time proportion on link $a \in A$ (green light timing in whole traffic lights cycle on link a), $\lambda = (\ldots, \lambda_a, \ldots)$; $t_a(x_a, \lambda_a)$ – travel time through congested signalized link $a \in A$; $\delta_{a,k}^w$ – indicator: $\delta_{a,k}^w = 1$ if link a is a part of path k connecting OD-pair w, and $\delta_{a,k}^w = 0$ otherwise. Now we are able to formulate the following bi-level programming problem:

$$\min_{x,\lambda} Z_1(x, \lambda) = \min_{x,\lambda} \sum_{a \in A} t_a(x_a, \lambda_a) x_a, \tag{1}$$

subject to

$$\sum_{a \in A_j} \lambda_a = 1 \quad \forall j \in J, \tag{2}$$

$$0 \le \lambda_a \le 1 \quad \forall a \in A, \tag{3}$$

when traffic flows are assumed to be assigned according to user-equilibrium of Wardrop

$$\min_x Z_2(x) = \min_x \sum_{a \in A} \int_0^{x_a} t_a(u, \lambda_a) du, \tag{4}$$

subject to

$$\sum_{k \in K^w} f_k^{rs} = F^w \quad \forall w \in W, \tag{5}$$

$$f_k^w \geq 0 \quad \forall k \in K^w, w \in W, \tag{6}$$

with definitional constraints

$$x_a = \sum_w \sum_{k \in K^w} f_k^w \delta_{a,k}^w. \tag{7}$$

In a similar way bi-level program was formulated for global optimal signal setting under queuing network equilibrium conditions [9]. For this purpose queuing delay influencing on storage capacity was introduced. Such constraints take into account queue length on congested links before signalized junctions. Eventually queues increasing on the approaches to an intersection was restricted so that upstream intersections were not blocked. Here we do not deal with such constraint because it could lead to empty set of possible decision in bi-level program when demand becomes too big.

III. User-equilibrium of Wardrop on the signalized network of parallel routes

Consider transportation network presented by digraph consisted of one OD-pair and n parallel (independent) routes. We use following notation: $N = \{1, \ldots, n\}$ – set of numbers of all routes; L_i – the set of sequentially numbered links of route i, $i = \overline{1,n}$, $|L_i| = l_i$; F – demand between OD-pair; f_i – traffic flow through route i, $i = \overline{1,n}$, $f = (f_1, \ldots, f_n)$; t_{il}^0, c_{il} and λ_{il} – free travel time, capacity and green time proportion of link l belonging to route i, $l = \overline{1, \ldots, l_i}$, $i = \overline{1,n}$; $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{il_i})$, $i = \overline{1,n}$; $t_{il}(f_i, \lambda_{il}) = t_{il}^0 \left(1 + \frac{f_i}{\lambda_{il} c_{il}}\right)$ – travel time through congested signalized link l of route i, $l = \overline{1, l_i}$, $i = \overline{1,n}$. We are modeling travel time as modified linear BPR cost function [7], [6], [9]. In this notation when signal control pattern is setted Wardrop user-equilibrium assignment of traffic flows could be formulated as follows:

$$\min_f z(f) = \min_f \sum_{i=1}^n \sum_{l=1}^{l_i} \int_0^{f_i} t_{il}^0 \left(1 + \frac{u}{\lambda_{il} c_{il}}\right) du, \tag{8}$$

subject to

$$\sum_{i=1}^n f_i = F, \tag{9}$$

$$f_i \geq 0 \quad \forall i = \overline{1,n}. \tag{10}$$

Lemma. *Assignment f^* is user-equilibrium of Wardrop if and only if there exists such ω (Lagrange multiplier) that*

$$\sum_{l=1}^{l_i} t_{il}^0 \left(1 + \frac{f_i}{\lambda_{il} c_{il}}\right) \begin{cases} = \omega & \text{if } f_i > 0, \\ \geq \omega & \text{if } f_i = 0. \end{cases} \tag{11}$$

Proof: To obtain (11) we use Kuhn-Tucher theorem. As soon as goal function is linear Kuhn-Tucker conditions are necessary and sufficient. Therefore Lagrangian of the problem (8)-(10) is

$$L = \sum_{i=1}^n \sum_{l=1}^{l_i} \int_0^{f_i} t_{il}^0 \left(1 + \frac{u}{\lambda_{il} c_{il}}\right) du + \\ + \omega \left(F - \sum_{i=1}^n f_i\right) + \sum_{i=1}^n \eta_i (-f_i).$$

Let us differentiate L with respect to f_i and equate it to zero then we obtain

$$\omega = \sum_{l=1}^{l_i} t_{il}^0 \left(1 + \frac{f_i}{\lambda_{il} c_{il}}\right) - \eta_i. \tag{12}$$

According to Kuhn-Tucker condition of complementary slackness $f_i \eta_i = 0$ $\forall i = \overline{1,n}$ it is clear that

$$\eta_i \begin{cases} = 0 & \text{if } f_i > 0, \\ \geq 0 & \text{if } f_i = 0. \end{cases} \tag{13}$$

Thus due to (12) and (13) we directly obtain (11). ■

Introduce following additional notation: $t_i^0 = \sum_{l=1}^{l_i} t_{il}^0$ (free travel time through whole route i) and $r_i(\lambda_i) = \left(\sum_{l=1}^{l_i} \frac{t_{il}^0}{\lambda_{il} c_{il}}\right)^{-1}$ $\forall i = \overline{1,n}$.

Corollary *Assignment f^* is user-equilibrium of Wardrop if and only if there exists such ω (Lagrange multiplier) that*

$$f_i = \begin{cases} \left(\omega - t_i^0\right) r_i(\lambda_i) & \text{if } t_i^0 < \omega, \\ 0 & \text{if } t_i^0 \geq \omega. \end{cases} \tag{14}$$

Proof: If $f_i > 0$ for some $i = \overline{1,n}$ then from (11)

$$f_i = \left(\omega - t_i^0\right) r_i(\lambda_i) > 0$$

and, consequently, we obtain first condition from (14). If $f_i = 0$ for some $i = \overline{1,n}$ then from (11) $t_i^0 \geq \omega$ and we obtain second condition from (14). ■

Without loss of generality, we assume that the routes are numbered as follows:

$$t_1^0 \leq \ldots \leq t_n^0. \tag{15}$$

Theorem 1. *When (15) holds user-equilibrium of Wardrop in the problem (8)-(10) is appeared under following assignment*

$$f_i = \begin{cases} r_i(\lambda_i) \frac{F + \sum_{j=1}^b t_j^0 r_j(\lambda_j)}{\sum_{j=1}^b r_j(\lambda_j)} - t_i^0 r_i(\lambda_i) & \text{if } i \leq b, \\ 0 & \text{if } i > b, \end{cases} \quad \forall i = \overline{1,n}, \tag{16}$$

where b is defined from

$$\sum_{i=1}^b r_i(\lambda_i)(t_b^0 - t_i^0) \leq F < \sum_{i=1}^b r_i(\lambda_i)(t_{b+1}^0 - t_i^0). \tag{17}$$

Proof: Introduce $b \in \{1, n\}$ such that when (15) holds $t_b^0 < \omega$ and $t_{b+1}^0 \geq \omega$. Substitute (14) into (9) then

$$\sum_{j=1}^n f_j = \sum_{j=1}^b (\omega - t_j^0) r_j(\lambda_j) = F.$$

Consequently,

$$\omega = \frac{F + \sum_{j=1}^{b} t_j^0 r_j(\lambda_j)}{\sum_{j=1}^{b} r_j(\lambda_j)}. \tag{18}$$

Now if we substitute (18) into (14) then we obtain (16).

To find b let us remind that $t_b^0 < \omega \le t_{b+1}^0$ and, hence, following inequality holds

$$\sum_{i=1}^{b} r_i(\lambda_i)(t_b^0 - t_i^0) \le \sum_{i=1}^{b} r_i(\lambda_i)(\omega - t_i^0) =$$
$$= F < \sum_{i=1}^{b} r_i(\lambda_i)(t_{b+1}^0 - t_i^0).$$

■

Theorem 1 offers explicit form of Wardrop user-equilibrium assignment on the network of parallel routes for any fixed signal control pattern.

IV. Optimal signal setting on intersecting networks of parallel routes

Consider network consisting of two intersecting networks of parallel routes (an example of such a network is shown in Fig. 1).

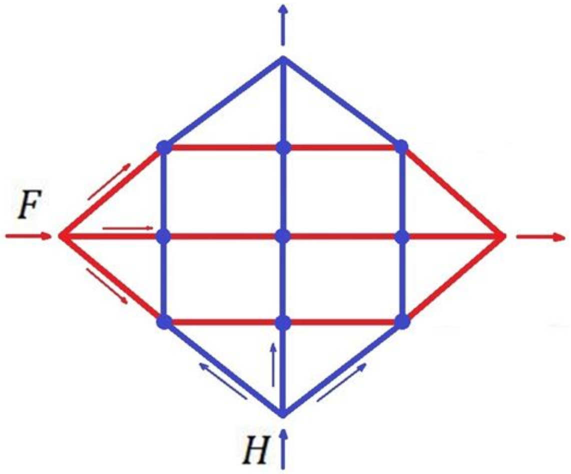


Figure 1. Intersecting networks of parallel routes

We suppose there are two independent traffic flows F and H between two independent OD-pairs. These traffic flows influence each other solely at signalized junctions because of alternative traffic lights and do not have any common traffic links. Introduce following notation: $N = \{1, \dots, n\}$ – set of numbers of all routes for F; $M = \{1, \dots, m\}$ – set of numbers of all routes for H; L_i – the set of sequentially numbered links of route i, $i = \overline{1,n}$, $|L_i| = l_i$; K_j – the set of sequentially numbered links of route j, $j = \overline{1,m}$, $|K_j| = k_j$; I – the set of pairs (l,k) for some $l \in L_i$ and $k \in K_j$, $i = \overline{1,n}$, $j = \overline{1,m}$ such that corresponding links approach in the same intersection of the network; f_i – traffic flow through route i, $i = \overline{1,n}$, $f = (f_1, \dots, f_n)$; h_j – traffic flow through route j, $j = \overline{1,m}$, $h = (h_1, \dots, h_n)$; t_{il}^0, c_{il} and λ_{il} – free travel time, capacity and green time proportion of link l belonging to route i, $l = \overline{1,\dots,l_i}$, $i = \overline{1,n}$; t_{jk}^0, c_{jk} and λ_{jk} – free travel time, capacity and green time proportion of link k belonging to route j, $k = \overline{1,\dots,k_j}$, $j = \overline{1,m}$; $t_{il}(f_i, \lambda_{il}) = t_{il}^0 \left(1 + \frac{f_i}{\lambda_{il} c_{il}}\right)$ – travel time through congested signalized link l of route i, $l = \overline{1,l_i}$, $i = \overline{1,n}$; $t_{jk}(h_j, \lambda_{jk}) = t_{jk}^0 \left(1 + \frac{h_j}{\lambda_{jk} c_{jk}}\right)$ – travel time through congested signalized link k of route j, $l = \overline{1,k_j}$, $j = \overline{1,m}$; $\lambda = \left(\{\lambda_{il}\}_{l=\overline{1,l_i}}^{i=\overline{1,n}}, \{\lambda_{jk}\}_{k=\overline{1,k_j}}^{j=\overline{1,m}}\right)$. Moreover by analogy with $r_i(\lambda_i)$, $i = \overline{1,n}$ we introduce $s_j(\lambda_j) = \left(\sum_{k=1}^{k_j} \frac{t_{jk}^0}{\lambda_{jk} c_{jk}}\right)^{-1}$.

In considered case the problem of optimal signal settings (1)-(7) has the following bi-level expression

$$\min_{f,h,\lambda} z_1(f,h,\lambda) = \min_{f,h,\lambda} \sum_{i=1}^{n} t_i^0 \left(1 + \frac{f_i}{r_i(\lambda_i)}\right) f_i +$$
$$+ \sum_{j=1}^{m} t_j^0 \left(1 + \frac{h_j}{s_j(\lambda_j)}\right) h_j, \tag{19}$$

subject to

$$\lambda_{il} + \lambda_{jk} = 1 \quad \forall (l,k) \in I, \tag{20}$$
$$0 \le \lambda_{il} \le 1 \quad \forall l = \overline{1,l_i}, i = \overline{1,n}, \tag{21}$$
$$0 \le \lambda_{jk} \le 1 \quad \forall k = \overline{1,k_j}, j = \overline{1,m}, \tag{22}$$

when traffic flows are assumed to be assigned according to user-equilibrium of Wardrop

$$\min_{f,h} z_2(f,h) = \min_{f,h} \sum_{i=1}^{n} \int_0^{f_i} t_i^0 \left(1 + \frac{u}{r_i(\lambda_i)}\right) du +$$
$$+ \sum_{j=1}^{m} \int_0^{h_j} t_j^0 \left(1 + \frac{u}{s_j(\lambda_j)}\right) du, \tag{23}$$

$$\sum_{i=1}^{n} f_i = F, \tag{24}$$

$$\sum_{j=1}^{m} h_j = H, \tag{25}$$

$$f_i \ge 0 \quad \forall i = \overline{1,n}, \tag{26}$$
$$h_j \ge 0 \quad \forall j = \overline{1,m}. \tag{27}$$

As soon as traffic links for flows F and H are independent their user-equilibrium assignments could be found separately in explicit forms due to Theorem 1. Then the problem (20)-(27) could be reformulated in such a way that user-equilibrium flow patterns for F and H are appeared as explicit condition of the lower programming level. Let us formulate the following

Theorem 2. *The global optimal signal control problem formulated as bi-level program (1)-(7) is appeared to have explicit reaction of users of network under any fixed signal setting for two intersecting independent networks of parallel links:*

$$\min_{\lambda} z_1(f,h,\lambda) = \min_{\lambda} \sum_{i=1}^{n} t_i^0 \left(1 + \frac{f_i}{r_i(\lambda_i)}\right) f_i +$$
$$+ \sum_{j=1}^{m} t_j^0 \left(1 + \frac{h_j}{s_j(\lambda_j)}\right) h_j, \tag{28}$$

subject to

$$\lambda_{il} + \lambda_{jk} = 1 \quad \forall (l,k) \in I, \tag{29}$$

$$0 \leq \lambda_{il} \leq 1 \quad \forall l = \overline{1,l_i}, i = \overline{1,n}, \tag{30}$$

$$0 \leq \lambda_{jk} \leq 1 \quad \forall k = \overline{1,k_j}, j = \overline{1,m}, \tag{31}$$

when traffic flows are assumed to be assigned according to user-equilibrium of Wardrop

$$f_i = \begin{cases} r_i(\lambda_i)\frac{F+\sum_{j=1}^{b} t_j^0 r_j(\lambda_j)}{\sum_{j=1}^{b} r_j(\lambda_j)} - t_i^0 r_i(\lambda_i) & \text{if } i \leq b, \\ 0 & \text{if } i > b, \end{cases} \tag{32}$$

$\forall i = \overline{1,n}$, *where b is defined from*

$$\sum_{i=1}^{b} r_i(\lambda_i)(t_b^0 - t_i^0) \leq F < \sum_{i=1}^{b} r_i(\lambda_i)(t_{b+1}^0 - t_i^0), \tag{33}$$

and

$$h_j = \begin{cases} s_j(\lambda_j)\frac{F+\sum_{p=1}^{v} t_p^0 s_p(\lambda_p)}{\sum_{p=1}^{v} s_p(\lambda_p)} - t_j^0 s_j(\lambda_j) & \text{if } j \leq v, \\ 0 & \text{if } j > v, \end{cases} \tag{34}$$

$\forall j = \overline{1,m}$, *where v is defined from*

$$\sum_{j=1}^{v} s_j(\lambda_j)(t_v^0 - t_j^0) \leq H < \sum_{j=1}^{v} s_j(\lambda_j)(t_{v+1}^0 - t_j^0). \tag{35}$$

Proof: The proof of this Theorem follows directly from bi-level formulation of signal control problem for two intersecting independent networks (19)-(27) and Theorem 1. ■

Therefore one can see that for network consisting of two intersecting subnetworks of parallel routes decision maker is able to manage signal timing sets directly. Indeed for such network bi-level optimization program could be reduced to the problem (28)-(35) when decision maker everytime knows reaction of users (32)-(35) on any fixed signal control setting λ. Hence his/her issue is to find such setting λ that offers minimum to the goal function (28).

V. Conclusion

In this paper the problem of traffic signal control in congested road networks was discussed. Bi-level programming approach was addressed to formulate signal setting problem under user-equilibrium traffic flow assignment with certain demand. For special type of networks strategies of the lower level are obtained explicitly. Due to explicit form of lower level strategies bi-level program could be transformed into decision-making tool with direct traffic signal influence on the networking performance. In further works the problem of signal control settings is planed to consider under competitive traffic assignment conditions [10], [11].

References

[1] *Allsop R.E., Charlesworth J.A.* Traffic in signal-controlled road network: an example of different signal timings inducing different routings. Traffic Engineering Control 18 (1977) 118–132.

[2] *Bard J.F.* Practical bi-level optimization: algorithms and applications. Dordrecht: Kluwer Academic Publishers (2002) 476 p.

[3] *Chiou S.-W.* Optimization of robust area traffic control with equilibrium flow under demand uncertainty. Computers and Operations Research 41 (2014) 399–411.

[4] *Clegg J., Smith M.J., Xiang Y., Yarrow R.* Bilevel programming applied to optimizing urban transportation. Transportation Research Part B 35(1) (2001) 41–70.

[5] *Dempe S.* Foundations of bilevel programming. Dordrecht: Kluwer Academic Publishers (2002) 305 p.

[6] *Smith M.J., Vuren T.* Traffic equilibrium with responsive traffic control. Trasportation science 27(2) (1993) 118–132.

[7] *Traffic Assignment Manual* In: U.S. Bureau of Public Roads (eds.) U.S. Department of Commerce. Washington, D.C. (1964)

[8] *Wong S.C.* Derivatives of the performance index for the traffic model from TRANSYT. Transportation Research Part B 29(5) (1995) 303–327.

[9] *Yang H., Yagar S.* Traffic assignment and signal control in saturated road networks. Transportation Research Part A 29(2) (1995) 125–139.

[10] *Zakharov V., Krylatov A.* Equilibrium Assignments in Competitive and Cooperative Traffic Flow Routing. IFIP Advances in Information and Communication Technology 434 (2014) 641–648

[11] *Zakharov V., Krylatov A., Ivanov D.* Equilibrium traffic flow assignment in case of two navigation providers. IFIP Advances in Information and Communication Technology 408 (2013) 156–163

THE DEGREE OF INFLUENCE OF CONSTRUCTIVE AND REGIME FACTORS ON THE CHARACTERISTICS TURBINE WHEEL STEPS SHOULDER WHO ARE MORE ANGLES OF ROTATION

Andrey Yu. Fershalov, Yuriy Ya. Fershalov, Lyudmila P. Tsigankova

Abstract — The issues dedicated to the results of various factors' and their combinations' impact upon the speed rate of rotor wheels with the vanes' turning angles $\Delta\beta=180°–(\beta 1\kappa+\beta 2\kappa)=151°…164°$. The analysis is based on the methodology of the simulation with the utilization of the regression models obtained as a result of the model experiment.

Keywords — speed ratio of the rotor wheel, rotor wheel, turbine stage, nozzle apparatus.

I. Introduction

Engine building industry is and has been the strategic field of Russian economy and takes exclusively important place in shipbuilding. The internal combustion engines and turbines are most often used as part of marine power. Besides modern internal combustion engines (ICE) are turbocharged, the latter is ensured by gas generator.

Both turbine and ICE have their advantages and weaknesses. The main advantages of turbines are: high efficiency; high aggregate power with small weight and dimensions; adaptability to automation; high reliability; simplicity of heat and kinematic scheme; simplicity of construction and maintenance; high technological effectiveness; possibility of aggregate repairs; simplicity of transportation and ease of assembly; minimum volumes of hazardous emissions into environment; high maneuverability and rate of load; most turbines have the capacity of short-term overload. Besides, recently there are great achievements both in turbo machinery aerodynamics and in the development of heat-resistant steels and alloys. The successes of aerodynamics and metallurgy allowed increasing turbines' thermal efficiency to the required level, and creating the conditions for their implementation into the industry.

Turbine installations are the combinations of a number of elements, in which complex processes take place, the ones which allow converting one form of energy to another. To create such installation it is important to have the complex of scientific knowledge being the result both of theoretical and experimental research in various fields. Therefore it becomes rather clear why modern turbine building is an innovative field with own approaches to the complicated problems solution with the use of theoretical and experimental methodology as well as mathematic methods [1], connected to the construction of mathematic models of real phenomena taking place inside the flow passages of turbines. Due to the complexity of the phenomena happening inside the turbine flow passages, these models, as a rule, are not the universal ones and can be used in calculations of turbine various parameters solely within the definite range of the parameters variations. In multistage turbines the greatest impact on the turbine efficiency has rotor wheels, because wrong assessment of aerodynamic properties of gas behind them can result in wrong profiling of the next stage.

By the information of the Kaluga turbine works the 1% increase of velocity ratio inside the full size turbine lattice increases the stage capacity by 0.73%.

In case of the small sized turbine stage, depending upon the lattice velocity ratio, the capacity increase is more due to the increase in the relative thickness of the boundary layer.

Therefore, reduction of power losses in the rotor wheels allows increase of the turbine stage efficiency, which is important relating to marine turbines, which operate autonomously with time-varying load. So, the issue of the efficiency improvement in case of operational modes variations is of great importance.

II. Results of Research

The research of the rotor wheels with large flow turning angle were based on the results of the turbine stage nozzle apparatus' operational mode simulation. Regression models, specifically designed for this purpose, were obtained in the simulation experiment results processing [2-4].

Optimization by one factor. Methodology of the analysis by this method is based on the determination of maximum and minimum values of the function under the study relative to each factor in its variations. Other factors take three values in turn: minimum, average and maximum. The discrepancies between maximum and minimum values, for each factor at every level, are ranked in ascending order. The positions (defined earlier) by the degree of the influence of each factor at each level are summarized, and the smaller this amount is, the stronger the effect on the tested model has a factor.

The degree of factors impact on the rotor wheel velocity ration (ψ) is ranked in the following way: β1 (angle of gas in-leakage to the rotor wheel inlet edges); Mw2t (Mach number at the rotor wheel passages' outlet in relative motion, calculated by theoretical parameters) and β1K (structural angle of rotor wheel vanes' inlet edges installation).

By the sum of difference values (ψmax-ψmin) the degree of factors impact on the rotor wheels velocity ratio is ranked in similar order. The degree of factors impact on the exit angle of flow from the rotor wheel passage (β2) is ranked in the following way: β2K; Mw2 (Mach number at the rotor wheel passages outlet in relative motion, calculated by the real parameters) and u/c2 (the ratio of the rotor wheel circumferential velocity at the average diameter to the real velocity of flow exit from rotor wheel passages). By the sum of difference values (β2max-β2min) the degree of factors impact on β2 is ranked in similar order.

To increase the reliability of the fact that by the degree of impact on the target function factors are placed in the mentioned order, there were applied methods given below.

Four-dimensional optimization. The aim of the analysis by this method was limited to finding maximum and minimum values of the functions under consideration by means of solving the task of the functions optimization with the fixed value of one factor being limited by the boundaries of the experiment done. Each factor takes in turn three values, i.e. minimum, maximum and average. Then the difference between maximum and minimum values of functions of each factor at every level is calculated. Then these values are ranked in descending order. The stronger the factor impact on function – the more the difference. The analysis done by this methodology hasn't statistically revealed (with the account of errors) various degree of impact on the studied factors' regression models.

Visual analysis of the factors' impact. Visual analysis implies the sensitivity analysis of the result obtained by calculations and showing its value alteration with the factors' values variation within the studied limitations. It can help to check the correctness of decisions and conclusions made concerning the connection between the function under research and variables studied. The numerical experiment has been done for the sake of the above with the use of the models of simulation based on the developed regression models. The results are shown as graph dependence.

Fig.1 shows that in case of minimum values of factors the most important impact on ψ has β1, which determines with the rotor wheel inlet edge angle of installation the angle of incidence and Mw2t.

Fig. 2 shows that with the average values of factors the main influence on ψ is exerted by β1 and Mw2t, β1K has less influence.

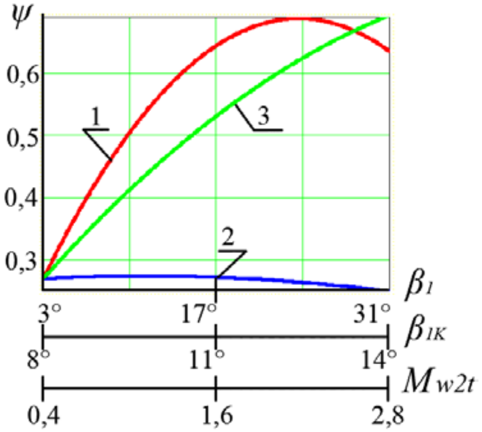


Fig. 1. Dependence of the coefficient ψ with minimum values of factors:
1 – ψ=ξψ(β1) when β1K=8°; Mw2t=0,4;
2 – ψ=ξψ(β1K) when β1=3°; Mw2t=0,4;
3 – ψ=ξψ(Mw2t) when β1=3°; β1K=8°

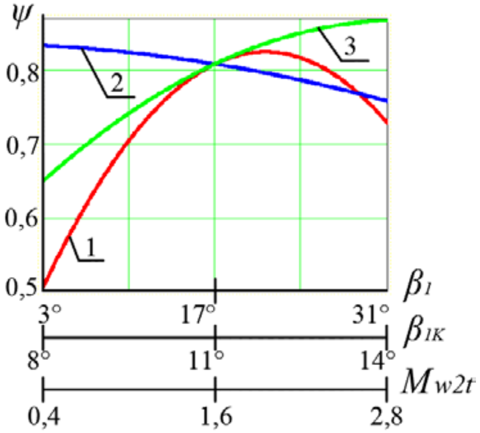


Fig. 2. Dependence of the coefficient ψ with of the average values of factors:
1 – ψ=ξψ(β1) when β1K=11°; Mw2t=1,6;
2 – ψ=ξψ(β1K) when β1=17°; Mw2t=1,6;
3 – ψ=ξψ(Mw2t) when β1=17°; β1K=11°

Fig. 3 shows that with maximum values of factors, the main influence on ψ is exerted by β1 and β1K, Mw2t has less influence.

Fig.4 demonstrates that with minimum values of factors main influence on β2 is exerted by Mw2 and u/c2, less influence has β2K despite its guiding impact on the flow. It is connected to the guiding effect of wheels fitted behind the rotor wheel and simulating the guiding apparatus.

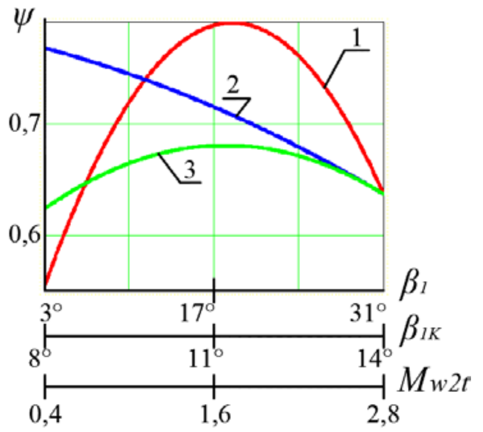


Fig. 3. Dependence of the coefficient ψ with maximum values of factors:
1 – ψ=ξψ(β1) when β1K=14°; Mw2t=2,8;
2 – ψ=ξψ(β1K) when β1=32°; Mw2t=2,8;
3 – ψ=ξψ(Mw2t) when β1=32°; β1K=14°

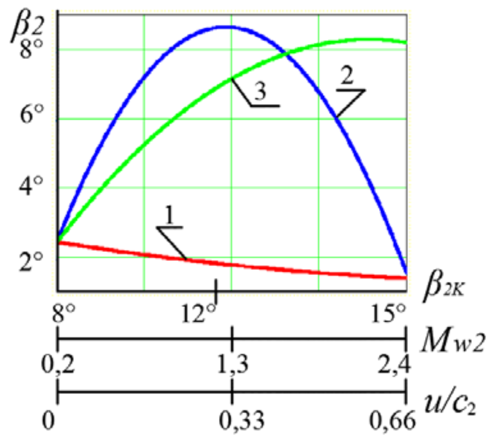


Fig. 4. Dependence angle β2 with minimum values of factors:
1 – β2=ξβ2(β2K) when Mw2=0,2; u/c2=0;
2 – β2=ξβ2(Mw2) when β2K=8°; u/c2=0;
3 – β2=ξβ2(u/c2) when β2K=8°; Mw2=0,2

Fig.5 shows that with the average values of factors the main influence on β2 is exerted by Mw2 and u/c2, β2K has less impact. The value of Mach number becomes the determining one for the flow deflection angle value at the rotor wheel passages outlet by the analogy with the nozzles of nozzle apparatus. Moreover, u/c2 value has an impact on the intensity of vortex, formed in the rotational motion of the working vanes' edges, ventilating gas between rotor wheel and the wheel simulating the guiding apparatus. The latter has also an impact on the flow exit angle value when outgoing from the rotor wheel passages.

Fig. 6 shows that with maximum values of factors main influence on β2 is exerted by β2K, Mw2 and u/c2 have less but not greatly different from each other impact. It is connected to high velocity of gas flow at the rotor wheel passages outlet, which reduces the degree of the opposite action of wheel fitted behind the rotor wheel.

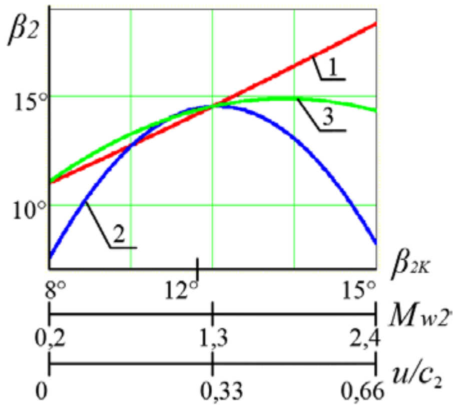


Fig. 5. Dependence angle β2 with of the average values of factors:
1 – β2=ξβ2(β2K) when Mw2=1,3; u/c2=0,33;
2 – β2=ξβ2(Mw2) when β2K=12°; u/c2=0,33;
3 – β2=ξβ2(u/c2) when β2K=12°; Mw2=1,3

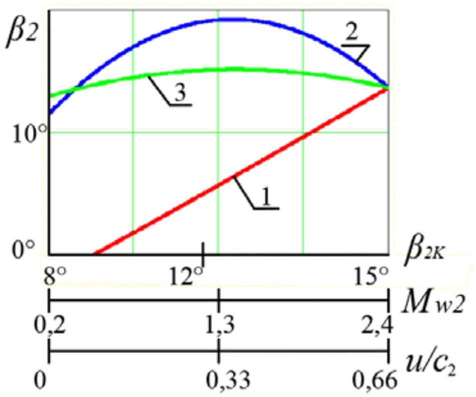


Fig. 6. Dependence angle β2 with maximum values of factors:
1 – β2=ξβ2(β2K) when Mw2=2,4; u/c2=0,66;
2 – β2=ξβ2(Mw2) when β2K=15°; u/c2=0,66;
3 – β2=ξβ2(u/c2) when β2K=15°; Mw2=2,4

III. Conclusions

Rotor wheels of the studied type showed rather high level of their power efficiency under the condition of correct determination of gas-dynamic parameters of flow passage of rotor wheel before them. To determine the values of gas-dynamic parameters of the flow outgoing from the nozzles of nozzle apparatus there has been designed testing beds [5, 6].

To consider gas parameters alterations between rotor wheel and nozzle apparatus it is necessary to use the results of work [7, 8]. Hereinafter, for the similar studies it is assumed to use acoustic methods [9, 10] to get more accurate results.

Due to the fact that the experiment has been carried out with the rotor wheels, having similar number of vanes of similar size, to transfer the results to the rotor wheel of different sizes or number of vanes it is necessary to use methodology given in work [11].

REFERENCES

[1] G.V. Alekseev, A.V. Lobanov, “Stability estimates for the solution to inverse extremal problems for the Helmholtz equation”, J. Appl. Ind. Math. 7 (2013) 1-13.

[2] A.Yu. Fershalov “Improving the efficiency rotor wheels of the ship of low-flow axial turbines”. PhD dissertation (technical). Far Eastern Federal University, Vladivostok, 2011. (in Russian).

[3] A. Yu. Fershalov, M. Yu. Fershalov, Yu. Ya. Fershalov, and T. V. Sazonov “Results of the Study Rotor Wheels Supersonic Microturbines with a Large Angle of Rotation of the Flow”, Applied Mechanics and Materials Vols. 752-753 (2015) pp 884-889.

[4] A. Yu. Fershalov, Yu. Ya. Fershalov, M. Yu. Fershalov, T. V. Sazonov, D. I. Ibragimov “Analysis and optimization of efficiency rotor wheels microturbines”, Applied Mechanics and Materials Vols. 635-637 (2014) pp 76-79.

[5] T. V. Sazonov, Yu. Ya. Fershalov, M. Yu. Fershalov, A. Yu. Fershalov, D. I. Ibragimov “Experimental installation for the study of nozzles microturbines”, Applied Mechanics and Materials Vols. 635-637 (2014) pp 155-158.

[6] Ju. Yu. Fershalov, T. V. Sazonov “Experimental research of the nozzles”, Advanced Materials Research Vols. 915-916 (2014) pp 345-348.

[7] M. Yu. Fershalov, Yu. Ya. Fershalov, A. Yu. Fershalov, T. V. Sazonov, D. I. Ibragimov “Microturbines degree of reactivity”, Applied Mechanics and Materials Vols. 635-637 (2014) pp 354-357.

[8] M. Yu. Fershalov, A. Yu. Fershalov, Ju. Yu. Fershalov “Calculation reactivity degree for axial low-account turbines with small emergence angles of nozzle devices”, Advanced Materials Research Vols. 915-916 (2014) pp 341-344.

[9] V.I. Korenbaum, M.A. Safronova, V.V. Markina, I.A. Pochekutova, A.I. D’yachenko “Study of the Formation Mechanisms of Forced Expiratory Wheezes in a Healthy Person When Breathing Gas Mixtures of Different Density”. Acoustical Physics, 2013, Vol. 59, No. 2, pp. 240–249. DOI: 10.1134/S1063771013020073.

[10] V.I. Korenbaum, A.A. Tagiltsev “Flow noise of an underwater vector sensor embedded in a flexible towed array”. Journal of Acoustical Society of America. 2012. 131(5):3755-3762. DOI:10.1121/1.3693647.

[11] Yu. Ya. Fershalov “Technique for physical simulation of gasodynamic processes in the turbomachine flow passages”. Russian Aeronautics (Iz.VUZ), 2010, Vol. 55, No. 1, pp. 424–429 (DOI: 10.3103/S1068799812040186). © Allerton Press, Inc., 2012. (Original Russian Text© Yu.Ya. Fershalov, 2012, published in Izvestiya VUZ. Aviatsionnaya Tekhnika, 2012, No. 1, pp. 71–74.

Evaluating the technical and scale efficiency of the large hospitals in Greece

Panagiotis Mitropoulos and Ioannis Mitropoulos

***Abstract*—** The aim of this study is to assess the performance of the large hospitals in Greece. More specifically, we measure how well the Greek public Hospitals use their resources, to serve as much as possible patients. We apply the method of Data Envelopment Analysis, to assess the efficiency of 24 large hospitals during the years 2009 to 2012. The input variables were the number of physicians, the number of nurses and other personnel, and expenditures of every hospital. The output variables were the number of inpatient admissions and the number of outpatient visits. The study identifies the inefficient hospitals and provides the magnitudes of specific input reductions or output increases needed to attain technical and scale efficiency. The analysis indicates that the overall efficiency has improved during the study period. The pure technical inefficiency (i.e., managerial inefficiency) is the main source of technical inefficiency while, the contribution of scale inefficiency has been observed to be smaller. The main finding of this paper is the need for better resource allocation in the Greek national health system, because some hospitals have resource surplus and in other hospitals it is observed a lack of resources.

***Keywords*—** Data Envelopment Analysis, Linear programming, Efficiency, Greek hospital performance.

I. Introduction

THE management of the health services are one of the most significant aspects of the public administration, since the health level of the population is a determinant for the economic growth and the social development. In addition, it is a discernible relation between the investments in public health care system, the improvement in living standard and the economic growth. Also, the improvement in health condition of a population increases the productivity, the income per capita and expands the period that a worker could be productive. Finally, a healthy population can improve the social well-being, the macroeconomic stability through the increase in tax income and the decrease in the public expenditures for the health system.

The resource allocation in health care units is now more important than ever. The outbreak of the recent economic crisis, has led to tightening public budgets. The austerity programs were implemented, as well, in healthcare and mainly focused in the hospital sector that comprises the majority of public total healthcare spending. Pressures for sector reform have stimulated interest in identifying and understanding the factors that can contribute to improve the hospital's performance In order to elevate the consequences of the recession the governments should allocate the resources in a way that minimize the expenditures while maximizing patient safety.

The Public hospitals’ expenses reached EUR 2.6 billion in 2010, decreased by EUR 500 million by the end of 2012 [1]. Although Greek hospitals are considered to be a public service with no profit goal, they still try to minimize costs by changing input allocation in order to maximize production and comply with the given budget. On the other hand, patients who previously used the private sector turned to the public sector, automatically increasing the public sector's spending during the economic downturn. The question is to what extent is the austerity related to increased productivity.

Data envelopment analysis (DEA) is a comparative evaluation method, based on linear programming, for measuring the relative efficiencies of a homogenous set of decision making units. DEA is receiving increasing importance as a tool for evaluating and improving the service operations [2].

Data envelopment analysis has been used extensively to address relative efficiency assessments in public sector activities with increased degree of complexity in recent years. Since the introduction of DEA methodology, a considerable number of researchers have applied DEA in evaluating the efficiency of different healthcare organizations. Several systematic reviews of health efficiency studies have been conducted over the last few years [2,3]. However, the majority of the reviewed studies are in the hospitals sector, reflecting its central role in the health care system. All of these above mentioned studies provide information about the growth of this research body and discussing the reliability of efficiency estimates, upon which relevant policy decisions were drawn. These studies also offer extensive overviews of the literature and some in-depth discussion on DEA’s applicability, strengths and limitations.

DEA is a linear programming technique, developed by Charnes, Cooper, and Rhodes [4], for estimating the relative efficiency of a homogeneous set of production units by considering multiple input and output factors. In the input-orientated case, the DEA method holds outputs constant and

This research is implemented within the framework of the Action «Supporting Postdoctoral Researchers» of the Operational Program ‘‘Education and Lifelong Learning’’ (Action’s Beneficiary: General Secretariat for Research and Technology), and is co-financed by the European Social Fund (ESF) and the Greek State.

P. Mitropoulos is with the Department of Business Administration Technological Education Institute of Western Greece, Patras, Greece (phone: 0030-2601369051; e-mail: pmitro@upatras.gr).

I. Mitropoulos is with the Department of Business Administration Technological Education Institute of Western Greece, Patras, Greece (e-mail: mitro@teiwest.gr).

Variables	2009 mean	 St.Dev	2010 mean	 St.Dev	2011 mean	 St.Dev	2012 mean	 St.Dev	2013 mean	 St.Dev
inputs										
Doctors	465	133	484	113	447	121	378	102	374	101
Other personnel	1.085	285	1.082	290	1.063	277	1.004	244	984	239
Operating cost (x 1.000 €)	69.928	34.818	61.341	29.066	56102	26973	50.147	24.395	45.007	23.299
outputs										
Inpatients	38613	15.619	41.276	14.653	44.821	16.583	44.155	16.295	44716	16.350
Outpatients	188504	80.690	1.73.555	74.151	1.64.814	56.715	1.64.655	50.948	174.413	53.679

Table1: Descriptive statistics concerning the hospitals of the study from 2009 to 2013.

seeks to identify inefficiency as a proportionate decrease in input production.

A hospital employs its resources (human, financial, and capital) to produce health care services. A basic economic problem is how to avoid waste in that process. Since the health care managers have more control over the utilization of resources rather than over the arriving patients either for outpatient visit or admissions, it is more appropriate to assume that an input-oriented DEA model should be adopted. We assume that there are n hospitals to be evaluated. DEA constant returns to scale (CRS) model considers the following optimization problem to determine relative efficiency of r_0 hospital among R units.

$$\begin{aligned} & \min \theta \\ s.t.\ & \sum_r X_{ir}\lambda_r - \theta X_{ir_0} \le 0 \qquad i=1,\dots,I \\ & \sum_r Y_{jr}\lambda_r - Y_{jr_0} \ge 0 \qquad j=1,\dots,J \qquad\qquad (1) \\ & \varphi\ free, \quad \lambda_r \ge 0, \quad r=1,\dots,R \end{aligned}$$

Where θ is the radial output contraction factor measuring the level of efficiency for the r0 hospital, Y_{jr} is the amount of the j-th output produced by r-th unit, X_{ir} is the amount of the i-th input produced by r-th hospital and λ_r is the non-negative input/output weights that determine the best practice for the hospital being evaluated.

The variable returns to scale (VRS) model is obtained by simply adding the convexity constraint $\sum\lambda_r=1$ to the above linear program.

Finally the scale efficiency score for each hospital was obtained by dividing the CRS efficiency score by the VRS efficiency score. A scale efficiency score of one implies that the hospital in question is operating at optimal scale or size. If the scale efficiency score is less than one, then the hospital is either too big or too small relative to its optimal size.

II. Data and sample

The present study has been based on data provided by the Greek Ministry of Health concerning 24 large hospitals for the years 2009 to 2013. The total number of Greek public hospitals is 134. In the original data set, hospital sizes range from 18 to 936 beds. However, hospitals with different sizes usually have different characteristics in terms of economies of scale, market share and access to advanced technologies [5]. In order to control the differences in size and make the samples more homogeneous, we select 24 hospitals with more than 250 beds for analysis. Considering the different operation characteristics of hospitals we can categorize them as university and non university hospitals with 7 and 17 units assigned to each group, respectively.

In this study the selection of input/output variables was designed according to the previous studies in the literature [5,6]. The input variables chosen for our analysis are: (1) the number of doctors; (2) the number of other personnel (which includes nurses, administrative and support staff); and (3) total operating cost (excluding the payroll expenses). The output variables consist of: (1) the number outpatient visits; and (2) the number of inpatient admissions. The sample means and standard deviation of the inputs/outputs that used in the analysis are presented in Table 1.

III. Results and discussion

We observe that all type of efficiencies have improved during the study period. The average CRS efficiency increased by 4% from 0,82 in 2009 to 0,86 in 2013. The average VRS efficiency was also increased 3% and at the same time the hospitals scale efficiency increased slightly by 1%.

We further examine the efficiency differences observed during the study period using nonparametric statistical procedures. The Friedman test for dependent samples is employed to test the hypothesis of no difference in efficiency scores obtained from 2009 to 2013 in each type of efficiency. The results of the Friedman test specify significant differences in CRS and scale efficiencies (with p-values of 0,001 and 0,0006 respectively) while the VRS differences were insignificant at any confidence level. Therefore, the observed differences in scale efficiencies of hospitals are based mostly to the deviations of the pure technical efficiency. Using again the Friedman test the differences in the efficiency scores between the consecutive years of the study, we observe that a significant change appeared in the period 2010-2011 for the CRS and scale efficiency scores (p-values: 0,007 and 0,0001). In addition the scale efficiency score significantly change in 2012-2013 (p-value=0,003).

The results pertaining to returns-to-scale in Greek hospitals highlight that the predominant form of scale inefficiency is the decreasing returns-to-scale imply that a hospital has an inefficient large size. More specifically, the average number of hospitals with decreasing returns to scale was 10 imply that these hospital has an inefficient large size. To decrease the consequently high unit cost and come back to their optimal

		CRS efficiencies					VRS efficiencies					Scale efficiencies				
	No	**2009**	**2010**	**2011**	**2012**	**2013**	**2009**	**2010**	**2011**	**2012**	**2013**	**2009**	**2010**	**2011**	**2012**	**2013**
All hospitals	24	0,82	0,82	0,87	0,87	0,86	0,88	0,90	0,92	0,91	0,92	0,93	0,92	0,95	0,96	0,94
Non university	17	0,78	0,81	0,86	0,86	0,85	0,86	0,90	0,92	0,91	0,92	0,90	0,90	0,93	0,94	0,92
University	7	0,92	0,87	0,89	0,91	0,91	0,93	0,90	0,90	0,92	0,92	0,98	0,96	0,99	0,99	0,99

Table 2: Efficiency results.

scale, such hospitals need to scale down their size (beds and staff). For those hospitals, it might be politically difficult to consider bed closures or decrease in their personnel, but at least they should not keep increasing them. However, because of the non competitive environment in the health public sector operating at decreasing returns to scale can be a choice dictated by the demand side.

In contrast, the average number of units with increasing returns to scale was only 2,4. These hospitals have not reached their optimal scale yet, which is mostly observed in small-size hospitals (average inputs bellow sample average). In the presence of increasing returns to scale, there is still room for expansion of outputs and reduction of unit costs. A hospital with increasing returns to scale will, therefore, benefit from augmenting its scale of operations. However, increasing the level of outputs requires an increase in demand for health, which often is beyond the hospital management's control.

Comparing the efficiencies between university and non-university hospitals we observe that the university hospitals perform better in all years of the study period. Regarding the scale efficiencies, it is evident that the university hospitals operate very close to their optimal scale. Hence the university hospitals operate with full utilization of inputs and optimal scale.

IV. Conclusion

This study has attempted to assess the technical and scale efficiency of the large Greek public hospitals in the recent time period. The efficiency scores are computed using DEA, which is a non-parametric method used to model the relationship between multiple inputs and outputs for a decision making unit (DMU), it then provides estimates of the potential improvement that can be made in inefficient units, in our case inefficient hospitals.

From our analysis of overall technical efficiency, pure technical efficiency, and scale efficiency we find that on average, the Greek public hospitals improve their operations during the study period. These figures are good indicators of the total costs that could be saved since technical and scale efficiency are necessary conditions for cost minimization.

References

[1] P. Mladovsky, D. Srivastava, J. Cylus, M. Karanikolos, T.Evetovits, S. Thomson, and M. McKee, "Health policy responses to the financial crisis in Europe," Geneva: World Health Organization, European Observatory on Health Systems and Policies, Policy Summary 5, 2012.

[2] B. Hollingsworth, "Non-parametric and parametric applications measuring efficiency in health care," Health Care Management Science, vol. 6, pp. 203-218, 2003.

[3] L. O'Neill, M. Rauner, K. Heidenberger, and M. Kraus, "A cross-national comparison and taxonomy of DEA-based hospital efficiency studies," Socio- Economic Planning Sciences, vol.42, no. 3, pp. 158-189, 2008.

[4] A. Charnes, .W.W. Cooper,and E. Rhodes, "Measuring efficiency of decision making units," *European Journal of Operations Research*, vol. 2, pp. 429-444, 1978.

[5] P. Mitropoulos, M. Talias, and I. Mitropoulos, "Combining stochastic DEA with Bayesian analysis to obtain statistical properties of the efficiency scores: An application to Greek public Hospitals," European Journal of Operational Research vol. 243, no. 1, pp. 302-311, 2015.

[6] P. Mitropoulos, I. Mitropoulos, and A. Sissouras, A., "Managing for efficiency in health care: The case of Greek public hospitals," European Journal of Health Economics, vol. 14, pp. 929-938, 2013.

The Unified Transform for a Reaction-Diffusion Brain Tumor Model that Incorporates Tissue Heterogeneity and Radiotherapy

A.G. Sifalakis, M.G. Papadomanolaki, E.P. Papadopoulou and Y.G. Saridakis

Abstract—Gliomas are primary brain tumors characterized by rapid growth and aggressive diffusive behavior. Radiation therapy, following extensive surgical resection, is included among the standard protocols for the treatment of this kind of malignant tumors. Several mathematical models have been developed to approximate the evolution of gliomas. In this paper we consider a linear reaction-diffusion tumor growth problem in $1 + 1$ dimensions that, except from the heterogeneity of the brain tissue, takes into consideration the effect of radiotherapy treatment. Extending recent results, our main objective is, by utilizing the unified transform, to obtain an integral representation of the solution that also incorporates the effect of radiotherapy. Among several advantages of the unified transform is the fact that one may recover the value of the solution at any point (x, t) directly, without prior knowledge of the solution at any previous time level other than the initial. Simple trapezoidal rule on appropriate hyperbolic contours leads to efficient numerical evaluation of the integral representation.

Keywords—Gliomas, Radiotherapy, Reaction-Diffusion PDEs, Fokas Unified Transform, Numerical Integration.

I. Introduction

GLIOMAS, one of the most common and aggressive forms of primary brain tumors, are well known for their rapid growth and highly diffuse invasion of adjacent normal tissue. To study the core properties of motile glioma cells, namely migration and proliferation, mathematical models (cf. e.g. [3], [5], [20], [23]; for a review see also [13], [11] and [10]) considered reaction-diffusion PDEs and, based on CT-scan data, calculated the values of the diffusion and proliferation parameters. Brain's tissue heterogeneity (gray and white matter) was incorporated later in the basic model by making use of a properly discontinuous diffusion coefficient (cf. [18], [19]).

Recently, in [15], [16] and [4], the model was further extended to also incorporate the effects of radiotherapy as, combined with surgery and chemotherapy, it is considered to be a standard treatment regime.

The unified transform, a new method for solving linear and integrable nonlinear PDEs, was introduced in [6]- [7], and since then has been studied and further developed by many researchers (see [8] for a review). The method is characterized by novel integral representations of the solution in the complex k-plane which are uniformly convergent and, via contour deformation, decay exponentially.

The implementation of the unified transform method for brain tumor models with discontinuous diffusion coefficient was introduced in [12] and further studied in [1] and [2]. Our main objective of the present work is to further extend our results on the unified transform method to include also heterogeneous brain tumor models that incorporate the effect of radiotherapy. Our results refer on the $1+1$ dimensions case as the work for higher dimensions is still in progress.

II. Methodology

A. Mathematical model

Assuming exponential tumor growth and a simple log-kill radiotherapy model, the core reaction-diffusion PDE of the mathematical model, considered here, takes the form (cf. [19], [16], [15]):

$$\frac{\partial \bar{c}}{\partial \bar{t}} = \nabla \cdot (\bar{D} \nabla \bar{c}) + \bar{\rho}\, \bar{c} - \bar{R}(\bar{t})\, \bar{c}, \tag{1}$$

where $\bar{c}(\bar{x}, \bar{t})$ denotes the tumor cell density at location $\bar{x} \in \mathbb{R}^n$ $(n = 1, 2, 3)$ and time $\bar{t}$, $\bar{\rho}$ stands for the net proliferation rate (cf. [3]), $\bar{D}$ is the diffusion coefficient representing the active motility of malignant cells (cf. [20]) and $\bar{R}(\bar{t})$ describes the effect of radiotherapy. The dimensions of the above variables are:

$$\left\{ \begin{array}{lll} [\bar{x}] = cm, & [\bar{t}] = day, & [\bar{c}] = \dfrac{cells}{cm^n}, \\ [\bar{D}] = \dfrac{cm^2}{day}, & [\bar{\rho}] = \dfrac{1}{day}, & [\bar{R}] = \dfrac{1}{day} \end{array} \right. \tag{2}$$

At the boundary we consider zero flux, which impose no migration of cells beyond the brain boundaries, and an initial condition $\bar{c}(\bar{x}, 0) = \bar{f}(\bar{x})$, where $\bar{f}(\bar{x})$ is the initial spatial distribution of malignant cells.

Due to the heterogeneity of the brain tissue the diffusion coefficient $\bar{D}$ is defined by (cf. [18], [19]) :

$$\bar{D} \equiv \bar{D}(\bar{x}) = \left\{ \begin{array}{ll} D_w, & \bar{x} \text{ in white matter } (\bar{x} \in \bar{\Omega}_w) \\ D_g, & \bar{x} \text{ in gray matter } (\bar{x} \in \bar{\Omega}_g) \end{array} \right. , \tag{3}$$

where D_w and D_g are scalars with $D_w > D_g$.

Manuscript received June 15, 2015; revised July 10, 2015.

This work was supported by the ESF and Greek national funds through the operational program *Education and Lifelong Learning* of the National Strategic Reference Framework (NSRF) THALES (Grant number: MIS-379416).

All authors are with the Applied Mathematics & Computers Laboratory, Technical University of Crete, 73100 Chania, Crete, Greece

Email of the corresponding author : y.saridakis@amcl.tuc.gr

Considering a low-dose-rate and fractionated radiotherapy, activated in the time interval $(\bar{T}_G, \bar{T}_R]$, the effect of radiotherapy is described by (cf. [15]):

$$\bar{R} \equiv \bar{R}(\bar{t}) = R_{\text{eff}} k_R(\bar{t}) \ , \tag{4}$$

where $k_R(\bar{t})$ denotes the temporal profile of the radiation schedule and, by using a time step of one day, takes the value one on the radiotherapy days and zero otherwise, that is

$$k_R(\bar{t}) = \begin{cases} 1, & \bar{t} \in (\bar{T}_G, \bar{T}_R] \\ 0, & \bar{t} \notin (\bar{T}_G, \bar{T}_R] \end{cases} . \tag{5}$$

R_{eff} denotes the effect of n fractions of radiation per day and is described by (cf. [15] and the references therein)

$$R_{\text{eff}} = \alpha(nd) + 2\beta nd^2 \left[g(\mu\tau) + 2\frac{\cosh(\mu\tau) - 1}{(\mu\tau)^2} h_n(\phi) \right],$$

with

$$g(\mu\tau) = \frac{\mu\tau - 1 + e^{-\mu\tau}}{(\mu\tau)^2} \text{ and } h_n(\phi) = \frac{(n - 1 - n\phi + \phi^n)\phi}{n(1-\phi)^2}$$

where α, β are sensitivity parameters, d is the dose rate at time $\bar{t}$, μ is the half time for repair of radiation-induced DNA damage, τ is the irradiation duration and $\phi = e^{-\mu(\tau + \Delta\tau)}$ with $\Delta\tau$ denoting the time interval between fractions. We point out that the values of the parameters used in all the above relations may be found, for example, in [15] (see the list of parameter values included in Table 1 of [15] and the corresponding references therein).

Working towards the direction of describing the model problem in $(\bar{x}, \bar{t})$ regions where $\bar{c}$ may be considered analytic inside and continuous on the boundary, let us first define $\bar{c}$ on three consecutive time regions $\bar{t}_\ell, \ \ell = 1, 2, 3$ as follows:

$$\begin{cases} \bar{c}\left(\bar{x}, \bar{t}_1\right) = \bar{c}\left(\bar{x}, \bar{t}\right) & , \quad \bar{t} \in (0, \bar{T}_G] \\ \bar{c}\left(\bar{x}, \bar{t}_2\right) = \bar{c}\left(\bar{x}, \bar{t} - \bar{T}_G\right) & , \quad \bar{t} \in (\bar{T}_G, \bar{T}_R] \\ \bar{c}\left(\bar{x}, \bar{t}_3\right) = \bar{c}\left(\bar{x}, \bar{t} - \bar{T}_R\right) & , \quad \bar{t} \in (\bar{T}_R, \bar{T}_F] \end{cases} . \tag{6}$$

Using, now, the above notation, the model problem in 1+1 dimensions is written as:

$$\begin{cases} \dfrac{\partial \bar{c}}{\partial \bar{t}_\ell} = \left(\bar{D}\bar{c}_{\bar{x}}\right)_{\bar{x}} + \bar{\rho}_\ell \, \bar{c} \ , \ \bar{x} \in [\bar{a}, \bar{b}] \ , \ 0 < \bar{t}_\ell \leq \bar{T}_\ell \\ \bar{c}\left(\bar{x}, 0\right) = \bar{c}_\ell(\bar{x}) \\ \bar{c}_{\bar{x}}\left(\bar{a}, \bar{t}_\ell\right) = \bar{c}_{\bar{x}}\left(\bar{b}, \bar{t}_\ell\right) = 0 \end{cases} \tag{7}$$

where

$$\begin{cases} \bar{\rho}_1 = \bar{\rho} & , \quad \bar{T}_1 = \bar{T}_G & , \quad \bar{c}_1(\bar{x}) = \bar{f}(\bar{x}) \\ \bar{\rho}_2 = \bar{\rho} - R_{\text{eff}} & , \quad \bar{T}_2 = \bar{T}_R - \bar{T}_G & , \quad \bar{c}_2(\bar{x}) = \bar{c}(\bar{x}, \bar{T}_1) \\ \bar{\rho}_3 = \bar{\rho} & , \quad \bar{T}_3 = \bar{T}_F - \bar{T}_R & , \quad \bar{c}_3(\bar{x}) = \bar{c}(\bar{x}, \bar{T}_2) \end{cases} . \tag{8}$$

B. Dimensionless Variables and Equivalence Transformations

The dimensionless form of the IBVPs in (7) is given by

$$\begin{cases} \dfrac{\partial c}{\partial t_\ell} = \left(D \, c_x\right)_x + \rho_\ell c \ , \ x \in [a, b] \ , \ 0 < t_\ell \leq T_\ell \\ c\left(x, 0\right) = c_\ell\left(x\right) \\ c_x\left(a, t_\ell\right) = c_x\left(b, t_\ell\right) = 0 \end{cases} \tag{9}$$

where (cf. [18], [2])

$$\begin{cases} x = \chi\bar{x}, \quad a = \chi\bar{a}, \quad b = \chi\bar{b}, \quad t_\ell = \bar{\rho}\bar{t}_\ell, \\ c\left(x, t_\ell\right) = \dfrac{1}{\chi N_0} \ \bar{c}\left(\chi\bar{x}, \bar{\rho}\bar{t}_\ell\right) \\ c_\ell\left(x\right) = \dfrac{1}{\chi N_0} \bar{c}_\ell\left(\chi\bar{x}\right) \\ D = \dfrac{\bar{D}}{D_w} \ , \ \rho_\ell = \dfrac{\bar{\rho}_\ell}{\bar{\rho}} \end{cases} \tag{10}$$

with

$$\chi = \sqrt{\frac{\bar{\rho}}{D_w}} \text{ and } N_0 = \int_{\bar{a}}^{\bar{b}} \bar{f}\left(\bar{x}\right) d\bar{x}, \tag{11}$$

and, obviously, $T_j = \bar{\rho}\bar{T}_j$. Also, observe that N_0 denotes the initial number of glioma cells in $[\bar{a}, \bar{b}]$.

Furthermore, upon immediate application of the corresponding result in [2], we also have that:

Lemma 1. *If $c(x, t_\ell), \ \ell = 1, 2, 3$ satisfies the IBVP in (9)-(11) and $u(x, t_\ell)$ is defined by*

$$u(x, t_\ell) = e^{-\rho_\ell t_\ell} c(x, t_\ell), \tag{12}$$

then $u(x, t_\ell), \ \ell = 1, 2, 3$ satisfies the IBVP

$$\begin{cases} \dfrac{\partial u}{\partial t_\ell} = \left(D \, u_x\right)_x \ , \ x \in [a, b] \ , \ 0 < t_\ell \leq T_\ell \\ u\left(x, 0\right) = u_\ell\left(x\right) \equiv c_\ell\left(x\right) \\ u_x\left(a, t_\ell\right) = u_x\left(b, t_\ell\right) = 0 \end{cases} . \tag{13}$$

C. The Unified Transform

To proceed, now, with the application of the Unified Transform for the solution of the IBVPs in (13), let us first fix notation with brain's heterogeneity regions of white Ω_w and gray Ω_g matter inside the interval $[a, b]$. Namely, as in [2], we shall consider $[a, b]$ partitioned into $n + 1$ sub-intervals $R_j := (w_{j-1}, w_j)$, with $a \equiv w_0 < w_1 < w_2 < \ldots < w_n < w_{n+1} \equiv b$, and if $R_j \subseteq \Omega_w$, for some j, then $R_{j-1} \subseteq \Omega_g$ and $R_{j+1} \subseteq \Omega_g$. With this notation, the diffusion coefficient D, defined in (10), takes the form

$$D(x) = \gamma_j = \begin{cases} \gamma, & \text{when } \ x \in \Omega_g \\ 1, & \text{when } \ x \in \Omega_w \end{cases} , \tag{14}$$

where $\gamma = D_g / D_w$. As, now, the parabolic nature of the problem directly implies continuity of both u and Du_x across each interface point w_j, for each $j = 1, 2, \ldots, n$ and $\ell = 1, 2, 3$, there holds

$$\begin{cases} \lim\limits_{x \to w_j^+} u(x, t_\ell) = \lim\limits_{x \to w_j^-} u(x, t_\ell) \\ \lim\limits_{x \to w_j^+} D(x) u_x(x, t_\ell) = \lim\limits_{x \to w_j^-} D(x) u_x(x, t_\ell) \end{cases} . \tag{15}$$

Let $u^{(j)}(x, t_\ell)$ denote the solution of the problem defined in Lemma 1 over the region $[w_{j-1}, w_j] \times [0, T_\ell]$. Observing that its analyticity and continuity properties in the interior and on the boundaries on this region allow Green's and Cauchy's theorems to be applied, immediate application of our analysis in [2] implies:

Proposition 1. *If $u^{(j)}(x, t_\ell)$, for each $j = 1, 2, \ldots, n$ and $\ell = 1, 2, 3$, denotes the solution of the IBVP defined in Lemma 1 over the region $[w_{j-1}, w_j] \times [0, T_\ell]$ and $k \in \mathbb{C}$, then*

$$\begin{aligned} u^{(j)}(x, t_\ell) \;=\; & \frac{c_j}{2\pi} \int_{-\infty}^{+\infty} e^{ic_j kx - k^2 t_\ell} \widehat{u}_\ell^{(j)}(c_j k) dk \\ & - \frac{1}{2\pi c_j} \int_{\partial\Gamma^+} e^{ic_j k(x - w_{j-1}) - k^2 t_\ell} \\ & \cdot [\widetilde{u}_x^{(j)}(w_{j-1}, k^2) + ic_j k \widetilde{u}^{(j)}(w_{j-1}, k^2)] dk \\ & - \frac{1}{2\pi c_j} \int_{\partial\Gamma^-} e^{ic_j k(x - w_j) - k^2 t_\ell} \\ & \cdot [\widetilde{u}_x^{(j)}(w_j, k^2) + ic_j k \widetilde{u}^{(j)}(w_j, k^2)] dk \ , \end{aligned} \tag{16}$$

where

- $c_j = 1/\sqrt{\gamma_j}$
- *Γ^+ and Γ^- denote the contours (see also Fig. 1)*

$$\Gamma^+ = \{k \in \mathbb{C} : arg(k) \in (\frac{\pi}{4}, \frac{3\pi}{4})\},$$

$$\Gamma^- = \{k \in \mathbb{C} : arg(k) \in (\frac{5\pi}{4}, \frac{7\pi}{4})\},$$

- *$u_\ell^{(j)}(x)$ are the initial data, defined in Lemma 1, restrained over region $[w_{j-1}, w_j]$, and $\widehat{u}_\ell^{(j)}(x)$ denotes its Fourier transform, defined by*

$$\widehat{u}_\ell^{(j)}(k) = \int_{w_{j-1}}^{w_j} e^{-ikx} u_\ell^{(j)}(x) dx \ . \tag{17}$$

The quantities $\widetilde{u}^{(j)}$ and $\widetilde{u}_x^{(j)}$ are given by the solution of the $(2n+2) \times (2n+2)$ complex linear system

$$GU = F, \tag{18}$$

where the nonzero elements of the matrix $G = \{G_{p,q}\}$ are defined by:

- *for $j = 1$:*

$$\begin{bmatrix} G_{1,1} & G_{1,2} & G_{1,3} \\ G_{2,1} & G_{2,2} & G_{2,3} \end{bmatrix} = \begin{bmatrix} A_1^{(1)} & A_3^{(1)} & A_4^{(1)} \\ A_5^{(1)} & A_7^{(1)} & A_8^{(1)} \end{bmatrix} \tag{19}$$

- *for $j = 2, 3, \ldots, n$:*

$$\begin{bmatrix} G_{2j-1,2j-2} & G_{2j-1,2j-1} & G_{2j-1,2j} & G_{2j-1,2j+1} \\ G_{2j,2j-2} & G_{2j,2j-1} & G_{2j,2j} & G_{2j,2j+1} \end{bmatrix} = \\ = \begin{bmatrix} A_1^{(j)} & A_2^{(j)} & A_3^{(j)} & A_4^{(j)} \\ A_5^{(j)} & A_6^{(j)} & A_7^{(j)} & A_8^{(j)} \end{bmatrix} \tag{20}$$

- *for $j = n + 1$:*

$$\begin{bmatrix} G_{2n+1,2n} & G_{2n+1,2n+1} & G_{2n+1,2n+2} \\ G_{2n+2,2n} & G_{2n+2,2n+1} & G_{2n+2,2n+2} \end{bmatrix} = \\ = \begin{bmatrix} A_1^{(n+1)} & A_2^{(n+1)} & A_3^{(n+1)} \\ A_5^{(n+1)} & A_6^{(n+1)} & A_7^{(n+1)} \end{bmatrix} \tag{21}$$

with

m	$A_m^{(j)}$	$A_{m+1}^{(j)}$
1	$ic_j\gamma_j k e^{-ic_j k w_{j-1}}$	$\gamma_{j-1} e^{-ic_j k w_{j-1}}$
3	$-ic_j\gamma_j k e^{-ic_j k w_j}$	$-\gamma_j e^{-ic_j k w_j}$
5	$-ic_j\gamma_j k e^{ic_j k w_{j-1}}$	$\gamma_{j-1} e^{ic_j k w_{j-1}}$
7	$ic_j\gamma_j k e^{ic_j k w_j}$	$-\gamma_j e^{ic_j k w_j}$

and

$$U = \begin{bmatrix} \widetilde{u}^{(1)}(a, k^2) \\ \widetilde{u}^{(1)}(w_1, k^2) \\ \widetilde{u}_x^{(1)}(w_1, k^2) \\ \vdots \\ \widetilde{u}^{(n)}(w_n, k^2) \\ \widetilde{u}_x^{(n)}(w_n, k^2) \\ \widetilde{u}^{(n+1)}(b, k^2) \end{bmatrix} , \ F = \begin{bmatrix} \widehat{f}^{(1)}(c_1 k) \\ \widehat{f}^{(1)}(-c_1 k) \\ \vdots \\ \widehat{f}^{(n+1)}(c_{n+1} k) \\ \widehat{f}^{(n+1)}(-c_{n+1} k) \end{bmatrix} .$$

D. Numerical Integration Contours and Integral Properties

It is known (cf. [21], [22]; see also [9], [14], [2]) that for the efficient numerical evaluation of the above integrals in (16) one may apply the trapezoid rule on suitable hyperbolic contours. For this, we deform (cf. e.g. [2]) the integration paths $\partial\Gamma^\pm$ to hyperbolas of the complex plane by the mapping:

$$k_\theta \equiv k(\theta) := i \sin(\beta - i\theta), \tag{22}$$

where the angle β is chosen to be $\beta = \pi/6$ and the curves $\pm k(\theta)$ are shown schematically in Fig. 1 that follows.

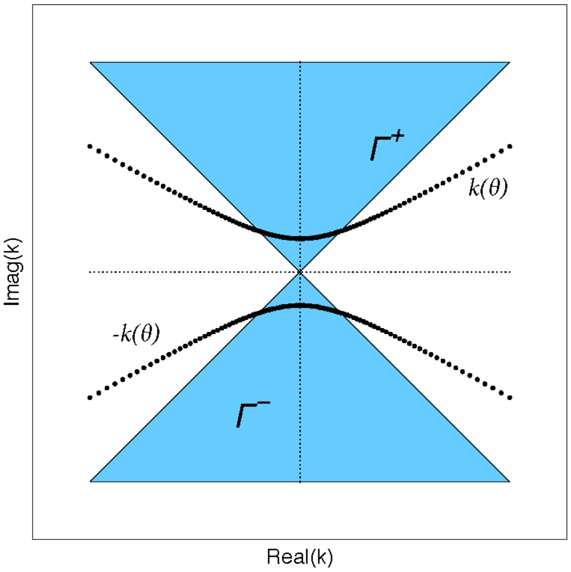


Fig. 1: The contours $\pm k(\theta)$ for numerical integration

The above mapping leads us in rewriting the solution (16) as

$$\begin{aligned} u_\ell^{(j)}(x,t_\ell) \;=\; & \frac{c_j}{2\pi}\int_{-\infty}^{+\infty} e^{ic_jkx-k^2t_\ell}\widehat{u}_\ell^{(j)}(c_jk)dk \\ & -\frac{1}{2\pi c_j}\int_{-\infty}^{+\infty} e^{ic_jk_\theta(x-w_{j-1})-k_\theta^2t_\ell} \\ & \cdot[\widetilde{u}_x^{(j)}(w_{j-1},k_\theta^2)+ic_jk_\theta\widetilde{u}^{(j)}(w_{j-1},k_\theta^2)]k'_\theta dk_\theta \\ & -\frac{1}{2\pi c_j}\int_{-\infty}^{+\infty} e^{-ic_jk_\theta(x-w_j)-k_\theta^2t_\ell} \\ & \cdot[\widetilde{u}_x^{(j)}(w_j,k_\theta^2)-ic_jk_\theta\widetilde{u}^{(j)}(w_j,k_\theta^2)]k'_\theta dk_\theta \;, \end{aligned} \tag{23}$$

where k'_θ denotes the derivative of $k(\theta)$.

For the efficient implementation of the numerical quadrature rules - in particular for the evaluation of the last two integrals - one has to take into consideration (cf. [2] for example) basic algebraic properties such as:

- The real parts of the integrands are *even* functions of θ.
- The imaginary parts of the integrands are *odd* functions of θ.
- The integrands are decaying functions of θ.

Application of the above properties directly implies that

$$\int\limits_{-\infty}^{\infty} U(\theta)d\theta = 2\int\limits_{0}^{\infty}\mathrm{Re}\left(U(\theta)\right)d\theta \approx 2\int\limits_{0}^{\Theta}\mathrm{Re}\left(U(\theta)\right)d\theta \;,$$

where $U(\theta)$ denotes any one of the last two integrands involved in (23) and Θ is a relatively *small* real number. For a good estimate of Θ one may require the dominant exponential term $e^{-k_\theta^2\tau}$, common in all integrals, to satisfy

$$\left|e^{-k_\theta^2\tau}\right| \le 10^{-M} \quad \text{for all} \quad \theta \ge \Theta \equiv \Theta(\tau;M)$$

for sufficiently large M, hence (cf. [12])

$$\Theta = \frac{1}{2}\ln\frac{4\tau+8M\ln 10}{\tau} \quad . \tag{24}$$

III. Numerical Experiments

Two different numerical experiments are included in this section to visually demonstrate the behavior of their semi-analytical solution by an effective combination of the unified transform and numerical quadrature rules on hyperbolic contours. We would like to clarify that said model problems are virtual cases and have no relevance with real patient data.

The per day radiotherapy protocol, followed in both models, is identical. Namely, we assumed that the administered per day radiation dose is $d = 1.8$Gy and, by using the parameter values (cf. [15]) $\alpha = 0.027$, $\beta = 0.0027$, $n = 1$, $\mu = 11.4$, $\tau = 0.0083$, $\Delta\tau = 1$, the radiation coefficient in both models satisfies $R_{\text{eff}} = 0.05707849$.

A. Model Problem I

Referring to the model problem in (7), consider the values:

$$\begin{cases} \bar{a} = -10 \text{ cm},\; \bar{b} = 10 \text{ cm},\; \bar{w}_1 = -5 \text{ cm},\; \bar{w}_2 = 5 \text{ cm} \\ \bar{\Omega}_g = [\bar{a},\bar{w}_1)\cup(\bar{w}_2,\bar{b}] \text{ and } \bar{\Omega}_w = [\bar{w}_1,\bar{w}_2] \\ D_g = 0.0013 \text{ cm}^2\text{day}^{-1},\; D_w = 0.0065 \text{ cm}^2\text{day}^{-1} \\ \bar{\rho} = 0.012 \text{ day}^{-1},\; N_0 = 100 \text{ cells} \end{cases} \tag{25}$$

The initial distribution of cells is considered to be

$$\bar{f}(\bar{x}) = N_0\delta(\bar{x}),$$

where $\delta(\bar{x})$ denotes Dirac's delta.

We considered a radiotherapy period of 35 days, started on $\bar{T}_G = 180$ day and finished on $\bar{T}_R = 215$ day, during which the total administered radiation dose is 63Gy.

The results from applying a simple trapezoidal rule, using 50 quadrature points, for the evaluation of each one of the integrals in relation (23) are depicted in Fig. 2 and Fig. 3.

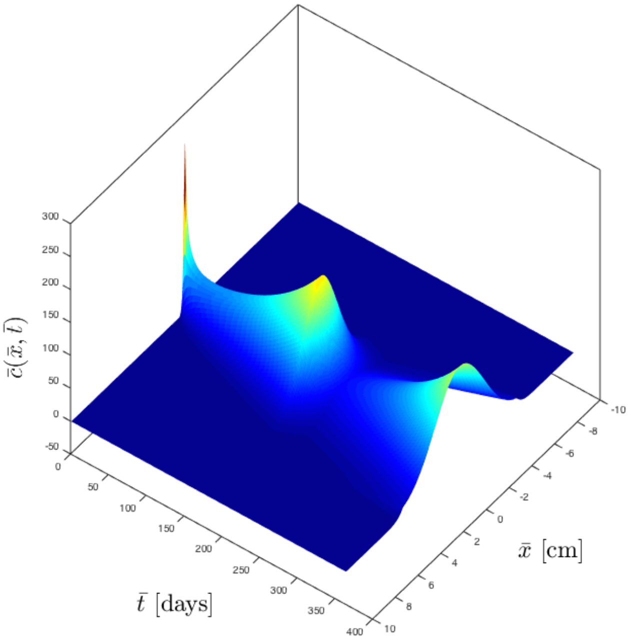


Fig. 2: Time evolution of the cell density $\bar{c}(\bar{x},\bar{t})$ for one glioma cell source. Radiotherapy period of 35 days (from 180 to 215).

More specifically, in Fig. 2, the evolution of the cell density function $\bar{c}(\bar{x},\bar{t})$ is depicted for a total period of $\bar{T}_F = 365$ days. By inspection, one may easily recognize the time periods the tumor grows without treatment, hence diffusion and proliferation dominate, as well as the time period of radiotherapy.

The radiotherapy effect on the total number of tumor cells $N(\bar{t})$, defined by

$$N(\bar{t}) = \int_{\bar{a}}^{\bar{b}} \bar{c}(\bar{x},\bar{t})d\bar{x} \ ,$$

is depicted in Fig. 3. The differential between treated and untreated tumor growth reveals that the administered radiotherapy extended survival by 166 days.

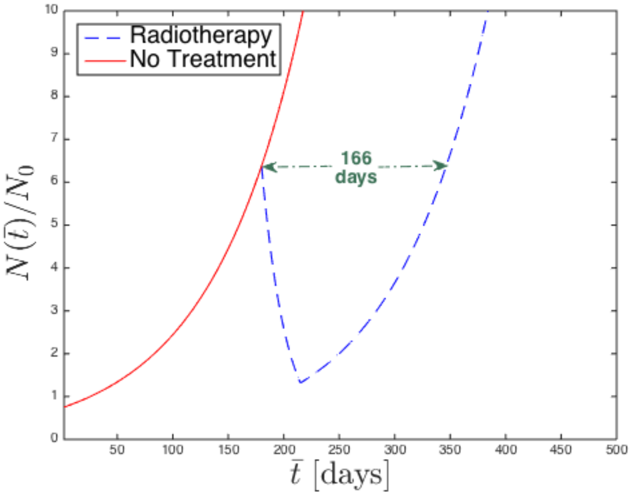


Fig. 3: The effect of radiotherapy on the total number of tumor cells.

B. Model Problem II

In this problem we consider four initial point sources of malignant cells. Referring, again, to the model problem in (7), we consider the values:

$$\begin{cases} \bar{a} = -10 \text{ cm}, \ \bar{b} = 10 \text{ cm} \\ \bar{w}_1 = -6 \text{ cm}, \ \bar{w}_2 = -5 \text{ cm}, \ \bar{w}_3 = 1 \text{ cm}, \ \bar{w}_4 = 7 \text{ cm} \\ \bar{\Omega}_g = [\bar{a},\bar{w}_1) \cup (\bar{w}_2,\bar{w}_3) \cup (\bar{w}_4,\bar{b}] \\ \bar{\Omega}_w = [\bar{w}_1,\bar{w}_2] \cup [\bar{w}_3,\bar{w}_4] \\ D_g = 0.0013 \text{ cm}^2\text{day}^{-1}, \ D_w = 0.0065 \text{ cm}^2\text{day}^{-1} \\ \bar{\rho} = 0.012 \text{ day}^{-1}, \ N_0 = 400 \text{ cells} \end{cases} \tag{26}$$

and the initial distribution of tumor cells is given by

$$\bar{f}(\bar{x}) = \frac{N_0}{4}\left[\delta(\bar{x}+8) + \delta(\bar{x}+3) + \delta(\bar{x}-4) + \delta(\bar{x}-6)\right].$$

We implemented a radiotherapy period of 40 days, started on $\bar{T}_G = 180$ day and finished on $\bar{T}_R = 220$ day, during which the total administered radiation dose is 72Gy.

The results from applying a trapezoidal rule, using 50 quadrature points, for the evaluation of each one of the integrals in relation (23) are depicted in Fig. 4 and Fig. 5.

As in model problem I, Fig. 4 depicts the evolution of the cell density function $\bar{c}(\bar{x},\bar{t})$ for a total period of $\bar{T}_F = 365$ days. Again, the time periods the tumor grows without treatment as well as the time period of radiotherapy are easily recognizable.

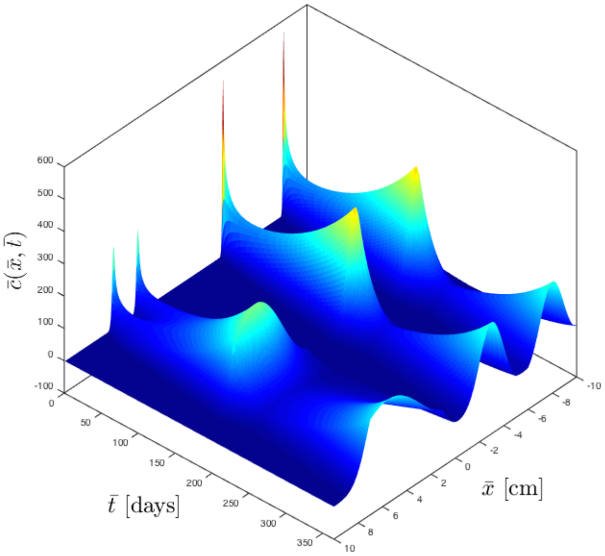


Fig. 4: Time evolution of the cell density $\bar{c}(\bar{x},\bar{t})$ for four glioma cell sources. Radiotherapy period of 40 days (from 180 to 220).

The radiotherapy effect on the total number of tumor cells $N(\bar{t})$ is depicted in Fig. 5. The differential between treated and untreated tumor growth reveals that the administered radiotherapy extended survival by 190 days.

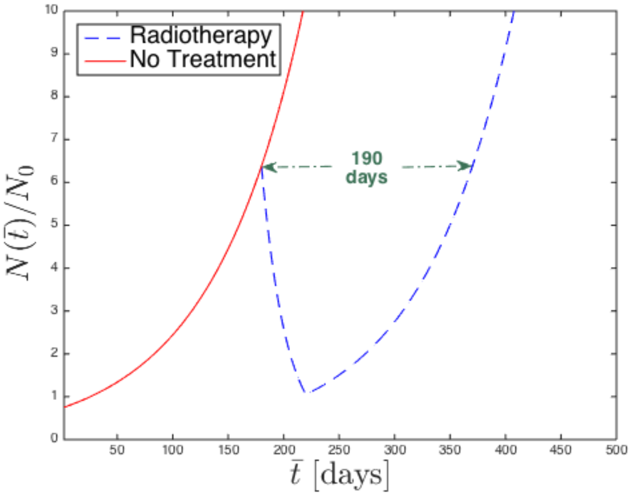


Fig. 5: The effect of radiotherapy on the total number of tumor cells.

IV. Conclusions

In the present work, extending recently produced results (cf. [12], [1], [2]), we used the Unified Transform, as well as the trapezoidal rule on hyperbolic contours, to derive and evaluate novel integral representations of a linear reaction-diffusion problem's solution that models the growth of aggressive brain tumors, in a heterogeneous environment, and incorporates the effect of radiotherapy. The results, although limited to the 1+1 dimension case, reveal the potential of the method to describe exactly the solution at any space-time point without depending on any other data apart from the initial data of all stages. And this, not only for domains that solutions remain smooth, but for multi-domain environments that include discontinuities of the solution's partial derivatives not only in space but in time as well, as we showed here. In this way we clearly enlight the path to effectively solve more general models that incorporate surgical resection and chemotherapy as well, and, of course, contribute to the solution of the problem in 2+1 dimensions.

Acknowledgement

The present research work has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES (Grant number: MIS-379416). Investing in knowledge society through the European Social Fund.

References

[1] M. Asvestas, A. G. Sifalakis, E.P. Papadopoulou and Y. G. Saridakis, *Fokas method for a multi-domain linear reaction-diffusion equation with discontinuous diffusivity*, IOP Science Journal of Physics: Conference Series, 490, 012143, 2014.

[2] M. Asvestas, E.P. Papadopoulou, A. G. Sifalakis and Y. G. Saridakis,, *The Unified Transform for a Class of Reaction-Diffusion Problems with Discontinuous Time Dependent Parameters*, Proceedings of the World Congress on Engineering 2015.

[3] J. Cook, D. E. Woodward, P. Tracqui and J. D. Murray, *Resection of gliomas and life expectancy*, J Neurooncol., 24, 131, 1995.

[4] D. Corwin, C. Holdsworth, R. C. Rockne, An. D. Trister, M. M. Mrugala, J. K. Rockhill, R. D. Stewart2, M. Phillips2 and K. R. Swanson, *Toward Patient-Specific, Biologically Optimized Radiation Therapy Plans for the Treatment of Glioblastoma*, PLOS ONE, 8(11), 1-9, 2013.

[5] G.C. Cruywagen, D.E. Woodward, P. Tracqui, G.T. Bartoo, J.D. Murray and E.C. Alvord Jr, *The modeling of diffusive tumors*, Journal of Biological Systems , vol.3, pp.937-945, 1995.

[6] A. S. Fokas, *A unified transform method for solving linear and certain nonlinear PDEs*, Proc.R.Soc. A, 453, 1411-1443, 1997.

[7] A. S. Fokas, *A new transform method for evolution PDEs*, IMA J. Appl. Math.,67(6), 559-590, 2002.

[8] A. S. Fokas, *A Unified Approach to Boundary Value Problems*, SIAM, Philadelphia, 2008.

[9] N. Flyer and A. S. Fokas, *A hybrid analytical-numerical method for solving evolution partial differential equations I: The half-line*, Proc. R. Soc. A, 464, 1823-1849, 2008.

[10] H. L. P. Harpold, E. C. Alvord Jr and K. R. Swanson, *The Evolution of Mathematical Modeling of Glioma Proliferation and Invasion*, J Neuropathol Exp Neurol, 66(1), 1-9, 2007.

[11] U. Ledzewicz, H. Schttler, A. Friedman and E. Kashdan, *Mathematical Methods and Models in Biomedicine*, Springer Science and Business Media, 2012.

[12] D. Mantzavinos, M. G. Papadomanolaki, Y. G. Saridakis and A. G. Sifalakis, *Fokas transform method for a brain tumor invasion model with heterogeneous diffusion in 1+1 dimensions*, Applied Numerical Mathematics (http://dx.doi.org/10.1016/j.apnum.2014.09.006), 2014.

[13] J.D. Murray, *Mathematical Biology I and II*, Springer-Verlag, 3rd Edition 2002.

[14] T.S. Papatheodorou and A. N. Kandili, *Novel numerical techniques based on Fokas transforms, for the solution of initial boundary value problems*, Journal of Computational and Applied Mathematics 227:75-82, 2009.

[15] G. Powathil, M. Kohandel, S. Sivaloganathan, A. Oza and M. Milosevic, *Mathematical modeling of brain tumors: effects of radiotherapy and chemotherapy*,Phys. Med. Biol. 52 :3291-3306, 2007.

[16] R. Rockne, E. C. Alvord Jr., J. K. Rockhill and K. R. Swanson, *A mathematical model for brain tumor response to radiation therapy*, J. Math. Biol., 58, 561578, 2009.

[17] D.A. Smith, *Well-posed two-point initial-boundary value problems with arbitrary boundary conditions*, Math. Proc. Camb. Philos. Soc. 152:473496, 2012.

[18] K. R. Swanson, *Mathematical modeling of the growth and control of tumors*, PHD Thesis, University of Washington, 1999.

[19] K. R. Swanson, E. C. Alvord Jr and J. D. Murray, *A quantitive model for differential motility of gliomas in grey and white matter*, Cell Proliferation, 33, 317-329, 2000.

[20] P. Tracqui, G. C. CruywagenG, D. E. Woodward, T. Bartoo, J.D. Murray and E. C. Alvord Jr, *A mathematical model of glioma growth: the effect of chemotherapy on spatio-temporal growth*, Cell Prolif, 28 1731, 1995.

[21] L. N. Trefethen, J. A. C. Weideman and T. Schmelzer, *Tablot quadratures and rational approximations*, *BIT Numerical Mathematics, 46:653-670*, 2006.

[22] J. A. C. Weideman and L. N. Trefethen, *Parabolic and hyperbolic contours for computing the bromwich integral*, Math. Comp., 76(259), 13411356, 2007.

[23] D. E. Woodward, J. Cook, P. Tracqui, G. C. Cruywagen, J. D. Murray and E. C. Alvord Jr, *A mathematical model of glioma growth: the effect of extent of surgical resection*, *Cell Proliferation*, vol.29, pp.269-288, 1996.

Heat transfer analysis in concrete slabs using a general purpose FEM computer code

Ioan Both, Frantisek Wald
Steel and Timber Structures
Czech Technical University in Prague
Prague, Czech Republic
ioan.both@upt.ro
wald@fsv.cvut.cz

Raul Zaharia
Steel Structures and Structural Mechanics
Politehnica University of Timisoara
Timisoara, Romania
raul.zaharia@upt.ro

Abstract**—**The effect of elevated temperatures on structural elements may be simulated using numerical models for heat transfer analysis. Function of the complexity of the simulated structures different procedures or element types should be properly selected. The paper presents a summary on the procedure of obtaining the temperature field for structural elements in Abaqus, a sensitivity study on mathematical formulation and element size and a practical application related to validation.

Keywords — heat transfer, numerical analysis, finite element, sensitivity study

I. Introduction

The increased interest in fire design engineering in the last decades led to performing more fire tests and, subsequently, to enlarged databases regarding the behavior of structures at elevated temperature. The comparison with the results obtained from the fire tests represents an approach to confirm the correctitude of the numerical models, when an advanced calculation model is chosen for fire design.

Both thermal and structural effects of elevated temperatures on structural elements may be obtained using general purpose finite element codes like Abaqus [1] or Ansys [2], or dedicated software like Safir [3] or Vulcan [4].

The Eurocodes [5], [6] accepted the advanced calculation models for the structural design in fire situations. According to Section 4.4 in EN 1994-1-2 [6], the results obtained using these tools should be validated against experimental results.

The paper presents the available options to perform a heat transfer analysis using the general purpose FEM computer code Abaqus [1], the results of a sensitivity study based on the finite element type and on the mesh density along the heat transfer direction. A comparison with the experimental results obtained from a fire test is also presented.

II. Analysis procedures

Abaqus defines seven heat transfer analysis procedures: uncoupled heat transfer analysis, sequentially coupled thermal-stress analysis, fully coupled thermal-stress analysis, fully coupled thermal-electric-structural analysis, adiabatic analysis, coupled thermal-electrical analysis and cavity radiation [1]. In structural engineering, the most relevant procedures are the first three.

The uncoupled heat transfer analysis is used for temperature field assessment in a steady-state or transient approach, when structural behavior is omitted. The requirements, in conjunction with the fire design codes [5], [6], [7], should include conduction of materials, convective and radiative interactions. The finite element type is specific for heat transfer analysis and cannot be used for subsequent stress analysis, having active only the temperature degree of freedom (referred in Abaqus as DOF 11).

The sequentially coupled thermal-stress analysis consists of two separate models: one for the heat transfer and one for the structural analysis. The temperatures determined in the first model are imported as a predefined field in the numerical model that performs the calculations for the stress analysis. In addition to the uncoupled heat transfer analysis, mechanical properties must be defined for the stress analysis procedure. The finite elements used in the first model must be modified with elements suitable for structural analysis [1].

The fully coupled thermal-stress analysis performs the heat transfer and stress analysis almost simultaneously, and therefore the finite elements must have both temperature and displacement degrees of freedom active. Particular types of finite elements are used for this procedure. When choosing Abaqus/Standard or Abaqus/Explicit, transient analysis may be performed with both procedures.

Both sequentially and fully coupled analysis give the results either for steady state or transient analysis. An advantage of all three procedures is the possibility to define temperature dependent material properties.

III. Sensitivity study

A. Analysis case

The reference structure for the analysis is chosen to be a square concrete slab with a thickness of 0.15 m and an edge of 1.0 m. Each edge is divided in 5 finite elements, which leads to square finite elements of 0.2 m, which is a common value used in the FEM analysis of such structures. The analysis was performed using shell elements, as shown in Fig. 1a), and three-dimensional elements, as shown in Fig. 1b).

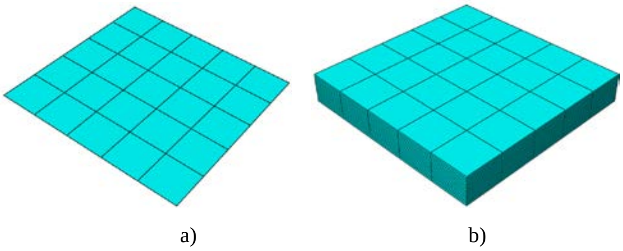


Fig. 1. Reference models a) 2D, b) 3D

The division of the elements perpendicular to the slab plane (on the thickness of the slab) was considered function of the integration points for the two-dimensional elements. The same division was considered for the three-dimensional elements.

The shells were defined by *Sections* characterized by Simpson integration rule and assigned 5, 9 or 15 integration points. For the thickness of the volume elements, the edge of the slab was divided into 5, 9 and 15 elements.

The element types used for comparison belong to both uncoupled heat transfer analysis and fully coupled thermal-stress analysis, as shown in Table I.

Concrete thermal properties are defined according to EN 1994-1-2 [5] as temperature dependent. For conductivity, the upper limit defined by the EN 1++4-1-2 was considered.

TABLE I. Finite element type

	Element type		
	Geometric order	***Heat transfer***	***Fully Coupled***
2D	Linear	DS4	S4RT
	Quadratic	DS8	S8RT
3D	Linear	DC3D8	C3D8RT
	Quadratic	DC3D20	C3D20RT

The elevated temperatures are defined by the nominal standard fire curve given in Section 3.2 of EN 1991-1-2 [14] and represented in Fig. 2.

The interactions to the environmental temperatures consider the heated (bottom) surface and the unheated (top) surface. For the heated surface the convective heat coefficient is defined by the value 25 W/m^{2}K and an emissivity of 0.7. The unheated surface interactions are 4 W/m^{2}K and 0.7 for the convective and radiative heat transfer, respectively.

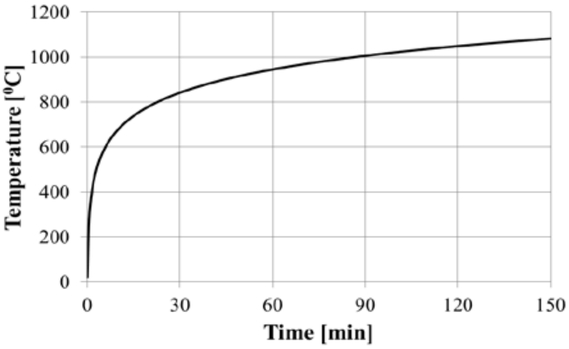


Fig. 2. Nominal standard fire curve

Both heat transfer and fully coupled analysis are defined by a time step of 7200 s with a maximum increment of 5 s, according to the recommendations given in [14]. In the numerical models, the initial temperature is defined in a predefined field at a value of 20 ^{0}C.

For the fully coupled thermal-stress analysis, the Young modulus and the boundary conditions must be specified. If the thermal expansion is not defined, the boundary conditions have no influence on the response of the structure to elevated temperatures, but, of course, the supports must be defined for the structural analysis purposes.

B. Results

For most of the analyses, the numerically determined temperatures follow the same pattern for the heated surface (Fig. 3) and for the unheated surface (Fig. 4). The exceptions are presented in Fig. 5-8. Slightly higher temperatures are obtained for the solid elements with linear geometric order and only four elements over the slab thickness (Fig. 5).

For the solid elements with linear geometric order, using only four elements over the slab thickness, leads to more scattered results for the heated and unheated surface of the slab, as shown Fig. 6 and Fig. 7. These scattered results are noticeable for the heated surface only in the first 30 minutes, in which the standard nominal fire presents the highest rate of temperature increase, see Fig. 6.

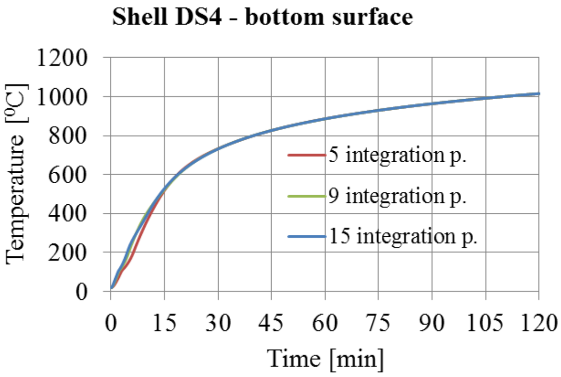


Fig. 3. Heated surface temperature for shell elements with linear geometric order – uncoupled heat transfer analysis

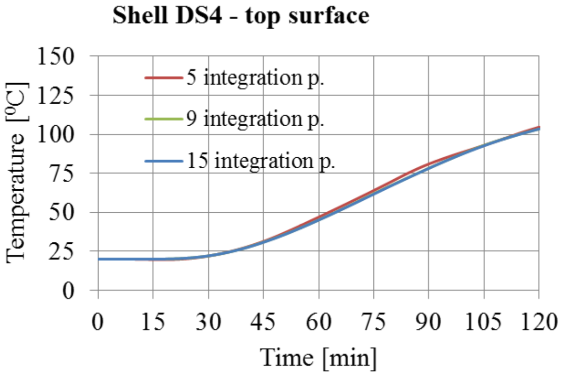


Fig. 4. Unheated surface temperature for shell elements with linear geometric order – uncoupled heat transfer analysis

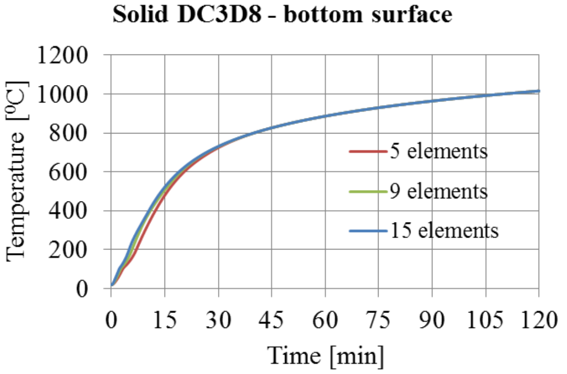


Fig. 5. Unheated surface temperature for solid elements with linear geometric order – uncoupled heat transfer analysis

The Quadratic elements show very close results for all mesh divisions, excepting for the fully coupled thermal-stress analysis with four elements over the slab thickness (Fig. 8). For this case the results show inconsistency, since values smaller than the predefined temperature are obtained between minute 30 and 45, as shown in Fig. 8.

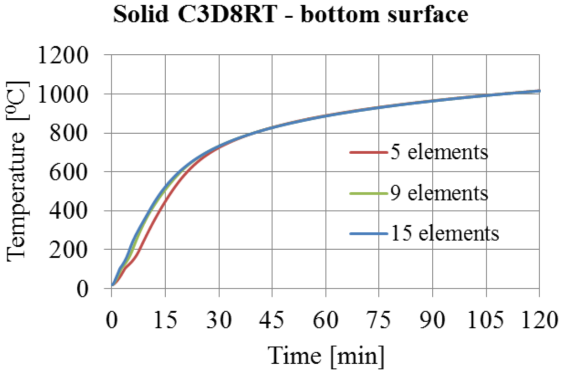


Fig. 6. Heated surface temperature for solid elements with linear geometric order – fully coupled analysis

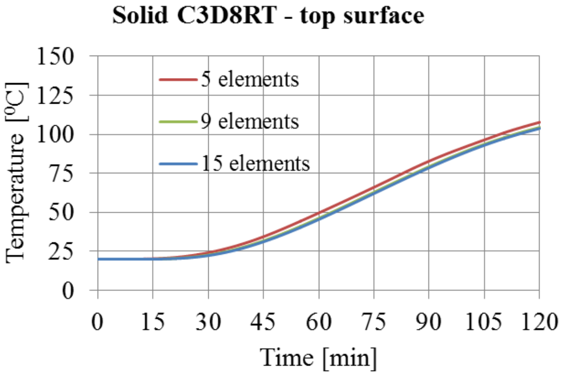


Fig. 7. Unheated surface temperature for solid elements with linear geometric order – fully coupled analysis

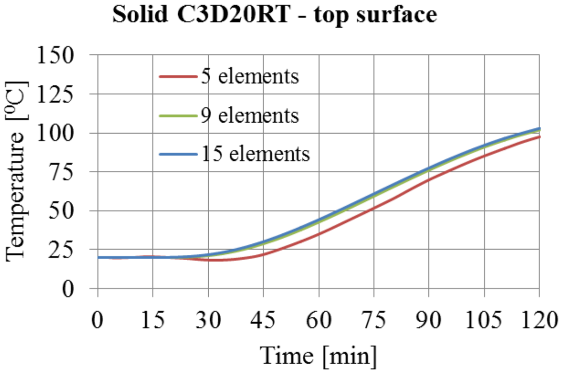


Fig. 8. Unheated surface temperature for solid elements with quadratic geometric order – fully coupled analysis

All the analyses involving linear elements are completed in 3 min and 1s with an additional 4 s for the fully coupled 3D model involving 15 elements over the thickness. The analyses performed with quadratic elements are completed within a similar time interval, 3 min and 8 s. Exceptions from this rule are represented by the less time efficient models of the coupled 3D model with 9 elements (8 min 41 s) and 15 elements (16 min 36s) along the slab thickness. The similar time interval for most analyses implies that the computational time is limited by the incrementation procedure and not by the mathematical formulation. The latter is important only for the quadratic elements, in a fully coupled analysis. The mentioned time intervals were obtained by performing the numerical calculation on a second generation I5 processor of 2.50 GHz (on one core) and allowing the analysis to go up to 90% of the 8 GB RAM.

From the data presented above, it results that the accuracy of the computation is influenced only by the number of slab division along the heat transfer direction. Similar results are obtained if a minimum of 9 integration point are defined for either linear of quadratic shell elements. The same conclusion can be stated for solid elements when 9 finite elements are defined on the thickness of the slab.

As Fig. 9 demonstrates, very small differences are recorded for the temperatures on the top surface, when comparing the results obtained in the analysis of shell elements (with 9 integration points on the thickness) and the analysis of solid elements (with 9 elements on the thickness). Unless solid elements are not required for detailed investigation, shell elements are suitable in the numerical analysis of this type of structures.

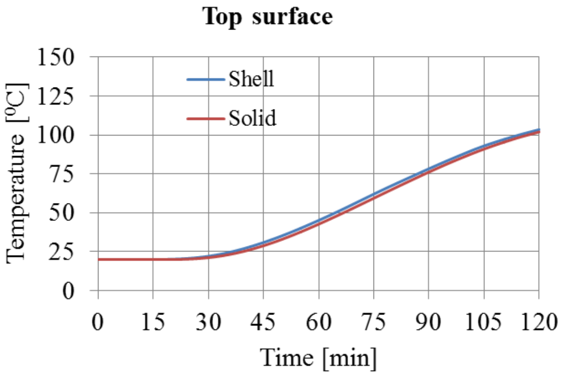


Fig. 9. Temperatures function of element type

IV. Practical Application

In order to validate the heat transfer procedure used above in Abaqus, the results of a fire test on a slab were considered. The test was conducted in 2008 in Mokrsko, Czech Republic under the supervision of the staff of Department of Steel and Timber Structures, Czech Technical University in Prague. A compartment fire using wooden cribs was simulated in an enclosure designed using different structural systems. Three quarter of the surface was covered by a steel and concrete composite floor, while the other quarter used prefabricated concrete hollow panels. For the validation of the numerical procedure used in Abaqus, the recorded temperatures in the steel and concrete composite slab, disseminated in Wald *et.al.*[8], were considered.

The section dimensions of the concrete slab are presented in Fig. 10. The trapezoidal ribs follow the geometry of the CofraPlus60 with a thickness of 0.75 mm.

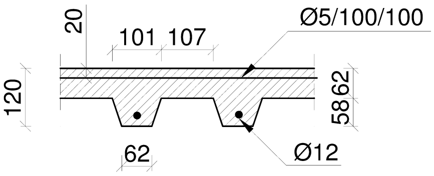


Fig. 10. Concrete slab profile [9]

The geometry, mesh and temperature field of the 3D numerical model of the trapezoidal slab are presented in Fig.11. DC3D8 elements are used to model the slab with an approximate size of 5 mm. The steel sheet deck is not defined in the numerical model. The considered numerical procedure and assumptions are the same as for the analyses performed in Section 3. Exceptions are the thermal load, the thermal conductivity of the concrete and the convective heat transfer coefficient for the heated surface. The evolution of the gas temperature introduced in the heat transfer analysis was considered the same as recorded in the experiment. The thermal conductivity of the concrete is considered to follow the lower limit values defined in EN 1994-1-2 [6]. For the convective heat transfer coefficient, according to Section 3.3 in EN 1991-1-2 [7], in case of a natural fire, a value of 35 W/m^{2}K is adopted.

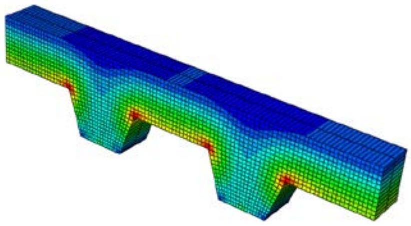


Fig. 11. Model of slab – temperature distribution

The comparison between the numerical results and the experimental results is presented in Fig. 12, at the unheated side of the slab.

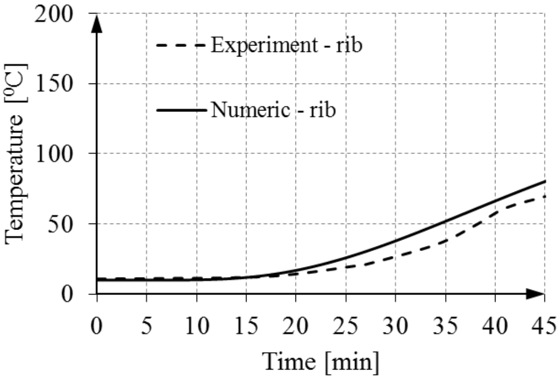


Fig. 12. Temperatures at the unheated side of the slab

For the numerical structural analysis of slabs subjected to elevated temperatures, the necessary resources are very high. EN 1994-1-2 [5] allows the use of an effective thickness of a composite slab either for trapezoidal or re-entrant profiles. The formulas are given in Annex D of the code. Using the formulas provided by the code, the effective thickness of the trapezoidal slab presented in Fig. 10 is equal with 85 mm A comparison between the top surface temperature obtained in the case of a trapezoidal slab profile and a constant thickness slab profile using the effective thickness is presented in Fig. 12.

For the common dimensions of slabs belonging to compartments subjected to fire, a numerical model involving solid elements requires a large number of finite elements. The time required for such an analysis is greatly shortened if shell elements are used. The temperatures obtained for the same studied thermal load of the fire test, using DS4 shell elements, are represented in Fig. 13. The section was defined with 9 integration point over the thickness of the slab. Both shell and solid models identify themselves in the same pattern of temperature.

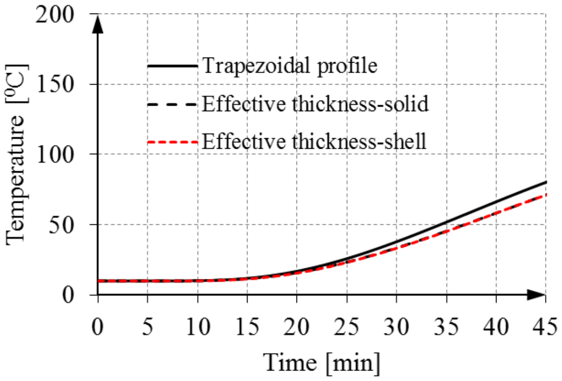


Fig. 13. Temperatures at the unheated side of the slab function of slab profile

V. Conclusion

The paper presents the results of a sensitivity study regarding the finite elements that can be used in a heat transfer analysis of slabs in Abaqus. The results are mainly dependent on the number of slab divisions along the heat transfer direction. A good agreement of temperatures is obtained if a minimum of 9 integration points are defined for shell elements or 9 finite elements are defined along the thickness of the solid slab.

The validation of the heat transfer procedure used in Abaqus was performed using the results of a real scale fire test. Comparing the top surface temperatures of the trapezoidal slab profile, the effective thickness slab modeled with solid elements and the effective thickness slab modeled with shell elements lead to similar results. The structural response of a numerical model may be influenced by the thermal analysis only if the discretization of the numerical model is of poor quality, otherwise the resulted temperatures have similar evolution with respect to time.

Acknowledgment

This publication was supported by the European social fund within the framework of realizing the project „Support of inter-sectoral mobility and quality enhancement of research teams at Czech Technical University in Prague“, CZ.1.07/2.3.00/30.0034.

This work was financially supported by the Research Grant GAČR 15-19073S.

References

[1] ABAQUS, Abaqus Analysis User’s Manual Volume II, Dassault Systèmes Simulia Corp., Providence, RI, USA: Dassault Systèmes, 2011,

[2] ANSYS. 2008. ANSYS® Academic Research, Release 14.0,

[3] J.-M. Franssen, “Safir- A thermal/structural program modelling structures under fire,” in Engineering Journal, vol. 42, issue 3 pp. 143-158, 2005,

[4] Vulcan, [Online], Available: http://www.vulcan-solutions.com/ , Accesed: 2 November 2014,

[5] C.E.N., EN 1993-1-2 Eurocode 3: Design of steel structures - Part 1-2: General rules - Structural fire design, April 2005, Brussels,

[6] C.E.N., EN 1994-1-2 Eurocode 4 - Design of composite steel and concrete structures - Part 1-2: General rules - Structural fire design, June 2004, Brussels,

[7] C.E.N., EN 1993-1-2 Eurocode 3: Design of steel structures - Part 1-2: General rules - Structural fire design, April 2005, Brussels,

[8] F. Wald, P. Kallerová, P. Chlouba, Z. Sokol, M. Strejček, J. Pospíšil, et.al., Fire Test on an Administrative Building in Mokrsko, Česká technika, Prague, May 2010,

[9] J. Vácha,P. Kyzlík,I. Both, F. Wald, “Beams with corrugated web at elevated temperature, experimental results”, Thin-Walled Structures, in press.

On Kolmogorov’s Theory of Local Isotropy and its Relation to Ordinary Hydrodynamic Turbulence

H.P. Mazumdar, S. Pramanik, C. Mamaloukas

***Abstract*—** In this paper, we present a straight forward review of Kolmogorov’s theory of Local Isotropy to turbulence and discuss its relation to ordinary hydrodynamic turbulence. In the first section, we describe briefly the well known theory of local isotropy to homogeneous turbulence. In the second section, we construct the relation between Kolmogorov’s theory and ordinary hydrodynamic turbulence. In the last section, we calculate skewness factor in relation to the theory of homogeneous and isotropic turbulence and compare its value to that obtained from turbulence measurements.

***Keywords*—** Ordinary Hydrodynamic Turbulence, Kolmogorov's theory, Local Isotropy, Homogeneous Isotropic Turbulence, skewness factor.

I. Introduction

A. Kolmogorov Hypothesis

LET us consider a simple domain *G* within the turbulent flow and specify a set of points $x^{(0)}, x^{(1)}, ..., x^{(n)}$ also within *G*. New coordinates and velocity difference are defined as (Ref. Turbulent Flow - by S.B.Pope).

$$y \equiv x - x^{(0)}$$

$$\upsilon(y) \equiv U(x,t) - U\left(x^{(0)},t\right)$$

The joint PDF of υ at the *N* points $y^{(1)}, y^{(2)}, ..., y^{(n)}$ is denoted by f_N. Now, the turbulence is said to be locally homogeneous in *G* if for every fixed *N* and $y^{(n)}\,(n = 1, 2, ..., N)$, f_N is independent of $x^{(0)}$ and $U\left(x^{(0)},t\right)$. The turbulence is said to be locally isotropic if in addition to its satisfying conditions of local homogeneity, f_N is invariant with respect to rotations and reflexions of the coordinate axes.

In any turbulent flow when the Reynolds number $\mathrm{Re}\left(=\frac{UL}{v}\right)$, the turbulent may approximately be locally isotropic if the domain *G* is sufficiently small i.e., $\left|y^{(n)}\right| << L$, for all *n* and it is not near the boundary of the flow.

B. First Similarity Hypothesis:

For locally isotropic turbulence, the *N*-point PDF f_N is uniquely determined by the viscosity *ν* and the energy dissipation rate *ε*.

C. Second Similarity Hypothesis:

If the moduli of $y^{(n)}$ and the differences $y^{(m)} - y^{(n)}, \left(m \neq n\right)$ are large compared to the Kolmogorov length scale η, then the N-point PDF f_N is uniquely determined by ε and independent of ν.

Kolmogorov used the structure functions like

$$D_{11}(r,t) = \overline{\left(u_1' - u_1\right)^2}$$
$$D_{111}(r,t) = \overline{\left(u_1' - u_1\right)^3} \tag{2.1}$$

in formulating his theory of Local Isotropy. If the ordinary isotropic random velocity field is also locally isotropic, the correlation functions $B_{11}(r,t)\left[=\overline{u_1u_1'}\right]$ and $B_{11,1}(r,t)\left[=\overline{u^2{}_1u_1'}\right]$ are related to the structure functions pertaining to local isotropic turbulence [Azad and Hummel, 1981].

For examples,

$$D_{11}(r,t) = 2\left[B_{11}(0,t) - B_{11}(r,t)\right] \tag{2.2}$$

$$D_{111}(r,t) = 6B_{11,1}(r,t) \tag{2.3}$$

$B_{11}(r,t)$ and $B_{11,1}(r,t)$ in ordinary isotropic turbulence satisfy Karman-Howarth equation:

H.P. Mazumdar, Physics and Applied Mathematics, Indian Statistical Institute, Kolkata, West Bengal, India, email: hpmi2003@yahoo.com
S. Pramanik, Department of Mathematics, University of Burdwan, Burdwan-713104, West Bengal, India, email: sambaran_math@yahoo.co.in
C. Mamaloukas, Department of Statistics, Athens University of Economics and Business, Athens, 10434, Greece, email: mamkris@aueb.gr

$$\frac{\partial}{\partial t}B_{11}\left(r,t\right)=\left(\frac{\partial}{\partial r}+\frac{4}{r}\right)\left[B_{11,1}\left(r,t\right)+v\frac{\partial}{\partial r}B_{11}\left(r,t\right)\right] \quad (2.4)$$

Length scale is

$$L=\int_0^\infty \frac{B_{11}\left(r\right)}{B_{11}\left(0\right)}dr \quad (2.5)$$

$$\frac{\partial}{\partial t}D_{11}\left(r,t\right)=0 \text{ for all } r<<L \quad (2.6)$$

$$\frac{\partial}{\partial t}B_{11}\left(0,t\right)=\frac{2}{3}\frac{d}{dt}\left(\frac{\overline{u_1u_1'}}{2}\right)=-\frac{2}{3}\overline{\varepsilon} \quad (2.7)$$

Using (2.6) and (2.7) in (2.4), Kolmogrov derived

$$D_{111}\left(r\right)-6v\frac{d}{dt}D_{11}\left(r\right)=-\frac{4}{5}\overline{\varepsilon}r \quad (2.8)$$

Obukov and Yaglom (1951) introduced:

$$S=\frac{D_{111}\left(r\right)}{\left[D_{11}\left(r\right)\right]^{\frac{2}{3}}} \quad (2.9)$$

S is the skewness factor.

For not too small values of $r=\overline{u}r$, the values of the velocity difference skewness factor is

$$S\left(r\right)=\frac{\overline{\left(u_1'-u_1\right)^3}}{\left[\overline{\left(u_1'-u_1\right)^2}\right]^{\frac{3}{2}}} \quad (2.10)$$

which approaches to the limiting value *S* = 0 at a very slow rate (Chen, 1969). Substituting (2.9) in (2.8), we obtain

$$6v\frac{d}{dt}D_{11}\left(r\right)-S\left[D_{11}\left(r\right)\right]^{\frac{3}{2}}=\frac{4}{5}\overline{\varepsilon}r \quad (2.11)$$

Monin and Yaglom (1975) suggested that equation (2.8) can be transformed to a Spectral equation containing energy spectrum *E*(*k*) and energy transfer function (spectrum) *T*(*k*) by using relations between $D_{11}(r)$ with *E*(*k*) and $D_{111}(r)$ with *T*(*k*).

They further suggest assuming in this case that the turbulence is also isotopic and use the spectral form of Karman-Howarth equation.

Using the relation between $D_{11}(r)$ and *E*(*k*) in equation (2.11), we derive a general formula for *S* in terms of *E*(*k*). We then use the self-preserving solutions for *E*(*k*), as obtainable by solving the equation governing the decay of *E*(*k*) in isotropic turbulence for very large Reynolds number.

II. Formula for skewness factor in terms of E(k):

Monin and Yaglom (1975) gave the relation between $D_{11}(r)$ and *E*(*k*) as

$$D_{11}\left(r\right)=4\int_0^\infty\left[\frac{1}{3}+\frac{\cos kr}{\left(kr\right)^2}-\frac{\sin kr}{\left(kr\right)^3}\right]E\left(k\right)dk \quad (2.12)$$

Substituting (2.12) in equation (2.11), we obtain

$$24v\left[\int_0^\infty\left(\frac{\sin kr}{kr}-\frac{3\cos kr}{\left(kr\right)^2}+\frac{3\sin kr}{\left(kr\right)^3}\right)\frac{1}{r}E\left(k\right)dk\right]- \quad (2.13)$$

$$-S\left[4\int_0^\infty\left(\frac{1}{3}+\frac{\cos kr}{\left(kr\right)^2}-\frac{\sin kr}{\left(kr\right)^3}\right)E\left(k\right)dk\right]^{\frac{3}{2}}=\frac{4}{5}\overline{\varepsilon}r$$

Finally,

$$S=\frac{24v\left[\int_0^\infty\left(\frac{\sin kr}{kr}-\frac{3\cos kr}{\left(kr\right)^2}+\frac{3\sin kr}{\left(kr\right)^3}\right)\frac{1}{r}E\left(k\right)dk\right]-\frac{4}{5}\overline{\varepsilon}r}{\left[4\int_0^\infty\left(\frac{1}{3}+\frac{\cos kr}{\left(kr\right)^2}-\frac{\sin kr}{\left(kr\right)^3}\right)E\left(k\right)dk\right]^{\frac{3}{2}}} \quad (2.14)$$

III. Application of the Formula (2.14) to Self-preserving Energy Spectrum:

Much speculations have been made regarding the self-preserving spectrum *E*(*k*) for isotropic turbulence. A part of this spectrum which in a certain way represents a quasi-equilibrium state was considered by Kolmogorov (1941), Ousagar (1945), Weizsacker (1948), Heisenberg (1948) and many others. Heisenberg obtained a self-preserving solution of the equation governing the decay of energy spectrum where in similarity relations of the quasi-equilibrium were extended to very small wave number *k*. The basic equation for the energy spectrum *E*(*k*,*t*) in isotropic turbulence is (Hinze, 1975):

$$\frac{\partial}{\partial t}E\left(k,t\right)=T\left(k,t\right)-2vk^2E\left(k,t\right) \quad (2.15)$$

or,

$$\frac{\partial}{\partial t}\int_0^k E\left(k',t\right)dk'=\int_0^k T\left(k',t\right)dk'-2v\int_0^k k'^2E\left(k',t\right)dk' \quad (2.16)$$

The equation (2.15) is obtainable by taking Fourier transform of Von-Karman-Howarth equation (Karman and Lin 1949).

Taking a three-dimensional Fourier-transform of the equation for the change of the double correlation tensor:

$$\frac{\partial}{\partial t}\left(u'^2R_{ik}\right)-u'^3\frac{\partial}{\partial\xi_j}\left(\tau_{ijk}+\tau_{kji}\right)=2vu'^2R_{ik} \quad (2.17)$$

where u'^2 is the mean square of the turbulent velocity, *t* is the time *v* is the kinematic viscosity, and $R_{ik}\left(\xi_l,t\right)$ and $\tau_{ijk}\left(\xi_l,t\right)$ are the double and triple correlation tensors defined by Karman-Howarth for two points P and P' separated by a space vector ξ_l, we proceed to determine spectral version of equation (2.17). By contracting the resulting equation and multiplying it with $\frac{4\pi k^2}{3}$, where k is the wave number, we obtain the equation for the change of spectrum

$$\frac{\partial}{\partial t}E\left(k,t\right)=T\left(k,t\right)-2\nu k^{2}E\left(k,t\right) \qquad (2.18)$$

where $E=\left(\dfrac{4\pi k^{2}}{3}\right)E_{nn}$, $T=\left(\dfrac{4\pi k^{2}}{3}\right)2ik_{j}T_{njn}$,

$$E_{ik}\left(k_{l},t\right)=\frac{u'^{2}}{\left(2\pi\right)^{3}}\iiint R_{ik}\left(\xi_{l},t\right)e^{i\left(k_{n}\xi_{n}\right)}d\tau\left(\xi\right), \qquad (2.19)$$

$$T_{ijk}\left(k_{l},t\right)=\frac{u'^{3}}{\left(2\pi\right)^{3}}\iiint \tau_{ijk}\left(\xi_{l},t\right)e^{i\left(k_{n}\xi_{n}\right)}d\tau\left(\xi\right) \qquad (2.20)$$

Evaluating these integrals in terms of spherical coordinates in the ξ-space, we obtain

$$\begin{aligned} E&=\frac{1}{3}\left[k^{2}E_{1}''(k)-kE_{1}'(k)\right],\\ E_{1}\left(k\right)&=2\frac{u'^{2}}{\pi}\int_{0}^{\infty}f\left(r,t\right)\cos krdr,\\ T&=\frac{2}{3}\left[k^{2}H_{1}''(k)-kH_{1}'(k)\right],\\ kH_{1}\left(k\right)&=2\frac{u'^{3}}{\pi}\int_{0}^{\infty}h\left(r,t\right)\sin krdr \end{aligned} \qquad (2,21)$$

where $f(r,t)$ and $h(r,t)$ are the double and triple correlation functions satisfying the Karman-Howarth equation:

$$\frac{\partial}{\partial t}\left(u'^{2}f\right)+2u'^{3}\left(\frac{\partial h}{\partial r}+\frac{4h}{r}\right)=2\nu u'^{2}\left(\frac{\partial^{2}f}{\partial r^{2}}+\frac{4}{r}\frac{\partial f}{\partial r}\right) \qquad (2.22)$$

Relations (2.21) which connect equation (2.15) and (2.22) were derived by Lin (1947).

Heisenberg modeled the term $W\left(k,t\right)=\int_{0}^{k}T\left(k',t\right)dk'$ as

$$\int_{0}^{k}T\left(k',t\right)dk'=-2C_{H}\int_{k}^{\infty}\left[\frac{E\left(k',t\right)}{k'^{3}}\right]^{\frac{1}{2}}dk'\int_{0}^{k}E\left(k',t\right)k'^{2}dk' \qquad (2.23)$$

and obtain solution of equation (2.16) as

$$E\left(k,t\right)\approx\frac{1}{\sqrt{t}}f\left(k\sqrt{t}\right) \qquad (2.24)$$

Tollmein (1952, 1953) also obtained solutions of equation (2.16) with Heisenberg's form (2.23) for the transfer spectrum $W(k,t)$ and normalized such solutions so that $E(k,t)$ for large k were as near time independent as possible and expressed the resulting values as functions of $k\lambda$, λ being the micro scale of turbulence due to Taylor (1938).

Sen (1951) showed first that the decay equation (2.15) or (2.16) for $E(k,t)$ admits of a self-preserving solutions, a family for the case when Reynolds number is large so that ν may be neglected compared to turbulent viscosity.

Sen obtained solutions of the decay equation (2.16) with Heisenberg's form for the transfer spectrum expressing $E(k,t)$ as a function of a single independent variable kl, where l is given by $l=\dfrac{1}{k_{0}}\left(\dfrac{t}{t_{0}}\right)^{c}$, where k_0 and t_0 are constants and c is a parameter which may have values between 0 and $\dfrac{2}{3}$. Sen assumed the general form of the self-preserving solution as

$$E\left(k,t\right)=\frac{1}{c^{2}k_{0}^{3}t_{0}^{2}}\left(\frac{t}{t_{0}}\right)^{3c-2}f\left[\frac{k}{k_{0}}\left(\frac{t}{t_{0}}\right)^{c}\right] \qquad (2.25)$$

In what follows, we shall assume the simple form for $T(k, t)$ due to Kovasznay (1948). Accordingly, equation (2.15) reduces to

$$\frac{\partial}{\partial t}E\left(k,t\right)=-2C_{k}\frac{d}{dk}\left[k^{\frac{5}{2}}E^{\frac{3}{2}}\left(k,t\right)\right]-2\nu k^{2}E\left(k,t\right) \qquad (2.26)$$

where C_k is a constant.

In the case when Reynolds number is sufficiently large, so that we may neglect the last term on the right hand side of (2.26), we substituted (2.25) in (2.26) and obtain

$$\left(3c-2\right)f\left(x\right)+cxf'\left(x\right)=-2\frac{d}{dx}\left[x^{\frac{5}{2}}f^{\frac{3}{2}}\left(x\right)\right]=g\left(x\right) \qquad (2.27)$$

where $x=\dfrac{1}{k_{0}}\left(\dfrac{t}{t_{0}}\right)^{c}$ and $g\left(x\right)=-2\dfrac{d}{dx}\left[x^{\frac{5}{2}}f^{\frac{3}{2}}\left(x\right)\right]$

It can be easily shown that

$$f\left(x\right)\approx x^{\frac{2-3c}{c}} \quad \text{as} \quad x\rightarrow 0 \qquad (2.28)$$

and

$$f\left(x\right)\approx x^{-\frac{5}{3}} \quad \text{as} \quad x\rightarrow\infty \qquad (2.29)$$

(inertial sub range law)

We obtain numerical solutions of (2.27) for the case c = 1⁄2 with the initial condition, given by

$$f\left(x\right)=4x-64x^{3}, \quad x\rightarrow 0, \quad c=1/2 \qquad (2.30)$$

Exactly, in the same manner, we calculate the cases for other values of c e.g., c = 2⁄5, 1⁄3, and 2⁄7 and obtained

$$f\left(x\right)=5x^{2}-122.9837x^{\frac{9}{2}}, \quad x\rightarrow 0, \quad c=2/5 \qquad (2.31)$$

$$f\left(x\right)=6x^{3}-205.7571x^{6}, \quad x\rightarrow 0, \quad c=1/3 \qquad (2.32)$$

and

$$f\left(x\right)=7x^{4}-314.8444x^{\frac{15}{2}}, \quad x\rightarrow 0, \quad c=2/7 \qquad (2.33)$$

These initial conditions are being derived, using power series expansion (keeping up to second contributing term) of f (x) for small values of x in the same manner as Reid and Harris (1959).

We plot f (x) vs. x for this case (Fig. 1). The computations of skewness factor corresponding to (x, f (x)) data are obtained here, by recasting the formula for *S* as given in equation (2.14).

Once the $f\,'s$ are found out, the corresponding $g\,'s$ may be calculated through equation (2.27). The behavior of *g*(*x*) for $x \to 0$ in each of the above cases may also be worked out keeping up to second contributing terms. To complete the analysis we now calculate *g*(*x*) from (2.27) for the cases $c = 1/2,\ c = 2/5,\ c = 1/3,\ c = 2/7$ with the initial condition and obtain

$$g(x) = -64x^3 + 1824x^5, \quad x \to 0, \quad c = 1/2 \quad (2.34)$$

$$g(x) = -122.9837x^{\frac{9}{2}} + 6600x^7, x \to 0, c = 2/5 \; (2.35)$$

$$g(x) = -205.7571x^6 + 15120x^9, x \to 0, c = 1/3 \; (2.36)$$

and

$$g(x) = -314.8444x^{\frac{15}{2}} + 29988x^{11}, x \to 0, c = \frac{2}{7} \; (2.37)$$

The curves for $g\,'s$ are displayed in Fig.2. It is to be mentioned that *k* is normalized by $k_0\left(\frac{t}{t_0}\right)^{-c}\left[l^{-1}\right]$.

The general form of self-preserving solution (2.25) under quasi-equilibrium condition may now be written as

$$E_{qe}(x) = \frac{1}{C_k^2 k_0^3 t_0^2} l^{\frac{3c-2}{c}} f(x) \qquad (2.38)$$

where $x = kl$.

Use of (2.38) for any particular relation of the turbulence decay process, depending on the choice of *c* and the corresponding unit of length *l*, would imply such self-preserving solution does not depend explicitly on time.

Further, when *k* is small, Batchelor (1959) has discussed that among the terms of equation (2.15), $T(k,t)$ and $-2vk^2 E(k,t)$ are both of smaller order than $\frac{\partial}{\partial t}E(k,t)$.

Now, substituting (2.38) in equation (2.15) and accepting the premise that similarity relation of the quasi-equilibrium situation may be extended to very small values of *k*, we obtain easily for the case when the left hand side of (2.15) predominates

$$E_{qe} \approx (kl)^{\frac{2-3c}{c}} \qquad (2.39)$$

When *c* = 1⁄2, we obtain Heisenberg's linear spectrum for small *k* , from (2.39) as

$$E_{qe} \approx lk \quad (k \to 0) \qquad (2.40)$$

Now permitting us to substitute (2.38) in the formula (2.14) for the skewness factor *S* and in the expression of $\bar{\varepsilon}$ e.g.,

$$\bar{\varepsilon} = -\frac{d}{dt}\left(\frac{u_0 u_1}{2}\right)\left[= -\frac{d}{dt}\int_0^\infty E(k,t)\,dk\right]$$

we obtain

$$\frac{S}{C_k}A = \frac{4}{5}D - 24\frac{1}{\mathrm{Re}}\left(\frac{t}{t_0}\right)^{1-2c} B \qquad (2.41)$$

where

$$A = \left[\frac{4}{3}\int_0^\infty f(x)\,dx - 4\int_0^\infty f(x)\left(\frac{\sin xy}{x^3y^3} - \frac{\cos xy}{x^2y^2}\right)dx\right]^{\frac{3}{2}}$$

$$B = \int_0^\infty f(x)\left(\frac{3\cos xy}{x^3y^3} - \frac{3\sin xy}{x^3y^3} + \frac{\sin xy}{xy}\right)\frac{1}{y}\,dx \qquad (2.42)$$

$$D = (2c-2)\,y\int_0^\infty f(x)\,dx, \quad \mathrm{Re} = \frac{1}{vk_0^2 t_0}$$

It is to be noted that we have normalized *r* by the unit of length *l* viz.,

$$y = \frac{r}{\frac{1}{t_0}\left(\frac{t}{t_0}\right)^c} = \frac{r}{l} \qquad (2.43)$$

in obtaining equation (2.41).

Inspection of (2.41) suggests that complete self-preservation is possible for the case *c* = 1⁄2 only or by making *Re* → ∞.

Finally, for calculation of skewness factor *S* corresponding to the numerical data (*x*, *f* (*x*)) obtained here with Kovaznay's form of energy transfer function, equation (2.41) simplified further to

$$\frac{S}{C_k} = -\frac{4}{5}\frac{y\int_0^\infty f(x)\,dx}{\left[\frac{4}{3}\int_0^\infty f(x)\,dx - 4\int_0^\infty f(x)\left(\frac{\sin xy}{x^3y^3} - \frac{\cos xy}{x^2y^2}\right)dx\right]^{\frac{3}{2}}} \qquad (2.44)$$

It is to be remembered that the formula (2.44) is applicable for all the other cases e.g., *c* = 2⁄5, *c* = 1⁄3 and *c* = 2⁄7 besides, the case (*c* = 1⁄2 and *Re* → ∞). We choose some values of *y* and calculate the corresponding values of *S* from equation (2.44), using (*x*,*f*(*x*)) data (view Fig. 1).

We plot −*S* vs. *y* (Fig. 3). An experimental values for C_k was suggested by Reid and Harris (1959) as C_k = 0.12. Townsend (1948) provided experimental evidence for the Local Isotropic Theory of Turbulence, developed by A.N. Kolmogorov

(1941). According to Townsend measurements of absolute constant values of the skewness factor is close to −0.39.

We present here a table-1 below, showing at different stationary *x*, the skewness factor −*S*, respectively for different cases viz., *c* = 1⁄2, *c* = 2⁄5, *c* = 1⁄3 and *c* = 2⁄7.

Thus, we may conclude that the theory of Local Isotropy is a great contribution of A.N. Kolmogorov to the development of the Theory of Statistical Fluid Mechanics.

Table 1

values of parameter c	stationary x (non-dimensional distance)	skewness factor (−S)
c = 1/2	48	0.3940
c = 2/5	49	0.3970
c = 1/3	52	0.3924
c = 2/7	55	0.3945

IV. REMARKS

1) The constants k_0 and t_0 appearing in self-preserving solution (2.25) may well taken as the representative wave-number of the energy containing eddies and its characteristics time.

2) The unit of length, given by $l = \frac{1}{k_0}\left(\frac{t}{t_0}\right)^c$ may also be considered of the order of $L\left[=\int_0^\infty \frac{B_{11}(r)}{B_{11}(0)}dr\right]$

3) In compliance with the condition $r << L$, the present analysis is valid for $y\left(=\frac{r}{l}\right)<<1$.

4) As we have neglected the effects of viscous dissipation in deriving (2.27) and (2.44), it would be appropriate to consider $r >> \eta_d$ where η_d is the Kolmogorov length scale.

5) The results agree very well with the working hypothesis introduced by Obukov (cf. Monin and Yaglom, 1975) that the skewness factor *S*(*r*) is negative. So the plottings −*S*(*r*) vs. *y* or |*S*| vs. *y* are the same.

6) The present results agree very well with the experimental observations (Stewart, 1951; Frenkiel and Klebanoff, 1965; etc.) that −*S* decreases at a very slow rate.

7) Townsend (cf. Panchev, 1971) obtained the value of *S* in the inertial interval −0.4 with oscillation between −0.36 and −0.42 at the limits of measurement errors. According to the present calculations, the value of *S* obtained −0.39 which may be considered highly satisfactory.

V. FIGURES

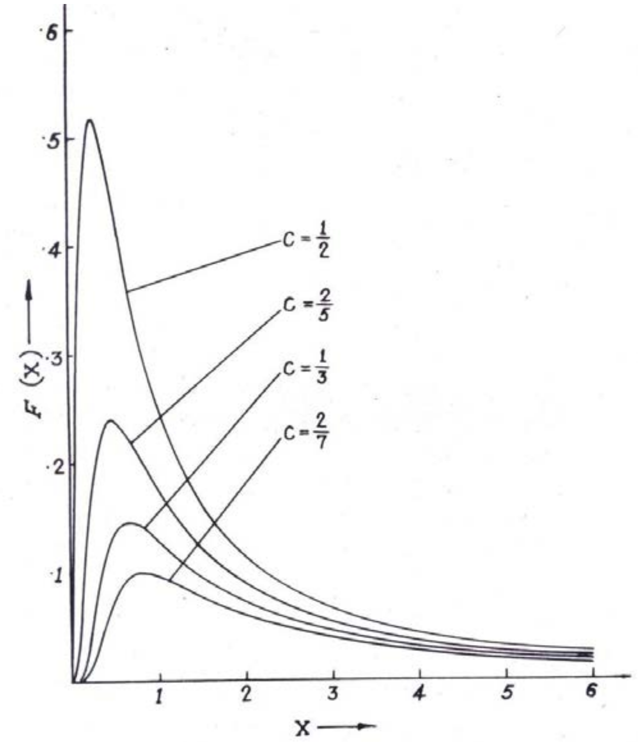


Fig. 1: The energy spectra during initial-period decay corresponding to the cases $c=\frac{1}{2}$, $c=\frac{2}{5}$, $c=\frac{1}{3}$, $c=\frac{2}{7}$, $\mathrm{Re}=\infty$

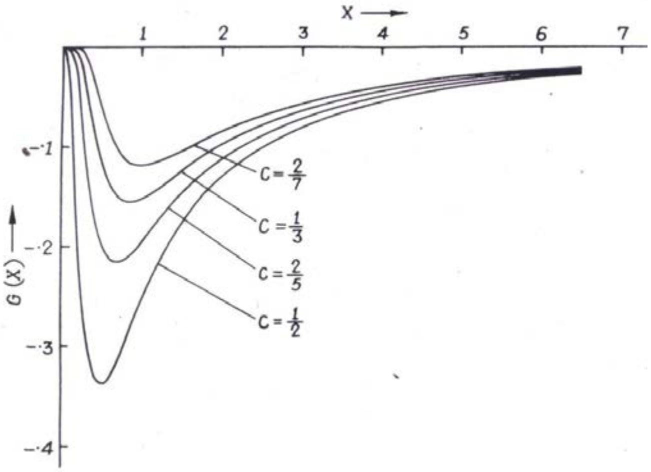


Fig. 2: The energy transfer functions during initial-period of decay corresponding to the cases $c=\frac{1}{2}$, $c=\frac{2}{5}$, $c=\frac{1}{3}$, $c=\frac{2}{7}$, $\mathrm{Re}=\infty$

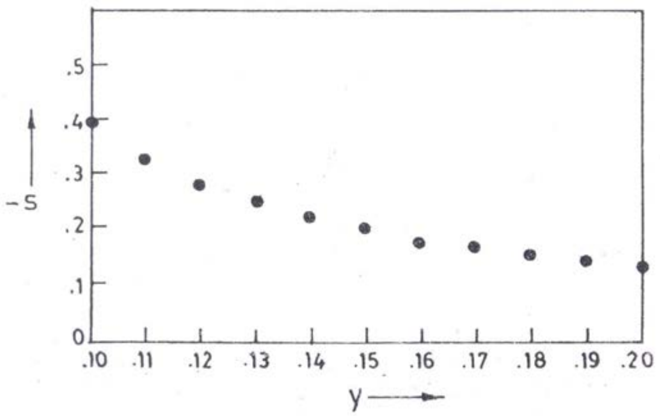


Fig. 3: Skewness factor S as function of y.

REFERENCES

[1] Pope, S.B., 1990, Turbulent Flow, Camb.Univ. Press, London.

[2] Azad, R.S. and Hummel, R.H., 1981, Phys. Fluids 24(10). P 1774

[3] Karman, T.Von. and Howarth, L., 1938, Proc. Roy. Soc., A 164. No.917, P. 192.

[4] Kolmogorov, A.N., 1941a, C.R. Acad, Sci., U.S.S.R. 30, P.302.1941c, C.R. Acad, Sci., U.S.S.R. 32, P. 16.

[5] Obukov, A.M., and Yaglom, A.M., 1951, Prikl. Mat. Mekh., 15, No.1, P.3.

[6] Von Atta, C.W. and Chen, W.Y., 1968, J. Fluid Mech., 34, No.3, P.497. 1969a, J. Fluid Mech., 38, No.4, P.743. 1969b, J. Fluid Mech., 12, Suppl. II, IV -264, II-269

[7] Monin, A.S. and Yaglom, A.M., 1975, Statistical Fluid Mechanics, 2, M.I.T. Press, Massachusetts.

[8] Ousager, L., 1945, Phys.Rev., 68, No. 11-12, P.286.

[9] Weizsacker, C.F.Von, 1948, Z.Physik, 124, No. 7-12, P.614.

[10] Heisenberg, W., 1948, Z.Phys., 124, No.7-12, P.628.

[11] Hinze, J.O., 1975, Turbulence, Mc-Graw Hill, P.215, New York.

[12] Tollmien, W., 1952-1953, Wissensch. Zs. Techn. Hochschule, Dresden, 2, No.3., P.443. 13. Taylor, G.I., 1938, Proc. Roy. Soc., A 164. No.919, P. 476.

[13] Sen, N.R., Bull, Cal.Math.Soc., 43, No.1., P.1.

[14] Kovasznay, L.S.G., 1948, J. Aeronaut. Sci., 15, No. 12, P.745.

[15] Reid, W.H., and Harris, D.L., 1959, Phys. Fluids, 2, No.2, P.139.

[16] Batchelor, G.K., 1959, The Theory of Homogeneous Turbulence, Camb.Univ. Press, London.

[17] Stewart, R.W., 1951, Proc. Camb. Phil. Soc., 47, No.1,P.146.

[18] Frenkiel, F.N., and Klebanoff, P.S., 1965a, C.R.Acad., Sci., 260, No.23, P.6026 1965b, Phys. Fluids, 8, No.12, P.2291. 1967b, Phys. Fluids, 10, No.3, P.507.

[19] Townsend,A.A., 1948, Proc. Camb. Phil. Soc., 44, No.4,P.560-565.

[20] Panchev, S., 1969, Phys. Fluids, 12, No.4, P.935.

Problems of choosing the optimal route for a public transport system with regular trips

A. M. Valuev

Abstract— The paper proposes a formulation of the problem of choosing the optimal route, carried out by means of public transport with regular trips. The methods of conversion of problems of the quickest route in the case of fixed schedule of public transport trip into the previously solved problem of finding the optimal route in a network with a given dynamics are proposed. The results of computational experiments for the rapid transit system in Moscow are presented.

Keywords—dynamic programming, oriented multigraph, public transport trip, time-optimal route.

I. Introduction

There are various types of networks with time-varying characteristics for which problems of optimization of routes emerge. For road networks within a city or an agglomeration there is some regularity in traffic intensity dynamics within a day or a week that gives the possibility to a driver to select the best route between two given points depending on the start time. But to use this possibility a highly developed informational support is needed that must be based on permanent monitoring of traffic on all roads altogether with systematic treatment of the obtained data. Besides, the available information shows only trends and probably the current state of traffic, the latter being estimated with some delay. Exact prognosis of the duration of a given route is principally impossible. Nevertheless, the determination of a rational, probably suboptimal route based on such information and estimates may be useful.

For a passenger of public transport the problem is more definite if the public transport schedule is available for him and buses, trams, underground and commuter trains and other means of public transport really do their trips according to it. However, the number of possible passenger routes between given start and destination points for a big city is great and their assessment demands a lot of calculations. It is shown below that under realistic assumptions even a relatively short passenger trip by the Moscow metro between points separated with the distance of ca 5 km may be effected with at least three rational ways, each being optimal for some intervals of the start time.

So the choice of the best route needs a regular computation method. The approach proposed in the paper includes two main items:

1) The way of representation of the entire set of passenger trips in the form of an oriented graph and simplification of the latter due to a definite route search problem. After that, the problem of finding the time-optimal route for a given start time from a given initial vertex to a given destination vertex (or all other vertices) may be solved with modifications of known methods for route optimization.
2) An algorithm for finding the entire set of start moments at which optimal routes change and its substantiation.

In the paper only the problem of time-optimal routing is considered that corresponds to situations where transport fare is not significant or when the passenger pays for transport services in advance, e.g. buying a monthly ticket.

In many cases, however, people take into account not only duration of the trip but its cost as well. To balance time and money, a weighted sum of them may be proposed as the most adequate criterion, as it is used for commercial air transport operations. Then the optimal route between depend on time and cost weights assigned by a certain passenger.

The solution of the last formulated problem is more complicated but may be reduced to the definition of Pareto set of routes with respect to two or more target indices may be considered. For it, modification of the respective methods, notably by Martin [1], may be developed in the same way as for single-criterion optimal routing on dynamical networks. In general, they will demand much more time-consuming calculations which amount depend on cardinality of the Pareto set.

II. Dynamic Networks and Problems of Route Optimization on Them

Since first works of 1950s and 1960s, notably by Bellman, Dijkstra, and Floyd (results of this period are systematically presented in [2]), a great deal of attention was attracted to problems of route optimization on static networks. Later Levit [3] proposed an algorithm for which computer experiments on road networks revealed its preference with respect to Dijkstra

This work was supported in part by Ministry of Education and Science of Russian Federation as a part of the mandatory section of the state task 2014/113 for National University of Science and Technology MISiS, project No. 952.

A. M. Valuev is with the National University of Science and Technology MISiS, Moscow, 119991, Leninskiy Prospekt, 4, Russia (phone: +7-499-230-2582; fax: 303-555-5555; e-mail: valuev.online@gmail.com).

algorithm. Further a well known A* search method was developed [4].

In fact, the significance of dynamic (time-varying) networks seems greater, especially with respect to transportation problems of any kind. Moreover, there exist many forms of network dynamics, both deterministic and stochastic, quantitative (change of characteristics of edges) and qualitative (change of the network topology). Nevertheless, much less attention was given to routing problems for dynamic networks.

However, massive study of route optimization problems for static networks yields many practical results that may be transferred to the domain of time-varying networks, especially as to efficient computer implementation and taking into account geometrical (geographical) aspect of networks and routes on them [5].

The most widely studied case of dynamic networks is a automobile road network. See, e.g., [6]. Finding the optimal flight plan for a civil aircraft according to the “free flight” concept, i.e., the path detouring immobile and moving obstacles may be represented as the problem of route optimization on a time-varying network too [7].

The problem of path choice for a passenger of public transport mostly is treated as a problem of stochastic optimization [8]. It means that transport schedule is not known by passengers or is usually violated. Even the problem of boarding an overcrowded bus may be considered. It was shown for such condition that the most useful behavior of a passenger is to adhere not to initially chosen route but to a certain routing strategy for which choices during the path are done depending on the current information.

Nevertheless, in many big cities quite opposite situation takes place: schedule is available at each stop of public transport routes and vehicles really adhere to it. The author himself was a witness of this situation as typical for Riga, Prague, Stockholm and Gdansk.

The most widely used assumption for a routing problem is that we deal with a FIFO network. On most roads, so to begin earlier the passage of uniform segment without junctions means to terminate its earlier. On a single-lane road vehicles of any kind in the exact sense form a moving FIFO queue. In this case it is taken for granted that that the vehicle entering the lane earlier exits it earlier as well. For more wide number of cases, for both for congested and free traffic the same general property takes place: time of reaching the end of a road network segment (edge, or arc, in graph-theoretical meaning) increases (or sometimes stays the same) as the time of reaching its beginning increases. The network with such kind of dynamics of passage time for each arc is named a FIFO network. For them, the most popular Dijkstra algorithm must be slightly modified to solve the problem.

III. Method of Finding the Quickest Path on an Oriented Multigraph with Variable Time of Passing Arcs

In that case the original algorithm by Dijkstra may be easily modified for the solution of the Quickest Path the Quickest Path problem.

Let us denote N_V the number of vertices and enumerate them be natural numbers from 1 to N_V. The set of arcs beginning in the vertex V and ending in the vertex W is denoted by $S_A(V,W)$. For a given V the sets of preceding and following adjacent vertices are determined respectively as

$$V_{IN}(V)=\{W \mid S_A(W,V)\neq\varnothing\},\ V_{OUT}(V)=\{W \mid S_A(V,W)\neq\varnothing\}.$$

For a given arc A, its beginning and end vertices are denoted $BEG(A)$, $END(A)$.

The time of passage of a given arc A depends on the moment T_B of reaching its initial vertex. We express this dependence as $T_{PA}(A,T_B)$.We assume that the moment of reaching the end of the arc

$$T_{EA}(A,T_B)=T_B+T_{PA}(A,T_B)$$

is a monotonously non-decreasing function of T_B. The function

$$T_E(W,V,T_B)=\min\ \{T_{EA}(A,T_B) \mid A\in S_A(W,V)\}$$

expresses the minimum time of reaching the vertex W by arcs directly connecting vertices V and W after reaching the vertex V at the time T_B. $T_E(W,V,T_B)$ is a monotonously non-decreasing function of T_B as well. The arc on which this minimum time is achieved is denoted by $A_{MIN}(W,V,T_B)$. After that, the modification of Dijkstra algorithm for the problem of finding quickest routes from the given vertex V_0 for start time T_0. Path to the

```
Step 1. Set TL={$V_0$}, PL=$\varnothing$, $T_1(V_0)= T_0$, NL={1,..., $N_V$}\TL, $A_{IN}(V_0)$.
Step 2. If TL=$\varnothing$, stop. Otherwise:
  2.1. Find $V$=arg min {$T_1(W) \mid W\in$TL}. Set $T_B=T_1(V)$.
  2.2. Set $T_1(W)=T_E(W,V,T_B)$, $A_{IN}(W)=A_{MIN}(W,V,T_B)$, TL=TL$\cup${$W$}, NL=NL\{$W$} for all $W\in$NL$\cap V_{OUT}(V)$.
  2.3. For all $W\in$TL$\cap V_{OUT}(V)$ if $T_1(W)>T_E(W,V,T_B)$ then set $T_1(W)=T_E(W,V,T_B)$, $A_{IN}(W)=A_{MIN}(W,V,T_B)$.
  2.4. Set PL=PL$\cup${$V$}, TL=TL\{$V$}. Return to Step2.
```

The algorithm as presented above gives the possibility to reconstitute paths to each vertex from its end, taking $A_{IN}(V)$ for the arc entering the vertex V and $BEG(A_{IN}(V))$ for the preceding vertex.

Levit algorithm [3] may be modified for time-varying multigraphs in the analogous way.

IV. Representation of the public transport system

A. *Oriented Multigraph of Routes*

We suppose that all possible passenger routes are successions of segments, each being either a PTR by a trip of a certain *public transport route* (PTR) between two stops or a pedestrian movement. As for movement by foot, we suppose that may take place between the departure point and one of nearest stops, between two close stops or between a stop in the neighborhood of the destination point and the latter.

So vertices of *oriented graph of routes via public transport* (OGR) are stops of PTR and points of departure and destination. If several PTR have a stop in the same place we do not distinguish them.

A passenger may go by a certain trip of a certain PTR between an arbitrary pair of its stops. So, in general, each moving between two stops on a certain PTR must be treated as an arc of the OGR. So each arc is linked to a certain PTR. The above mentioned *possible pedestrian movements* (PPM) are treated as arcs too. So if a route has N stops, it yields $N(N\text{-}1)/2$ arc. But if we reduce the search to the set of *rational* routes with the given departure and destination points then we must conclude that arcs corresponding to PTRs may belong to any of them if and only if its

1) the first stop is connected by a PPM with a departure point;
2) the last stop is connected by a PPM with a destination point;
3) the first or the last stop is connected by a PPM to the stop of another PTR;
4) the first or the last stop may be *reasonably* used to change to another PTR.

The fourth case demands explanation. If two PTRs cross in the only stop then it may be reasonably used to change between them. If two PTRs have a succession of common stops then as a rule the points of reasonable change between them may be reduced to the first and the last of them. If both PTR have the same speed, it does not matter where to change. If one of PTR is faster, then the change from it to the slower PTR must be done on the last common stop and the change from the slower PTR to the faster on the first common stop (we assume that the change stop does not alter the fare).

We attribute to each vertex of OGR a definite natural number and furthermore do not distinguish a vertex and its number.

B. Schedules of PTRs

We suppose that all trips of a certain PTR have the same succession of stops, the same duration of movement between a certain pair of stops and the same fare for this movement. If it not true, then each set of trips with the same characteristics is treated as a separate PTR.

For the K-th PTR the succession of stops is denoted by $\{ISM(K,J),\ J=1,\ldots,\ NSM(K)\}$. As a rule. a public transport route has two opposite directions but circular routes and routes including a loop exist as well. We will treat each PTR as a cycle (to be more exact, as a contour on an oriented multigraph). With the use of remainder function MOD, we express the next stop on the route as

$$ISM(K,\ \mathrm{MOD}(J\text{-}1,NSM(K)\text{-}1)+1)).$$

To determine time of passing of route segments the following characteristic is introduced: the time $TS(K,J)$ of reaching the J-th stop of the K-th PTR (counted from the beginning of the trip). Analogously the fare of travel between the J-th and the M-th stops $FS(K,J,M)$ may be introduced. We denote $TST(K,S)$ the start time of the S-th trip of the K-th PTR; so it arrives to the J-th stop at the time

$$T_{\mathrm{SA}}(K,S,J)=TST(K,S)+TS(K,J).$$

Let A be the number of arc corresponding the travel by the K-th PTR between stops $ISM(K,J)$ and $ISM(K,I)$. Then the trip to leave the vertex $ISM(K,J)$ by this PTR after entering it at the time moment T_{B} is determined from

$$S(A,T_{\mathrm{B}})=\arg\min\ \{S \mid T_{\mathrm{SA}}(K,S,J)\geq T_{\mathrm{B}}\}$$

and the desired $T_{\mathrm{EA}}(A,T_{\mathrm{B}})$ as $T_{\mathrm{SA}}(K,\ S(A,T_{\mathrm{B}}),I)$. Therefore, the public transport system with the given system of PTRs and given schedule of their trips is represented in the above form of time-varying oriented multigraph.

V. Properties of the Quickest Route in the Public Transport System

Lemma 1. For each route on a given OGR between two given vertices the time of its termination is a non-decreasing stepwise function of the start time.

Proof of the lemma results from the FIFO property for each arc constituting the route.

Lemma 2. The time-optimal route between any pair of vertices of OGR is acyclic (contourless) for any start time. If there exist more than one time-optimal route then at least one of them is acyclic.

Proof. If path P contains a cycle (to be more exact, a contour), then it contains a vertex V entered twice. Canceling the contour, we begin the final part of the path earlier. Then. according to Lemma 1, the time of the modified path termination will be not greater than for P.

Lemma 3. For a limited period of start times the time of termination of any acyclic path has a finite number of .

Theorem 1. Within each given period there is a finite succession of time moments such that for start time between any of them the set of time-optimal paths between any pair of vertices stays the same.

Proof. The set all acyclic paths on a given OGR is finite. According to Lemma 2, it suffices to reduce to acyclic paths. Time of termination of each path

Theorem 2. The succession $\{T_{1\mathrm{PC}},\ldots,\ T_{N\mathrm{PC}}\}$of time moments of change of the time-optimal paths between two given vertices may within period $[T_0,T_1]$ be found with the following algorithm:

Step 1. Arrange the set of values of $T_{\mathrm{SA}}(K,S,J)$ for all PTR, their trips and stops in the ascending order and find the minimal difference ΔT between two subsequent values.

Step 2. Find the optimal route $P_{\mathrm{C}}=P(T_0)$ for the start time.

Set N=0.

Step 3. Determine the start time T_{BC} after which $T_{\mathrm{E}}(P_{\mathrm{C}}, T_{\mathrm{BC}})$ changes. If $T_{\mathrm{BC}}>T_1$, halt.

Step 4. Set N=N+1, $T_{N\mathrm{PC}}=T_{\mathrm{BC}}$. Find the optimal route $P_{\mathrm{C}}=P(T_{\mathrm{BC}}+\Delta T/2)$ and return to Step 3.

The theorem yields the constructive way to determine optimal paths for any start time. The below example shows, however, that optimal paths may alter frequently in public transport systems incorporating fast and intensive subsystem such as the underground in greatest cities. For this case we can propose an alternative to repeating the calculation of optimal path for numerous moments of start forming $\{T_{1\mathrm{PC}},..., T_{N\mathrm{PC}}\}$. It consists in calculation of the entire set of suboptimal routes [9]. We name the path (for a given T_{B}) *suboptimal* if the increase its duration with respect to the duration on the optimal path is not greater than the given value ΔT_0. For the proper value of ΔT_0 the optimal paths for any start time $T_{\mathrm{B}}\in[T_0,T_1]$ will belong to the set of suboptimal paths for T_0. Calculating it with the method [9] may be more efficient than calculating optimal paths for all $T_{\mathrm{B}}\in\{T_{1\mathrm{PC}},..., T_{N\mathrm{PC}}\}$. After determination of the set of suboptimal paths for a given T_{B} finding the suboptimal path for any other start time consists in calculation duration for each suboptimal paths for this start time and comparison of them.

VI. Example of Finding the Time-Optimal Paths in Moscow Metro

The following example somewhat schematic (in fact, the train schedule has only internal usage and is not available for passengers). But it is realistic with respect to the scheme of lines, duration of their passage and intervals between trains. In the city center, in fact, there exist several rational ways to reach one station from another one. Besides, in these conditions other types of public transport are not competitive. We consider a small fragment of the whole network and indicate only directions of motion that may yield the optimal routes for a certain start time.

The network comprising all possible rational routes in the problem are shown on the Fig. 1. Here black arrows indicate trips by metro trains and white arrows pedestrian trips on the surface and underground passages between adjacent stations. Black dotted line means travel between stations “PC” and “KC” without leaving the train. In fact, it is the same motion as from “PC” to “TC” and then immediately from “TC” to “KC”, but in the model used they differ for the above reasons.

Duration of both pedestrian travels and metro train trips is presented in Tables I and II.

There are three rational routes, namely:

1) Route 1: SP–NO–TE–OR–KR–EP.
2) Route 2: SP–NO–PR–PC–KC–EP.
3) Route 2: SP–TR–MA–TC–KC–EP.

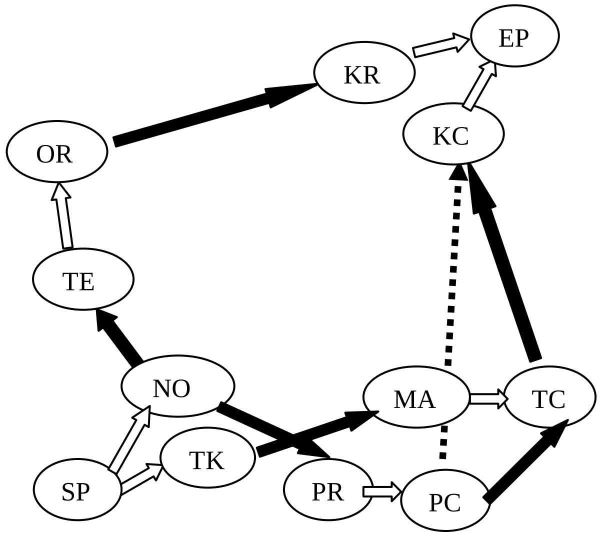


Fig 1. Network of trips by the metro and pedestrian trips

Table I.

Pedestrian passage	Duration, min
SP-NO	5
SP-TK	4
TE-OR	3
KR-EP	3
PR-PK	2
MA-TC	2
KC-EP	6

Table II. Characteristic of metro train trips

Train trips	Duration of a trip	Interval between trips	Arrival time of the 1st train
NO-TE	3 min	1 min 45 s	6 min
OR-KR	6 min 30 s	2 min 30 s	6 min
NO-PR	1 min 30 s	1 min 45 s	4 min
PC-KC	7 min	2 min	8 min
TC-KC	5 min 30 s	2 min	9 min 30 s
TR-MA	2 min 30 s	4 min	2 min 30 s

Computations show they all frequently alternate as optimal routes, often existing simultaneously more than one optimal route. It results from the fact that arrival times when measured in quarters of a minute are integer. In fact, it is not possible to determine the schedule with much greater accuracy, so the situation is typical for transport systems with frequent trips. For conditions presented with Tables I and II and Fig. 1 main results are presented in Table III.

Table III. Moments of change of optimal routes and duration of the optimal and suboptimal routes for them

$T_{i\mathrm{PC}}$	Route 1	Route 2	Route 3	Optimal route
0 min 30s	23min	23min	23min	23min
0 min 45s	23min	25min	23min	23min
1 min 30s	25min 30s	25min	23min	23min
4 min 15 s	25min 30s	29min	27min	25min 30 s
6 min 15 s	30min 30s	31min	27min	27min
9 min 15s	30min 30s	33min	31min	30 min 30 s
10 min 15 s	33min	33min	31min	31min
11min 15 s	33min	35min	35min	33min
12 min 15s	35min 30s	35min	35min	35 min
14 min 15s	38min	37min	35min	35 min
16 min 15s	38min	39min	39min	38 min
17min 15 s	40min 30s	39min	39min	39 min

In this example within a short period of less than 20 min there was one subperiod for which all above mentioned routes were optimal, 3 periods of optimality of Route 1, 3 periods of optimality of Route 2. For other subperiods either Route 1 and Route 3 or Route 2 and Route 3 were optimal simultaneously.

REFERENCES

[1] E. Martins, “On a multicriteria shortest path problem”, *European Journal of Operational Research*, vol.16, pp. 236–245, 1984.

[2] D. T. Phillips, and A. Garcia-Diaz, *Fundamentals of network analysis*, Englewood Cliffs, NJ: Prentice Hall, 1981, pp. 27–213.

[3] E. M. Vasil'eva, and B. Yu. Levit, *Nonlinear transportation problems on networks*, Moscow: Finansy i statistika, 1972, pp. 3–99, (In Russian).

[4] A. V. Goldberg., H. Kaplan, and R. F. Werneck, “Reach for A*: Efficient Point-to-Point Shortest Path Algorithms”, In Workshop on Algorithm Engineering & Experiments, 2006, pp. 129–143.

[5] I. Abraham, D. Delling, A. V. Goldberg, and R. F. Werneck, “A hub-based labeling algorithm for shortest paths in road networks”, in *Experimental Algorithms*, Berlin-Heidelberg: Springer, 2011, pp. 230-241.

[6] R. K. Ahuja, J. B Orlin, S. Pallottino, and M. G. Scutellà, “Dynamic shortest paths minimizing travel times and costs”, *Networks*, vol. 41, no. 4, pp. 197–205, July 2003

[7] A.M. Valuev, and V.V. Velichenko, “On the Problem of Planning a Civil Aircraft Flight along a Free Route”, *Journal of Computer and Systems Sciences International*, vol. 41, no. 6, pp. 979–987, 2002.

[8] V. Trozzi, I. Kaparias, M. G. Bell, and G. Gentile, “A dynamic route choice model for public transport networks with boarding queues”, *Transportation Planning and technology*, vol. 36, no. 1, pp. 44–61, 2013.

[9] A.M. Valuev, “On the Problem of Efficient Search of the Entire Set of Suboptimal Routes in a Transportation Network”, in *Proceedings of the 12th International Conference on Computational and Mathematical* Methods in Science and Engineering, CMMSE2012. La Manga, Murcia, Spain, 2-5 July 2012. Vol. 4. P. 1560–1563.

Andrey M. Valuev graduated from the Moscow Institute of Physics and Technology (MIPT) in 1978 and received a degree of Candidate of Physical and Mathematical Sciences in Computational Mathematics from MIPT in 1984. Doctor of Sciences in Mathematical Modeling, Numerical Methods and Software Packages (2008).

Professor of Department of Public and Municipal Administration in Industrial Regions of the National University of Science and Technology MISiS (formerly in the Moscow State Mining University) and senior researcher of the Laboratory of Man-Machine Dynamics in Mechanical Engineering Research Institute named by A.A.Blagonravov of RAS. Since 2009 to 2014 he was a professor of Department of Computational Modeling of Technological Processes of MIPT. For 30 years participated in many research projects related with mathematical modeling and application of control theory and computational methods, mainly in mining industry and transport systems.

Member of Organizational Bureau of monthly seminar “Scientific and practical problems of automobile-road complex of Russia” under leadership of vice president of the Russian Academy of Sciences V. V. Kozlov.

Some refinement of the conception of symmetry of Volterra integral equations and constructing symmetrical methods for solving them.

G. Mehdiyeva, V. Ibrahimov, M. Imanova

***Abstract*—** Since ancient times, people mainly use the concept of symmetry for the approval of the greatness of the creator, to illustrate of the beauty of the object or of the approval the talent. Therefore, the concept of symmetry can be considered as a well-known notation for all people. Obviously, for the showing of the symmetry of the object it is not necessary to use the exact definition of the "symmetry". But for proving of asymmetry of the object uses his mathematical definition. Therefore, we here have tried to clarify the definition of the concept of symmetry for Volterra integral equations and multistep hybrid methods. To illustrate the results obtained here, are constructed concrete more accurate hybrid methods that are applied to solving the model integral equations of Volterra type. Are proposed methods using the approximate values of the solution the investigated problem only at one mesh point and having an order of accuracy $p \le 8$. And also are used a simple algorithm for using of the constructed here methods.

Keywords— integral equation, notation of symmetry, symmetrical methods.

I. Introduction

In the world there exist many symmetrical objects. Usually the symmetrical objects exist longer, than their peers. For example, Sheikh Safi al-din Khānegāh and Shrine Ensemble in Ardabil, this is at the good condition until present time. This object is built exclusively by using of the laws of symmetry. There are many such examples. Generally speaking, our world has a symmetrical shape. Note that there are objects for which to determine the symmetry are difficult. There are such objects in Samarkand, which is the one of the scarce museum-cities in the world. Consequently, people have faced with the concept of symmetry for a long time. Therefore, many people consider the question of the determining the symmetry of objects is exhausted. However, we try to show that it is not always valid.

The aim of this work is to disseminate the concept of symmetry to the Volterra integral equations. However, in the scientific literature one can find the definition for a symmetric Fredholm integral equation (see e.g. [1], p.124). Here, to solve the above described problems is offered mainly to use the inductive methods. However, in some cases appears necessary to use the deductive methods.

Consider the following nonlinear integral equation of Volterra type:

$$y(x) = f(x) + \int_{x_0}^{x} k(x,s,y(s))ds, \quad x \in [x_0, X]. \quad (1)$$

There is a notion of symmetry to the next integral (see e.g. [1]-[5]):

$$y(x) = f(x) + \int_{a}^{b} b(x,s)y(s)ds, \quad x \in [a,b]. \quad (2)$$

It is a linear integral equation of Fredholm type. The notation of symmetry of the integral equation (2) is determined by the help of symmetry of the kernel $b(x,s)$ for the integral. Consequently, in the nonlinear case we are needed to use different schemes. For example, in the case of

$$b(x,s)y(s) \equiv a(x,s,y(s)).$$

In this case, to determine the notation of symmetry of the integral equation (2) one can be used the symmetry of function $a(x,s,z)$ .Obviously, the integrals involved in equations (1) and (2) are functions of the variable x. Therefore, one can be write the following

$$v(x) = \int_{x_0}^{x} k(x,s,y(s))ds.$$

The notation of symmetry of the integral (1) can be determined by using $v(x)$ in the following form:

Definition 1. If the functions $v(x)$ and $f(x)$ are symmetrical, then (1) is called a symmetric integral equation of Volterra type.

Take into account given that there is no uniform definition for the notation of symmetry for the function of one variable, let us consider to the following definition.

G.Yu.Mehdiyeva - doctor of science, PhD, professor, head of chair of Computational mathematics of Baku State University, Baku, Azerbaijan (corresponding author to provide phone 994125106048 e-mail: imn_bsu@mail.ru)

V.R.Ibrahimov - doctor of science, PhD, professor of the department of Computational mathematics of Baku State University, Baku, Azerbaijan (e-mail: ibvag@mail.com)

M.N.Imanova is with the Baku State University, PhD, teacher of department of Computational mathematics, Baku, Azerbaijan (e-mail: imn_bsu@mail.ru).

Definition 2. If the equality $v(-x) = \pm v(x)$, is hold, then the function $v(x)$ will be called the symmetric. Note that if in equation (1), the integral is as follows:

$$\varphi(x,t) = \int_t^x k(x,t,s,y(s))ds, \qquad (3)$$

then the notion of symmetry for the function $\varphi(x,t)$ can be determined by the classical definition of a symmetric function of two variables.

Note that from the symmetry of the function $b(x,s)$ is not consequent the symmetric of function $a(x,s,y(s))$. Indeed, if suppose that $b(x,s) = \exp(x+s)$ and $y(s) = \exp(-s)$ then from the equality is consequence $a(x,s,y(s)) = \exp(s)$ that the function $a(x,s,y(s))$ is not symmetrical.

In view of the above mentioned it would be logical to determine the notion of symmetry of the integral in the following form:

Definition 3. The Volerra integral equation (1) or the following Fredholm integral equation

$$y(x) = f(x) + \int_a^b k(x,s,y(s))ds, \quad x \in [a,b]$$

is symmetric, if the kernel $k(x,s,y)$ of the integral is symmetric in this area as the function $\psi(x,s) = k(x,s,y(s))$ of two variables.

However, use of such determination in the study of the integral equation (1) sometimes may be accompanied by certain difficulties. This is related with the solution of the integral equation. In such cases, you can use definition 1.

In the study of symmetric integral equations interesting to compare the results obtained by the symmetry and asymmetry of numerical methods. Therefore, in the next section we consider the construction of symmetric numerical methods take into account, that there is a class of asymmetric methods.

II. Some ways of constructing symmetrical multi-step methods

It is known that scientists for the solving of integral equations of Volterra type in basically use the methods, which successfully applied in the solving of ordinary differential equations (ODE) (see e.g. [6] - [16]). That is related to the existence of wide classes of such methods, applied to the solution of the ODE. Among these methods are both symmetrical and asymmetrical methods.

For this aim, here consider refinement of the concept of symmetry of the numerical methods and integral equations. It is known that in the study of some processes for solving various equations are commonly used approximation methods. It follows the symmetry of the method desirable research with the problem to the solution of that problem is applied. It should be noted that at present time not exist the precise mathematical definition for the integral concept of symmetry for the integral equations. However, there are some attempts to determine the symmetry of the integral equations. For example, the symmetric for linear Fredholm integral equation is determined by the concept of symmetry for the kernel of the integral equations. There are several definitions for the concept of symmetry for the Runge-Kutta, Adam and others methods (see. [17 str.232, 354], [18, p.106]. In the paper (see [19]) Dalkvista. Proposed to define of the symmetry for the next difference method

$$\sum_{i=0}^{k} \alpha_i y_{n+i} = h \sum_{i=0}^{k} \beta_i y'_{n+i} \qquad (4)$$

in the following form method is symmetrical if there holds:

$$\alpha_j = -\alpha_{k-j} \quad \beta_j = \beta_{k-j} \quad (j = 0,1,2,...,k).$$

In this case appear some questions. For example, how one can be to determine of the symmetry methods in the case when the order of the stable methods of type (4), the value of the variable k - is odd or the method is explicit.

It is known that from formula (4) one can be obtained forward jumping methods, which are usually more accurate than the implicit methods of type (4). Therefore, the study of the symmetry for these methods has both theoretical and practical interest. Consider the following explicit method

$$y_{n+2} = y_n + 2hy'_{n+1} \,. \qquad (5)$$

This method has a degree $p = 2$ and usually is called the midpoint method. Some of scientists suppose that the method (5) is symmetrical and truncation error for the following sequence of methods:

$$\bar{y}_{n+1} = y_n + hf(x_n, y_n),$$
$$\hat{y}_{n+1} = y_n + h(f(x_n, y_n) + f(x_{n+1}, \bar{y}_{n+1}))/2,$$
$$y_{n+2} = y_n + 2hf(x_{n+1}, \hat{y}_{n+1}) \,, \qquad (6)$$

here $y' = f(x,y)$.

These facts compel us to consider the definition the concept of symmetric for multi-step methods in more general form. The concept of symmetry for the midpoint method is mainly associated with the submission of its expansions of truncation error in powers on h (see e.g. [20, p.44]). Consider the following forward-jumping method

$$y_{n+1} = y_n + h(5f_n + 8f_{n+1} - f_{n+2})/12 \qquad (7)$$

which is stable and has the degree $p = 3$ (see for exam. [20, p.104-105]).

The method of (7) has a degree $p = 3$ of order accuracy and can be applied to the solving of the initial value problem for the ODE. Remark that the forward-jumping methods have constructed almost simultaneously construction of the Runge-Kutta methods. However, for their use appears the necessary to determine the values of the solution of the considering problem on the next mesh points. For example, in method (7), these values are quantity of y_{n+2}. The method of (7) can be formally considered as a one-step, since with the help of knowing y_n one can be compute the values of variable of y_{n+1}. But for the definition of the values must be known the values y_n and y_{n+2}, by using the method (7) may be considered as a two-step method. Application of the forward-jumping methods to the solving of equation (1) was investigated in [7].

Consider the following symmetrical method

$$y_{n+1} = y_n + h(f(x_n, y_n) + f(x_{n+1}, y_{n+1}))/2 \quad (8)$$

Let us commonly is called trapezoidal method. This method is symmetrical (see [17] - [19]). By using the following functions:

$$\begin{aligned}\varphi(y_n, y_{n+1}) = y_n - y_{n+1} + \\ + h(f(x_n, y_n) + f(x_{n+1}, y_{n+1}))/2\end{aligned}. \quad (9)$$

Can be is determined the symmetry of the trapezoidal method. Obviously, the function (9) is not symmetric, since

$$\begin{aligned}\varphi(y_{n+1}, y_n) = y_{n+1} - y_n + \\ + h(f(x_n, y_{n+1}) + f(x_{n+1}, y_n))/2\end{aligned}.$$

But if we replace h by the (-h), it becomes symmetric.

If the above mentioned scheme is applied to the midpoint method, then we have:

$$\varphi(y_{n+2}, y_{n+1}, y_n) = y_n - y_{n+2} - 2h(f(x_{n+1}, y_{n+1}) . \quad (10)$$

Therefore, the midpoint method is symmetric.

By using of the definition of the some properties of numerical methods from the values of their coefficients here is proposed to determine the symmetry of the method in the following form.

Determination 4. The multistep methods with the constant coefficients, is called symmetric if this presentation take part in the values of the solutions of the problem, as in the previous and subsequent points. And the amount of it used in previous and subsequent points of the same.

Obviously, the method (6) may be replaced by the following:

$$y_{n+1} = y_n + h(ly'_n + my'_{n+1} - ly'_{n+2})/m . \quad (11)$$

We can prove that in the class of the methods of (11) there is no method with the degree $p > 2$. In order that the method (11) had a degree $p = 2$ between the values of the quantity l and m should satisfy the relation of type $m = 4l$. In this case, the accuracy of the method (11) and the trapezoidal method are coincide. For comparison, these methods consider their application to solve the following equations:

$$y(x) = \frac{1}{2}\int_{-x}^{x} \cos(s)ds, \ x \in [0,1], \quad (12)$$

$$y(x) = \sin(-x) + \int_{-x}^{x} \cos(s)ds, \ x \in [0,1] \quad (13)$$

the exact solution are equal $y(x) = \sin(x)$.

Recently, experts for constructing more accurate methods for solving Volterra integral equations, use hybrid methods (see. e.g. [12] - [16]). To this end, we consider the application of the following hybrid method:

$$y_{n+1} = y_n + h(f_{n+1/2-\alpha} + f_{n+1/2+\alpha})/2, \ \ (\alpha = \sqrt{3}/6), \quad (14)$$

to the solving of the next Volterra integral equation:

$$y(x) = f(x) + \int_{-x}^{x} k(x, s, y(s))ds, \ x \in [0,1] . \quad (15)$$

Note that when using the method of (14), the main difficulty is to calculate the values of quantities $y(x_n + h/2 \pm 2h)$. To this end, here has used of the explicit Euler and the trapezoidal method. To illustrate the results, we consider the use of the above mentioned methods to solve the problem (12) and (13).

The results of applying the methods (11) for $m = 4$ and the trapezoidal method for solving of the equation (12) posted in table 3. Note that the results of applying have tabulated in Table 1 and 2 of the method (7) and method (14) to the solving of the equations (12) and (13) respectively.

Table 1

x_n	method (7)	method (14)
h=0.05		
0.10	3.554E-09	1.444E-10
0.40	4.482E-08	5.634E-10
0.7	1.286E-07	9.862E-10
1.0	2.255E-07	1.217E-9
h=0.01		
0.10	2.234E-10	2.310E-13
0.40	3.348E-09	9.013E-13
0.7	9.896E-09	1.491E-12
1.0	1.893E-08	1.947E-12

Table 2

x_n	method (7)	method (14)
h=0.05		
0.10	7.208E-07	4.623E-09
0.40	7.762E-06	1.803E-08
0.7	2.154E-05	2.983E-08
1.0	3.389E-05	3.896E-08
h=0.01		
0.10	4.468E-10	4.621E-13
0.40	6.697E-09	1.802E-12
0.7	1.972E-08	2.982E-12
1.0	3.786E-08	3.895E-12

Table 3

x_n	method (8)	method (11)
h=0.1		
0.10	8.3204E-05	1.653E-04
0.40	3.245 E-04	6.374E-04
0.7	5.369E-04	1.041E-03
1.0	7.013E-04	1.341E-03
h=0.05		
0.10	2.079E-05	4.149E-05
0.40	8.113E-05	1.609E-04
0.7	1.342E-04	2.645E-04
1.0	1.753E-04	3.432E-04
h=0.01		
0.10	8.319E-07	1.663E-06
0.40	3.245E-06	6.480E-06
0.7	5.368E-06	1.070E-05

1.0	7.012E-06	1.396E-05

For comparison of some methods we apply them to the solving two different integral equations with variable symmetrical boundaries the solutions of which coincide. According to the results of calculations, we obtain that the hybrid method (14) is more accurate than other. This result can be considered natural, since the hybrid method is more accurate than another. This view can be extended to the other methods. However, as can be seen from the above mentioned equations, the hybrid method has higher order of accuracy from the remaining methods. Remark that the order of accuracy for the methods of (8) and (11) (for $l=1$) are same. But when applied that method to the solving of the equation (12), in some cases we received non identical results which show once again that comparison of methods using only their accuracy is not available. It is desirable to use the region of stability or some properties of the solution of the original problem**.**

III. Conclusion

Here we describe some of the ways to determine the symmetry of the numerical methods and Volterra integral equations. And to contract a more accurate hybrid method that have been applied to the solution of integral equations with variable boundaries. For simplicity, numerical methods have constructed them as a finite-difference formula. This approach allowed us to apply methods that are constructed as solutions of ODE and to the solution of integral equations of Volterra type. The study of numerical solutions of integral equations with symmetric boundary conditions shows the relevance of the study of various aspects of the symmetry of the object. In the application of numerical methods to the solving of these integral equations necessary to calculate the approximate value of the solution at the point x_m and $-x_m$. Therefore, the process uses multi-step methods for the solution of integral equations with symmetric variable boundaries in some sense can be considered symmetrical.

In the end, we want to thank all readers and will be happy if we get feedback from them.

Acknowledgment

The authors wish to express their thanks to academician Ali Abbasov for his suggestion to investigate the computational aspects of our problem and for his frequent valuable suggestion.

References

[1] Manzhirov A.V. Polyanin A.D. Handbook of Integral Equations: Methods of solutions. Moscow: Publishing House of the "Factorial Press", 2000, 384 p.

[2] V. Volterra. Theory of functional and of integral and integro-differensial equations, Dover publications. Ing, New York, Nauka, Moscow, 1982 p.304 (in Russian).

[3] T.A.Aliev. A.M. Abbasov, G.A. Guluyev, F.H. Pashayev, U.E. Sattarova. System of robust noise monitoring of anomalous seismic processes, Soil Dynamics and Earthquake Engineering, Vol. 53, October 2013, pp. 11-15.

[4] T.A.Aliev. Digital Noise Monitoring of Defect Origin Series Springer, London, 2007, 235 p.

[5] A.N. Guz. About continuum theory of materials with small-scale distortions in the structure, Dokl. ANSSR in 1983, Vol.268, № 2, pp. 307-313.

[6] V.Volterra. Theory of functionals and integral and integro-differential equations. M., Nauka, 1982, 304p. .

[7] Mehdiyeva G.Yu., Imanova M.N., Ibrahimov V.R. On one application of forward jumping methods. Applied Numerical Mathematics, Volume 72, October 2013, p. 234–245.

[8] P.Linz Linear Multistep methods for Volterra Integro-Differential equations, Journal of the Association for Computing Machinery, Vol.16, No.2, April 1969, pp.295-301.

[9] G. Yu Mehdiyeva, M. N. Imanova, V. R. Ibrahimov Application of the hybrid method with constant coefficients to solving the integro-differential equations of first order. 9th International conference on mathematical problems in engineering, aerospace and sciences, AIP, Vienna, Austria, 10-14 July 2012, 506-510.

[10] Mehdiyeva G., Imanova M., Ibrahimov V. On a Research of Hybrid Methods. Numerical Analysis and Its Applications, Springer, 2013, p. 395-402.

[11] V.Ibrahimov V.Aliyeva The construction of the finite-difference method and application Proceedings of the International Conference on Numerical Analysis and Applied Mathematics 2014 (ICNAAM-2014) AIP Conf. Proc. 1648, 850049-1–850049-5;

[12] Makroglou A.A. Block - by-block method for the numerical solution of Volterra delay integro-differential equations, Computing 3, 1983, 30, №1, p.49-62.

[13] A, Feldstein, J.R Sopka. Numerical methods for nonlinear Volterra integro differential equations // SIAM J. Numer. Anal. 1974. V. 11. P. 826-846.

[14] Ibrahimov V.R, Imanova M.N. On a Research of Symmetric Equations of Volterra Type International journal of mathematical models and methods in Applied sciences Volume 8, 2014, 434-440.

[15] Imanova M.N. One the multistep method of numerical solution for the Volterra integral equation Transactions issue mathematics and mechanics series of physical -technical and mathematical science, , XXBI, 2006,№1.

[16] Makroglou A. Hybrid methods in the numerical solution of Volterra integro-differential equations. Journal of Numerical Analysis 2, 1982, pp.21-35.

[17] E. Hairier, S.P.Norsett, G.Wanner. Solving ordinary differential equations. (Russian) M., Mir, 1990, 512 p.l

[18] Современные численные методы решения обыкновенных дифференциальных уравнений // Редакторы Дж.Холл и Дж. Уатт, Изд-во «Мир», Москва, 1979.

[19] G.Dahlquist Convergence and stability in the numerical integration of ordinary differential equations. Math. Scand. 1956, №4, p.33-53

[20] Mehdiyeva G., Ibrahimov V. On the research of multistep methods with constant coefficients, LAP LAMBERT Academic Publishing, 2013, 314 p., (Russian).

About some communications between methods of applications to the solving ODE and integral equations of Volterra type

G. Mehdiyeva, V. Ibrahimov, M. Imanova

Abstract— The classic approach at solving integral equations with the variable boundaries consists of some methods of the theory of difference methods or differential equations. In contrast from the approach here are considered the relationship between methods of application to the solution of differential equations and to the solution of integral equations of Volterra type. To this end, we consider the application of forward-jumping methods and hybrid methods for the solution of the integral equations with the symmetric variable boundaries. Here are constructed effective methods of taking into account the special structure of these equations. And are proposed specific methods for the solving of the above equations. And for illustration of the results obtained here are used the model equations.

Keywords— integral equation, notation of symmetry, symmetrical methods.

I. Introduction

IT is known that in the research of the solving of integral equations Volterra type, commonly are used communication between the ordinary differential equations and the integral equations of Volterra type. For example, there are the methods of quadrature. This method depends on the area of application in different forms. Naturally arise the question of finding the connection between the specified methods. It is clear that by using a comparison of these methods can be effective methods to build. One such method is a hybrid method, which are constructed on the intersection of the Runge-Kutta and Adams (see. for example [1]- [6]). Based on this, here are considering the application of hybrid methods to the solution of integral equations with variable boundaries. In the construction of the method for the solving integral equations with variable boundaries, we reduce them to equations with one fixed boundary. Then we use known methods in a new production.

Let us consider the following integral equation with variable borders:

$$y(x) = \varphi(x) + \int_0^x F(x,s,y(s))ds, \quad x \in [0, X]. \qquad (1)$$

Let us assume that (1) has a unique solution defined on the interval $[0, X]$. Segment $[0, X]$ using the constant step $0 < h$ -split into N equal parts, and the split point will take in the form: $x_i = x_0 + ih \ (i = 0,1,...,N)$. By means of y_i and $y(x_i) \ (i = 0,1,...,N)$ denote the approximate and exact values of the solution of equation (1) at the point x_i, respectively.

Considering that the integral equation (1) is investigated at a high level (see [7]-[20]), determine the relationship between the equations (1) and integral equations with variable symmetric boundaries.

It is obvious that the following integral equation:

$$y(x) = f(x) + \int_{-x}^x K(x,s,y(s))ds$$

can be expressed as:

$$y(x) = f(x) + \int_0^x (K(x,s,y(s)) + K(x,-s,y(-s)))ds.$$

Denoting by

$$\varphi(x,s,y(s)) = K(x,s,y(s)) + K(x,-s,y(-s))$$

rewrite the last integral in form:

$$y(x) = f(x) + \int_0^x \varphi(x,s,y(s))ds,$$

which coincides with the integral equation (1).

Therefore, choosing the kernel of the integral in the form:

$$\varphi(x,s,y(s)) \equiv K(x,s,y(s)) + K(x,-s,y(-s));$$

obtain the following symmetric integral equation:

$$y(x) = f(x) + \int_{-x}^x K(x,s,y(s))ds, \quad x \in [0, X]. \qquad (2)$$

Consider the application of the following methods to the solution of equation (1) (see [5], [15]):

G.Yu.Mehdiyeva - doctor of science, PhD, professor, head of chair of Computational mathematics of Baku State University, Baku, Azerbaijan (corresponding author to provide phone 994125106048 e-mail: imn_bsu@mail.ru)

V.R.Ibrahimov - doctor of science, PhD, professor of the department of Computational mathematics of Baku State University, Baku, Azerbaijan (e-mail: ibvag@mail.com)

M.N.Imanova is with the Baku State University, PhD, teacher of department of Computational mathematics, Baku, Azerbaijan (e-mail: imn_bsu@mail.ru).

$$y_{n+1} = y_n + f_{n+1} - f_n + h(F(x_n, x_n, y_n) + \\ + F(x_{n+1}, x_n, y_n) + 2F(x_{n+1}, x_{n+1}, y_{n+1}))/4\,, \qquad (3)$$

$$y_{n+1} = y_n + f_{n+1} - f_n + h(F(x_n, x_n, y_n) + F(x_{n+1}, x_n, y_n) \\ + 2F(x_{n+1}, x_{n+1}, y_{n+1}) - 2F(x_{n+2}, x_{n+2}, y_{n+2}) + \\ + 2F(x_{n+2}, x_{n+1}, y_{n+1}))/4 \; . \qquad (4)$$

The accuracy of these methods are the same and equal $p = 2$. The first of these methods is the one-step and the second is two-step, note that the second method is symmetric.

Note that if the integral equation (1) is symmetric, it becomes necessary to parallel computation y_{-m} with computation y_m. Generally, methods are specially adapted to solving symmetric integral equations are more efficient than the methods of constructing to the solution of the integral equations of type (1).

Easily, we can prove that the accuracy of the following method

$$y_{n+1} = y_n + h(ly'_n + my'_{n+1} - l_2 y'_{n+2})/(m+d)$$

when $d = 0, m = 4l$ and $l_1 = l_2$ equally $p_{\max} = 2$. However, exist methods with the degree $p > 2$ when $d \neq 0$ and $l_1 \neq l_2$. For example, the following method

$$y_{n+1} = y_n + h(5y'_n + 8y'_{n+1} - y'_{n+2})/12\,,$$

which has a degree of $p = 3$ (see [21]).

The purpose of this work is to apply the method of constructing solutions to equation (1), the study of the solution of integral equations of Volterra type with symmetrical borders. Note that if the decision of the above mentioned problems of application such a numerical method, using a scheme (3) and (4), it formally be considered the question of solutions of integral equations with symmetric boundaries of the Volterra is solved $y(-x_m)$ and $y(x_m)$ $(m = 0,1,2,...)$ solution of the integral equation. To get more accurate results it is necessary to use implicit methods, but their application to the solution of specific equations have difficulty choosing appropriate methods of forecasting. Consequently, application to solving of integral equations with symmetric variable boundaries method for constructing solutions to the equation (1) does not take place mechanically. According to this we consider the use of multi-step methods with constant coefficients to solving integral equations with symmetric boundaries of Volterra type.

II. Application of multi-step methods to the solution of integral equations with special structure

Suppose that the kernel of the integral can be represented as a linear combination of some functions:

$$K(x,s,y) = \sum_{i=1}^{m} a_i \varphi_i(x,s,y) \qquad (5)$$

or as follows:

$$K(x,s,y) = \prod_{i=1}^{m} b_i \psi_i(x,s,y)\,. \qquad (6)$$

If we consider the case $m = 1$, then regardless of the use of representations the kernel of the integral at the form (5) or (6) in equation (1), we obtain an integral equation of Voltaire-Uryson type. But with $m = 2$ by using the decomposition (6) in the kernel of the integral equation (1) we obtain integral equation of Gammarshteyna type. Let us consider applications the expansions (5) in the research of the numerical solution of equation (1). Suppose $a_1 = a_2 = 1$ и $\varphi_2(+x, -s, y(-s)) = \varphi_1(x, s, y(s))$. Then, from the equation (1) we have:

$$y(x) = f(x) + \int_0^x (\varphi_1(x,s,y(s)) + (\varphi_1(x,-s,y(-s)))ds\,.$$

It follows that

$$y(x) = f(x) + \int_{-x}^{x} \varphi_1(x,s,y(s))ds\,. \qquad (7)$$

In the papers [8]-[10], [18], [22] consider the application the multi-step method to the solution of the integral equation (1). However, in these works in construction of methods applied to the solution of equation (1) used to replace the integral by integral sum as a result of what computational work increases at the transition from a single integrated point to another. In order to eliminate this disadvantage of the quadrature method in the works [5], [11], [14], [19] the methods of the following type:

$$\sum_{i=0}^{k} \alpha_i y_{n+i} = \sum_{i=01}^{k} \alpha_i f_{n+i} + h \sum_{j=0}^{k} \sum_{i=0}^{j} \gamma_i^{(j)} K(x_j, x_i, y_i) \qquad (8)$$

$$(\alpha_k \neq 0)$$

From the proposed method can obtain as explicit as well as implicit methods. Among the implicit methods the most interesting are methods with higher accuracy. It is known that among the forward-jumping methods of type (8) exist methods are more accurate than implicit methods of type (8). For construction the forward-jumping methods it is possible to use the formula (8) with no restrictions $\alpha_k \neq 0$. Then, in the simplest case forward-jumping method can be written in the following form (see e.g. [5]):

$$\sum_{i=0}^{k-m} \alpha_i y_{n+i} = \sum_{i=0}^{k-m} \alpha_i f_{n+i} + h \sum_{j=0}^{k} \sum_{i=0}^{j} \gamma_i^{(j)} K(x_j, x_i, y_i)\,.$$

$$(m > 0,\ \sum_{j=0}^{k} \gamma_k^{(j)} \neq 0) \qquad (9)$$

Here the quantity m - receives integer values. Note that if formally in relationship (9) we put $m = 0$, we obtain the method (8). However, the method of (9) is the forward-jumping method to the positive values of the quantity m and $\gamma_k^{(0)} + ... + \gamma_k^{(k)} \neq 0$. Consequently, many methods such as (8) and (9) do not intersect. The coefficients in the methods of (8) and (9) are determined by the same way. For example, the coefficients of the method (8) $\gamma_i^{(j)} \quad (i,j=0,1,...,k)$ can be defined as a solution following linear algebraic system of equations:

$$\sum_{j=0}^{k} \gamma_i^{(j)} = \beta_i \quad (i=0,1,...,k), \tag{10}$$

where the coefficients $\alpha_i, \beta_i \quad (i=0,1,...,k)$ are the solution of the following homogeneous system of linear algebraic equations:

$$\sum_{i=0}^{k} \alpha_i = 0; \ \sum_{i=0}^{k} i\alpha_i = \sum_{i=0}^{k} \beta_i;$$
$$\sum_{i=0}^{k} i^l \alpha_i = l \sum_{i=0}^{k} i^{l-1} \beta_i \quad (l=2,3,...,p). \tag{11}$$

It is known, that order of the accuracy of stable methods, the coefficients which are defined as the solutions of system (11) is satisfy the next condition $p \le k+2.$ Since the coefficients $\gamma_i^{(j)} \quad (i,j=0,1,...,k)$ are determined from the linear-algebraic equations (10), it possible to assume that the stable methods of the type (8) are more accurate.

Therefore the order of the accuracy of the constructed stable methods such as (8) are satisfied the condition $p \le k+2$. But the stable methods such as (8) with the order of accuracy $p > k+2$ have not constructed. In the last time for the construction methods with the higher order of accuracy are used of hybrid methods. These methods are constructed in the middle of the XX century at the junction of the Runge-Kutta and Adams method, which have applied to the solving of the ordinary differential equations. In [] these methods are called the methods of the fractional steps. Here we consider to construction and application of some hybrid methods to the solving of the integral equations of Volterra type with the symmetric boundaries. For the construction of the methods to the solving of integral equations usually are use is the multistep methods or forward-jumping otherwise hybrid methods. In this regard, consider construction a hybrid methods, which can be written are as follows:

$$\sum_{i=0}^{k} \alpha_i y_{n+i} = h\sum_{i=0}^{k} \beta_i y'_{n+i} + h\sum_{i=0}^{k} \gamma_i y'_{n+i+v_i} \quad . (12)$$
$$(|v_i| < 1; \ i=0,1,...,k)$$

If in this formula we put $y' = f(x,y)$, then the formula (12) turns to the methods, which have applied to solving of the ordinary differential equations. Consider the determination of the coefficients of the method (12). For this purpose, we use the method with undetermined coefficients. Than we can write the follows:

$$\sum_{i=0}^{k} \alpha_i = 0, \quad \sum_{i=0}^{k} i\alpha_i = \sum_{i=0}^{k} (\beta_i + \gamma_i), \tag{13}$$

$$\sum_{i=0}^{k} \frac{i^m}{m!} \alpha_i = \sum_{i=0}^{k} \left(\frac{i^{m-1}}{(m-1)!} \beta_i + \frac{l_i^{m-1}}{(m-1)!} \gamma_i \right),$$
$$(m = 2,3,..., p; \ l_i = i + \nu_i, \ |\nu_i| < 1, \ i = 0,1,...,k)$$

Obviously, when $\gamma_i = 0 \ (i=0,1,...,k)$ or $\nu_i = 0 \ (i=0,1,...,k)$ from the (12) we obtain the known multi-step methods, among which the most popular are the methods of the trapezoidal and midpoint differences. One of them is an implicit and other is explicit method. Usually these methods are used as a prediction formula. Application of these methods to solving of the initial problem for ODE investigated thoroughly enough. Therefore, here we consider of the application of some modifications of these methods to solving of integral equations of the type (2). It is known that the methods of the trapezoidal and midpoint tare stable and have the same order of accuracy $p = 2$. Note that for the construction of more accurate methods can be used methods with the fractional steps or hybrid methods. One of the simple hybrid method $k = 1$ can be written in the following form:

$$y_{n+1} = y_n + h(y'_{n+2} + y'_{n+1-2})/2 \quad (\alpha = (3-\sqrt{3})/6) \ (14)$$

For illustrating of the advantages of the hybrid methods, considered here the application of the trapezoidal, forward-jumping and the hybrid methods to solving of the following simple examples: 1. $y(x) = -x + \int_{-x}^{x} \frac{1+y^2(s)}{1+s^2}$ (the exact solution $y(x) = x$)

2. $y(x) = \frac{x}{6} - x + \int_{-x}^{x} (\frac{s^3}{6} + s + 1 - y(s))ds$ (the exact solution $y(x) = x^3/6 + x$)

The results received by above mentioned methods are the same.

III. Conclusion

Here are mainly hybrids methods are compares with others. We show the advantages of these methods. And also are shown some disadvantages of hybrid methods that are related with their use. As is known, one of the most popular schemes for use of implicit methods it is predictor-corrector scheme (see. Eg). Such schemes are constructed and applied in the study of some differential problems by using of hybrid methods. With the help of specific problems we have shown that the quality of the result of the methods of the predictor and corrector is descended from the selection from the formula of predictor in the using of the predictor-corrector methods. Selection formula of predictor to using in hybrid methods is complicated by the fact that it is necessary to determine the values of the solution of the considering problem in the irrational mesh points. For example in our method, those values are $y_{n+\alpha}$ and $y_{n+1-\alpha}$. . Usually these disadvantages of the methods are using in constructing algorithms for them application. As shown above, hybrid methods or methods with fractional steps size are more accurate and have the extended stability region. For example, consider the following method of Simpson

$$y_{n+2} = y_n + h(y'_{n+2} + 4y'_{n+1} + y'_n)/3$$

which is stable and has the order of accuracy $p = 4$. In this method, h we replace through $h/2$. Then we have

$$y_{n+1} = y_n + h(y'_{n+1} + 4y'_{n+1/2} + y'_n)/6$$

Constructed different methods with the different properties for calculating the values $y_{n+1/2}$ is easier than constructed the methods to calculate values $y_{n+\alpha}$. For the illustration of the advantages of the hybrid method here considered to solving of the following problem $y' = \cos x,\ y(0) = 0,\ x \in [0,1]$ by using the Simpson methods and its modifications. Consequently, the stability region of the modification of Simpson's methods is extended than for the Simpson's method. Note that the above has shown that hybrid methods are more accurate than the corresponding implicit methods. Trapezoidal method is constructed by the using of two values of the solution of the original problem and has the order of accuracy $p = 2$. However, the method (14) is constructed by the using of two values of the solution of the original problem in the mesh and hybrid points. Therefore, when constructing an algorithm for using the hybrid method (14) the values of the solution at hybrid points are replaces by the corresponding formulas so that the algorithm has been used only two values of the solution of the original problem. Thus we see that hybrid methods are more promising.

.

ACKNOWLEDGMENT

The authors wish to express their thanks to academician Ali Abbasov for his suggestion to investigate the computational aspects of our problem and for his frequent valuable suggestion.

REFERENCES

[1] Butcher J.C. A modified multistep method for the numerical integration of ordinary differential equations. J. Assoc. Comput. Math., v.12, 1965, pp.124-135.

[2] Gear C.S. Hybrid methods for initial value problems in ordinary differential equations. SIAM, J. Numer. Anal. v. 2, 1965, pp. 69-86

[3] Mehdiyeva G., Ibrahimov V. Nasirova I.I. Transactions issue mathematics and mechanics series of physical-technical and mathematical science, 2005, 5, 55-62.

[4] L.M.Skvortsov. Explicit two-step Runge-Kutta methods. Math. modeling 21, 9 (2009), 54-65.

[5] Mehdiyeva G.Yu., Imanova M.N., Ibrahimov V.R. On one application of forward jumping methods. Applied Numerical Mathematics, Volume 72, October 2013, p. 234–245.

[6] V. Volterra. Theory of functional and of integral and integro-differensial equations, Dover publications. Ing, New York, Nauka, Moscow, 1982 p.304 (in Russian).

[7] Mehdiyeva G., Imanova M., Ibrahimov V. A way to construct an algorithm that uses hybrid methods. Applied Mathematical Sciences, HIKARI Ltd, Vol. 7, 2013, no. 98, p.4875-4890.

[8] P.Linz Linear Multistep methods for Volterra Integro-Differential equations, Journal of the Association for Computing Machinery, Vol.16, No.2, April 1969, pp.295-301.

[9] A, Feldstein, J.R Sopka. Numerical methods for nonlinear Volterra integro differential equations // SIAM J. Numer. Anal. 1974. V. 11. P. 826-846.

[10] Makroglou A.A. Block - by-block method for the numerical solution of Volterra delay integro-differential equations, Computing 3, 1983, 30, №1, p.49-62.

[11] Mehdiyeva G., Imanova M., Ibrahimov V. Some application of the hybrid methods to solving Volterra integral equations Advances in Applied and Pure mathematics, Proceedings of 2 Intern.Conf. on Math.Comp and Aqtatist.Science (MCSS), 2014, 352-356.

[12] A.F. Verlan, V.S. Sizikov. Integral equations: methods, algorithms, programs. Kiev, Naukovo Dumka, 1986, 384 p.

[13] Makroglou A. Hybrid methods in the numerical solution of Volterra integro-differential equations. Journal of Numerical Analysis 2, 1982, pp.21-35.

[14] Mehdiyeva G., Imanova M., Ibrahimov V. On a Research of Hybrid Methods. Numerical Analysis and Its Applications, Springer, 2013, p. 395-402.

[15] G. Yu Mehdiyeva, M. N. Imanova, V. R. Ibrahimov Application of the hybrid method with constant coefficients to solving the integro-differential equations of first order. 9th International conference on mathematical problems in engineering, aerospace and sciences, AIP, Vienna, Austria, 10-14 July 2012, 506-510.

[16] Manzhirov A.V. Polyanin A.D. Handbook of Integral Equations: Methods of solutions. Moscow: Publishing House of the "Factorial Press", 2000, 384 p.

[17] Mehdiyeva G., Imanova M., Ibrahimov V. Application of the hybrid methods to solving Volterra integro-differential equations. World Academy of Science, engineering and Technology, Paris, 2011, 1197-1201

[18] H.Brunner. Imlicit Runge-Kutta Methods of Optimal oreder for Volterra integro-differential equation. Methematics of computation, Volume 42, Number 165, January 1984, pp. 95-109.

[19] Mehdiyeva G., Imanova M., Ibrahimov V. Solving Volterra Integro-Differential Equation by the Second Derivative Methods, Applied Mathematics & Information Sciences, Volume 9, No. 5 (Sep. 2015), PP:2521-2527

[20] Henrici P. Discrete variable methods in ordinary differential equation Wiley, New York, 1962.

[21] Mehdiyeva G., Ibrahimov V. On the research of multistep methods with constant coefficients, LAP LAMBERT Academic Publishing, 2013, 314 p., (Russian).

[22] Yanenko N.N. Метод дробных шагов решения многомерных задач математической физики, Издательство «Наука»- Сибирское отделение Новосибирск-1967

Modern information and communication approaches to traffic monitoring

Mikhail Volkov, *Fellow, MTUCI, MADI*, Marina Yashina, *Proffesor, MTUCI, MADI*

***Abstract*—Mathematical modeling of traffic flows on complex city networks usually use initial average data of space and time. Processing of real data capturing and problem of information verifications are very important for promotion of infocommunication technology.**

Macro modeling of transportation on networks often uses data collection of Origin-Destination Matrix type. Usually this construction is formed by statistical methods of approximate measurements.

In the article we suggest modern methods of data collection for the use of traffic flows modeling on complex city network, for example crossroads. Algorithm of recover the Origin-Destination Matrix on real-time data is presented.

***Index Terms*—traffic monitoring, Origin-Destination Matrix, flow characteristics, CV, motor transport flow**

I. Introduction

Recent advances in info communication require to specify what is correspondence. Nowadays technologies allow to generate route protocol for every road user with common means of navigation. Real correspondence describes not only the source and outflow, but route, velocity, part of street-road network and current transport movement.

Potential correspondence is only a draft with not evident initial conditions. Most travel motivations are generalized. Modern economy and manufacturing process are becoming more flexible and computerized.

On the other hand traffic network is commonly unpredictable in the case of heavy traffic, despite the navigators, which recognize real time traffic conditions and mass media that informs users about traffic state, trying to optimize their possible route. These methods are not always successful as real time correspondence optimization limited by heavy traffic. So, assume that info communication way of potential correspondences processing and modeling is needed, or what is the same: potential network users site (PNUS). Formalization request on network usage including movement with fixed starting and (or) ending points, parking space usage and other correspondence characteristics enable to synchronize demand, specify correspondence at the estimate planning trip stability, as primary demand network scripts are possible. These new technologies enable to structure correspondences in accordance with classification and get higher precision, compare with request method. Moreover, despite the great interest to this subject strict recovery correspondence technology request model is missing. In addition, it is difficult to recognize real time movement information accuracy or plan future behavior. Process measurement by interfering its functioning seems to be old fashioned.

On the other hand actual correspondence in the first approximation can be estimated with geo information technologies. For instance sensors in position control system Yandex jams set and transmit information voluntarily. This allows to assess traffic network condition. However, current experience shows that solution of the same problems which are considered in web search engines there is no need in totally Glonassalization. The problem is real quality of clients information, so called smartphones. The aim is to generate additional information about movement conditions. The second problem is the training of tracer elements which the partly considered in the article (general tracks, traced society management etc).

Finally, we consider rapidly growing video observation methods, i.e. intelligence observation. The main idea is that parts of a network are fixed on a video sequence, incoming transport units are captured and observed (or any other activity) until the border of a fixed area. So that generalized correspondence matrix of the fixed area can be assessed. Two possible ways can be introduced: (1) flow image analysis from above or so called panoramic video sequence. (2) IT system which recognizes car registration number and is installed with sufficient frequency within the network. Nevertheless these approaches seem to disturb privacy policy, but what privacy is in a long term traffic jam?

That's why the generalized correspondence matrix is a complex of potential and real correspondence. Every component has significant meaning in management and synchronization of a traffic flow.

II. Monitoring systems

Active growth of car ownership from the 30s of the last century and the evolution of computer technology have dramatically changed the standards of human life, speed up information exchange. Various traffic and network flows are actively evolved. On the other hand real technologies allow to keep huge data in databases. For instance traffic flow monitoring system Yandex jams.

III. Problem description

System for intense traffic flow monitoring and movement matrix through the routing graph nodes (cross-roads) is developed . It consists of data collect server with subsequent

Mikhail Volkov is with the Department of Math Cybernetics and Information Technology, Moscow Technical University of Communication and Informatics (MTUCI), Moscow, Aviamotornaya 8a, 111024 Russia e-mail: mtuci@mtuci.ru

Marina Yashina is with the Department of High Mathematics, Moscow Automobile and Road State Technical University (MADI), Moscow, Leningradskiy prosp. 64, 125319 Russia e-mail: madi@madi.ru

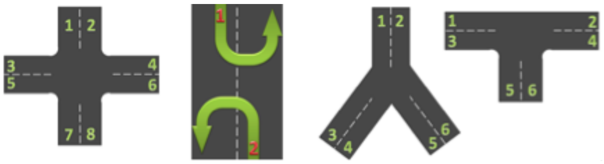


Fig. 1. Types of intersections

corresponding matrix generalization, client application for smartphones (Android OS). Developed and tested on Samsung Galaxy S4 zoom, client application compatible with any smartphones based on Android 4.0.4 or higher (API 16+). Following improvements and features are available:

- Traffic data mining
- Synchronization for experiments
- Routing messages with access matrix
- Data storage
- Data geolocation display
- Selection of a cross-road type and position
- Experiment results storage on a database
- Previous experiments data extraction from DB for processing
- Results processing for generating correspondence matrix

IV. Crossroads types

Area of roads or streets intercrossing (in a plane) is called crossroad. Depending on amount of intercrossing streets (roads) and crossing angle vary following crossroads: a) four sided rectangular (cross shaped), b - sharp-angled quadrilateral (X-shaped); a triangular-rectangular (T-shaped); g - sharp-angled triangular (V-shaped); d - multilateral; e - Square. Square - is a crossroad including large area which is consume road dwelling area (on a picture road dwelling area is white). The boundaries of the intersection considered imaginary lines connecting the corners of the perimeter of the buildings overlooking the intersection. At the crossroads of two constituents defined border lines, mentally drawn from the angles of the buildings perpendicular to the side opposite the side road. The program generates 4 types of intersections: a four-sided rectangular (cross-shaped), a reversal, an acute-angled three-way (U-shaped), three-sided rectangular (T-shaped).

Area - a crossing that is different from the usual considerable size of the occupied territory, which also goes beyond the mind of extended lines of building streets. The picture area that extends beyond the extension of the line construction, - white. The boundaries of the intersection considered imaginary lines connecting the corners of the perimeter of the buildings overlooking the intersection. At the crossroads of two constituents defined border lines, mentally drawn from the angles of the buildings perpendicular to the side opposite the side road. The program features 4 types of intersections: a four-sided rectangular (cross-shaped), a reversal, an acute-angled three-way (U-shaped), three-sided rectangular (T-shaped).

V. Restoration of matrix mixing

Fig. 2. A - matrix of inputs; B - Matrix stirring; C - matrix outputs

$$A = B * C \tag{1}$$

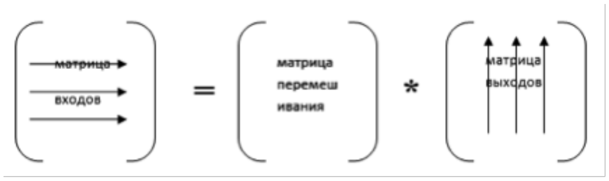


Fig. 2. A - matrix of inputs; B - Matrix stirring; C - matrix outputs

$$B = A * C^{-1} \tag{2}$$

Mixing matrix is the product of a matrix inverse matrix inputs to outputs. The inverse matrix is calculated programmatically using LU decomposition. The matrix equation $AX = I_n$ for the inverse matrix X can be considered as a set of n systems of the form $Ax = b$. Let the i-th column of the matrix X by X_i; Then $AX_i = e_i, i = 1, ...n$, because the i-th column of the matrix is I_n the unit vector e_i. in other words, of finding of an inverse matrix is reduced to solving n equations in one matrix and different right-hand sides. After the LUP-expansion (time $O(n^3)$) to address each of the n equations takes time $O(n^2)$), so that this part of the operation requires time $O(n^3)$).

If the matrix A is non-degenerate, then it is possible to calculate the LUP-decomposition $PA = LU$. Let $PA = B, B^{-1} = D$. Then the properties of the inverse matrix can be written: $D = U^{-1}L^{-1}$. If we multiply this equation by U and L you can get two kinds of equality $UD = L^{-1}$ and $DL = U^{-1}$. The first of these equations is a system of linear equations for n^2 of which are known right-hand sides $\frac{n(n+1)}{2}$ (of the properties of triangular matrices). The second is also a system of linear equations for n^2 of which are known right-hand sides $\frac{n(n-1)}{2}$ (as of the properties of triangular matrices). Together they represent a system of equations n^2. With the help of these equations can be recursively n^2 identify all elements of D. Then the equation $(PA)^{-1} = A^{-1}P^{-1} = B^{-1} = D$ obtain the equality $A^{-1} = DP$.

In the case of LU-decomposition requires permutation of columns D but the solution can disperse even if the matrix A is nonsingular. The complexity of the algorithm - $O(n^3)$).

To work with the matrix program was added to the class Matrix. The program is retrieving data from the database and calculation matrix of mixing.

VI. The architecture of the system hardware

VII. Technologies and Solutions

The server application is written on Java, using the Servlet technology and based on the cloud OpenShift on the web server Apache Tomcat application as it provides a good development speed, portability (Under software portability at the source code refers to the ability to compile the source code and build an executable file of working on more than one hardware or software platform. For example, the Javaprogram with minimal changes can be run under x86 and x64 architecture Windows, OSX and Linux. To serialize data using technology JSON. JSON (Eng. JavaScript Object Notation) - text format data exchange based on JavaScript and are usually used with this language. Like many other text formats, JSON

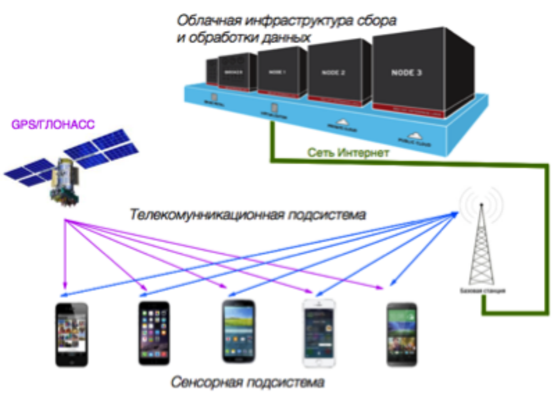
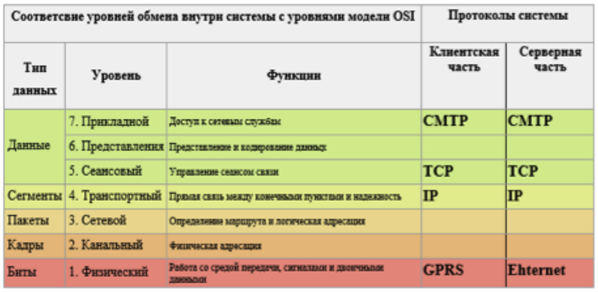


Fig. 3. System hardware



Соответсвие уровней обмена внутри системы с уровнями модели OSI			Протоколы системы	
Тип данных	Уровень	Функции	Клиентская часть	Серверная часть
Данные	7. Прикладной	Доступ к сетевым службам	CMTP	CMTP
	6. Представления	Представление и кодирование данных		
	5. Сеансовый	Управление сеансом связи	TCP	TCP
Сегменты	4. Транспортный	Прямая связь между конечными пунктами и надежность	IP	IP
Пакеты	3. Сетевой	Определение маршрута и логическая адресация		
Кадры	2. Канальный	Физическая адресация		
Биты	1. Физический	Работа со средой передачи, сигналами и двоичными данными	GPRS	Ehternet

Fig. 4. Communication protocols and the OSI Model

is easily read by people. JSON format was developed by Douglas Crockford. To work with the GPS use GPS Android standard driver and a library for NetBrowser interaction with GPS/ GNSS receiver JAVAD Gismore. The interaction with the server implemented in sotvetsvii methodology REST-api. For client-server interaction protocol was developed CMTP (Crossroads Monitoring Transfer Protocol), based on the protocol TCP / IP (Transport Control Protocol over Internet Protocol). The description of this protocol is on. Using this protocol allows different clients, realizing minimal functionality CMTP interact with the server, regardless of hardware and software platforms.

The interaction with the server implemented in sotvetsvii methodology REST-api.

For client-server interaction protocol was developed CMTP (Crossroads Monitoring Transfer Protocol), based on the protocol TCP / IP (Transport Control Protocol over Internet Protocol). The description of this protocol is on. Using this protocol allows different clients, realizing minimal functionality CMTP interact with the server, regardless of hardware and software platforms.

VIII. Conclusion

Algorithm of recover the Origin-Destination Matrix on real-time data is presented. We propose modern methods of data collecting for using of traffic flows modeling on complex city network, for example at a crossroads. Our system can collect data for math models of traffic processes.

Acknowledgment

The authors would like to thank prof. Buslaev A.P. for discussion and problem statement.

References

[1] Bugaev A.S., Buslaev A.P., Kozlov V.V., Yashina M.V. Distributed Problems of Monitoring and Modern Approaches to Traffic Modeling, p. 6, 14th International IEEE Conference on Intelligent Transportation Systems (ITSC 2011), Wash-ington, USA, 5-7.10.2011. DOI: 10.1109/ITSC.20116082805. (2011) 477 - 481.

[2] Buslaev A.P., Volkov M.M., Provorov A.V., Yashina M.V. SSSR system in Nonde-structive Measurement. Ninth Interna-tional Conference on Traffic and Granu-lar Flow 2011. Book of abstracts. , M. T -Comm, 2011, 341-342 p.

[3] Provorov A.V., Buslaev A.P., Yashina M.V., Volkov M.M., SSSR - system in Nondestructive Measurement, , T-Comm - Telecommunications and Transport, Media Publisher, Moscow, 2011

[4] Buslaev A.P., Volkov M.M., Optimization and control of transport processes in the distributed systems, DepCos 2014, Brunow, Poland, DOI 10.1007978-3-319-07013-112 pp.123-132

[5] Alexander P. Buslaev, Marina V. Yashina, Mikhail Volkov, Algorithmic and Software Aspects of Information System Implementation for Road Maintenance Management, Proceedings of the Tenth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX, June 29 July 3 2015, Brunow, Poland, DOI 10.1007978-3-319-19216-17, pp 65-74

[6] Buslaev A.P.,Provorov A.V., Yashina M.V. Infocommunication systems of satu-rated traffic control in megapolis Proceed. of ICOMP, 2013 , Las-Vegas, USA, 2013

Computer network traffic models: research, hypotheses, results

Alexander Buslaev, Alexander Zernov, Pavel Sokolov, Marina Yashina,

***Abstract*—Dynamical system on regular network of type Chainmail was introduced by Buslaev A.P. et al. In the paper we consider two approaches to the study of flows on networks. The first approach is a classical model of cellular automata. The novelty consists of a gradual increase in number of parameters of the problem and exploring the links between the regular networks we consider. The second approach is based on a simplified model of continual traffic, i.e. cluster model that has developed from the follow-the-leader model and hydrodynamic models. Regular composite networks of Chainmail type were studied**

Some results of estimation of the system behavior were obtained.

***Index Terms*—Regular networks, Traffic flow, Cluster model, Chainmail.**

I. Introduction

A problem of traffic control is a hot issue in the Russian Federation due to the rapid growth of motorization, because now there are much more cars than roads. The speed of construction of the road network in Russia significantly lags behind the pace of increase in the number of cars. In spite of some progress made in addressing the problems of traffic in developed countries it appears that there is still no satisfactory theoretical base for traffic modelling and optimization, especially in big cities with complex road network.

It is explicitly confirmed by a difficult traffic situation on the streets of many large cities in the world, as well as confirmed indirectly by attempts to solve typical traffic optimization problems within the framework of international conferences, eg, Traffic and Granular Flow Conferences [1]. One of key challenges is accounting for a substantial human component in traffic flows modelling, as the traffic is a complex socio-technical system.

II. Information revolution

Despite of the complexity of traffic modelling, rapid development of information technology and increase in the capacity of computing equipment provide certain reasons for optimism in addressing traffic modelling and optimization problems. The purpose of information technology and computing hardware consists of a gradual and reliable replacement of direct human control and intervention in traffic flows and its regulation. Substantial progress is particularly notable in development of intelligent systems for a single vehicle and its response to the changes in road situation and accidents prevention on a local road section. However, there is substantially less progress in addressing the task of optimization of traffic flow optimization on larger segments of road network. Still, there is no reliable solution to the problem of optimization of a tactics of an individual participant in order to contribute to traffic flow optimization in a whole. The significant progress in technologies of data gathering and traffic monitoring has been recently achieved, but it only emphasizes the lack of a common theoretical approach to its processing and interpretation.

III. The agent-based model as an advanced ”coin toss”

Agent-based models that are based on the synthesis of cellular automata theory, theory of graphs and networks, and computer equipment, only provide an illusion of a theoretical solution of the traffic problem. In fact, the agent-based approach is well-formulated method of a ”coin toss” for a respective problem. As we know from the history of the theory of probability, an experiment with a coin resulted in the discovery of Bernoulli distribution, which was followed by the central limit theorem. Unfortunately,there was no progress in methods, applicable for reliable traffic modelling yet. But this is natural: a coin has only two faces, compared with the wide array of parameters of a road network. These two particular properties of the traffic problems: the complexity of modelling of the socio-technical system of traffic and the experience of successful solution of some specific aspects of traffic modelling, create a favorable environment for maintaining interest in its solution.

IV. Expansion of the parameters list beyond ”speed, density, intensity”

Below we consider two approaches to the study of flows on networks. The first approach is a classical model of cellular automata. The novelty consists of a gradual increase in number of parameters of the problem and exploring the links between the regular networks we consider. The second approach is based on a simplified model of continual traffic - cluster model that has developed from the follow-the-leader model and hydrodynamic models. The point is that the objects of interaction are not the particular areas of traffic, but their stable components - packs, clusters, which move synchronously and interact with other clusters according to certain rules.

A. Buslaev is with the Moscow Automobile and Road STU, Moscow, Russia, 125319, Leningradsky prosp., 64, E-mail: apal2006@yandex.ru

A. Zernov is with the Moscow Automobile and Road STU, Moscow, Russia, 125319, Leningradsky prosp., 64, E-mail: alekzernov@mail.ru

P. Sokolov is with the Moscow Automobile and Road STU, Moscow, Russia, 125319, Leningradsky prosp., 64, E-mail: user7824@gmail.com

M. Yashina is with the Moscow TU of Communications and Informatics , Moscow, Russia, 111024, Aviamotornaja Street, 8a, E-mail: yashmarina@yandex.ru

V. Model of total connected incompressible clusters on networks

A. Problem formulation

We investigate the cluster model of the flow at the rings. We describe the scheme of cluster movement at rings. As a result of the simulation we obtained velocity characteristic of the system. We created software that simulates cluster movement at the ring, and calculates the characteristics of the system. In order to obtain limit values of velocity, a calculating experiment has to be run.

B. Definition of a cluster

A cluster is a steady movement of two or more particles at equal distances from each other as a limit of a follow-the-leader model. In this model, a cluster is shown as a geometrical figure, visualizing a flow of particles moving in the same direction with identical speed. The square of the figure determines the number of particles in the cluster or the cluster mass M. The mass of the cluster remains constant. The speed V of the cluster is a monotonic function, which depends on its density y [2]. For example function $u(y)$ has is the following ,

$$u(y) = u_0 \frac{y_{max} - y}{y_{max}}$$

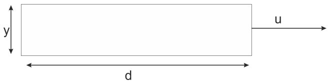


Fig. 1. Cluster

C. Movement support description

A ring is a closed strip with length ”L”. Clusters are moving at the rings.

Network is a number of rings with intersection points.

D. Movement Rules

If a moving cluster is at a intersection point while another cluster on the neighboring ring is about to reach the same intersection point, it stops until a confliction cluster leaves the intersection point (on first in-first out basis).

E. The results of a simulation study for the regular network

We have a regular network of $M \times N(M, N-$even numbers), made up of equal, intersecting rings. Each ring has four points of intersection with the adjacent rings. The clusters with the equal length move in the same direction around the rings[3], [4].

The direction of movement is defined in this way: each ring has two indices, which are numbers of the row and the column. If the sum of the indices is even, the cluster moves within this ring counterclockwise, otherwise the cluster moves clockwise.

As a result of mathematical modeling we get the following data:

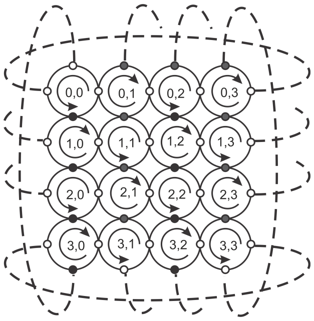


Fig. 2. Proper network co-directional traffic

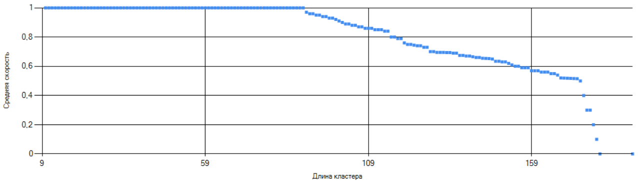


Fig. 3. Plot velosity of cluster length

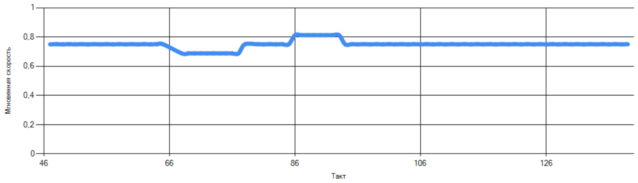


Fig. 4. Plot velosity when a partially dynamic traffic $L = 90$

Step 1: If the cluster $L = 180$, we have a full congestion. The system velocity decreases monotonically with the increase of the length of a cluster.

Step 2: If a partial dynamical jam is formed, the instantaneous speed is fluctuating. If the jam is not dynamic, the instantaneous speed becomes constant, but does not reach the maximum value.

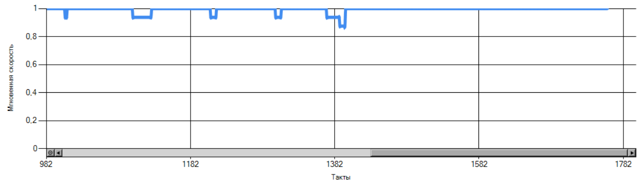


Fig. 5. Cluster velocity plot at a $L = 89$

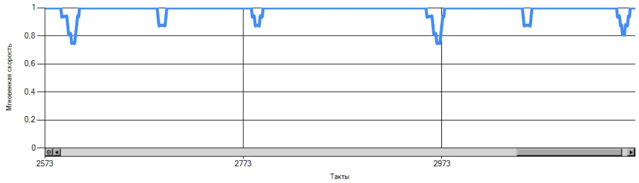


Fig. 6. Plot fluctuation at the $L = 90$

Step 3: The system has the synergy state if cluster $L < 90$, and swings at cluster L of 90.

F. Software Description

The Software consists of functions, user interface, a set of parameters.

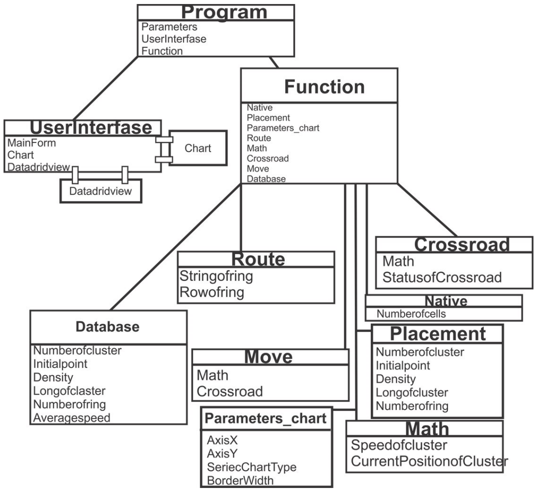
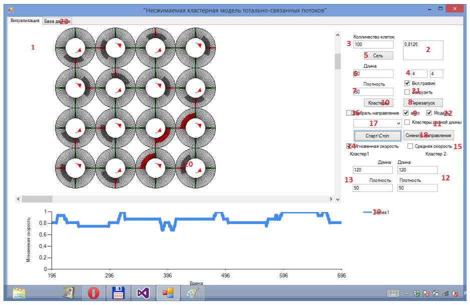


Fig. 7. Scheme of software

This figure shows the functions of the user interface and a set of parameters which are included in the software.

1) Interface software:


"Несжимаемая кластерная модель тотально-связанных потоков"

Количество клеток: 100

0,8125

Длина: 50

4 4

Плотность: 50

☑ Вкл.график

☐ Загрузить

☐ Выбрать направление

☑ авто

☑ Модель

☐ Кластеры разной длины

☑ Мгновенная скорость

☐ Средняя скорость

Кластер1 — Длина: 120, Плотность: 50

Кластер 2 — Длина: 120, Плотность: 50

Fig. 8. Interface software

Area 1 shows cluster movement.

Field 2 shows system velocity.

Field 3 shows the number of particles at every support.

Fields 4 shows the size of the network (the number of rows and columns).

Button 5 is designed to build a network according to parameters defined above.

Field 6 shows the length of the cluster.

Fields 7 shows cluster density.

Button 8 restarts the program.

Button 9 enables or disables automatic cluster placement according to the given parameters.

Button 10 is show for cluster placement (if the automatic cluster placement is disabled, the software will place a cluster at each ring, the ring at which the current cluster will be placed is chosen randomly).

Button 11 enables or disables ”male/female” rule. If this function is enabled, the cluster settings on the secondary diagonal are to be set in the fields 12 and the fields of clusters on the main diagonal are to be set in the fields 13.

Button 14 enables / disables the calculation and chart of instantaneous velocity of the entire system.

Button 15 enables / disables the calculation and chart of the average velocity of the entire system.

Button 16 enables / disables the change of direction function.

Field 17 sets the movement direction.

Button 18 changes the current direction.

Button 19 starts or stops the movement of the clusters.

Field 20 is intended for displaying charts.

Button 21 loads parameters from the database.

Button 22 switches on and off models

Tab 23 contains database.

VI. Nagel - Schrekkenberk cellular automata on morphing of Manhattan network and the type of chainmail

A. Finite-parametric network

We consider a network with a finite set of parameters. Let us call the plain networks with a periodic structure as **regular networks**.

For a given rectangular area we generate a cell decomposition. Cell number $M \times N$ and the size $a \times b = \Delta x \times \Delta y$.

Rectangular area is stretched on a torus, i.e. we assume the left and right vertical borders equal, as well as upper and lower limits, and thus, we get a Manhattan network Fig.9 on the torus [2].

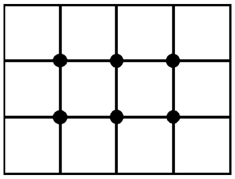


Fig. 9. Manhattan network on the torus

In Fig. 9 we replace each node by a circle of radius r on the network.

If the radius of the circles reaches the maximum value, $r = \frac{a}{2}$, the length of the linear segments is equal to 0. In this case, the network is a **chainmail**. Fig.10 shows a network corresponding to these parameters. Such a network was investigated in [3].

We note, that the network nodes are the points of tangency.

We assume that movement on each traffic channel is possible only in one direction. We set a movement direction on the created network.

in one direction only a linear segment length is comparable to the radius, so that we obtain network type Fig.11.

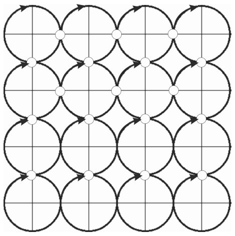


Fig. 10. Chainmail on the torus

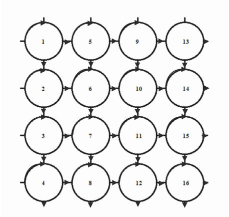


Fig. 11. Right two-network on the torus, with the direction of movement

B. *Measurement of linear size of the network*

We create a uniform partition by cells of the created two-component network, Fig.12.

Let k be the number of cells in a circle partition, and let l be the number of cells on the segments. The number of cells in the network is equal to K.

$$K = k \cdot (M \cdot N) + 2 \cdot l \cdot M \cdot N.$$

Note that

$$1 \leq l \leq \frac{k}{\pi}.$$

If $l = 1$ we obtain a network of chainmail type, [3].

We assume Fig.12 $k = 24$, $l = 6$ and size network

$$K = 16K + 32L = 576.$$

C. *Generation of the initial data*

At every moment in the same cell there is no more that one particle. The initial arrangement of particles is created randomly.

We fix network density ρ.

a) For each particle a randomly selected item number is chosen (a ring or a segment). Then a position of a particle on the element is randomly assigned.

b) We select a periodic unit. We place particles on this unit uniformly. The placement of particles is periodically extended to the whole network.

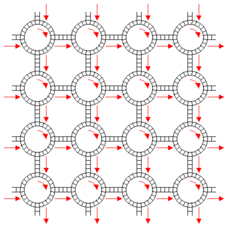


Fig. 12. Right two-network on the torus, with the direction of movement

D. *Rules of particle movement*

1) Moving forward: Moving a particle one cell forward is carried with the probability $p = 1$ provided that the corresponding neighbor cell is vacant. This motion is called **individual movement**.

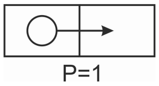


Fig. 13. Particle moving to one cell forward is carried with the probability p

2) Selection of movement: If there is a choice of trajectory of further movement of a particle, is a particle is on the intersection point of a ring and a liner segment, Fig.14. Movement of a particle is determined by the parameter α. Then α is the probability of the particle remains on the ring, respectively, $(1 - \alpha)$ is the probability of the particle lefts the ring. If $\alpha = \frac{1}{2}$, a random movement of particles occurs over the network.

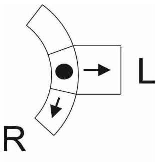


Fig. 14. Motion of a particle at the junction of the ring and the segment, with the probability α

3) Competition: For the network in Fig.12 and a set field of directions **conflicting points** are cells in which the particles switch from legs to rings. At the same time a particle may move by the ring to the same point.

Competition rules is resolved with the probability $\beta = \frac{1}{2}$, which is called equal movement. An example of competition resolution is shown in Fig.15.

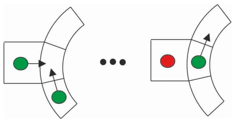


Fig. 15. Example permits competition with the probability $\beta = \frac{1}{2}$

E. *Particle state. Software implementation*

The state of the particles is indicated by colors in the program.

Green color indicates that the next cell is vacant and thus the particle will not stop.

Red color indicates that the particle is not moving, waiting for the next cell to get vacant. This situation can occur on the ring or on the linear segment or in case of a particle movement to a conflict cell and its entering into competition with another particle.

Yellow color indicates that the particle has lost competition for the vacant cell and is not moving.

F. Instant and average velocity

The instantaneous and average velocity of the particles are calculated by the following formulas:

$$v_m = \frac{1}{C}\sum_{i=1}^{C} v_i, \text{ where C is the particles number}$$

$$v_{cp} = \frac{1}{N}\sum_{i=1}^{N} v_{m_i}, \text{ where N is the measurements number}$$

G. Morphing of network by L, $K = const$

Morphing is a seamless transition from one network to another, provided that the number of cells in both networks is identical. An example of morphing is shown in Fig.16.

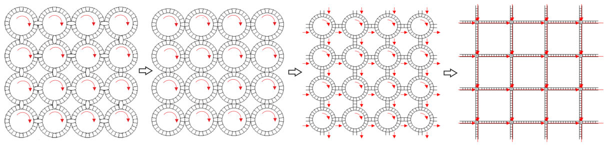


Fig. 16. Morphing

H. Geometry

H.A. *Chanmail parameters $k = 28$(there is a common cell), $L = 1$, Fig.17.*

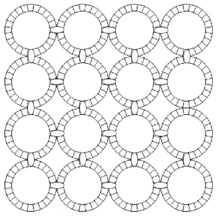


Fig. 17. Chainmail on the torus

H.B. *Correct network parameters $k = 28$, $L = 2$, Fig.18.*

H.C. *Right two-Network parameters $k = 24$, $L = 4$, Fig.19.*

H.D. *Manhattan network parameters $k = 4$, $L = 14$, Fig.20.*

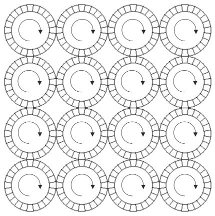


Fig. 18. Correct network on the torus

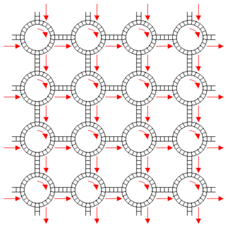


Fig. 19. Right two-Network on the torus

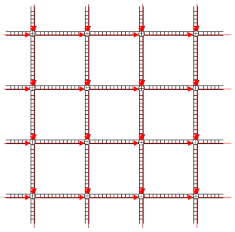


Fig. 20. Manhattan network on the torus

I. The results of a simulation study

Particles are placed randomly over all cells, while there can't be more than one particle in a cell.

The motion of each particle is individual. The movement of a particle to a vacant cell occurs with the probability $p = 1$.

The control of particles movement from the ring to a linear segment is carried out by the parameter. The parameter α varies.

Resolution of particles competition at intersection points occurs with the probability of $\frac{1}{2}$.

By changing the parameters k (the number of cells on the rings), L (distance between the rings in the cells), we simulate a number of different networks with the same number of cells equal to 448. We get morphing Fig.16 from multiple networks. The number of rings in all networks is identical and equals to 16.

Comparing the obtained graphs of the dependence of average velocity on the density at different values α we get the following results shown in the figures.

In Fig.21 it is shown network of Chainmail type. The following results are obtained.

1. *In case of a global movement of all particles for the networks types C and D, the flow behavior is characterized by the movement of the linear portion of the closed circuit [3].*

2. *At global movement of all particles for the network Mail results converged with the results of the article [4].*

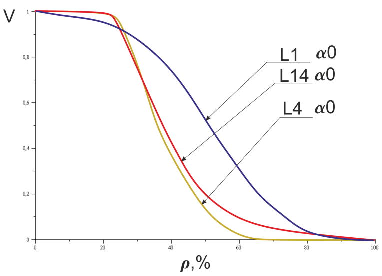


Fig. 21. Networking **A, C, D** $\alpha = 0$

3. *In the global motion of the particles for networks A, B and C formed by the collapse of 90% flux.*

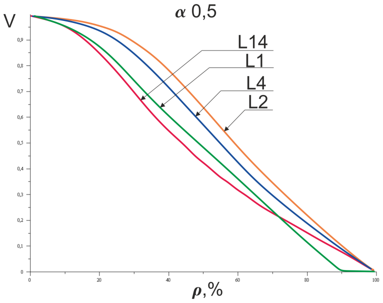


Fig. 22. Networking **A, B, C, D** $\alpha = 0, 5$

In Fig.22 we can see the following result.

4. *When the random walk of the particles for networks A, B, C and D on the flow behavior of the network is similar to the behavior of the flow on a multilane road.*

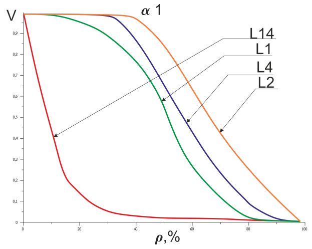


Fig. 23. Networking **A, B, C, D** $\alpha = 1$

In Fig.23 we can see the following result.

5. *At local motion of the particles, for networks B, C and D collapse occurs faster with the decrease in the length of the ring.*

For a network chainmail graph comparing the density of the number of competitions, for different values α we get the following results Fig.24.

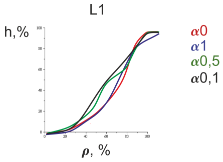


Fig. 24. h - the number of competitions

In Fig.24 we can see the following result.

6. *Network chainmail regardless of local or global motion does not change the amount of competition.*

VII. Conclusion and Future works

We have introduced and studied composite networks of Chainmail type. Results of system behavior were obtained.

Future works will be connected with the following.

A. Model totally related incompressible cluster in networks

1) Quasi regular network: *In this network rings which sum of the indices is even have a radius of 2 tines greater than the rings which sum of the indices is odd. This if $i + j = 2k$, then $R_{i,j} = 2r$.*

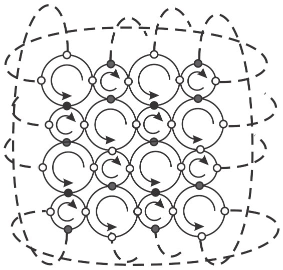


Fig. 25. Chainmail on the torus

To spend the same study as well as for the correct network.

B. Generation networks. Motion study

Further development of the program will be in the bellow direction:

1. *Further study of traffic on two networks, by changing the parameters of a new configuration, yielding a two-network.*

2. *Generating wrong networks. Studying the behavior of traffic on the wrong network.*

References

[1] *V. V. Kozlov, A. P. Buslaev, A. S. Bugaev, M. V. Yashina, A. Schadschneider, M. Schreckenberg. Traffic and Granular Flow '11. Eds.: Proc. of Int. Conf. TGF'11, Mocow, Russia, Springer, 2013.*

[2] *A.P. Buslaev, A.G. Tatashev, M.V. Yashina. Cluster flow models and properties of appropriate dynamic system // Journal of Applied Functional Analysis, vol. 8, no. 1, pp. 54-76 (2013)*

[3] *V.V. Kozlov, A.P. Buslaev, A.G. Tatashev. Monotonic walks on a necklace and coloured dynamic vector // International Journal of Computer Mathematics (2014). DOI: 1080 00207150.2014/915964*

[4] *V.V. Kozlov, A.P. Buslaev, A.G. Tatashev. Behavoir of pendulums on a regular polygon // Journal of Communication and Computer, vol. 11, 2014, pp. 30-38.*

[5] *V.V. Kozlov, A.P. Buslaev, A.G. Tatashev, M.V. Yashina. Monotonic walks of particles on a chainmail and coloured matrices // Proceedings of the 14th International Conference on Computational and Mathematical Methods in Science and Engineering, CMSSE 2014, Cadiz Spain, June 3-7 2014, vol. 3, pp. 801-805.*

[6] *V.V. Kozlov, A.P. Buslaev, A.G. Tatashev. A dynamical communication system on a network // Journal of Computational and Applied Mathematics, vol. 275 (2015), pp. 247-261.*

The use of MPI technology for solving problems of agent-based modeling of traffic flows

Vitalii V. Shiriaev, Grigory M. Chernyak

Abstract—**Agent-based traffic microsimulation is inherently a resource-intensive task, but at the same time it is well suited for parallelization. In this work we show an application of MPI parallelization techniques to an existing agent-based traffic model written in Java. We apply geometric decomposition of graph-based calculation domain combined with the ghost cell pattern and then study the efficiency of this approach.**

Index Terms—**Geometric decomposition, MPI, Parallel computing, Traffic microsimulation**

I. Introduction

ONE of the common approaches to the traffic flow modeling problem is microsimulation. A model for microsimulation is comprised of a large set of individual entities and a number of rules for their interaction. These entities are also called agents. In traffic flow microsimulation each vehicle is represented by a separate agent. The state of a vehicle at any moment depends on its previous state and the state of its surrounding agents: other vehicles, traffic lights, etc.

Microsimulation often requires a lot of computational resources due to the large number of agents being simulated. It is often desirable to minimize computation time, so one may use parallel computing techniques for that purpose. One of them is the Message Passing Interface (MPI) - the technique that is used in distributed memory environments. When using MPI all the agents are divided into several groups and groups are assigned to separate processes. This grouping is called a decomposition of calculation domain. Each process performs calculations with its group of agents and has no direct access to the data of the other processes. However, agents are not completely independent, so each group usually contains agents that depend not only on the agents from this group, but on the ones from the other groups as well. So MPI processes have to exchange some data on the border of their domain parts and the amount of the data exchanges significantly affects overall performance. Hence, domain decomposition method should be chosen as to minimize these data dependencies.

Here we describe MPI parallelization of the traffic micro model, introduced in [1].

II. Original model

The traffic model in hand is a car-following model that works in terms of vehicle positions and velocities on a quasi-one-dimensional road network. The road network is represented by a directed graph with its edges standing for one-way roads with one or more lanes and its vertices standing for crossroads. Two-way roads are treated as pairs of one-way roads. Each vehicle is assigned a position on one of the edges of the graph, a lane number and a speed. During simulation vehicle acceleration and lane changing are determined with a composite driver algorithm that aims to model the behavior of a real driver. The algorithm makes decisions based on the perceptions that would be available to a real driver: positions and velocities of the vehicles in the range of vision of the driver. A vehicle is considered as being in the range of vision of another vehicle if the distance between them measured along the graph is within some predefined constant.

In this way each vehicle-agent depends on a number of vehicle-agents on the same edge of the graph and possibly on the adjoining edges. This model peculiarity suggests a choice of geometric domain decomposition for MPI calculations.

III. Domain decomposition

The main goal of domain decomposition is to split agents into uniform groups in terms of calculation time. In other words, simulation of each group of agents should take approximately the same time, so that the MPI processes would not wait for each other. In our case all the vehicle-agents are processed with the same algorithm, whereas the processing time of non-vehicle-agents can be neglected, which makes calculation time proportional to the number of vehicles in the group. So it is desirable to split the calculation domain into parts that have equal amount of vehicles.

On the other hand it is important to minimize data exchanges between the processes. The model allows us to make a geometric decomposition, that is to assign all the agents that currently are located on a number of adjoining graph edges to a single group. In this way the observation range for most vehicle-agents will be within the edges that are in the group and data dependencies will be kept low. In order to further reduce the dependencies it is desirable to minimize the amount of border edges between the domain parts.

Therefore a geometric decomposition should aim for the minimization of border edges as well as for keeping the number of vehicles in each part approximately the same. However, vehicles move from one edge to another during the simulation. One option is to perform a dynamic decomposition that evolves during simulation to adapt to changes in vehicle distribution. Another option is to do a static decomposition using some kind of prediction of the vehicle distribution. Below we describe a static decomposition with the most rough estimation of vehicles being evenly distributed along all the edges of the graph. In this approximation the requirement for

Vitalii V. Shiriaev and Grigory M. Chernyak are with the Systems Simulation Laboratory, Moscow, Russia.

This work was supported by the RFBR (grant 13-08-12241)

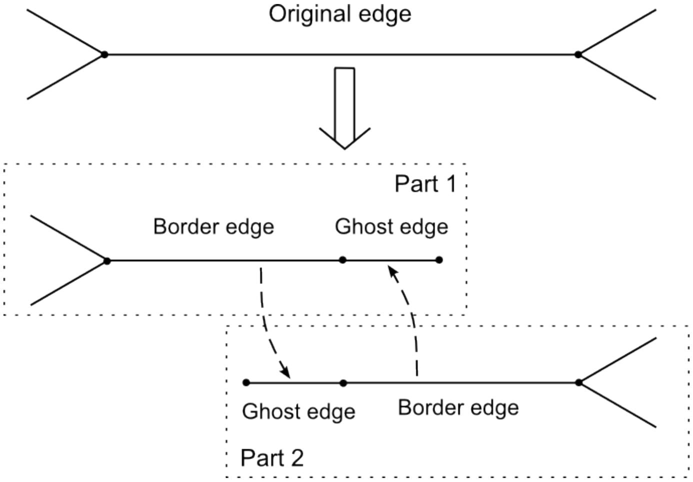


Fig. 1. Edge splitting

equal vehicle numbers amounts to the requirement for equal total edge lengths.

In this way the domain decomposition task is reduced to a static graph partitioning task. We preferred to split the graph in the middle of its edges. The length of these border edges should be at least twice the observation distance of a driver. This allows the observation range of the vehicles at the end of each part to lie completely within the split edge and not to go over to adjacent edges thus reducing the complexity of data exchange.

So each vertex of the graph is assigned a weight equal to half of the sum of the lengths of all edges connected to this vertex. Then the vertices are filtered: if two vertices are connected with an edge that is shorter than twice the observation distance of a driver then these two vertices are merged into one and all edges between them are removed. Then the graph is partitioned into parts with equal weights while keeping the number of edge cuts to the minimum. The graph partitioning task was performed with the METIS software [2].

IV. Border exchanges

Having partitioned the graph into parts we need to establish communication between the MPI processes so that every process can obtain the dependency data for the agents on the border edges of its part. A vehicle-agent on the inner edges of the domain part needs no external dependencies. A vehicle-agent on the border edge that is close enough to the edge cut needs parameters of the vehicles that are located within its observation distance on the other half of the border edge in the adjoining part.

Consider a single border edge. Observation area of every vehicle on this edge goes no farther into the other half of the edge than the observation area of a vehicle that would be located precisely at the position of the edge cut. So if one gets data on all the vehicles within an observation distance from the other side of the edge cut, the data dependency of any vehicle-agent on a border edge would form its subset.

In this way one may transfer a single data set and determine all the dependencies as its subsets instead of transferring these subsets individually. This approach is analogous to the ghost cell pattern [3] that is commonly used in structured grid computations with MPI. The main idea is to collect and store all the dependency data as a bulk to reduce data exchanges.

We attach a special ghost edge to the cut of each of the border edges. A ghost edge contains a number of ghost vehicles that serve only as a source of the dependency data.

Apart from holding the dependency data, ghost edges facilitate vehicle transition. When a vehicle goes out of the domain part through a border edge it naturally enters the ghost edge connected to the border edge. Then this vehicle-agent is moved from its domain part onto the adjacent one, and from the responsibility area of one process to that of another through the data exchange. Ghost edges simplify tracking of these transitioning vehicles.

In this way the data exchanges on the border edges are done in an unobtrusive manner: the main calculation algorithm works without any knowledge on the distributed memory environment, it simply treats the ghost edges and ghost agents as real ones.

The overall algorithm works as follows. First, each process collects the exchange data for its adjacent processes. The exchange data consist of the data on vehicles on border edges which will serve as ghost vehicles for adjacent processes and the data on transitioning vehicles that moved from border edges onto ghost edges. Then it removes all transitioning vehicles and all ghost vehicles and asynchronously sends the collected data to the adjacent processes. At the same time it receives the data collected by adjacent processes, fills ghost edges with ghost vehicles and adds vehicles that transitioned onto its domain part. In effect these data exchanges synchronize all processes. Finally, each process independently performs a step of the main calculation on its domain part.

V. Results

The algorithm for parallelization was implemented using mpiJava library that is an MPI binding for Java.

Tests of the described MPI parallelization were performed on a set of synthetic road networks. All road networks were generated as regular grid networks of different size. Several tests were conducted with different number of processes on different sizes of grid networks.

In these tests we measured the acceleration achieved with the parallelization, that is the ratio of total simulation time without parallelization to the total simulation time with parallelization.

For all test networks the achieved acceleration grows less than linearly with the number of processes. This can be attributed to load imbalance between processes. The larger the number of processors, the more it will be affected by the same load imbalance since more processing time will be lost during synchronization when all processes must wait for an overloaded process to complete its simulation step. Also large numbers of processes result in large amounts of border edges, increasing data exchange time.

It should be noted that the observed load imbalance is not caused by partitioning non-uniformity, but rather by non-uniformity of the traffic distribution. It was found that the

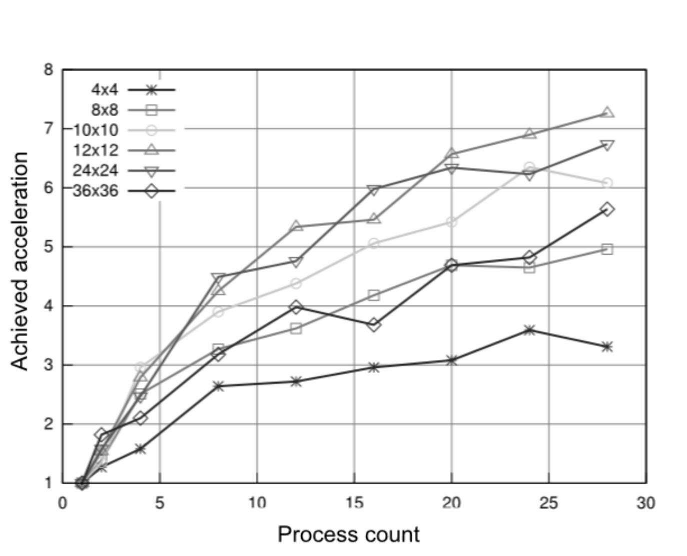


Fig. 2. Acceleration achieved for different road networks

algorithm with METIS partitioning gives result of 2% relative standard deviation of part size if the network has at least 20-30 edges per process. However the observed load imbalance was of 20-100% relative standard deviation. Moreover, the overloaded parts were not the geometrically largest ones. Hence the load imbalance is caused by a non-uniformity of vehicle distribution between the parts of the graph.

Also, we note that for a fixed number of processes, 24 in our test, the best acceleration is achieved on medium-sized road networks. We suppose that small-sized networks suffer from a large ratio of communication time to actual calculation time. Large networks on the other hand are thought to exhibit more load imbalance than the middle-sized ones because they are more affected by traffic distribution inhomogenities.

Also, we observed that synchronization and data exchange times were of the same order as the main calculation time in these tests.

VI. Conclusion

Traffic microsimulation tasks are well suited for parallelization in distributed memory environments. However, in order to achieve good parallelization efficiency it is important to balance the load among the processes to reduce synchronization times.

We plan to adopt a dynamic domain space partitioning approach to our model and to study how dynamic partitioning performed during simulation can benefit the load balance.

Apart from load balancing, we consider reducing synchronization costs by adopting the deep halo pattern [3] and by building additional levels of ghost edges, which will allow to use less frequent synchronization.

Also, we plan to minimize data exchange costs by overlapping communication with computation since the state of all the agents on the inner edges of the graph part can be computed while the communication for border edges is running.

References

[1] Panasyuk Ya., G. M. Chernyak, Dudinov I. K., Helvas A. V. *Agent-based modeling of microscopic traffic flow in COS.SIM* - Proceedings of the ninth international conference “Traffic flow and Granular Protection 2011” (Traffic and Granular Flows), 2011

[2] George Karypis, Vipin Kumar *A Fast and Highly Quality Multilevel Scheme for Partitioning Irregular Graphs* - SIAM Journal on Scientific Computing, Vol. 20, No. 1, pp. 359392, 1999

[3] Fredrik Berg Kjols, Marc Snir *Ghost Cell Pattern*, 2010

On modelling of traffic on multilane intersection

Andrew Yaroshenko, *Fellow, MADI,*Dmitriy Lopanov, *Postgraduate student, MADI,*

***Abstract*—In paper, we consider investigation of model of traffic flow on unregulated multi-lane intersection. Simulation model was developed. Characteristics of traffic and numerical estimates of parameters of stationary states were studied.**

***Index Terms*—simulation models, intersections, heterogeneous flow, numerical characteristics, deterministic-stochastic approach, multi-lane flow.**

I. Introduction

An actual task for the research of traffic in cities is the problem of estimating the velocity on the roads and intersections in the saturated flow mode [1]. Under these conditions, the velocity of vehicles is about 25-30 km/h (or 7.3 m/sec) [2], [3]. We research the estimates of the average velocity of the heterogeneous flow on the unregulated multi-lane intersection. The research is conducted through the simulation model, built on the basis of deterministic-stochastic approach [4], in which the vehicle's velocity is determined by the sum: $v_{det.} + v_{st.}$. The flow rate is set to zero, because in saturated flows at intersections motion is determined by individual movement of each vehicle. In this formulation, linear motion of vehicle is modeled similar to problem that is described in [5] with parameter $v_{max} \equiv 1$. To construct a simulation model the plane of the road is divided into cells such that each band is represented by a sequence of cells of equal size, [6]. Based on the traffic conditions cell's length is assumed to be the length of the cells of 4 meters - the average length of the car, [7]. Therefore tact simulation is about 0.5 seconds; such value allows to simulate the states in which the vehicle occupies one cell of intersection.

II. Traffic modeling

We describe a general model of the traffic flows on an unregulated multi-lane intersection. Since the traffic flow is *heterogeneous*, it means that there are different types of vehicles (cars, trucks and etc.) and, consequently, they have a different velocity. Let the vehicles be divided into two types: fast and slow. If movement is possible, then fast vehicle moves with probability 1. Slow vehicle moves with probability γ (if movement is possible).

We define the following rules of the movement and lane-changing:

1. Behavior of vehicle outside the intersection:

The vehicle1 moves forward on the road, if there isn't other vehicle2. If there is the vehicle2 ahead, then the vehicle1 will try to change the lane to adjacent one. The vehicle1 can change lane only when it won't be a hindrance to other vehicle on adjacent left lane. If the vehicle can't move forward or can't change the lane, then it waits, i.e., it is *idle*.

2. Behavior of vehicle at the intersection and in front of the intersection:

There is a basic rule doesn't prohibits any lane changing on the intersection.

The rules for Forward movement are the same as described above. In controversial situations the vehicles that are moving along a secondary road will pass vehicles from the main road.

We illustrate forward movement and lane-changing for vehicles on each road:

If there is a vehicle in some place of the road, then we paint the cell.

1. Forward movement on the main road:

1.1. Outside the intersection (Fig. 1):

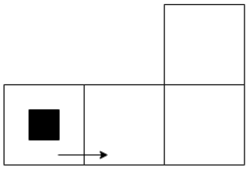


Fig. 1. Forward movement on the main road outside the intersection.

1.2. On the intersection or in front of the intersection (Fig. 2):

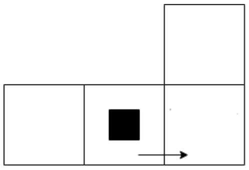


Fig. 2. Forward movement on the main road in front of the intersection (or on the intersection).

2. Lane changing on the main road (Fig. 3):

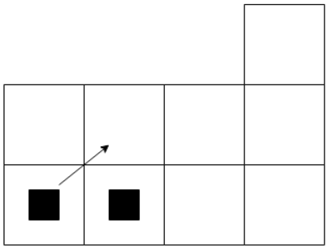


Fig. 3. Lane changing on the main road

3. Forward movement on the secondary road:

3.1. Outside the intersection (Fig. 4):

A. Yaroshenko is with the Department of High Mathematics, Moscow Automobile and Road State Technical University (MADI), Moscow, Leningradskiy prosp. 64, 125319 Russia e-mail: http://en.madi.ru

D. Lopanov is with the Department of High Mathematics, Moscow Automobile and Road State Technical University (MADI), Moscow, Leningradskiy prosp. 64, 125319 Russia e-mail: http://en.madi.ru

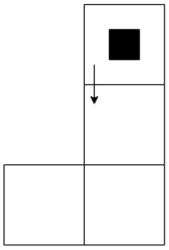


Fig. 4. Forward movement on the secondary road outside the intersection.

3.2. On the intersection or in front of the intersection (Fig. 5):

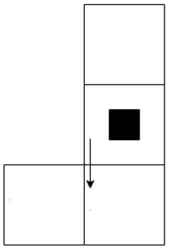


Fig. 5. Forward movement on the secondary road in front of the intersection (or on the intersection).

4. Lane changing on the secondary road (Fig. 6):

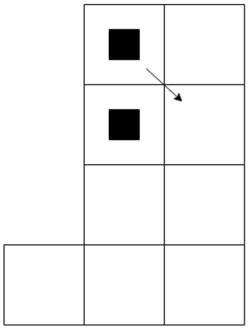


Fig. 6. Lane changing on the secondary road.

III. NUMERICAL CHARACTERISTICS

The two main characteristics, which we are studying, are the velocity and density of the vehicles.

The calculation of the *velocity* in the each subsequent time step t is as follows:

$$v_{t+1} = v_t + \frac{v_t - v_0}{t} \tag{1}$$

The *density* in each time step t is calculated according to the formula:

$$\rho = \frac{m}{l}, \tag{2}$$

where m is amount of particles in each time step, and l - is the length of the road.

IV. SOFTWARE IMPLEMENTATION OF THE MODEL

The research problem of the velocity and density of the unregulated multi-lane intersection also raises the problem of software implementation:

- The cumulation of data of the velocities of the moving vehicles.
- The cumulation of data of the density on the road.
- Display the data about velocity and density on the chart.
- Automatic the creation of a statistical database of the conducted experiments.

In the simulation we should take into consideration the following parameters:

1) Incoming flow q_i on each lane.
2) the proportion of the slow vehicles: α.
3) The number of lanes on each road.
4) The length of the road before entering the intersection.
5) The slow vehicles' velocity.
6) The simulation velocity.

We have conducted a series of experiments that vary the initial conditions as follows:

1) The intensity on the main road $\in [0,1;0,9]$.
2) The intensity on the secondary road $\in [0,1;0,9]$.
3) The proportion of the slow vehicles $\in [0,1;0,9]$.
4) The slow vehicles' velocity - $0,8$.
5) Number of lanes $\{2,3,4\}$.
6) The length of the road before entering the intersection - $\{7,13\}$.
7) The number of time steps in each experiment: 1000.

A special case of the general model with the closed canalized traffic with parameter α that is set to zero researched. The traffic is *canalized*, if the vehicles don't change lanes. The traffic is *closed*, if the beginning and the ending of road's lane are connected, so there are no incoming flows and number of vehicles is permanent during a simulation.

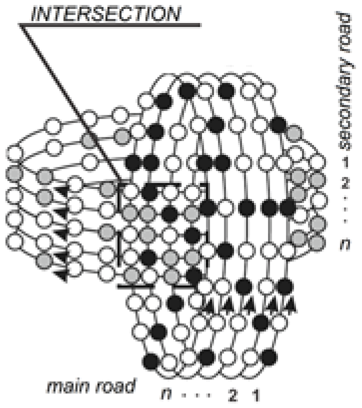


Fig. 7. Closed canalized traffic model.

V. STATISTICAL RESEARCHES

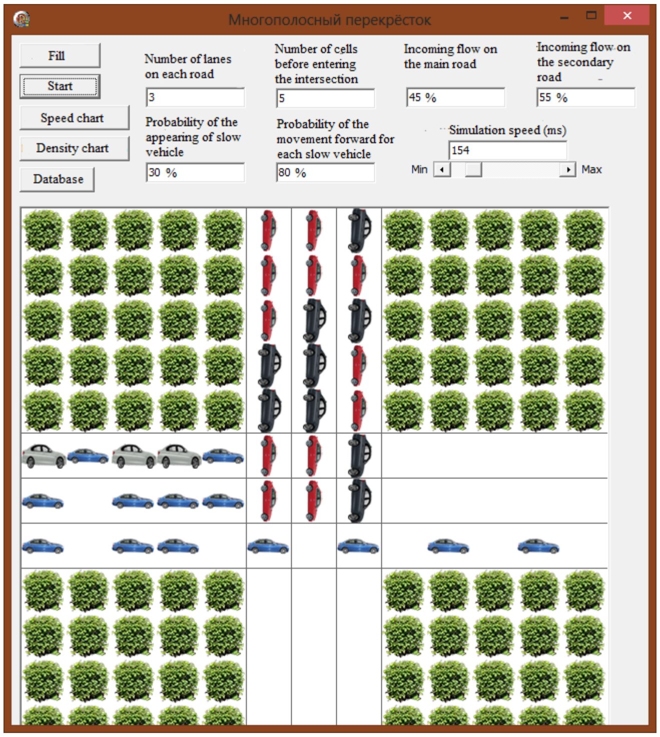
Based on conducted experiments (4374 experiments) in general heterogeneous traffic model, we may make the following conclusions:

✓the velocity on the main road is close to the velocity of the slow vehicles.

✓the velocity on the secondary road tends to zero when the intensity on the main road is greater than or equal to 0.3. The intensity on the secondary road can take any value (Fig. 9, Fig. 10).

✓the velocity doesn't depend on the length of segment of the road before the intersection.

✓the presence of the slow vehicles resolves the engagement situation for any value of the density (Fig. 8).



Number of lanes on each road: 3

Number of cells before entering the intersection: 5

Incoming flow on the main road: 45 %

Incoming flow on the secondary road: 55 %

Probability of the appearing of slow vehicle: 30 %

Probability of the movement forward for each slow vehicle: 80 %

Simulation speed (ms): 154

Fig. 8. Engagement situation, in which vehicles from both roads can't continue to move forward.

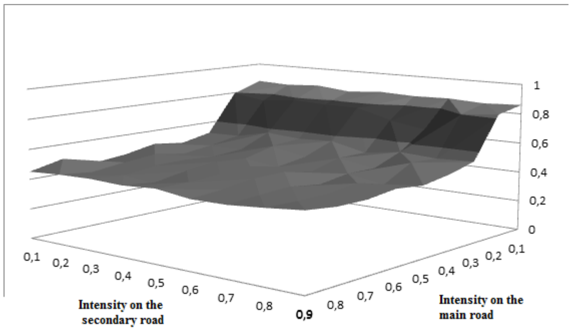


Fig. 9. A plot of velocity as a function of the intensity on the main road.

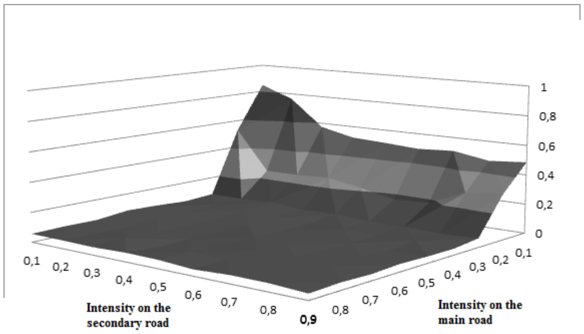


Fig. 10. A plot of the velocity as a function of the intensity on the secondary road.

Results in case of $\alpha = 0$: In the case of canalized traffic average velocity estimates were obtained in addiction of density, number of contours and parameter p (Fig. 11, Fig. 12):

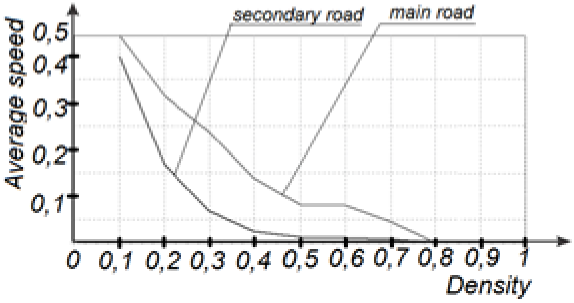


Fig. 11. A plot of the velocity as a function of the density $p = 0.5$ and $n = 5$.

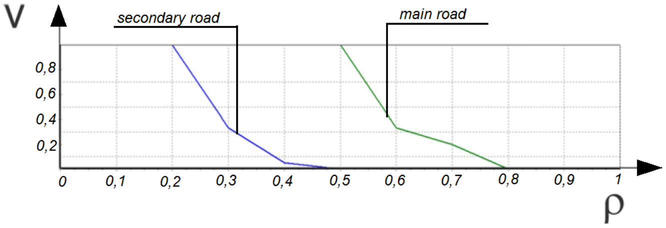


Fig. 12. A plot of the velocity as a function of the density $p = 1$ and $n = 5$.

Software implementation of the simulation model was developed by means of an object-oriented programming language Delphi in programming environment RAD Studio 2010 with the auxiliary software shell AlphaControls.

VI. Conclusion

The simulation model was researched. Based on conducted experiments in a software implementation of the simulation model the following results have been obtained:

1) A simulation model of one-way traffic flow on the unregulated multi-lane intersection of two roads was developed.
2) Estimates of the average velocity of the vehicles' movement at the unregulated intersections were obtained.
3) Estimates of the average velocity as a function of the proportion of the slow vehicles in the traffic flow were obtained.
4) In the case of the canalized traffic:
   a) If $\rho_1 > 0.5$ and $\rho_2 > \frac{(n-k)k}{2}$, the contour's engagement becomes the total and appears for any initial conditions (parameters).
   b) If $0.15 < \rho_1 < 0.5$ and $\rho_2 > 0.15$, the engagement in the region of intersection appears periodically.
   c) If $\rho_1 < 0.15$ and $\rho_2 < 0.15$, the engagements disappear for any initial conditions.
5) If $\rho_1 > 0.8$ $(i = 1, 2)$, there is total engagement, i.e. traffic is stopped, in the closed canalized traffic model (Fig. 12).
6) In the case of the closed canalized traffic and the general traffic model the existence of a steady state and station-

ary value of average vehicle velocity is experimentally proved.

VII. Future Works

In further work is planned to investigate the elaboration of the described model:

1) velocity estimation with two-way traffic.
2) Comparative analysis of the velocity estimates in a discrete model multi-lane intersection with and without regulation.

Acknowledgment

The authors would like to thank their supervisor prof. Buslaev A.P. for discussion and problem statement.

References

[1] V.F. Babkov. Road conditions and traffic safety, Moscow. Transport. pp. 271. 1993. (in russian)

[2] Statistics and information portalYandex.Maps. Service available at http://maps.yandex.com/

[3] V.V. Silyanov. Traffic theory in road design and traffic control. Moscow. Transport, pp. 303, 1977. (in russian)

[4] A.P. Buslaev, A.V. Novikov, V.M. Prokhodko, A.G. Tatashev, M.V. Yashina. Stochastic and simulation approaches to optimization of road traffic. Ed. Corresponding Member of RAS VM Prikhodko. Mir. pp. 368. 2003. (in russian)

[5] Nagel K., Schreckenberg M. A Cellular automation model for freeway traffic // J. Phys. I France. - 1992. - V. 2, N. 2221.

[6] M. Rickert, K.Nagel, M.Schreckenberg, A. Latour. Two lane Traffic Simulations using Cellular Automata. LANL Report No. LA-UR 95-4367, 2008

[7] REGULATION (EC) No 139/2004, MERGER PROCEDURE, Article 6(1)(b) NON-OPPOSITION, Date: 25/09/2008. http://ec.europa.eu/competition/mergers/cases/decisions/m5219_20080925_20310_en.pdf

[8] B.N. Chetverushkin, M.A. Trapesnikova, I.R. Furmanov, N.G. Churbanova. Macro- and micromodels in description of traffic on multilane roads. Moscow, Trudi MFTI, N 4, V 2. 2010. (in russian)

Discontinuous Hermite Collocation and IMEX Runge-Kutta for a Treated Quasi-linear Heterogeneous Brain Tumor Model

I. E. Athanasakis[‡], E. P. Papadopoulou[†] and Y. G. Saridakis[‡]

Abstract—Over the past few years several mathematical models have been developed to simulate and study the growth of treated or untreated aggressive forms of brain tumors. Encouraged by our recent results on the development of fourth order Discontinuous Hermite Collocation (DHC) numerical schemes to approximate the classical solution of parabolic evolution problems, in the present work we consider employing the DHC method for the solution of a quasi-linear tumor growth model which, apart from proliferation and diffusion, incorporates as well the effects from radiotherapy and chemotherapy. The model is also being characterized by a discontinuous diffusion coefficient to incorporate the heterogeneity of the brain tissue. To study the spatiotemporal dynamics of the model problem, the DHC spatial discretization is coupled with Implicit-Explicit (IMEX) Runge-Kutta (RK) third order schemes for the time discretization. The effectiveness of the resulting DHC-RK method is being demonstrated through several numerical experiments.

Keywords—High-grade Gliomas, Radiotherapy, Chemotherapy, Reaction-Diffusion PDEs, Discontinuous Hermite Collocation, Implicit-Explicit Runge-Kutta.

I. Introduction

HIGH-GRADE GLIOMAS are among the most common and aggressive forms of primary brain tumors. The most typical problem in diagnosis and treatment of patients with high-grade glioma, even after an extensive surgical procedure, is the rapid infiltration of tumor cells in adjacent normal tissue. Postoperative therapeutic treatment, such as radiotherapy and chemotherapy, is considered absolutely necessary to reduce tumor expansion.

As gliomas are known to consist of motile cells able to proliferate as well as migrate, well known and successful mathematical models, such as [11], [27], [28] and [9] (for a review see [13]), have been using reaction-diffusion evolution equations to describe the core spatiotemporal model's dynamics. The incorporation of brain's tissue heterogeneity (white-grey matter) was achieved in [19], [25] and [26] by introducing an appropriately discontinuous diffusion coefficient.

Recently, in [20], the effects of low-dose-rate radiotherapy, as a generalized linear quadratic model, and chemotherapy, as a simple log-kill model, were incorporated into a logistic growth reaction-diffusion model and several different schedules of sequential or combined therapy were studied in detail. A very interesting approach, as it pertains to radiotherapy modeling, was also presented in [10] (see also [21]) where a patient-specific, biologically optimized radiotherapy plan was presented.

Collocation (cf. [22], [8]) is an easily implemented spatial discretization method for BVPs that requires no numerical integration as it does not rely on a variational formulation. Combined with third degree finite element basis function, such as Hermite cubic or Spline elements, produces fourth order approximations to sufficiently smooth solutions.

Since the introduction of a class of discontinuous Hermite elements and their combined usage with the Collocation method (cf. [15], [16]), for the treatment of linear reaction-diffusion problems with discontinuous diffusion coefficients, the method has been also coupled (cf. [3], [4]) with high order Runge-Kutta to increase performance and stability.

Following our recent results in [3] and [4], the main objective in this work is to study the performance of the DHC method, combined with Implicit-Explicit Runge-Kutta schemes, as it pertains to the solution of the logistic quasi-linear heterogeneous brain tumor invasion model that also incorporates the effects from radiotherapy and chemotherapy. In the present study we include the results from the 1+1 dimension case, while the results for higher dimensions will be presented elsewhere (cf. [5]).

II. Methodology

A. The Mathematical Model

The core model PDE, that describes heterogeneous brain tumor invasion (cf. [19]) and incorporates the effects from radiotherapy and chemotherapy (cf. [20], [21]), is given in the form:

$$\frac{\partial \bar{c}}{\partial \bar{t}} = \nabla \cdot \left(\bar{D}(\bar{x}) \nabla \bar{c} \right) + \rho \bar{c}\left(1 - \frac{\bar{c}}{c_k}\right) - \bar{R}(\bar{c}, \bar{t}) - \bar{G}(\bar{c}, \bar{t}) \ , \quad (1)$$

where $\bar{c}(\bar{x}, \bar{t})$ denotes the tumor cell density, ρ denotes the net proliferation rate, c_k denotes the carrying capacity and $\bar{D}(\mathbf{x})$ is the diffusion coefficient representing the active motility of malignant cells satisfying

$$\bar{D}(\bar{x}) = \begin{cases} D_g & , \quad \bar{x} \in \bar{\Omega}_g \text{ Grey Matter} \\ D_w & , \quad \bar{x} \in \bar{\Omega}_w \text{ White Matter} \end{cases} \quad , \quad (2)$$

with D_g and D_w scalars and $D_w > D_g$, since glioma cells migrate faster in white than in grey matter.

Manuscript received May 31, 2015; revised June 30, 2015.

[‡]School of Production Engineering and Management, Applied Math & Computers Laboratory, Technical University of Crete, 73100 Chania, Greece

[†]School of Mineral Resources Engineering, Applied Math & Computers Laboratory, Technical University of Crete, 73100 Chania, Greece

Email of the corresponding author : y.saridakis@amcl.tuc.gr

The term $\bar{R}(\bar{c},\bar{t})$ denotes low-dose-rate and fractionated radiotherapy, and is defined by (cf. [14], [20] and the relevant references therein):

$$\bar{R}(\bar{c},\bar{t}) = \begin{cases} R_{\text{eff}} k_R(\bar{t})\bar{c} & , \quad \bar{t} \in \left(\bar{T}_{R_0}, \bar{T}_{R_1}\right] \text{ (therapy on)} \\ 0 & , \quad \bar{t} \notin \left(\bar{T}_{R_0}, \bar{T}_{R_1}\right] \text{ (therapy off)} \end{cases} \tag{3}$$

where $k_R(\bar{t})$ denotes the temporal profile of the radiation schedule and, by using a time step of one day, is simply one on therapy days and zero otherwise. R_{eff} denotes the effect of n fractions per day and is given by

$$R_{\text{eff}} = nd\left\{\alpha + 2\beta d\left[g(\mu\tau) + 2\left(\frac{\cosh(\mu\tau)-1}{(\mu\tau)^2}\right)h_n(\phi)\right]\right\} \tag{4}$$

with

$$g(\mu\tau) = \tfrac{\mu\tau - 1 + e^{-\mu\tau}}{(\mu\tau)^2} \quad \text{and} \quad h_n(\phi) = \tfrac{(n-1-n\phi+\phi^n)\phi}{n(1-\phi)^2} \ , \tag{5}$$

where α and β are sensitivity parameters, d is the dose rate, μ is the half time for repair of DNA damage, τ is the irradiation duration and $\phi = e^{-\mu(\tau+\Delta\tau)}$ with $\Delta\tau$ denoting the time interval between fractions.

In analogy to the radiotherapy equation in (3), the term $\bar{G}(\bar{c},\bar{t})$ denotes the effect of chemotherapy and, assuming a simple log-kill mode, is defined by (cf. [20] and the relevant references therein) :

$$\bar{G}(\bar{c},\bar{t}) = \begin{cases} k_G(\bar{t})\bar{c} & , \quad \bar{t} \in \left(\bar{T}_{G_0}, \bar{T}_{G_1}\right] \text{ (therapy on)} \\ 0 & , \quad \bar{t} \notin \left(\bar{T}_{G_0}, \bar{T}_{G_1}\right] \text{ (therapy off)} \end{cases} \tag{6}$$

and $k(\bar{t})$ is proportional to the drug concentration.

On the anatomy boundaries zero flux boundary conditions are imposed while for $\bar{t} = 0$ an initial spatial distribution of malignant cells $\bar{c}(\bar{x}, 0) = \bar{f}(\bar{x})$ is assumed.

Using the dimensionless variables:

$$x = \sqrt{\frac{\rho}{D_w}}\bar{x} \ , \quad t = \rho\bar{t} \ , \quad c(x,t) = \frac{1}{c_k}\bar{c}\left(\sqrt{\frac{\rho}{D_w}}\bar{x}, \ \rho\bar{t}\right) \ ,$$

$$D = \frac{\bar{D}}{D_w} \ , \ R = R(t) = \frac{R_{\text{eff}} k_R(\rho\bar{t})}{\rho} \ , \ G = G(t) = \frac{k_G(\rho\bar{t})}{\rho}$$

$$\text{and} \ \ f(x) = \frac{1}{c_k}\bar{f}\left(\sqrt{\frac{\rho}{D_w}}\bar{x}\right)$$

with $N_0 = \int \bar{f}(\bar{x})d\bar{x}$ to denote the initial number of tumor cells in the brain, the model equation in (1) becomes

$$\frac{\partial c}{\partial t} = \nabla \cdot (D\nabla c) + c(1-c) - Rc - Gc \ . \tag{7}$$

We remark that the parabolic nature of the above equation implies continuity of c as well as of both $\partial c/\partial t$ and $D\nabla c$. Therefore, in view of the jump discontinuities of the diffusion, radiotherapy and chemotherapy parameters, described in relations (2), (3) and (6) respectively, appropriate compatibility conditions have to be imposed on the interface between white Ω_w and gray Ω_g matter regions, as well as a proper time schedule has to be followed in order to distinguish and properly implement time intervals with no or any kind of therapy protocol, especially if it is to follow a time step other than the time step of one day.

To be more precise and in order to fix notation let us assume that radiotherapy and chemotherapy are respectively administered in the time intervals

$$T_1 < t \leq T_3 \text{ and } T_2 < t \leq T_4$$

with

$$0 = T_0 < T_1 < T_2 \leq T_3 < T_4 < T_5 = T.$$

Then, the dimensionless IBVP in 1+1 dimensions takes the form:

$$\begin{cases} \dfrac{\partial c}{\partial t} = \dfrac{\partial}{\partial x}\left(D\dfrac{\partial c}{\partial x}\right) + \rho_\ell c - c^2 \ , \ x \in [a,b], \ T_{\ell-1} < t \leq T_\ell \\ \dfrac{\partial c}{\partial x}(a,t) = \dfrac{\partial c}{\partial x}(b,t) = 0 \\ c(x,0) = c_\ell(x) \end{cases} \tag{8}$$

where

$$\rho_\ell = \rho_\ell(t) = \begin{cases} 1 & , \quad T_0 < t \leq T_1 \\ 1 - R & , \quad T_1 < t \leq T_2 \\ 1 - R - G & , \quad T_2 < t \leq T_3 \\ 1 - G & , \quad T_3 < t \leq T_4 \\ 1 & , \quad T_4 < t \leq T_5 \end{cases} \tag{9}$$

and

$$c_\ell(x) = \begin{cases} f(x) & , \quad T_0 < t \leq T_1 \\ c(x,T_1) & , \quad T_1 < t \leq T_2 \\ c(x,T_2) & , \quad T_2 < t \leq T_3 \\ c(x,T_3) & , \quad T_3 < t \leq T_4 \\ c(x,T_4) & , \quad T_4 < t \leq T_5 \end{cases} . \tag{10}$$

Furthermore, let us also assume that there are K interface points w_k in the region $[a,b]$ that distinguish white from gray matter. To be more specific, assume that

$$a = w_0 < w_1 < \cdots < w_k < \cdots < w_K < w_{K+1} = b,$$

and, without any loss of the generality, define

$$\Omega_g = \bigcup_{k=1}^{\lceil K/2 \rceil} \mathcal{W}_{2k-1} \quad \text{and} \quad \Omega_w = \bigcup_{k=1}^{\lfloor K/2 \rfloor} \mathcal{W}_{2k} \tag{11}$$

with

$$\mathcal{W}_k = (w_{k-1}, w_k) \ , \quad k = 1, \ldots, K+1 \ . \tag{12}$$

Then, the required compatibility conditions across each interface point w_k , $k = 1, \ldots, K$, take the form:

$$\lim_{x \to w_k^-} c(x,t) = \lim_{x \to w_k^+} c(x,t) \tag{13}$$

and

$$\lim_{x \to w_k^-} D(x)c_x(x,t) = \lim_{x \to w_k^+} D(x)c_x(x,t) \ . \tag{14}$$

Finally, we remark that the diffusion coefficient D in (7), is described by:

$$D = D(x) = \begin{cases} \gamma, & \text{when} \quad x \in \Omega_g \\ 1, & \text{when} \quad x \in \Omega_w \end{cases} , \tag{15}$$

where $\gamma = D_g/D_w$.

B. Derivative Discontinuous Hermite Collocation (DHC)

Let us consider a uniform partition of each one of the $k = 1, \ldots, K+1$ regions $\overline{\mathcal{W}}_k = [w_{k-1}, w_k]$ into N_k subintervals of length

$$h_k := \frac{w_k - w_{k-1}}{N_k} \ . \tag{16}$$

Therefore

$$[a,b] = \bigcup_{j=1}^{N+1} I_j \ , \quad I_j = [x_{j-1}, x_j] \tag{17}$$

with

$$x_j = a + j\ h_j(k) \quad , \quad j = 0, \ldots, N+1 \ , \tag{18}$$

where

$$N = \sum_{k=1}^{K+1} N_k \quad \text{and} \quad h_j(k) = h_k \text{ when } I_j \subseteq \overline{\mathcal{W}}_k \ , \tag{19}$$

for $k = 1, \ldots, K+1$.

The DHC method (cf. [16], [3]) seeks an approximate solutions $u(x,t) \sim c(x,t)$ in the form

$$u(x,t) = \sum_{j=0}^{N+1} \left[\alpha_{2j}(t)\phi_{2j}(x) + \alpha_{2j+1}(t)\phi_{2j+1}(x) \right] \tag{20}$$

where the *derivative discontinuous Hermite cubic basis functions* $\phi_{2j}(x)$ and $\phi_{2j+1}(x)$, centered at the node x_j, are defined by

$$\phi_{2j}(x) = \begin{cases} \phi\left(\dfrac{x_j - x}{h_j(k)}\right) & , \quad x \in I_j \\ \phi\left(\dfrac{x - x_j}{h_{j+1}(k)}\right) & , \quad x \in I_{j+1} \\ 0 & , \quad \text{otherwise} \end{cases} , \tag{21}$$

and

$$\phi_{2j+1}(x) \ = \ \begin{cases} -\dfrac{h_j(k)}{\gamma_j}\psi\left(\dfrac{x_j - x}{h_j(k)}\right) & , \quad x \in I_j \\ \dfrac{h_{j+1}(k)}{\gamma_{j+1}}\psi\left(\dfrac{x - x_j}{h_{j+1}(k)}\right) & , \quad x \in I_{j+1} \\ 0 & , \quad \text{otherwise} \end{cases} . \tag{22}$$

The functions $\phi(s)$ and $\psi(s)$ are the generating Hermite cubics over $[0,1]$, that is, for $s \in [0,1]$,

$$\phi(s) = (1-s)^2(1+2s) \quad , \quad \psi(s) = s(1-s)^2 \tag{23}$$

and

$$\gamma_j = \begin{cases} \gamma, & \text{when} \quad I_j \subseteq \Omega_g \\ 1, & \text{when} \quad I_j \subseteq \Omega_w \end{cases} . \tag{24}$$

It can, now, readily be verified that

$$u(x_j, t) = a_{2j}(t), \tag{25}$$

$$u_x(x_j,t) = \begin{cases} a_{2j+1}(t)/\gamma & , \quad \text{if } x_j \in \Omega_g \bigwedge x_j \neq w_k \quad \forall k \\ a_{2j+1}(t) & , \quad \text{if } x_j \in \Omega_w \bigwedge x_j \neq w_k \quad \forall k \end{cases} , \tag{26}$$

while, whenever $x_j = w_k$, for some k, there holds

$$\lim_{x \to w_k^-} \gamma_j u_x(x,t) = \lim_{x \to w_k^+} \gamma_{j+1} u_x(x,t) \tag{27}$$

hence, the compatibility condition (14) is satisfied.

For the evaluation of the unknown parameters $\alpha_i \equiv \alpha_i(t)$, $i = 0, \ldots, 2(N+1)$ the Collocation method produces a system of ordinary differential equations (ODEs) by forcing the approximate solution $u(x,t)$ to vanish at $2N+2$ *interior collocation* points and the 2 *boundary collocation* points. Collocation at the Gauss points (cf. [8]) adopts the two roots of the Legendre polynomial of degree 2 in each element I_j , $j = 1, \ldots, N+1$ to produce the needed interior collocation points. Namely, the interior Gaussian collocation points for each element I_j are given by

$$\sigma_{2j-1} = \frac{x_{j-1} + x_j}{2} - \frac{h_j}{2\sqrt{3}} \text{ and } \sigma_{2i} = \frac{x_{j-1} + x_j}{2} + \frac{h_j}{2\sqrt{3}}. \tag{28}$$

Substituting, now, $u(x,t)$ of (20) into the equation of the IBVP in (8), observing that in each I_j is an element of four degrees of freedom and noticing that in the interior of each I_j there are no interface points, the two *elemental* collocation equations are written as

$$\begin{aligned} \sum_{L=2j-2}^{2j+1} \dot{\alpha}_L(t)\phi_L(\sigma_i) \ &= \ \gamma_j \sum_{L=2j-2}^{2j+1} \alpha_L(t)\phi_L''(\sigma_i) \\ &+ \ \rho_\ell \sum_{L=2j-2}^{2j+1} \alpha_L(t)\phi_L(\sigma_i) \\ &- \ \left(\sum_{L=2j-2}^{2j+1} \alpha_L(t)\phi_L(\sigma_i) \right)^2 \end{aligned} \tag{29}$$

for $i = 2j-1$, $2j$ and where, of course, $\dot{\alpha}_L(t) = \dfrac{d}{dt}\alpha_L(t)$ and $\phi_L'(x) = \dfrac{d}{dx}\phi_L(x)$.

Working as in [6], the above elemental equations (29) are expressed in matrix form by:

$$\sum_{L=2j-2}^{2j+1} \alpha_L(t)\phi_L^{(m)}(\sigma_i) = C_j^{(m)}\boldsymbol{\alpha}_j \ , \ i = 2j-1, 2j \ , \tag{30}$$

where

$$C_j^{(m)} = \left[\begin{array}{cc} A_j^{(m)} & B_j^{(m)} \end{array} \right] \ , \quad m = 0, 2 \tag{31}$$

$$\boldsymbol{\alpha}_j = \left[\begin{array}{cccc} \alpha_{2j-2}(t) & \alpha_{2j-1}(t) & \alpha_{2j}(t) & \alpha_{2j+1}(t) \end{array} \right]^{\mathrm{T}} \tag{32}$$

with

$$A_j^{(m)} = \left[\begin{array}{ll} \phi_{2j-2}^{(m)}(\sigma_{2j-1}) & \phi_{2j-1}^{(m)}(\sigma_{2j-1}) \\ \phi_{2j-2}^{(m)}(\sigma_{2j}) & \phi_{2j-1}^{(m)}(\sigma_{2j}) \end{array}\right] \tag{33}$$

$$= \frac{1}{h_j^m}\left[\begin{array}{ll} s_1^{(m)} & \frac{h_j(k)}{\gamma_j}s_2^{(m)} \\ s_3^{(m)} & -\frac{h_j(k)}{\gamma_j}s_4^{(m)} \end{array}\right] \ , \quad m=0,2$$

$$B_j^{(m)} = \left[\begin{array}{ll} \phi_{2j}^{(m)}(\sigma_{2j-1}) & \phi_{2j+1}^{(m)}(\sigma_{2j-1}) \\ \phi_{2j}^{(m)}(\sigma_{2j}) & \phi_{2j+1}^{(m)}(\sigma_{2j}) \end{array}\right] \tag{34}$$

$$= \frac{1}{h_j^m}\left[\begin{array}{ll} s_3^{(m)} & \frac{h_j(k)}{\gamma_j}s_4^{(m)} \\ s_1^{(m)} & -\frac{h_j(k)}{\gamma_j}s_2^{(m)} \end{array}\right] \ , \quad m=0,2$$

and

	$m=0$	$m=2$
$s_1^{(m)}$	$\frac{9+4\sqrt{3}}{18}$	$-2\sqrt{3}$
$s_2^{(m)}$	$\frac{3+\sqrt{3}}{36}$	$-1-\sqrt{3}$
$s_3^{(m)}$	$\frac{9-4\sqrt{3}}{18}$	$2\sqrt{3}$
$s_4^{(m)}$	$-\frac{3-\sqrt{3}}{36}$	$-1+\sqrt{3}$

.

Using, now, the symbol $\circ$ to denote the Hadamard matrix product, the matrix form of the elemental equations in (29) may be written as (see also [6]):

$$\begin{aligned} C_j^{(0)}\dot{\boldsymbol{\alpha}}_j &= \gamma_j C_j^{(2)}\boldsymbol{\alpha}_j + \rho_\ell C_j^{(0)}\boldsymbol{\alpha}_j \\ &- \left(C_j^{(0)}\boldsymbol{\alpha}_j\right) \circ \left(C_j^{(0)}\boldsymbol{\alpha}_j\right) \end{aligned} \tag{35}$$

Moreover, observe that combination of the relations in (26) and the Neumann boundary conditions in (8) immediately implies

$$\alpha_1(t) = \alpha_{2N+3}(t) = 0 \ \ , \tag{36}$$

hence, also,

$$\dot{\alpha}_1(t) = \dot{\alpha}_{2N+3}(t) = 0 \ \ . \tag{37}$$

The above elemental and boundary collocation equations lead to the non-linear Collocation system of ODEs, described by:

$$C_0\dot{\boldsymbol{\alpha}} = \boldsymbol{\gamma}C_2\boldsymbol{\alpha} + \rho_\ell C_0\boldsymbol{\alpha} - (C_0\boldsymbol{\alpha} \circ C_0\boldsymbol{\alpha}) \tag{38}$$

where the $(2N+2)\times(2N+2)$ matrices $C_m, \ m=0,2$ and $\boldsymbol{\gamma}$ are defined by:

$$C_m = \left[\begin{array}{cccccc} \tilde{A}_1^{(m)} & B_1^{(m)} & & & & \\ & A_2^{(m)} & B_2^{(m)} & & & \\ & & \ddots & \ddots & & \\ & & & A_N^{(m)} & B_N^{(m)} & \\ & & & & A_{N+1}^{(m)} & \tilde{B}_{N+1}^{(m)} \end{array}\right]$$

and

$$\boldsymbol{\gamma} = \text{diag}\left(\gamma_1 \ \gamma_2 \ \gamma_2 \ \cdots \ \gamma_N \ \gamma_N \ \gamma_{N+1}\right),$$

while the $2N+2$ vectors $\boldsymbol{\alpha} \equiv \boldsymbol{\alpha}(t)$ and $\dot{\boldsymbol{\alpha}} \equiv \dot{\boldsymbol{\alpha}}(t)$ are described by

$$\boldsymbol{\alpha} = \left[\begin{array}{ccccc} \alpha_0(t) & \alpha_2(t) & \cdots & \alpha_{2N+1}(t) & \alpha_{2N+2}(t) \end{array}\right]^{\mathrm{T}}$$

$$\dot{\boldsymbol{\alpha}} = \left[\begin{array}{ccccc} \dot{\alpha}_0(t) & \dot{\alpha}_2(t) & \cdots & \dot{\alpha}_{2N+1}(t) & \dot{\alpha}_{2N+2}(t) \end{array}\right]^{\mathrm{T}} .$$

The vectors $\tilde{A}_1^{(k)}$ and $\tilde{B}_{N+1}^{(k)}$ denote the first columns of the matrices $A_1^{(k)}$ and $B_{N+1}^{(k)}$ respectively, as their second columns have been omitted due to the zero boundary conditions.

Concluding this section we point out that the linear independence of the derivative discontinuous Hermite cubic basic functions yields the non-singularity of the matrix C_0 of the Collocation ODE system in (38) implying the existence, of course, of the inverse C_0^{-1}.

C. Implicit-Explicit Runge-Kutta schemes

Implicit-Explicit (IMEX) Runge-Kuttta schemes (cf. [18], [2], [12] and the references therein) are based on implementing an implicit scheme for the stiff part and an explicit scheme for the non or mildly stiff part of a spatial discretized system of ODEs. Here, based on the effective coupling of the Hermite Collocation method with Runge-Kutta schemes we've reported in [3] and [4], we implement an IMEX Runge-Kutta scheme (cf. [17]) that operates on the Collocation system of ODEs in (38) and applies Diagonally Implicit Runge-Kutta (DIRK; cf. [1]) for the linear part and Strong Stability Preserving Runge-Kutta (SSPRK; cf. [23], [24]) for non linear part.

To be more specific, let us write the Collocation system of ODEs in (38) at time level $t=t_n=n\Delta t$ as

$$C_0\dot{\boldsymbol{\alpha}}^{(n)} = \boldsymbol{\gamma}C_2\boldsymbol{\alpha}^{(n)} + \rho_\ell C_0\boldsymbol{\alpha}^{(n)} - \left(C_0\boldsymbol{\alpha}^{(n)} \circ C_0\boldsymbol{\alpha}^{(n)}\right) \tag{39}$$

or, equivalently, as

$$\dot{\boldsymbol{\alpha}}^{(n)} = \mathcal{L}\left(\boldsymbol{\alpha}^{(n)}\right) + \mathcal{N}\left(\boldsymbol{\alpha}^{(n)}\right) \tag{40}$$

where

$$\mathcal{L}\left(\boldsymbol{\alpha}^{(n)}\right) = \boldsymbol{\gamma}C_0^{-1}C_2\boldsymbol{\alpha}^{(n)} + \rho_\ell\boldsymbol{\alpha}^{(n)} \tag{41}$$

and

$$\mathcal{N}\left(\boldsymbol{\alpha}^{(n)}\right) = -C_0^{-1}\left(C_0\boldsymbol{\alpha}^{(n)} \circ C_0\boldsymbol{\alpha}^{(n)}\right) \tag{42}$$

with

$$\dot{\boldsymbol{\alpha}}^{(n)} = \left[\dot{\alpha}_0^{(n)} \ \ \dot{\alpha}_2^{(n)} \cdots \dot{\alpha}_{2N+2}^{(n)}\right]^{\mathrm{T}}$$

and

$$\boldsymbol{\alpha}^{(n)} = \left[\alpha_0^{(n)} \ \ \alpha_2^{(n)} \cdots \alpha_{2N+2}^{(n)}\right]^{\mathrm{T}} .$$

Then, the IMEX Runge-Kutta scheme for the solution of the system in (40) is expressed as (cf. [17]):

$$\begin{aligned} \boldsymbol{\alpha}^{(1)} &= \boldsymbol{\alpha}^{(n)} + \lambda\Delta t\mathcal{L}\left(\boldsymbol{\alpha}^{(1)}\right) \\ \boldsymbol{\alpha}^{(2)} &= \boldsymbol{\alpha}^{(n)} + \Delta t\mathcal{N}\left(\boldsymbol{\alpha}^{(1)}\right) + \Delta t\left(1-2\lambda\right)\mathcal{L}\left(\boldsymbol{\alpha}^{(1)}\right) + \\ &+ \lambda\Delta t\mathcal{L}\left(\boldsymbol{\alpha}^{(2)}\right) \\ \boldsymbol{\alpha}^{(3)} &= \boldsymbol{\alpha}^{(n)} + \frac{\Delta t}{4}\left(\mathcal{N}\left(\boldsymbol{\alpha}^{(1)}\right) + \mathcal{N}\left(\boldsymbol{\alpha}^{(2)}\right)\right) + \end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{\Delta t(1-2\lambda)}{2}\mathcal{L}\left(\boldsymbol{\alpha}^{(1)}\right) + \lambda\Delta t\mathcal{L}\left(\boldsymbol{\alpha}^{(3)}\right)\\
\boldsymbol{\alpha}^{(n+1)} \quad &= \quad \boldsymbol{\alpha}^{(n)} + \Delta t\left[\mathcal{N}\left(\boldsymbol{\alpha}^{(1)}\right) + \mathcal{N}\left(\boldsymbol{\alpha}^{(2)}\right) + \right.\\
&+ \quad 4\mathcal{N}\left(\boldsymbol{\alpha}^{(3)}\right) + \mathcal{L}\left(\boldsymbol{\alpha}^{(1)}\right) + \mathcal{L}\left(\boldsymbol{\alpha}^{(2)}\right) + \\
&+ \quad \left. 4\mathcal{L}\left(\boldsymbol{\alpha}^{(3)}\right)\right]
\end{aligned}
$$

Finally, we remark that the convergence and stability properties of the above scheme have been studied in [17].

III. Numerical Simulations

In this section, we report the results from the numerical investigation of the performance of the IMEX-DHC method on two virtual model problems.

For both model problems the values of the radiotherapy and chemotherapy parameters used are given by (cf. [20]) $G = 0.0571$ day^{-1} and $R = 0.0196$ day^{-1}, respectively. The radiotherapy protocol followed included equal doses of 1.8Gy per day for 35 days, from day 170 to day 205, while the chemotherapy protocol, starting from day 205, included six cycles of daily treatment for 5 consecutive days followed by a 20 day recess.

A. Model Problem I

For the first single source model, centered at $\bar{x} = 1$, we consider the values:

$$
\begin{cases}
\bar{a} = -10 \text{ cm},\ \bar{b} = 10 \text{ cm},\ \bar{w}_1 = -6 \text{ cm},\ \bar{w}_2 = 8 \text{ cm}\\
\bar{\Omega}_g = [\bar{a}, \bar{w}_1) \cup (\bar{w}_2, \bar{b}] \text{ and } \bar{\Omega}_w = [\bar{w}_1, \bar{w}_2]\\
D_g = 0.0013 \text{ cm}^2\text{day}^{-1},\ D_w = 0.0065 \text{ cm}^2\text{day}^{-1}\\
\bar{\rho} = 0.012 \text{ day}^{-1}\ ,\ N_0 = 2 \times 10^4 \text{ cells}
\end{cases} .
$$

The results form the numerical simulation are depicted in Figs. 1 and 2, as well as in Table I.

More specifically, Fig. 1 depicts the evolution of the cell density function $\bar{c}(\bar{x}, \bar{t})$. One may easily identify periods of untreated and treated tumor growth.

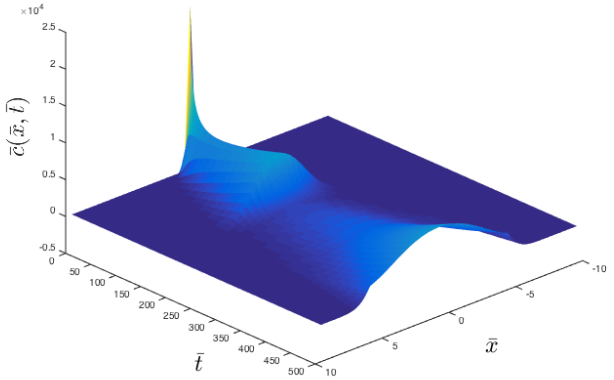


Fig. 1: Time evolution of the cell density $\bar{c}(\bar{x}, \bar{t})$

The radiotherapy effect on the total number of tumor cells $\bar{N}(\bar{t})/N_0$, where $N(\bar{t}) = \int_{\bar{a}}^{\bar{b}} \bar{c}(\bar{x}, \bar{t})d\bar{x}$, is depicted in Fig. 2.

Finally, Table I summarizes the performance of the DHC-IMEX method. One may easily observe the 4-th order of convergence of the DHC method.

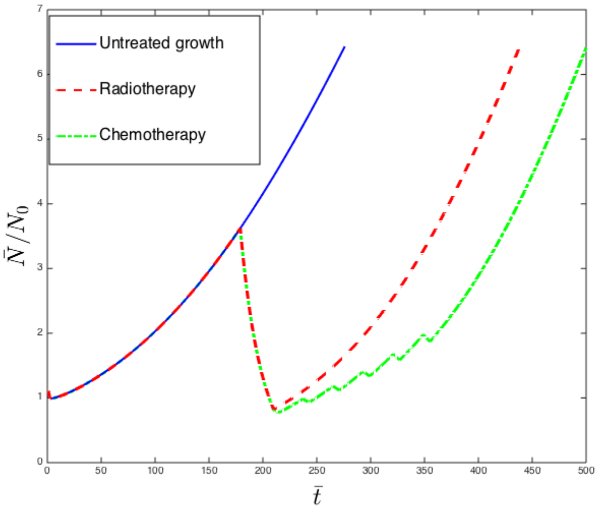


Fig. 2: The effect of radiotherapy on the total number of tumor cells.

Table I DCH-IMEX Performance

h	Error	O.o.c.	Time (sec)
1/8	3.5687e-06	-	0.24
1/16	2.3357e-07	3.93	0.30
1/32	1.4760e-08	3.98	0.42
1/64	9.2474e-10	3.99	0.88
1/128	5.6156e-11	4.04	1.55

B. Model Problem II

For the triple source model we consider the values:

$$
\begin{cases}
\bar{a} = -10 \text{ cm},\ \bar{b} = 10 \text{ cm},\ \bar{w}_1 = -4 \text{ cm},\ \bar{w}_2 = 6 \text{ cm}\\
\bar{\Omega}_g = [\bar{a}, \bar{w}_1) \cup (\bar{w}_2, \bar{b}] \text{ and } \bar{\Omega}_w = [\bar{w}_1, \bar{w}_2]\\
D_g = 0.0013 \text{ cm}^2\text{day}^{-1},\ D_w = 0.0065 \text{ cm}^2\text{day}^{-1}\\
\bar{\rho} = 0.012 \text{ day}^{-1}\ ,\ N_0 = 2 \times 10^4 \text{ cells}
\end{cases}
$$

All results are summarized in Figs. 3 and 4 as well as Table II and are completely similar to the corresponding ones of the previous model case.

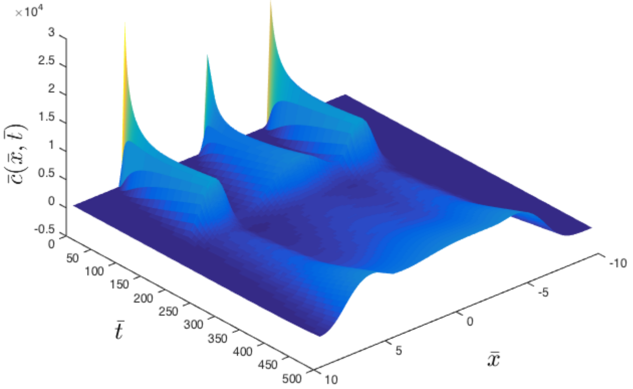


Fig. 3: Time evolution of the cell density $\bar{c}(\bar{x}, \bar{t})$.

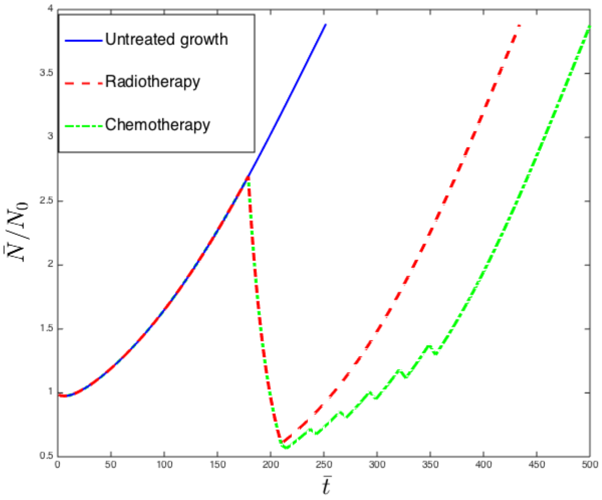


Fig. 4: The effect of radiotherapy and chemotherapy on the total number of tumor cells.

Table II DCH-IMEX Performance

h	Error	O.o.c.	Time (sec)
1/8	5.3380e-06	-	0.22
1/16	3.4585e-07	3.94	0.28
1/32	2.1802e-08	3.98	0.40
1/64	1.3655e-09	3.99	0.90
1/128	8.5010e-11	4.00	1.52

IV. Conclusion

We have developed and investigated the performance of a high order Derivative Discontinuous Hermite Collocation, coupled with an IMEX Runge-Kutta scheme, for the solution of a quasi-linear reaction diffusion IBVP that models the brain tumor growth taking into consideration brain's heterogeneity and the effects of radiotherapy and chemotherapy. The results obtained justify and encourage further analysis as well as implementation in higher dimensions.

Acknowledgment

The present research work has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program Education and Lifelong Learning of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALES (Grant number: MIS-379416). Investing in knowledge society through the European Social Fund.

References

[1] R. Alexander, *Diagonally Implicit Runge-Kutta Methods for stiff ODE's*, SIAM Num. Anal., vol. 14, no. 6, pp. 1006-1021, 1977.

[2] U. M. Ascher, S. J. Ruuth and R. J. Spiteri, *Implicit-Explicit Runge-Kutta Methods for Time-Dependent Partial Differential Equations*, Applied Numerical Mathematics, 25, 151-167, 1997

[3] I. E. Athanasakis, M. G. Papadomanolaki, E. P. Papadopoulou and Y. G. Saridakis, *Discontinuous Hermite Collocation and Diagonally Implicit RK3 for a Brain Tumor Invasion Model*, Procs WCE-ICAEM, 1, 241-246, 2013.

[4] I. E. Athanasakis, E. P. Papadopoulou and Y. G. Saridakis, *Runge-Kutta and Hermite Collocation for a biological invasion problem modeled by a generalized Fisher equation*, Journal of Physics: Conference Series, 490, 012133, 2014.

[5] I.E. Athanasakis, E.P. Papadopoulou and Y.G. Saridakis, *Discontinuous Tensor Product Collocation and Runge-Kutta schemes for heterogeneous brain tumor invasion models*, (in Preparation) 2015.

[6] I.E. Athanasakis, E.P. Papadopoulou and Y.G. Saridakis, *Hermite Collocation and SSPRK Schemes for the Numerical Treatment of a Generalized Kolmogorov-Petrovskii-Piskunov Equation*, Procs WCE-ICAEM 2015.

[7] S. M. Blinkov and I. Glezer, *The Human Brain in Figures and Tables*, New York: Plenum Press, 1968.

[8] C. de Boor and B. Swartz, *Collocation at Gaussian points*, SIAM Num. Anal., vol.10, pp. 582-606, 1973.

[9] P.K. Burgess, P.M. Kulesa, J.D. Murray and E.C. Alvord Jr, *The interaction of growth rates and diffusion coefficients in a three-dimensional mathematical model of gliomas*, Journal of Neuropathology and Experimental Neurology, vol.56, no. 6, pp.704-713, 1997.

[10] D. Corwin, C. Holdsworth, R. C. Rockne, An. D. Trister, M. M. Mrugala, J. K. Rockhill, R. D. Stewart2, M. Phillips2 and K. R. Swanson, *Toward Patient-Specific, Biologically Optimized Radiation Therapy Plans for the Treatment of Glioblastoma*, PLOS ONE, 8(11), 1-9, 2013

[11] G.C. Cruywagen, D.E. Woodward, P. Tracqui, G.T. Bartoo, J.D. Murray and E.C. Alvord Jr, *The modeling of diffusive tumors*, Journal of Biological Systems , vol.3, pp.937-945, 1995.

[12] T. Koto, *IMEX RungeKutta schemes for reactiondiffusion equations*, J Comp. App. Math., 215, 182 195, 2008

[13] J.D. Murray, *Mathematical Biology I and II*, Springer-Verlag, 3rd Edition 2002.

[14] P. Nilsson, H. D. Thames and M. C. Joiner, *A generalized formulation of the incomplete-repair model for cell survival and tissue response to fractionated low dose-rate irradiation*, Int. J. Radiat. Biol., 57, 12742, 1990.

[15] M.G. Papadomanolaki, *The collocation method for parabolic differential equations with discontinuous diffusion coefficient: in the direction of brain tumor simulations*, PhD Thesis, Technical University of Crete, 2012 (in Greek)

[16] M.G. Papadomanolaki and Y.G. Saridakis, *Hermite-Collocation for one dimensional tumor invasion model with heterogeneous diffusion*, HERMIS-$\mu\pi$, vol. 11, pp.63-68, 2010.

[17] L. Pareschi and G. Russo, *ImplicitExplicit RungeKutta Schemes and Applications to Hyperbolic Systems with Relaxation*, J. Scient. Comp., 25(1), 129-155, 2005

[18] S. J. Ruuth, *Implicit-Explicit Methods For Reaction-Diffusion Problems In Pattern Formation*, J. Math. Biology, 34, 148-176, 1995

[19] K.R.Swanson, *Mathematical modelling of the growth and control of tumor*, PHD Thesis, University of Washington, 1999.

[20] G. Powathil, M. Kohandel, S. Sivaloganathan, A. Oza2 and M Milosevic, *Mathematical modeling of brain tumors: effects of radiotherapy and chemotherapy*, Phy. Med. Biol. 52, 3291-3306, 2007.

[21] R. Rockne, E. C. Alvord Jr., J. K. Rockhill and K. R. Swanson, *A mathematical model for brain tumor response to radiation therapy*, J. Math. Biol., 58, 561578, 2009.

[22] R. D. Russel and L.F. Shampine, *A collocation method for boundary value problems* Numer. Math. , 19(1) pp. 128, 1972.

[23] C. W. Shu, *Total-variation-diminishing time discretizations*, SIAM J. Sci. Stat. Comput., **9**, 1073-1084, 1988

[24] C. W. Shu and S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes*, J. Comput. Phys., **77**, 439-471, 1988

[25] K.R.Swanson, E.C.Alvord Jr and J.D.Murray, *A quantitive model for differential motility of gliomas in grey and white matter*, Cell Proliferation, vol.33, pp.317-329, 2000.

[26] K. R. Swanson,C. Bridge, J. D. Murray and E. C. Alvord Jr, *Virtual and real brain tumors:using mathematical modeling to quantify glioma growth and invasion*, J.Neurol.Sci, vol.216, pp.1-10, 2003.

[27] P.Tracqui,G.C.CruywagenG,D.E.Woodward,T.Bartoo, J.D.Murray and E.C.Alvord Jr, *A mathematical model of glioma growth:The effect of chemotherapy on spatio-temporal growth*, *Cell Proliferation*, vol.28, pp.17-31, 1995.

[28] D.E.Woodward,J.Cook,P.Tracqui,G.C.Cruywagen,J.D.Murray,and E.C.Alvord Jr, *A mathematical model of glioma growth: the effect of extent of surgical resection*, *Cell Proliferation*, vol.29, pp.269-288, 1996.

Fuzzy sets theorem and cause event

Jelenka Savkovic-Stevanovic

Faculty of Technology and metallurgy Belgrade University, Karnegijeva 4, 11000 Belgrade,Serbia
stevanoviccace@gmail.com

Abstract. In this paper fuzzy sets theorem was derived. The qualitative and quantitative approach to causal modeling were considered. Fuzzy sets are modeled as the possible behavior of a system, and as the causal networks in which the nodes represent the variables. In this paper multilevel fuzzy functions were examined to causality parameters select. Fuzzy sets for causality selection were derived and conditional parameters were considered.

Keywords: Fuzzy set, theorem, multilevel function, membership higher order**.**

1 Introduction

The scientists view toward causality differs considerably from that of philosophers. Scientists are interested in discovering functional relationships among physical phenomena in order to explain their behavior. Over the years, scientists have studied two aspects of causality: isolation of the variables which represent cause phenomena and those which represent effect phenomena, and determination of the magnitude and direction of change in effect phenomena corresponding to a change in cause phenomena [1]-[5]. The identified variables along with their functional relationships can serve as a useful computational model for making inference.

Causal structure can be establish by examining some statistics associated with the variables of interest. The question "Is correlation proof of causation?" is very important.

The quantitative approach to causal modeling and inference involves computing the path coefficients between the cause and the effect variables and using the resulting equation to predict the change in the cause. Historically, this is the most important approach for analyzing causality and has already been explored to a great extent in sociology, economics and etc. [5]-[10].

In this paper fuzzy set and cause- effects functional relationships have studied.

2 Fuzziness

Fuzzy set is set in which members are presented as ordered pairs that include information on degree of membership. Let introduce a fuzzy subset A of the traditional set,

$$U(u_1,u_2,u_3,....u_k)\,. \qquad (1)$$

$$u_i = e\{A\} \qquad (2)$$

where $\mu_A(u_i)$ is degree of membership u_i in the subset A, and $\mu_A(u_i) = \in \{0,1\}$. If $\mu_A(u_i) = 0$ then u_i is not member of the subset *A*, and if $\mu_A(u_i) = 1$ then u_i is a member of the subset *A*, full membership.

A classical set of say, *k* elements, is a special case of a fuzzy set, where each of those *k* elements has one, for the degree of the membership, and every other element in the classical set has a degree of membership zero, for each reason do not bother to list it.

The usual operations can perform or ordinary sets are union, in which take all the elements that are in one set or the other, and intersection, in which take the elements that are in both sets.

In the case of fuzzy sets, taking a union is finding the degree of membership that an element should have in the new fuzzy set which is the union of two fuzzy sets.

Consider a union of two traditional sets and an element that to only one of those sets. If these sets are treated as fuzzy sets this element has degree of membership of 1 in one case and 0 in the other, since it belongs to one set and not the other. Let go to put this element in the union. Should be to look at the two degree of membership namely, 0 and 1, and pick the higher value of the two namely 1. In other words, what want for the degree of membership of an element when listed in the union of two fuzzy sets, is the maximum value of its degree of membership within the two fuzzy sets forming a union.

$$x + y = \max(x, y) \qquad (3)$$

For example,

$$0 + 1 = \max(0,1) = 1$$

$$1 + 1 = \max(1,1) = 1$$

Analogously, the degree of membership of an element in the intersection of two fuzzy sets is the minimum or the smaller value of its degree of membership individually in the two sets forming the intersection. For example,

$$xy = \min(x, y) \qquad (4)$$

For example,

0 x 1 = min(0,1)=0

1 x 0.8=min(1,0.8)=0.8.

In the fuzzy recording method the regeneration process is called defuzzification.

2 Conditional parameters

Although the quantitative approach has proven very useful for dealing with many real-world problems, it is neither sufficient nor necessary under some circumstances. Furthermore, because in reality, there may not exist enough quantitative knowledge to permit full quantitative modeling, abstract qualitative models are worthwhile to explore. Qualitative causal modeling has become one major line of research toward the representation of deep models in knowledge based systems. Two well qualitative causal simulation techniques will be describe here.

The first approach refers to the technique which predicts the possible qualitative behaviors of a system on the basis of the model comprising the predefined physical parameters and constraint predicates.

In the second approach causal knowledge is modeled as causal networks in which the nodes represent propositions or variables, the arcs signify direct dependencies between the linked propositions, and the strengths of these dependences are quantified by conditional probabilities. Stochastic simulation is a method of computing probabilities by counting the fraction of time an event occurs in a series of simulation run. If a causal model is available, it can be used to generate random samples of hypothetical scenarios that are likely to develop in the domain. The probability of an event or any combination of events can be computed by recording the fraction of time it registers true in the samples generated.

3 Parameters selection

An approach to parameters selection by multilevel fuzzy functions is developed on the basis of historical data and experience on a considered process. Fizzy set theory is a step toward an approchement between the precision of classical mathematics and the pervasive imprecision of the real world. Fuzziness of a phenomena steams from the lack of clearly defined boundaries.

Causality parameters can be selected by multilevel functions estimation, too.

Let consider set A and subsets $\breve{A}$

$$A\ , \tilde{A}_j\,(j = 1,2..m) \qquad \tilde{A} \subset A \qquad (5)$$

be to output, global observation, set and subsets, which contain various states to be diagnosed. Since output states in complex processes are often inconclusive, fuzzy set and fuzzy subsets A and $\tilde{A}_j$ are assumed to describes in practice. Assume that the observed field is a measurable output vector space consisting of n vectors (Fig.1).

$$X = (X_1, X_2, X_3, ...X_n) \qquad (6)$$

where X_i is *i*th vector with which A_j can be ambiguously predicted, i.e. $\tilde{A}_j$ can be determined according to the values of

$$X_i\,(i = 1,2,...n;\ j = 1,2,..m).$$

Suppose that *m* fuzzy subsets are divided into *k* groups by various characteristics such as the kinds of cause, where p_i have sum equal one.

$$A = (\tilde{A}_1, \tilde{A}_2, ...\tilde{A}_{pi}), ..., (\tilde{A}_{j+1}, \tilde{A}_{j+2}, ..., \tilde{A}_{j+p_i})$$

where p_i have the following relationship

$$\sum_{i=1}^{k} p_i = 1 \qquad (8)$$

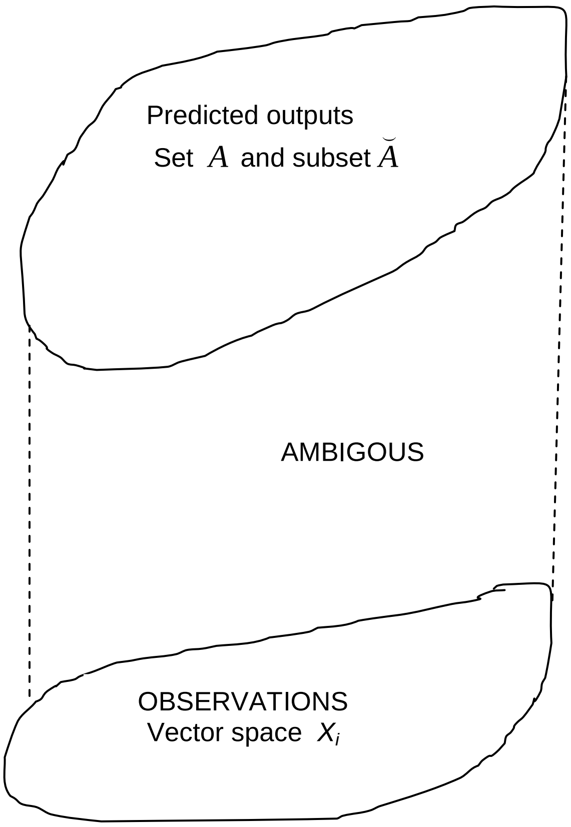


Fig.1 Parameters selection in various level

Any fuzzy subset *A'*$_j$ of X_i is characterized by a membership functions μ_{A_j} which associates with every member x_i of X_i , i.e., $\mu'_{A_j}(x_i)$ representing the degree of membership of x to fuzzy subset A'_j . A definition for the construction of a membership function is defined as follows.

4 Theorem

Statement 1. For every fuzzy subset $(\tilde{A}_{q1}, ..., \tilde{A}_{qp_q})$ $(1 \le q \le k)$ in the fuzzy set A , given by

$$(\tilde{A}_{1i}, \tilde{A}_{2i}, ..., \tilde{A}_{pi}) ......(\tilde{A}_{ki}, ..., \tilde{A}_{kpi}) \qquad (9)$$

a limited sequence *fq* with *f*-cut is determined , i.e.

$$0 < f_{q_1} < f_{q_2} < ... < f_{q_{p_{q-1}}} < 1 \qquad (10)$$

The corresponding membership function must have the following relationship:

IF $f_{qp_q-1} \le \mu_{qp_q}(x_i) \le 1$ THEN $x_i \subset \tilde{A}_{qp_q}$ (11)

IF $f_{qp_q-2} \le \mu_{qp_q-1}(x_i) \le f_{qp_q-1}$ THEN $x_i \subset \tilde{A}_{qPq-1}$ (12)

IF $0 \le \mu_{q1}(x_i) \le f_{q1}$ THEN $x_i \subset \tilde{A}_{q1}$ (13)

where $x_i \subset \tilde{A}_q$ means that it satisfied by the condition in which the state $\tilde{A}_q$ appears. Therefore, the membership functions are divided into several levels, such as μ_q having p_q levels. This is the membership function of the first order and can be denoted by $\mu^1_{A'_j}(x_i)$.

Statement 2. The membership function of the second order can be structured by the composition by several membership function of the first order:

$$\mu^2(x_i) = \sum_{i=d_1}^{d_n} w_i \mu(x_i) \qquad (14)$$

where w_i is a weight factor whose value depends on the degree of the membership between x_i and A_j, and sum of w_i equal one.

$$\sum_{i=d}^{g} p_i = 1 \qquad (15)$$

It is obvious that the membership functions of the second order have some similar characteristics with the membership functions of the first order, i.e. corresponding to the *qth* group of subsets

$(A_{q1}, ..., A_{qp_q})$ $(1 \le q \le k)$.

There are relationships given by

IF $f_{qp_q-1} \le \mu^2{}_{qp_q}(....) < 1$ THEN $x_i \subset \tilde{A}_{qp_q}$ (16)

IF $f_{qp_q-2} \le \mu^2{}_{qp_q-1}(....) \le f_{qp_q-1}$ THEN $x_i \subset \tilde{A}_{qPq-1}$ (17)

IF $0 \le \mu^2{}_{q1}(.....) \le f_{q1}$ THEN $x_i \subset \tilde{A}_{q1}$ (18)

where x represent x_i , $i= d_1,\dots, g.$

Statement 3. The membership function of the third order can be structured by the composition of several membership functions of the second order:

$$\mu^3(x_i) = \sum_{i=l_1}^{l_n} w_i \mu^2(x_i) \tag{19}$$

where w_i is a weight factor whose value depends on the degree of the membership between x_i and A_j, and sum of w_i equal one.

$$\sum_{i=l}^{m} \boldsymbol{p}_i = \mathbf{1} \tag{20}$$

It is obvious that the membership functions of the second order have some similar characteristics with the membership functions of the first order, i.e. corresponding to the *qth* group of subsets

$$(A_{q1},\dots\dots, A_{qp_q}) \quad (1 \le q \le k).$$

There are relationships given by

IF $f_{qp_q-1} \le \mu^3{}_{qp_q}(....) < 1$ THEN $x_i \subset \tilde{A}_{qp_q}$ (21)

IF $f_{qp_q-2} \le \mu^3{}_{qp_q-1}(....) \le f_{qp_q-1}$ THEN $x_i \subset \tilde{A}_{qPq-1}$ (22)

IF $0 \le \mu^3{}_{q1}(.....) \le f_{q1}$ THEN $x_i \subset \tilde{A}_{q1}$ (23)

where x represent x_i , $i= l_1,\dots,l_n, m.$

Statement 4. The membership function of the *n-1* order are structured by the composition of several membership functions of the (*n-2*) order:

$$\mu^{(n-1)}(x_i) = \sum_{i=r_1}^{r_n} w_i \mu^{(n-2)}(x_i) \tag{24}$$

where w_i is a weight factor whose value depends on the degree of the membership between x_i and A_j, and sum of w_i equal one.

$$\sum_{i=r_1}^{t} \boldsymbol{p}_i = \mathbf{1} \tag{25}$$

It is obvious that the membership functions of the (n-1) order have some similar characteristics with the membership functions of the (n) order, i.e. corresponding to the *qth* group of subsets

$$(A_{q1},\dots\dots, A_{qp_q}) \quad (1 \le q \le k) \tag{26}$$

Then, there are relationships given by

IF $f_{qp_q-1} \le \mu^{(n-1)}{}_{qp_q}(....) < 1$ THEN $x_i \subset \tilde{A}_{qp_q}$ (27)

IF $f_{qp_q-2} \le \mu^{(n-1)}{}_{qp_q-1}(....) \le f_{qp_q-1}$ THEN $x_i \subset \tilde{A}_{qPq-1}$ (28)

IF $0 \le \mu^{(n-1)}{}_{q1}(.....) \le f_{q1}$ THEN $x_i \subset \tilde{A}_{q1}$ (29)

where x represent x_i , $i= r_1,\dots..,r_n,t.$

Statement 5. The membership function of the *n* order are structured by the composition of several membership functions of the (*n-1*) order:

$$\mu^{(n)}(x_i) = \sum_{i=v_1}^{v_n} w_i \mu^{(n-1)}(x_i) \quad (30$$

where w_i is a weight factor whose value depends on the degree of the membership between x_i and A_j, and sum of w_i equal one.

$$\sum_{i=v_1}^{w} p_i = 1 \tag{31}$$

It is obvious that the membership functions of the (n-1) order have some similar characteristics with the membership functions of the (n) order, i.e. corresponding to the *qth* group of subsets

$$(A_{q1},.....,A_{qp_q}) \quad (1 \le q \le k).$$

Then, there are relationships given by

IF $f_{qp_q-1} \le \mu^{(n)}{}_{qp_q}(....) < 1$ THEN $x_i \subset \tilde{A}_{qp_q}$ (32)

IF $f_{qp_q-2} \le \mu^{(n)}{}_{qp_q-1}(....) \le f_{qp_q-1}$ THEN $x_i \subset \tilde{A}_{qPq-1}$ (33)

IF $0 \le \mu^{(n)}{}_{q1}(.....) \le f_{q1}$ THEN $x_i \subset \tilde{A}_{q1}$ (34)

where *x* represent x_i , $i= v_1,.....,v_n,w,$

If there membership function of the (n-2) order, then exists the multilevel fuzzy function (n-1) order:

In limited range :

If $f_{qp_q-1} \le \mu^{n-1}_{qp_q}(....) < 1$ **then**

$cause \quad i$ belongs $\breve{A}_{qp_q}$. (35)

For the terms between ranges causality is classified:

If $f_{qp_q-2} \le \mu^{n-1}_{qp_q-1}(....) < f_{qp_q-1}$ **then**

$cause \quad i$ belongs $\breve{A}_{qp_q-1}$. (36)

At the end

If $f_{qp_q-1} \le \mu^{n-1}_{q1}(....) < f_{q1}$ **then**

$cause \quad i$ belongs $\breve{A}_{q1}$. (37)

5 Application

Let consider every fuzzy subset in the fuzzy set a limited sequence f_q with f cut, i.e.:

$$0 < f_{q1} < f_{q2} < ... < f_{qp_{q-1}} < 1$$

For two parameters selection, the membership function of the second order is:

$$\mu^2(x_i) = \sum_{i=d}^{g} w_i \mu^1(x_i), \quad \sum_{i-d}^{g} w_i = 1$$

and

$$\mu^2(x_i) = \mu^2(x_d,...,x_g) \; 1 \le d < g \le n$$

For instance, for two design parameters selection in a control system gain/phase margin, in regard to the *qth* group of parameter follows:

IF $f_1 \le \mu^2(x_i) \le 1$

THEN **gain** is **setting**

IF $0 \le \mu^2(x_i) \le f_1$

THEN **phase** is **setting** **(**38)

For the proper allocation of the equivalent gain/phase margin contour the following statement is structured.

IF $f_1 \le \mu^3(x_i) \le 1$ (39)

THEN **gain/phase margin is accurate.**

where

$$\mu^3(x_i = \sum_{l=1}^{h} w_i \mu^2(x_i), \quad \sum_{l=1}^{h} w_i = 1$$

and

$$\mu^3(x_i) = \mu^3(x_1,....,x_h), \quad 1 \le l < h \le m$$

6 Conclusion

In this paper theorem of the fuzzy sets for causality selection is formulated and defined. For two parameters classification the membership function of the second order and third order were derived.

The appropriate production rules are derived. If hypothesis is true, each of the condition in the rule must be true.

References

[1]Savkovic-Stevanovic J., An advances learning and discovering system, *Transaction on Information Science and Applications*,**7** (7) July, 1005-1014, 2010, 1790-0832.

[2] J.B.Savkovic--Stevanovic, Causes-effects functions, *Comput. Ecol. Eng.*,8(2) 95-101, 2012, ISSN 1452-0729.

[3] J.B.Savkovic--Stevanovic, Fuzzy theory and fuzzy systems, *Comput. Ecol. Eng.*,8(2) 47-58, 2012, ISSN 1452-0729.

[4] J.Savković-Stevanović , Causality phenomena investigation, The Inter. Symposium on Science 2.0 Expansion of Science-S2ES,Orlando, Florida U.S.A,June 29-July2,2010.

[5]Savkovic-Stevanovic J., Information and knowledge representation, *Comput. Ecol.Eng*.,vol.5, No.1,2009,35-40,2009, ISSN 1452-0729.

[6]Savković-Stevanović J., Decision support system extraction, *11th WSEAS International Conference on Mathematical and Computational methods in Science and Engineering, MACMSE09*,,ID 639-126,pp.44-49, Morgan State University, USA, November 7-9, Baltimore, 2009.

[7]Savković-Stevanović J., A knowledgeable intelligent system, *AIKED10- International Conference on Artificial Intelligence, Knowledge, Engineering and Data bases*, Cambridge, London, pages 6, Feb.20-22, 2010.

[8] J.B.Savkovic-Stevanovic, Fuzzy theory and fuzzy systems, *Comput. Ecol. Eng.,*8(2) 47-58, 2012, ISSN 1452-0729.

[9]J.B.Savkovic-Stevanovic, Causes-effects functions, *Comput. Ecol. Eng.,*8(2) 95-101, 2012, ISSN 1452-0729.

[10] Savkovic-Stevanovic, Fuzzy control clustering system, EUROSIM2013, 8th Congress on Modeling and Simulation, 10-13 Sept., Cardif, Wales,U.K.,2013.

Authors Index