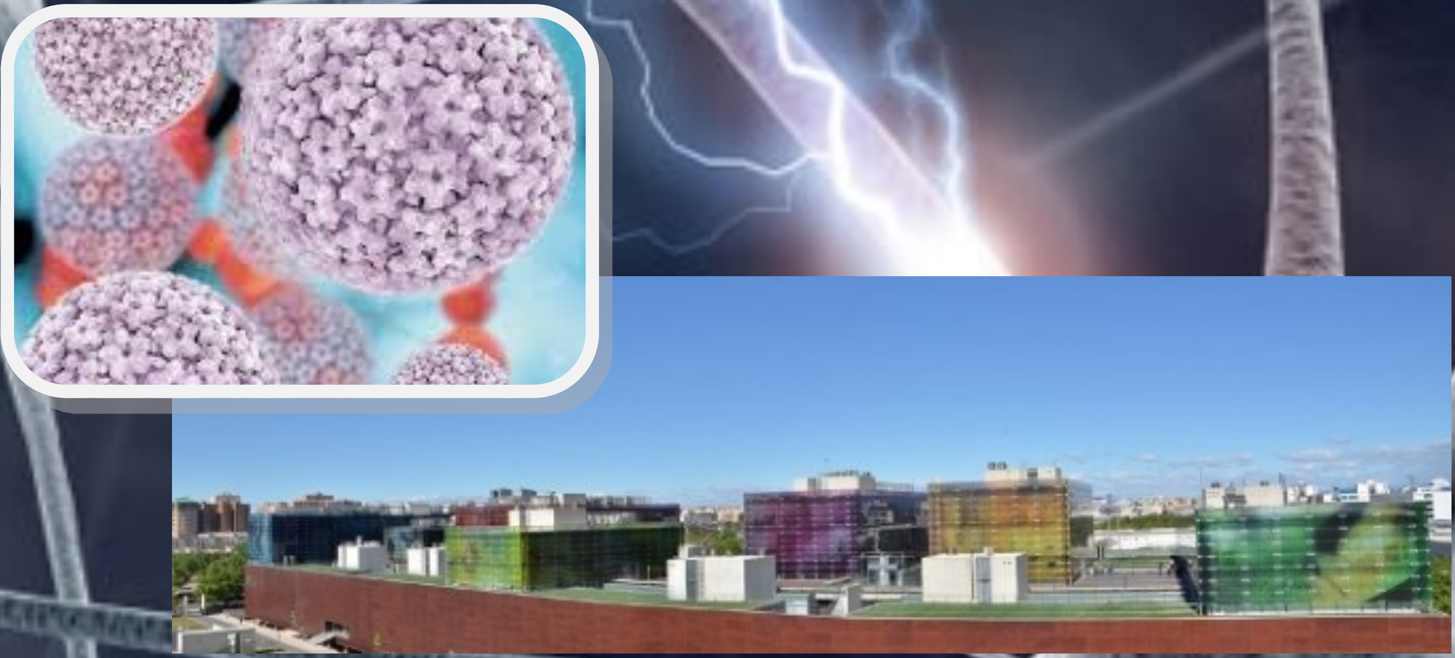
MODELLING FOR ENGINEERING AND HUMAN BEHAVIOUR 2016

Instituto Universitario de Matemática Multidisciplinar



L. Jódar, J. C. Cortés and L. Acedo(Editors)

Instituto Universitario de
Matemática Multidisciplinar

im²
Instituto de Matemática Multidisciplinar

UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Ciudad Politécnica de la Innovación

MODELLING FOR ENGINEERING, & HUMAN BEHAVIOUR 2016

Instituto Universitario de Matemática Multidisciplinar
Universitat Politècnica de València
Valencia 46022, SPAIN

Edited by
Lucas Jódar, Juan Carlos Cortés and Luis Acedo
Instituto Universitario de Matemática Multidisciplinar
Universitat Politècnica de València
I.S.B.N.: 978-84-617-7205-6

CONTENTS

HPV transmission and control in lifetime sexual partner networks

R. J. Villanueva[♭], V. Sánchez-Alonso[♭], L. Acedo[♭] [*],
J. A. Moraño[♭], and J. Díez-Domingo[†]

(♭) Instituto de Matemática Multidisciplinar

Universitat Politècnica de València,

Building 8G (2° Floor)

Camino de Vera, 46022 Valencia (Spain),

(†)Centro Superior de Investigación en Salud Pública (CSISP),

Valencia, Spain.

November 30, 2016

1 Introduction

The structure and properties of networks of sexual contacts in human populations is a topic of key interest in connection with the spread of sexually transmitted diseases (STD). However, this problem has received scarce attention and the modelling of STD epidemiology is usually based upon theoretical proposals in terms of the network structure usually unvalidated. The goal of this paper is to provide a method to build a reasonable network structure from statistical data from the Health and Sexual Habits Survey in Spain. In particular, we seek to satisfy the constraints imposed by the distribution of the number of partners for both males and females. We show that such a network can be obtained by a matching method of the bipartite graph of males and females which takes into account the preassigned degree of connectivity. In order to perform the pairing we apply the principle of psychological

[*] e-mail: luiacrod@imm.upv.es

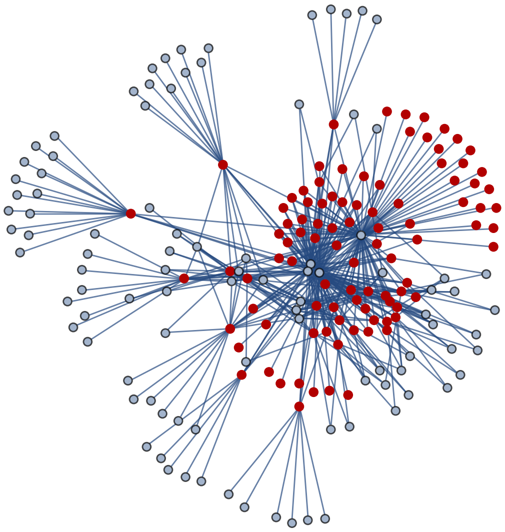


Figure 1: Lifetime sexual network for 200 individuals generated by using the greedy algorithm[13] and the statistical data for Spain [9, 10].

similarity by considering that people with a given tendency to acquire a certain number of partners tend to form relationships with other people with the same habits. This quantity is measured by a distance function $\pi(x, y)$. The method is applied to infer the structure of networks with up to $50,000$ people, which is larger than any other one analyzed in previous field studies.

Sexually transmitted diseases have been a major public health threat for a long time in human history. Modern concerns about STD began with the pandemic of syphilis which spread over Europe in the early sixteenth century. Nowadays, syphilis still affects twelve million people all around the world every year, causing 113,000 deaths in 2010 [1]. Gonorrhea spreads at a rate of AIDS 88 million cases each year [2], while human papillomavirus is thought to be the direct cause of 561,200 new cervical cancer cases only in 2002 [3]. The global pandemic of caused by the lentivirus HIV is perhaps the most acute and widespread in human history since it has already caused 36 million deaths worldwide and it has a pool of 35.3 million people infected by HIV in 2012.

This kind of diseases are more likely to produce large-scale pandemics than other transmissible diseases, respiratory or other, because the efficacy of sexual contacts for the infection is large and the infectious agent has long latency periods as in the case of HIV. Consequently, nor the carrier neither his/her partner is not aware of their exposure to it. For example, it has been estimated that around 40-50 % contacts are capable of transmitting HPV [5]. Moreover, some STDs are caused by oncoviruses such as Hepatitis B or HPV which increase the death rate of people who develop the disease.

2 Network building

In order to understand the evolution of these diseases we need a reliable model of the social network substrate in which the pandemic builds up. Individuals who change partners or have several partners simultaneously are the hubs favouring the spread of STD, but the distribution of degrees of the nodes in the network and the average chemical path from an infected individual to a susceptible one, are important parameters controlling the final extension of a new STD in a population and the speed of its spread. However, most models are based upon some assumptions which could not be valid for certain populations. Some studies claimed that the web of human sexual contacts is a scale-free network characterized by a power-law distribution for the number of individuals with a certain degree of connectivity, k: $P(k) \propto 1/k^{\alpha}$ with a value of α in the range $2 < \alpha < 3$, and slightly smaller for males than for females [6]. Although, $P(k)$ provides some valuable information about the network, it is not a sufficient prescription on how to build it for a given population size. Moreover, a power-law distribution of contacts could not be representative of some populations or vary from country to country. In particular, in the Jefferson High School's network it has been found that a densely connected core appears without the need of a high connectivity degree [7].

Some field studies have ascertained the structure of moderate size real networks of sexual contacts: In 2004, Bearman et al. published the results for a set of 800 adolescents in a midsized town of the United States [7]. They showed that a big cluster with a ring and extended filaments contained most of the adolescents implying that, potentially, the infection of an individual could propagate to the whole of the population given sufficient time and infectivity. A similar study was performed in 2007 at the Likoma Island in Malawi with the idea of predicting and explaining the expansion of HIV in sub-saharian populations [8]. That study disclosed that the sexual network contained many cycles, in contrast with the single cycle at Jefferson High

School. For that reason, it was speculated that superimposed cycles could be the explanation of the high prevalence of HIV infection in small populations of Africa.

In this work we propose a method to build a network of sexual contacts derived from real statistical data for the distribution of the number of partners without any assumption about its global structure. For the time being, only small networks of sexual contacts have been ascertained in detail from real data and, consequently, our method fills the gap between the small communities of individuals and the large scale. Validation of the resulting structure is a problem that cannot be solved directly but our method provides a substrate in which simulate sexually transmitted diseases and obtain some indirect evidence of its validity.

In our model we consider a population of N individuals, M males and F females, $N = M + F$. Let LSP_i be the number of lifetime sexual partners for node i. The challenge in building the sexual network is to find an efficient method to assign the edges to the graph in such a way that those incident to the male nodes match those incident with the edges attached to the female nodes.

In mathematical terms:

$$\sum_{i=1}^{M} LSP_i = \sum_{j=1}^{F} LSP_j \ . \tag{1}$$

For example, some authors say that in USA, the median number of female LSP between 15-44 years old in the period 2006-2008 is 3.2 [11, 12]. Analogously, the median number of male LSP between 15-44 years old in the period 2006-2008 is 5.1. Ideally, if we multiply the number of males between 15-44 years old by 5.1, the result should be approximately the number of females between 15-44 years old by 3.2. Taking into account that males and females are around 50% of the total population, the ideal situation becomes difficult to match. Sociologists say that males tend to overestimate the number of their sexual partners and females tend to underestimate it. Therefore, if we want to build a network where the number of male and female LSP coincide, we will have to make a decision about the average number of male LSP and to assign LSP to females in such a way that the sum of LSP of males and females coincide, or vice versa. Thus, we decide to consider the average LSP male value, $\langle k \rangle_m$, and build the network from it.

A typical network for 200 individuals is shown in Fig. 1. In a second phase we have used these networks to propagate the propagation of HPV and its control through vaccination programmes.

References

[1] R. Lozano et al., Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010, The Lancet 380 (2012) 2095-2128.

[2] World Health Organization, Emergence of multi-drug resistant *Neisseria Gonorrhoeae*. Threat of global rise in untreatable sexually transmitted infections. WHO/RHR/11.14, 2011.

[3] D. M. Parkin, The global health burden of infection-associated cancers in the year 2002, Int. J. Cancer 118 (12) (2006) 303044.

[4] UNAIDS factsheet, `http://www.unaids.org/en/resources/campaigns/globalreport2013/factsheet` (Accessed on December, 12, 2014).

[5] A. Burchell, H. Richardson, S. M. Mahmud, H. Trottier, P. P. Tellier, J. Hanley, F. Coutlée, E. L. Franco, Modeling the Sexual Transmissibility of Human Papillomavirus Infection using Stochastic Computer Simulation and Empirical Data from a Cohort Study of Young Women in Montreal, Canada, American Journal of Epidemology 169 (3) (2006) 534543.

[6] F. Liljeros, C. R. Edling, L. A. Nunes, H. E. Stanley, Y. Åberg, The web of human sexual contacts, Nature 411 (2001) 907-908.

[7] P. S. Bearman, J. Moody, K. Stovel, Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks, American Journal of Sociology, 110(1) (2004) 44-91.

[8] S. Helleringer, H. P. Kohler, Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi, AIDS 21 (2007) 2323-2332.

[9] Valentian Institute of Statistics, `http://www.ive.es`

[10] Encuesta de salud y hábitos sexuales 2003. Instituto Nacional de Estadstica, http://www.ine.es

[11] A. Chandra, W. D. Mosher, C. Copen, Sexual Behavior, Sexual Attraction, and Sexual Identity in the United States: Data From the 20062008. National Survey of Family Growth, National Health Statistic Reports, No. 36, March 3, 2011. http://www.cdc.gov/nchs/data/nhsr/nhsr036.pdf

[12] Key Statistics from the National Survey of Family Growth, http://www.cdc.gov/nchs/nsfg/key_statistics/n.htm

[13] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to Algorithms, MIT Press and McGraw-Hill, 1990.

A Soft Set Approach for Multiple Attributes Decision Making Problems under Incomplete Information

José Carlos R. Alcantud[♭][†] [*]and Gustavo Santos-García[†]

(♭) BORDA Research Unit, University of Salamanca, Spain,

(†) Facultad de Economía y Empresa and IME, University of Salamanca, Spain,

Campus Unamuno. E37007 Salamanca, Spain.

November 30, 2016

1 Introduction

Decisions on real life problems usually depend on vague, imprecise, uncertain or unknown data. In this paper we apply the theory of soft sets to solve a decision-making problem with incomplete information. We propose an algorithm that efficiently evaluates problems without a priori assumptions about the distribution of incomplete data. A well-founded solution is obtained by a combinatorial analysis of unknown data in each alternative. We provide performance analyses and evaluations of the proposed algorithm, which we compare with earlier solutions in the literature. In this context, a practical application illustrates the detailed implementation process of our approach and demonstrates its potential applications in decision-making problems with incomplete information.

[*] e-mail: jcr@usal.es

2 Mathematical Methods

In this paper we address the soft set based decision making problem under incomplete information as addressed by Han *et al.* [10], Qin *et al.* [15] and Zou and Xiao [20], from a different mathematical approach.

Many problems require the use of imprecise or uncertain data in real-life situations. To the purpose of their analysis, applications of soft computing techniques capable of capturing these features should be made. This possibility was boosted by the theory of fuzzy sets, which meant a paradigmatic change in Mathematics by allowing partial membership.

Since the seminal Zadeh [19], there is a vast literature on fuzzy sets and a large number of successful generalizations (v., [4] for notions and relationships). Among these extensions we follow the Molodtsov approach of soft set theory [14] given the extensive arguments about its applicability to several fields. Further relevant references include Maji *et al.* [13], Aktaş and Çağman [1]. Concerning generalizations, Maji, Biswas and Roy [11] introduce fuzzy soft sets (v., [2, 3] for decision making criteria in this model), and Wang, Li and Chen [16] introduce hesitant fuzzy soft sets.

Maji, Biswas and Roy [12] were the pioneers of soft set based decision making theory. They established the criterion that in order for an object to be selected it must maximize the choice value of the problem. In this sense, medical diagnosis constitutes a successful field of application (cf., e.g., [5, 6, 7, 18]). Thanks to its technical versatility, soft set theory has also been extensively applied to decision support in many fields such as engineering and economics (v., [9]).

Relatedly, Zou and Xiao [20] argued that in the process of collecting data there may be unknown, missing or inexistent data. Therefore standard soft sets under incomplete information must be investigated, which suggests the notion of incomplete soft sets (v., Han *et al.* [10] and Qin *et al.* [15] for additional analyses). An extension of this notion to incomplete fuzzy soft sets was developed by Deng and Wang [9] who predict unknown data in fuzzy soft sets. While a shortcoming of this paper was presented by Yang *et al.* [17], this problem was subsequently clarified by Deng and Chen [8].

3 Novelty of the Approach

The works of Han *et al.* [10], Qin *et al.* [15] and Zou and Xiao [20] present interesting approaches to incomplete soft set based decision making. These authors used averages, probabilities or any other specific evaluations in order to estimate the real value of missing data in a general way and afterwards, they made a decision based on the complementary data.

In this paper we contribute to decision making in the context of incomplete soft sets from an altogether different perspective. Rather than filling the incomplete data table, we propose a combinatorial study through all possible filled tables that can be produced from the original incomplete table. Then the alternatives are ranked by the proportion of filled tables where they achieve the highest choice value. In other words, a final indicator for each of the objects by our algorithm is defined as the value of this ratio. And our decision making procedure consists of selecting alternatives that maximize this indicator.

Our proposal is justified by a classical Laplacian argument from probability theory. In general there is perfect uncertainty about the real value of missing data. Therefore, we cannot support the idea that other aspects would let us faithfully estimate these unknown values. Under Laplace's principle of indifference, due to our complete ignorance we are entitled to assume that all possible tables where the missing data are replaced with either 0 or 1 are equiprobable.

A real application illustrates the detailed implementation process of our approach and shows evidence of its potential applications in decision-making problems with incomplete information.

On the other hand, we show differences with other algorithms. The proposed approach is quantitatively compared with other approaches in existing literature. Thus we show that this algorithm achieves decision values that are indeed different from the existing methods.

Finally, we include performance analyses and evaluations of the proposed algorithm. The effectiveness of the proposal is verified by many examples with increasing numbers of missing values.

References

[1] H. Aktaş and N. Çağman. Soft sets and soft groups. *Inf Sci*, 177:2726–35, 2007.

[2] J. C. R. Alcantud. Fuzzy soft set decision making algorithms: some clarifications and reinterpretations. In O. L. et al., editor, *Advances in Artificial Intelligence. 17th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2016*, pages 479–488. Springer-Verlag, 2016.

[3] J. C. R. Alcantud. A novel algorithm for fuzzy soft set based decision making from multiobserver input parameter data set. *Information Fusion*, 29:142–148, 2016.

[4] J. C. R. Alcantud. Some formal relationships among soft sets, fuzzy sets, and their extensions. *Int J Approx Reason*, 68:45–53, 2016.

[5] J. C. R. Alcantud, G. Santos-García, and E. H. Galilea. Glaucoma diagnosis: A soft set based decision making procedure. In J. M. Puerta, J. A. Gámez, B. Dorronsoro, E. Barrenechea, A. Troncoso, B. Baruque, and M. Galar, editors, *Advances in Artificial Intelligence*, volume 9422 of *LNCS*, pages 49–60. Springer, 2015.

[6] Y. Çelik and S. Yamak. Fuzzy soft set theory applied to medical diagnosis using fuzzy arithmetic operations. *J Inequal Appl*, 2013(1):82, 2013.

[7] B. Chetia and P. K. Das. An application of interval valued fuzzy soft sets in medical diagnosis. *Int J Contemp Math Sci*, 38(5):1887–94, 2010.

[8] T. Deng and Y. Chen. Comments from the author of "an object-parameter approach to predicting unknown data in incomplete fuzzy soft sets" [appl math model 37 (2013) 4139–4146]. *Appl Math Model*, 39(23–24):7744–7745, 2015.

[9] T. Deng and X. Wang. An object-parameter approach to predicting unknown data in incomplete fuzzy soft sets. *Appl Math Model*, 37(6):4139–4146, 2013.

[10] B.-H. Han, Y. Li, J. Liu, S. Geng, and H. Li. Elicitation criterions for restricted intersection of two incomplete soft sets. *Knowl-Based Syst*, 59:121–131, 2014.

[11] P. Maji, R. Biswas, and A. Roy. Fuzzy soft sets. *J Fuzzy Math*, 9:589–602, 2001.

[12] P. Maji, R. Biswas, and A. Roy. An application of soft sets in a decision making problem. *Comput Math Appl*, 44:1077–83, 2002.

[13] P. Maji, R. Biswas, and A. Roy. Soft set theory. *Comput Math Appl*, 45:555–62, 2003.

[14] D. Molodtsov. Soft set theory - first results. *Comput Math Appl*, 37:19–31, 1999.

[15] H. Qin, X. Ma, T. Herawan, and J. Zain. Data filling approach of soft sets under incomplete information. In N. Nguyen, C.-G. Kim, and A. Janiak, editors, *Intelligent Information and Database Systems*, volume 6592 of *LNCS*, pages 302–11. Springer, 2011.

[16] F. Wang, X. Li, and X. Chen. Hesitant fuzzy soft set and its applications in multicriteria decision making. *J Appl Math*, article ID 643785, 2014.

[17] Y. Yang, J. Song, and X. Peng. Comments on “An object-parameter approach to predicting unknown data in incomplete fuzzy soft sets” [Appl. Math. Modell. 37 (2013) 4139–4146]. *Appl Math Model*, 39(23–24):7746–7748, 2015.

[18] S. Yuksel, T. Dizman, G. Yildizdan, and U. Sert. Application of soft sets to diagnose the prostate cancer risk. *J Inequal Appl*, 2013(1), 2013.

[19] L. Zadeh. Fuzzy sets. *Inf Control*, 8:338–53, 1965.

[20] Y. Zou and Z. Xiao. Data analysis approaches of soft sets under incomplete information. *Knowl-Based Syst*, 21(8):941–5, 2008.

The generalized inverses of tensors and an application to linear models

Hongwei Jin[*] Minru Bai[†] Julio Benítez[‡] Xiaoji Liu[§]

November 30, 2016

1 Introduction

In this paper, we recall and extend some tensor operations. Then, the generalized inverse of tensors is established by using tensor equations. Moreover, we investigate the least-squares solutions of tensor equations. An algorithm to compute the Moore-Penrose inverse of an arbitrary tensor is constructed. Finally, we apply the obtained results to higher order Gauss-Markov theorem.

It is a well known definition that the Moore-Penrose inverse (see e.g. [1]) of a matrix $A \in \mathbb{C}^{m\times n}$ is a matrix $X \in \mathbb{C}^{n\times m}$ which satisfies

$$(1)\ AXA = A \qquad (2)\ XAX = X \qquad (3)\ (AX)^* = AX \qquad (4)\ (XA)^* = XA.$$

The Moore-Penrose inverse of A is unique and it is denoted by $A^\dagger$.

Operations with tensors, or multiway arrays, have become increasingly prevalent in recent years. A tensor can be regarded as a multidimensional array of data [2], which takes the form

$$\mathcal{A} = (a_{i_1\dots i_m}) \in \mathbb{R}^{n_1\times n_2\times\cdots\times n_p}. \tag{1.1}$$

Kilmer et al. [3] explore an alternate representation based on matrix slices and the functions unfold$(\cdot)$ and fold$(\cdot)$ on the third-order tensor, which permits to define several concepts (tensor transpose, inverse, and identity, especially the multiplication of tensors). The multiplication of tensors is a framework for tensor operations, which also leads to the notion of orthogonal tensors, norm of a tensor and factorizations of tensors. Later, Kilmer et al. [4] extended these results in [3] to p order tensors.

Now, a question is natural. Can we extend the Moore-Penrose inverse of matrices to tensors? By using the definitions given in [3, 4] we will see that the answer to the aforementioned question is "yes".

[*] College of Mathematics and Econometrics, Hunan University, 410082, Changsha, P.R. China and Departamento de Matemática Aplicada, Instituto de Matemática Multidisciplinar, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022, Valencia, Spain. *E-mail address*: hw-jin@hotmail.com.

[†] College of Mathematics and Econometrics, Hunan University, 410082, Changsha, P.R. China. *E-mail address*: minru-bai@163.com.

[‡] Corresponding author. Departamento de Matemática Aplicada, Instituto de Matemática Multidisciplinar, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022, Valencia, Spain. *E-mail address*: jbenitez@mat.upv.es.

[§] Faculty of Science, Guangxi University for Nationalities, 530006, Nanning, P.R. China. *E-mail address*: xiaojiliu72@126.com.

2 The Generalized Inverse of Tensors

Definition 2.1 *[4] Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times l \times n_3 \times \cdots \times n_p}$. Then the t-**product** $\mathcal{A} * \mathcal{B}$ is the $n_1 \times l \times n_3 \times \cdots \times n_p$ order-p tensor ($p \geq 3$) defined recursively as*

$$\mathcal{A} * \mathcal{B} = \mathrm{fold}(\mathrm{circ}(\mathrm{unfold}(\mathcal{A})) * \mathrm{unfold}(\mathcal{B})). \tag{2.1}$$

The t-product of two tensors presented above allows us to obtain the Moore-Penrose inverse of an arbitrary tensor $\mathcal{A}$.

Definition 2.2 *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$. If there exists a tensor $\mathcal{X} \in \mathbb{R}^{n_2 \times n_1 \times n_3 \times \cdots \times n_p}$ such that*

$$(1)\ \mathcal{A}*\mathcal{X}*\mathcal{A} = \mathcal{A} \qquad (2)\ \mathcal{X}*\mathcal{A}*\mathcal{X} = \mathcal{X} \qquad (3)\ (\mathcal{A}*\mathcal{X})^T = \mathcal{A}*\mathcal{X} \qquad (4)\ (\mathcal{X}*\mathcal{A})^T = \mathcal{X}*\mathcal{A}, \tag{2.2}$$

*then $\mathcal{X}$ is called the **Moore-Penrose inverse** of the tensor $\mathcal{A}$ and is denoted by $\mathcal{A}^\dagger$.*

In the following, we will show the existence and uniqueness of the Moore-Penrose inverse of a tensor $\mathcal{A}$.

Theorem 2.1 *The Moore-Penrose inverse of an arbitrary tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$ exists and is unique.*

The following lemma is proved in [4, Theorem 4.1] and called T-SVD of a tensor.

Lemma 2.1 *[4, Theorem 4.1] Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$. Then $\mathcal{A}$ can be decomposed as*

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T \tag{2.3}$$

where $\mathcal{U}$, $\mathcal{V}$ are orthogonal $n_1 \times n_1 \times n_3 \times \cdots \times n_p$, $n_2 \times n_2 \times n_3 \times \cdots \times n_p$ tensors, respectively, and $\mathcal{S}$ is an $n_1 \times n_2 \times \cdots \times n_p$ f-diagonal tensor.

By using Lemma 2.3, the following is straightforward.

Corollary 2.1 *Let $\mathcal{A}$ be a tensor and factorized as $\mathcal{A} = \mathcal{U}*\mathcal{S}*\mathcal{V}^T$, where $\mathcal{U}$, $\mathcal{V}$ are orthogonal tensors and $\mathcal{S} = (s_{i_1 \ldots i_p})$ is f-diagonal tensor. Then,*

$$\mathcal{A}^\dagger = \mathcal{V} * \mathcal{S}^\dagger * \mathcal{U}^T.$$

In the following, we will state some properties of the Moore-Penrose inverse of tensors and some representations of $\{1\}$-inverses, $\{1,3\}$-inverses and $\{1,4\}$-inverses of tensors.

Theorem 2.2 *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$. Then, the following statements are true:*

(a) $(\mathcal{A}^\dagger)^\dagger = \mathcal{A}$.

(b) $(\mathcal{A}^T)^\dagger = (\mathcal{A}^\dagger)^T$.

(c) $(\mathcal{A} * \mathcal{A}^T)^\dagger = (\mathcal{A}^T)^\dagger * \mathcal{A}^\dagger, \quad (\mathcal{A}^T * \mathcal{A} * \mathcal{A}^T)^\dagger = (\mathcal{A}^T)^\dagger * \mathcal{A}^\dagger * (\mathcal{A}^T)^\dagger$.

(d) $\mathcal{A}^\dagger = \mathcal{A}^T * (\mathcal{A} * \mathcal{A}^T)^\dagger = (\mathcal{A}^T * \mathcal{A})^\dagger * \mathcal{A}^T$.

(e) $\mathcal{X} \in \mathcal{A}^T\{1\}$ *if and only if* $\mathcal{X}^T \in \mathcal{A}\{1\}$.

Corollary 2.2 *Let* $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$, $\mathcal{A}^{(1)} \in \mathcal{A}\{1\}$, $\mathcal{A}^{(1,3)} \in \mathcal{A}\{1,3\}$ *and* $\mathcal{A}^{(1,4)} \in \mathcal{A}\{1,4\}$. *Then, the following statements are true:*

(a) $\mathcal{A}\{1\} = \{\mathcal{A}^{(1)} + \mathcal{Z} - \mathcal{A}^{(1)} * \mathcal{A} * \mathcal{Z} * \mathcal{A} * \mathcal{A}^{(1)} : \quad \mathcal{Z} \in \mathbb{R}^{n_2 \times n_1 \times n_3 \times \ldots \times n_p}\}$.

(b) $\mathcal{A}\{1,3\} = \{\mathcal{A}^{(1,3)} + (\mathcal{I} - \mathcal{A}^{(1,3)} * \mathcal{A}) * \mathcal{Z} : \quad \mathcal{Z} \in \mathbb{R}^{n_2 \times n_1 \times n_3 \times \cdots \times n_p}\}$.

(c) $\mathcal{A} * \mathcal{A}^{(1,3)} = \mathcal{A} * \mathcal{A}^{\dagger}$.

(d) $\mathcal{A}\{1,4\} = \{\mathcal{A}^{(1,4)} + \mathcal{Z} * (\mathcal{I} - \mathcal{A} * \mathcal{A}^{(1,4)}) : \quad \mathcal{Z} \in \mathbb{R}^{n_2 \times n_1 \times n_3 \times \cdots \times n_p}\}$.

(e) $\mathcal{A}^{(1,4)} * \mathcal{A} = \mathcal{A}^{\dagger} * \mathcal{A}$.

(f) $\mathcal{A}^{\dagger} = \mathcal{A}^{(1,4)} * \mathcal{A} * \mathcal{A}^{(1,3)}$.

3 The Least-squares Solutions of Tensor Equations

In this section, we will study the least-squares solution of the tensor equation.

Theorem 3.1 *Let* $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$, $\mathcal{X}_0 \in \mathbb{R}^{n_2 \times 1 \times n_3 \times \cdots \times n_p}$, $\mathcal{B} \in \mathbb{R}^{n_1 \times 1 \times n_3 \times \cdots \times n_p}$. *Let* $\mathcal{A}^{(1,3)}$ *be an arbitrary element of* $\mathcal{A}\{1,3\}$. *Then* $\mathcal{X}_0$ *is a least-squares solution of* $\mathcal{A} * \mathcal{X} = \mathcal{B}$ *if and only if*

$$\mathcal{A} * \mathcal{X}_0 = \mathcal{A} * \mathcal{A}^{(1,3)} * \mathcal{B}.$$

Theorem 3.2 *Let* $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$, $\mathcal{G} \in \mathcal{A}\{1\}$, $\mathbb{H} = \{\mathcal{A} * \mathcal{Z} | \mathcal{Z} \in \mathbb{R}^{n_2 \times 1 \times n_3 \times \cdots \times n_p}\}$. *Then, for all* $\mathcal{B} \in \mathbb{H}$, $\mathcal{X}_0 = \mathcal{G} * \mathcal{B}$ *is the minimum-norm solution of the consistent system* $\mathcal{A} * \mathcal{X} = \mathcal{B}$ *if and only if* $\mathcal{G} \in \mathcal{A}\{1,4\}$.

Theorem 3.3 *Let* $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}$, $\mathcal{G} \in \mathbb{R}^{n_2 \times n_1 \times n_3 \times \cdots \times n_p}$. *Then, for all* $\mathcal{B} \in \mathbb{R}^{n_1 \times 1 \times n_3 \times \cdots \times n_p}$, $\mathcal{X}_0 = \mathcal{G} * \mathcal{B}$ *is the minimum-norm least-squares solution of* $\mathcal{A} * \mathcal{X} = \mathcal{B}$ *if and only if* $\mathcal{G} = \mathcal{A}^{\dagger}$.

4 Computing the Moore-Penrose Inverse of a Tensor

According to Theorem 2.1, we propose the following algorithm to compute the Moore-Penrose inverse of an arbitrary tensor.

```
Algorithm 4.1: Compute the Moore-Penrose inverse of a tensor $\mathcal{A}$
  Input: $n_1 \times n_2 \times \cdots \times n_p$ tensor $\mathcal{A}$
  Output: $n_2 \times n_1 \times n_3 \times \cdots \times n_p$ tensor $\mathcal{X}$
    1. for $i = 3, \ldots, p$
           $\mathcal{D} = \mathrm{fft}(\mathcal{A}, [\,], i)$;
       end
    2. $N = n_3 n_4 \cdots n_p$
       for $i = 1, \ldots, N$
           $\mathcal{G}(:,:,i) = \mathrm{pinv}(\mathcal{D}(:,:,i))$; where $\mathrm{pinv}(\mathcal{D}(:,:,i))$ is the Moore-Penrose inverse of
           $\mathcal{D}(:,:,i)$,
       end
    3. for $i = p, \ldots, 3$
           $\mathcal{X} = \mathrm{ifft}(\mathcal{G}, [\,], i)$;
       end
```

5 Applications to Higher Order Gauss-Markov Theorem

In the following, we will state the higher order Gauss-Markov theorem for tensors.

Theorem 5.1 *Let $(\mathcal{Y}, \mathcal{X} * \mathcal{P}, \mathcal{V}^2)$ be a linear model. Suppose that the multilinear rank of $\mathcal{X}$ satisfies*

$$\big(rank(X_1), rank(X_2), \ldots, rank(X_\rho)\big) > \big(\max\{n_1, n_2\}, \max\{n_1, n_2\}, \ldots, \max\{n_1, n_2\}\big).$$

Then:

(a) The linear functional $\mathcal{D} * \mathcal{P}$ has a unique best linear unbiased estimator $\mathcal{D} * \widetilde{\mathcal{P}}$, where

$$\widetilde{\mathcal{P}} = \mathcal{X}^\dagger * (\mathcal{I} - (\mathcal{V} - \mathcal{V} * \mathcal{X}^\dagger * \mathcal{X})^\dagger * \mathcal{V})^T * \mathcal{Y}.$$

(b) $\widetilde{\mathcal{P}} \in \mathbb{K}$, where $\mathbb{K} = \{\mathcal{X}^T * \mathcal{Z} | \mathcal{Z} \in \mathbb{R}^{n_2 \times 1 \times n_3 \times \cdots \times n_p}\}$, and if $\mathcal{P}^*$ is any other linear unbiased estimators that belongs to $\mathbb{K}$, then

$$Cov(\mathcal{P}) \leq Cov(\mathcal{P}^*).$$

References

[1] A. Ben-Israel, T.N.E. Greville, Generalized Inverses: Theory and Applications, 2nd Edition, Springer Verlag, New York, 2003.

[2] H.A. Kiers, Towards a standardized notation and terminology in multiway analysis, J. Chemom. 14 (2000), pp. 105-122.

[3] M. E. Kilmer and C. D. Martin, Factorization strategies for third-order tensors, Linear Algebra Appl., 435 (2011), pp. 641-658.

[4] C. D. Martin, R. Shafer and B. LaRue, An Order-p Tensor Factorization with Applications in Imaging. SIAM J. Scientific Computing, 35 (2013), pp. 474-490.

A spectral method for the steady state of the 2 energy-group neutron diffusion equation

A. Bernal$^\flat$ *, R. Miró$^\flat$ and G. Verdú$^\flat$

($\flat$) Universitary Institute for Industrial, Radiophysical and Environmental Safety,
Universitat Politècnica de València,
Cami de Vera s/n, 46022, Valencia.

November 30, 2016

1 Introduction

The neutron diffusion equation is used to determine the neutron distribution inside nuclear reactors, which is related to power production, and consequently it is vital to evaluate the safety of nuclear reactors.

The neutron diffusion equation is a differential equation depending on time, space and energy variables[1]. The energy multigroup approach is normally used to deal with the energy dependence, and the 2 energy-group approach is the most commonly used in nuclear reactors.

Moreover, separation of variables is habitually used to simplify the solution of the equation. By means of this method, the time dependent neutron diffusion equation is transformed into an eigenvalue problem depending on spatial derivatives terms.

Finally, numerical methods are applied to this eigenvalue problem to discretize these spatial derivatives terms.These methods discretize the geometry and then use some kind of functions or polynomials to determine the final solution.

There is a great variety of methods which may be grouped into two classes depending on how they improve the accuracy of the solution. The first ones

*e-mail:abernal@iqn.upv.es

improve the accuracy by reducing the size of the mesh. The last ones enhance the accuracy by increasing the number of functions or polynomial degree. Overall, those methods increasing the number of functions have better convergence rates than those reducing the size mesh. One of the best of these methods applied to the neutron diffusion equation is the nodal collocation method[2], but it can be only applied to parallelepipeds. On the other hand, the finite element method can be applied to any mesh, but it uses the weak formulation of the neutron diffusion equation[3].

In this work, a spectral method applied to any kind of geometry is proposed, which is based on a construction of a polynomial basis from basic 3D monomials. The final solution will be a linear combination of the polynomials of this basis, which have a biorthonormal relation with the basic 3D monomials. Thus, the discretization of the equation is performed by multiplying the equation by the basic 3D monomials and integrating it in the defined volume.

The outline of the paper is as follows. Section 2 shows and analyzes the results. Section 3 highlights the major conclusions.

2 Results

The method is applied to a homogeneous reactor to validate it. This reactor is a parallelepiped of the following dimensions: 99 cm x 60 cm x 180 cm. It is composed of one material, whose cross sections and diffusion coefficients are defined in Table 1.

Table 1: Cross sections of the homogeneous reactor

D_1	D_2	Σ_{a1}	Σ_{a2}	$\nu\Sigma_{f1}$	$\nu\Sigma_{f2}$	Σ_{s12}
(cm)	(cm)	(cm^{-1})	(cm^{-1})	(cm^{-1})	(cm^{-1})	(cm^{-1})
1.28205	0.66667	0.01	0.1	0.01	0.1090176	0.075

Boundary conditions of zero flux are imposed and five eigenvalues were calculated. Eigenvalues and fluxes, which are the eigenvectors, will be evaluated to assess the method. As this reactor is homogeneous and parallelepiped, it has analytical solution, and therefore it will be the reference solution.

The following variables are used for the evaluation: eigenvalue error and flux error, which are defined in Equations 1 and 2, respectively.

$$EE(pcm) = \frac{|\mathbf{k} - \mathbf{k}_{ref}|}{\mathbf{k}_{ref}} \cdot 10^5 \tag{1}$$

$$FE_i(\%) = \frac{|\phi_i - \phi_{i_{ref}}|}{\phi_{i_{ref}}} \cdot 100 \tag{2}$$

The reference eigenvalues are calculated with Equation 3. The reference flux of the first energy group is calculated with Equation 4, in which x,y,z are coordinates defined with respect to the centroid of the parallelepiped. The flux is evaluated at the centroids of 3x3x6 nodes of the reactor. The flux of the second energy group is not evaluated, because it is proportional to that of the first energy group, so it is redundant.

$$\mathbf{k} = \frac{\nu\Sigma_{f1} + \frac{\nu\Sigma_{f2}\Sigma_{s12}}{D_2\left(\left(n_x\frac{\pi}{L_x}\right)^2+\left(n_y\frac{\pi}{L_y}\right)^2+\left(n_z\frac{\pi}{L_z}\right)^2\right)+\Sigma_{a2}}}{D_1\left(\left(n_x\frac{\pi}{L_x}\right)^2+\left(n_y\frac{\pi}{L_y}\right)^2+\left(n_z\frac{\pi}{L_z}\right)^2\right)+\Sigma_{a1}+\Sigma_{s12}} \tag{3}$$

$$\phi_1(x,y,z) = \cos\left(n_x\frac{\pi}{L_x}x\right)\cos\left(n_y\frac{\pi}{L_y}y\right)\cos\left(n_z\frac{\pi}{L_z}z\right) \tag{4}$$

The five highest eigenvalues calculated with Equation 3 are shown in Table 2, and their respective flux is shown in Figures 1-5.

The reactor was not spatial discretized, but high order plynomials were used instead. Several polynomial expansions of different orders were performed to analyze its sensitivity. In this work, results for orders 6, 8 and 10 are shown. For each of these expansions with order n, all monomials $x^iy^jz^k$, with $i+j+k \leq n$, are used. The eigenvalue errors (Equation 1) and computational time of the calculation are displayed in Table 3. Flux errors (Equation 2) are displayed for orders 6, 8 and 10 in Figures 6-8, respectively.

Table 2: Reference eigenvalues

k	n_x	n_y	n_z
0.993880	1	1	1
0.975991	1	1	2
0.947306	1	1	3
0.937644	2	1	1
0.921347	2	1	2

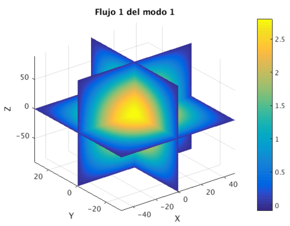


Figure 1: Flux 1 of mode 1

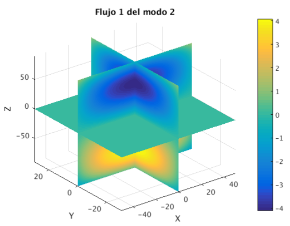


Figure 2: Flux 1 of mode 2

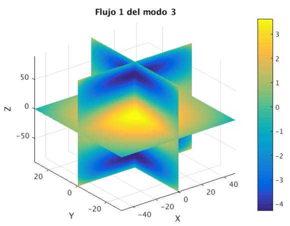


Figure 3: Flux 1 of mode 3

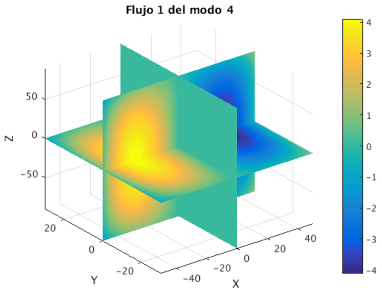


Figure 4: Flux 1 of mode 4

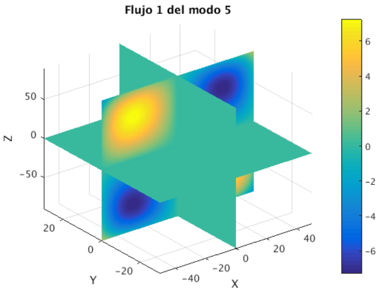


Figure 5: Flux 1 of mode 5

One can appreciate that the polynomial expansion of order 6 does not give accurate results, since eigenvalues errors are higher than 500 pcm and flux errors could be 7 %. However, results obtained with the polynomial expansion of order 8 are very accurate and those obtained with order 10 are excellent. It is important to highlight that the errors in this last case are almost zero. As regards the computational time, it should be pointed out that this calculations was performed with Matlab; so the computational will be highly reduced, if the algorithm is programed in languages such as Fortran.

Table 3: Eigenvalue results

Order	Time (s)	Matrix size	EE-1 (pcm)	EE-2 (pcm)	EE-3 (pcm)	EE-4 (pcm)	EE-5 (pcm)
6	1.4	168	24.29	4628.85	5576.71	5711.99	8734.63
8	5.9	330	0.03	105.81	354.92	134.4	321.24
10	30.2	572	0	0.8	6.51	0.22	5.07

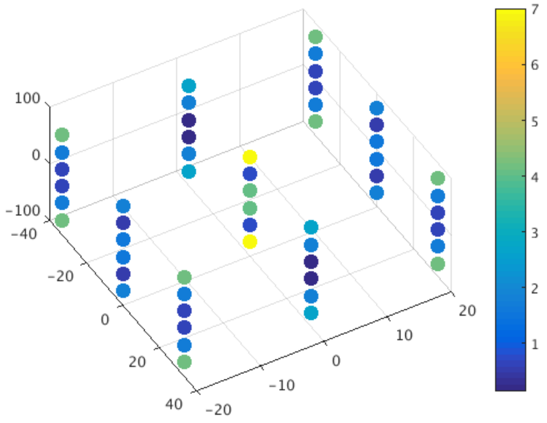


Figure 6: Flux errors for order 6

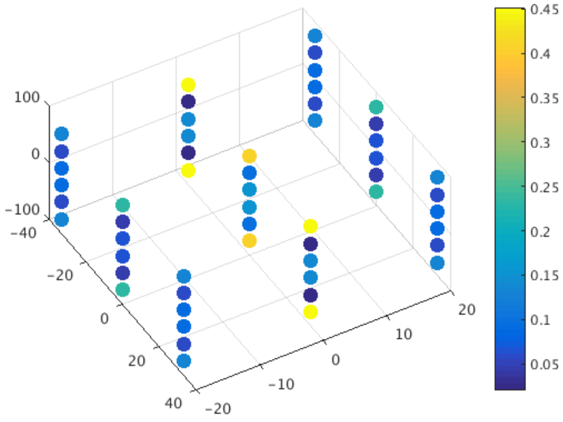


Figure 7: Flux errors for order 8

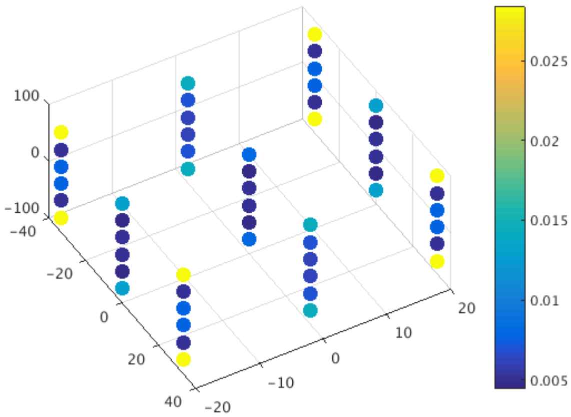


Figure 8: Flux errors for order 10

Finally, the authors did not show the results of odd polynomial orders, because they are the same as the even polynomial orders. This makes sense, since the reference solution is an even function, and consequently it does not contains odd polynomial terms.

3 Conclusions

A spectral method has been developed which can be applied to any kind of geometry.

The spectral method is based on a polynomial expansion of the neutron flux, which is orthonormal to basic monomials.

This method can be applied to any geometry without discretization. Only volume and surface integrals of the polynomials are required.The equation is discretized by multiplying by the moments.

The method has been tested in a homogeneous reactor. A sensitivity analysis of the polynomial expansion has been performed. A polynomial expansion of order 6 is required at least to obtain valid results. A polynomial expansion of order 10 gives excellent results. Polynomial expansions of odd order give the same results as the even order ones in the tested homogeneous reactor.

For future work, the method will be applied to discretized geometries. More physical applications will be added, shuch as transient state, thermal-hydraulic coupling and generic multigroup neutron diffusion equation.

References

[1] W. M. Stacey, Nuclear Reactor Physics. New York, John Wiley and Sons, 2001.

[2] A. Hébert. Development of the nodal collocation method for solving the neutron diffusion equation *Annals of Nuclear Energy*, Volume(10):527–541, 1987.

[3] A. Vidal-Ferrandiz, R. Fayez, D. Ginestar and G. Verd. Solution of the Lambda modes problem of a nuclear power reactor using an h-p finite element method *Annals of Nuclear Energy*, Volume(72):338–349, 2014.

Cocaine consumption in Spain: A mathematical modelling approach

C. Anaya[(a)] [*], C. Burgos[(a)] [†],
J.C. Cortés[(a)] [‡], R.-J. Villanueva[(a)] [§]

(a) Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València

November 30, 2016

1 Introduction

The aim to this work is capturing the data uncertainty changes in the cocaine consumption in Spain during the period $1995 - 2011$. By this way we have chosen a variation of the model started in [1], assuming that the model parameters are time-dependent functions.

In order to collect the data uncertainty the probabilistic fitting technique developed in [2] has been used. Also by this technique we have provided an estimation to the probability distribution of the model parameters.

In this study we use data given by the Survey on Alcohol and Drugs in Spain (EDADES), which is part of the Spanish National Drug Plan [3].

2 Model building

The Spanish old population ($15 - 65$ years) has been divided into four subpopulations:

[*] e-mail: anayalatr@gmail

[†] e-mail: clabursi@posgrado.upv.es

[‡] e-mail: jccortes@imm.upv.es

[§] e-mail: rjvillan@imm.upv.es

Survey dates	Non-consumer	Occasional Consumer	Regular Consumer	Habitual Consumer
$t_1 = 1995$	0.944	0.034	0.018	0.004
$t_2 = 1997$	0.948	0.032	0.015	0.005
$t_3 = 1999$	0.948	0.031	0.015	0.006
$t_4 = 2001$	0.911	0.049	0.026	0.014
$t_5 = 2003$	0.903	0.059	0.027	0.011
$t_6 = 2005$	0.884	0.070	0.030	0.016
$t_7 = 2007$	0.874	0.080	0.030	0.016
$t_8 = 2009$	0.860	0.102	0.026	0.012
$t_9 = 2011$	0.879	0.088	0.022	0.011

Table 1: Survey results. % of people who belong to non-consumers, occasional consumers, regular consumers or habitual consumers of cocaine in Spain from 1995 to 2011.

1. N_t: number of people who have never consumed cocaine at year t.

2. O_t: number of people who have consumed cocaine at least once in their lives at year t.

3. R_t: number of people who consume cocaine regularly (at least once in their past year) at year t.

4. H_t: number of people who consume cocaine habitually (at least once in their past month) at year t.

Furthermore, we have the total time-dependent population

$$T_t = N_t + O_t + R_t + H_t. \tag{1}$$

The following non-linear system of difference equations describes the cocaine consumption for the $15-65$ years old Spanish population

$$\begin{aligned}
N_{t+1} &= N_t + \mu T_t - d_N N_t - \beta_t \frac{N_t}{T_t}\left(O_t + R_t + H_t\right) + \epsilon_t H_t, &(2)\\
O_{t+1} &= O_t - d_O O_t + \beta_t \frac{N_t}{T_t}\left(O_t + R_t + H_t\right) - \gamma_t O_t, &(3)\\
R_{t+1} &= R_t - d_R R_t + \gamma_t O_t - \sigma_t R_t, &(4)\\
H_{t+1} &= H_t - d_H H_t + \sigma_t R_t - \epsilon_t H_t, &(5)
\end{aligned}$$

where

- μ is the birth rate,
- d_N is the death rate for non-consumers,
- d_O is the death rate for occasional consumers,
- d_R is the death rate for regular consumers,
- d_H is the death rate for habitual consumers,
- β_t is the time-dependent transmission rate for cocaine consumption,
- γ_t is the time-dependent transition rate between occasional and regular cocaine consumers,
- σ_t is the time-dependent transition rate between regular and habitual cocaine consumers,
- ϵ_t is the time-dependent rate at which habitual consumers enter and complete drug-treatment therapy.

3 Probabilistic analysis

This technique is referred to as probabilistic fitting and it consists of sampling data values and fit the model to the sampled data. Thus, we find adequate model parameters that best fit the data for each sample. These model parameters will allow the model to capture the data uncertainty survey error.

3.1 Data 95% confidence intervals (95% CI)

For simplicity, we will assume that the survey outputs are independent. Let us denote by $X^j = (X_1^j, X_2^j, X_3^j, X_4^j)$, $0 \leq X_i^j \leq n_j$, $i = 1, 2, 3, 4$, $j = 1, \ldots, 9$, where $n_1 = 15000$, $n_2 = 15000$, $n_3 = 15000$, $n_4 = 15000$, $n_5 = 15000$, $n_6 = 27934$, $n_7 = 23715$, $n_8 = 20109$, and $n_9 = 22128$ are the sample sizes of surveys and

- X_1^j is the non-consumer population,
- X_2^j is the occasional consumers population,

- X_3^j is regular consumers population,
- X_4^j is habitual consumers population.

Each random vector X^j follows a multinomial probability distribution,

$$P_{n_j}^j(x_1, x_2, x_3, x_4) = \frac{n_j!}{x_1!x_2!x_3!x_4!}(\theta_1^j)^{x_1}(\theta_2^j)^{x_2}(\theta_3^j)^{x_3}(\theta_4^j)^{x_4}, \quad j = 1, \dots, 9,$$

where $x_1+x_2+x_3+x_4 = n_j$ and θ_1^j, θ_2^j, θ_3^j and θ_4^j are the percentages collected in Table 1 for each survey j ($j = 1, \dots, 9$).

The Table 2 shows the quantiles 2.5 and 97.5 (95% CI) using probabilistic fitting.

Survey dates	Non-consumer	Occasional Consumer	Regular Consumer	Habitual Consumer
$t_1 = 1995$	$[0.940, 0.947]$	$[0.031, 0.037]$	$[0.016, 0.020]$	$[0.003, 0.005]$
$t_2 = 1997$	$[0.944, 0.952]$	$[0.029, 0.035]$	$[0.013, 0.017]$	$[0.004, 0.006]$
$t_3 = 1999$	$[0.944, 0.952]$	$[0.028, 0.034]$	$[0.013, 0.017]$	$[0.005, 0.007]$
$t_4 = 2001$	$[0.906, 0.916]$	$[0.046, 0.053]$	$[0.023, 0.029]$	$[0.012, 0.016]$
$t_5 = 2003$	$[0.898, 0.908]$	$[0.055, 0.063]$	$[0.024, 0.030]$	$[0.009, 0.013]$
$t_6 = 2005$	$[0.880, 0.888]$	$[0.067, 0.073]$	$[0.028, 0.032]$	$[0.015, 0.018]$
$t_7 = 2007$	$[0.870, 0.878]$	$[0.076, 0.083]$	$[0.028, 0.032]$	$[0.014, 0.018]$
$t_8 = 2009$	$[0.855, 0.865]$	$[0.098, 0.106]$	$[0.024, 0.028]$	$[0.011, 0.014]$
$t_9 = 2011$	$[0.875, 0.883]$	$[0.084, 0.092]$	$[0.020, 0.024]$	$[0.010, 0.012]$

Table 2: 95% CI of the EDADES surveys data using the joint multinomial probability function of each survey.

4 Results and Conclusions

A graphical representation of values of the Table 2 can be seen in Figure 1.

As we see in Figure 1, the model works good enough until 2009. This issue can be explained because the effect of the economic crisis and the implantation of the Spanish National Drug Plan, introduced a change in the behaviour of the individuals.

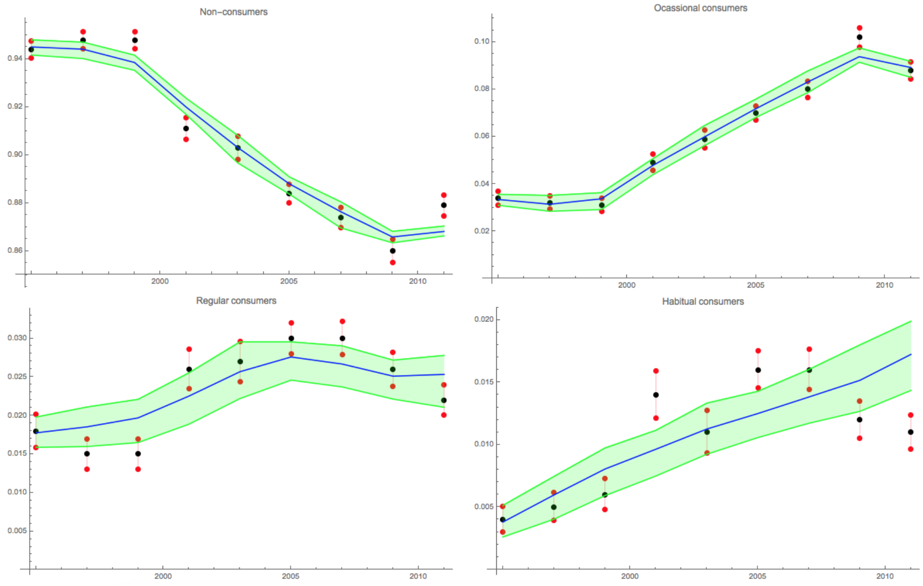


Figure 1: The green band corresponds to 95% CI model output and the blue line its mean. The red points are the 95% CI data given in Table 2 and the black points their means.

References

[1] E. Sánchez, R.-J. Villanueva, F.-J. Santonja, M. Rubio, Predicting cocaine consumption in Spain: A mathematical modelling approach. Drugs: Education Prevention and Policy, 18 (2), 108–115, 2011.

[2] Juan-Carlos Cortés, Francisco-J. Santonja, Ana-C. Tarazona, Rafael-J. Villanueva, Javier Villanueva-Oller, A probabilistic estimation and prediction technique for dynamic continuous social science models: The evolution of the attitude of the Basque Country population towards ETA as a case study, Applied Mathematics and Computation, 264, 13–20, 2015.

[3] Ministerio de Sanidad, Servicios Sociales e Igualidad (Ministry of Health, Social Services, and Equality), Observatorio Español Sobre la Droga y las Toxicomanas, Informe 2015 (Observatory in Spain Concerning Drugs and Addiction, 2015 Report).

[4] N.A. Christakis, J.H., Fowler, Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives,. Brown and Company, Hachette Book Group, 2009.

[5] INE, Spanish Statistic Institute.

[6] C. Jacob, N. Khemka, Particle Swarm Optimization in *Mathematica* An exploration kit for evolutionary optimization, IMS'04, Proc. Sixth International Mathematica Symposium, Banff, Canada, 2004.

[7] J. A. Nelder, R. Mead, A simplex method for function minimization, Computer Journal 7, 308–313, 1964.

Jordan forms of irreducible TP matrices

R. Cantó[*] and A.M. Urbano

Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, 46022 Valencia, Spain.

November 30, 2016

1 Introduction

A matrix $A \in \mathbb{R}^{n\times n}$ is called *(strictly) totally positive* if all its minors are (positive) nonnegative and it is abbreviated as (STP) TP. These classes of matrices arise in a wide variety of important applications, such that statistics, mathematical biology, combinatorics, dynamics, differential equations, approximation theory, operator theory and geometry.

The theory of totally positive matrices originated from the works of Gantmacher and Krein [1] on oscillations of vibrating systems, Schoenberg [2] about real zeros of polynomials, spline function with applications to mathematical analysis and approximation, and Karlin [3] on integral equations, kernels and statistics.

From then until now there has been a considerable amount of work accomplished on total positivity, some of which is contained in the exceptional survey paper by Ando [4]. In the last ten years, Koev [5, 6] has obtained advances in accurate eigenvalue computation and connections with singular value decompositions on TP matrices. Recently, two books on total positivity have appeared in [7, 8].

A is called *irreducible* if there is no permutation matrix P such that

$$PAP^T = \begin{bmatrix} B & C \\ O & D \end{bmatrix}$$

[*] e-mail: rcanto@mat.upv.es

where O is an $(n-r) \times r$ zero matrix $(1 \leq r \leq n-1)$.

We recall that the *rank* of A, denoted by rank(A), is the size of the largest invertible square submatrix of A. The *principal rank* of A, denoted by p-rank(A), is the size of the largest invertible principal submatrix of A. Note that the inequality $0 \leq p\text{-rank}(A) \leq \text{rank}(A) \leq n$ holds. If A is an irreducible TP matrix, then p-rank(A) is the number of nonzero eigenvalues of A. In addition, for the zero eigenvalue of A, $n-(p\text{-rank}(A))$ is its algebraic multiplicity and $n - \text{rank}(A)$ is its geometric multiplicity.

Fallat and Gekhtman [9] have proved that

$$\text{rank}\left(A^{p\text{-rank}(A)}\right) = p\text{-rank}(A),$$

that is, the maximal length of the Jordan chains associated with the zero eigenvalue of A is, at most, p-rank(A). This equality allows us to obtain the following relation between the order n of A, its rank r and its p-rank p

$$\left\lceil \frac{n}{n+1-r} \right\rceil \leq p \leq r \qquad \text{or} \qquad p \leq r \leq n - \left\lceil \frac{n-p}{p} \right\rceil .$$

We define that a triple (n, r, p) is *realizable* if there exists an irreducible TP matrix A of size $n \times n$ such that rank$(A) = r$ and p-rank$(A) = p$. In this work, from a realizable triple $(n, \text{rank}(A), p\text{-rank}(A))$, we characterize all possible Jordan canonical forms of A associated with its zero eigenvalue.

2 Main results

Given the matrices F and G, the Flanders theorem [10] proves that the difference between the Jordan blocks sizes associated with the eigenvalue zero of matrices FG and GF is -1, 0 or 1 for all blocks. In [11] we prove that this difference is always equal to 1 in the following case.

Theorem 1 *Let $C = FG \in \mathbb{R}^{n\times n}$ and $D = GF \in \mathbb{R}^{r\times r}$ be matrices with $F \in \mathbb{R}^{n\times r}$, $G \in \mathbb{R}^{r\times n}$ and rank(F) = rank$(G) = r$. Then C and D have the same elementary divisors with nonzero roots. Moreover, if $k_1 \geq k_2 \geq \cdots \geq k_p$ (resp. $k'_1 \geq k'_2 \geq \cdots \geq k'_p$) are the Jordan blocks sizes associated with the eigenvalue zero in FG (resp. GF), then $k_i - k'_i = 1$ for all i.*

We consider a realizable triple $(n, \mathrm{rank}(A), p\text{-rank}(A))$. The factorization $A = LU$ is a *LU full rank factorization* of A, if $L \in \mathbb{R}^{n\times r}$ is a TP lower echelon matrix, $U \in \mathbb{R}^{r\times n}$ is a TP upper echelon matrix and $\mathrm{rank}(L) = \mathrm{rank}(U) = r$. If we denote $A_1 = UL$, then we note that A_1 is an irreducible TP matrix with $\mathrm{rank}(A_1) \leq r$ and we prove that $p\text{-rank}(A) = p\text{-rank}(A_1) = p$. By using Theorem 1, we have $\mathrm{rank}(A^2) = \mathrm{rank}(A_1)$. In the case where A_1 is singular, we apply Theorem 1 again and we obtain a new matrix A_2 such that $\mathrm{rank}(A^3) = \mathrm{rank}(A_1^2) = \mathrm{rank}(A_2)$. Proceeding in this way, we construct a sequence of matrices $A, A_1, A_2, \ldots, A_w$ such that

$$\mathrm{rank}(A^i) = \mathrm{rank}(A_{i-1}), \qquad i = 2, 3, \ldots, w + 1$$

with A_w nonsingular. If we know the Jordan structure of A_w, and taking into account the rank of the irreducible TP matrices A_j, $j = 1, 2, \ldots, w$, we can obtain the different Jordan structures of A depending on the relations between the size, the rank and the p-rank of each matrix in this sequence. Next procedure computes this process.

Procedure:
Given the realizable triple (q_0, q_1, p), where

$$1 \leq q_1 \leq q_0 - 1, \qquad p \leq q_1 \leq q_0 - \left\lceil \frac{q_0 - p}{p} \right\rceil$$

the following steps compute all the Jordan forms associated with this triple. For $i = 1, 2, \ldots$

Step 1. Obtain the triples (q_i, q_{i+1}, p) where

$$\max\{p, 2q_i - q_{i-1}\} \leq q_{i+1} \leq q_i - \left\lceil \frac{q_i - p}{p - i} \right\rceil$$

For each q_{i+1} do

Step 2. If $q_{i+1} = p \longrightarrow$ end.

Step 3. If $q_{i+1} > p \longrightarrow$ go to Step 1.

Example: Let A be an irreducible TP matrix with the triple $(q_0, q_1, p) = (10, 7, 4)$. Then, the algebraic and geometric multiplicities of the zero eigenvalue of A are

$$\text{a.m.}(\lambda = 0)_A = q_0 - p = 10 - 4 = 6, \qquad \text{g.m.}(\lambda = 0)_A = q_0 - q_1 = 10 - 7 = 3.$$

Now, we study all possible Jordan canonical forms of A. Applying the above procedure to the triple $(10,7,4)$ we obtain the triples $(7,q_2,4)$, where q_2 satisfies the following inequalities

$$\max\{p, 2q_1 - q_0\} \leq q_2 \leq q_1 - \left\lceil \frac{q_1 - p}{p-1} \right\rceil \quad \longrightarrow \quad 4 \leq q_2 \leq 6.$$

Therefore, there are 3 possibilities:
If $q_2 = 4$ we just know the Jordan structure of A, that is, we obtain 3 Jordan chains of length 2,

$$\begin{array}{ccc} A & & A_1 \\ (10,7,4) & \longrightarrow & (7,4,4) \end{array} \qquad \Longrightarrow \qquad \begin{array}{cccc} \bullet & \bullet & \bullet & \mathrm{Ker}(A^2) \\ \downarrow & \downarrow & \downarrow & \\ \bullet & \bullet & \bullet & \mathrm{Ker}(A) \end{array}$$

If $q_2 = 5$, since it is greater that $p = 4$, we apply again the procedure and we obtain that A has 3 Jordan chains of lengths 3, 2 and 1,

$$\begin{array}{ccccc} A & & A_1 & & A_2 \\ (10,7,4) & \longrightarrow & (7,5,4) & \longrightarrow & (5,4,4) \end{array} \qquad \Longrightarrow \qquad \begin{array}{cccc} \bullet & & & \mathrm{Ker}(A^3) \\ \downarrow & & & \\ \bullet & \bullet & & \mathrm{Ker}(A^2) \\ \downarrow & \downarrow & & \\ \bullet & \bullet & \bullet & \mathrm{Ker}(A) \end{array}$$

If $q_2 = 6$, we apply the procedure three times and we obtain 3 Jordan chains of lengths 4, 1 and 1, that is

$$\begin{array}{ccccccc} A & & A_1 & & A_2 & & A_3 \\ (10,7,4) & \longrightarrow & (7,6,4) & \longrightarrow & (6,5,4) & \longrightarrow & (5,4,4) \end{array}$$

$$\begin{array}{cccc} \bullet & & & \mathrm{Ker}(A^4) \\ \downarrow & & & \\ \bullet & & & \mathrm{Ker}(A^3) \\ \downarrow & & & \\ \bullet & & & \mathrm{Ker}(A^2) \\ \downarrow & & & \\ \bullet & \bullet & \bullet & \mathrm{Ker}(A) \end{array}$$

Acknowledgements: This work has been partially supported by the Ministerio de Economia y Competitividad under MTM2013-43678-P.

References

[1] Gantmacher F. R., and Krein M. G. Sur les matrices complètement non-negatives et oscillatoires, *Comp. Math.*, Volume(4):445–476, 1937.

[2] Schoenberg I. J. Über variationsvermindernde lineare Transformationen *Math. Z.*, Volume(32):321–328, 1930.

[3] S. Karlin, Total Positivity, volume I. Stanford University Press, Stanford, California, 1968.

[4] Ando T. Totally positive matrices, *Linear Algebra and its Application*, Volume(90):165–219, 1987.

[5] Koev P. Accurate eigenvalues and SVDs of totally nonnegative matrices, *SIAM J. Matrix Anal. Appl.*, Volume(27):1–23, 2005.

[6] Koev P. Accurate computations with totally nonnegative matrices, *SIAM J. Matrix Anal. Appl.*, Volume(29):731–751, 2007.

[7] A. Pinkus, Totally Positive Matrices, volume 181 of Cambridge Tracts in Mathematics, Cambridge University Press, UK, 2010.

[8] S. M. Fallat, and C. R. Johnson, Totally Nonnegative Matrices, Princeton University Press, New Jersey, 2011.

[9] Fallat S. M., and Gekhtman M. I., Jordan Structures of Totally Nonnegative Matrices, *Canad. J. Math.*, Volume(57):82–98, 2005.

[10] Flanders H., Elementary divisors of AB and BA, *Proceedings of the American Mathematical Society*, Volume(2):871-874, 1951.

[11] Cantó R., Ricarte B., and Urbano A.M., Full rank factorization and the Flanders theorem, *Electronic Journal of Lin. Algebra*, Volume(18):352–363, 2009.

An equation to calculate the availability for the k-out-of-n system

S. Carpitella$^\flat$ [*], A. Certa$^\flat$, G. Galante$^\flat$, and J. Izquierdo†

($\flat$) Dipartimento dell'Innovazione Industriale e Digitale (DIID)
Università degli Studi di Palermo, Palermo, Italy,
($\dagger$) FluIng-Instituto Matemático Multidisciplinar
Universitat Politècnica de València, València, Spain.

November 30, 2016

1 Introduction

Reliability and availability (R&A) analyses are primary phases in management of complex systems and play a fundamental role in products and services quality [1]. Management of maintenance activities has to be based on optimisation to simultaneously improve the operative conditions of the system [2]. According to [3], maintenance optimization can be effectively pursued by the RCM (Reliability-centered maintenance) [4] approach. The authors demonstrate the global improvement of R&A by applying the RCM to the results derived by the FMECA (Failure Modes Effects and Criticality Analysis) [5]. Many efforts have been made to improve traditional methodologies of analysis, mainly tending to reduce the imprecision on data related to the occurrence of failures. [6] underlines how to investigate the behavior of a reparable system by using a combined measure of RAM (Reliability, Availability and Maintainability) [7]. R&A analyses help the analyst to improve the operative conditions to achieve higher levels of technical and economic performance. The literature presents cases [8] in which such functions are

[*] e-mail: silvia.carpitella@unipa.it

expressed in fuzzy terms, to reduce the uncertainty derived from reliability analyses. However, for evaluating R&A the components that constitute the system must be perfectly known. As highlighted by [9], a complex system could often be represented as a diagram whose blocks represent the components. In particular, simple configurations are related to series and parallel systems. To increase the overall level of reliability, it is suggested to use redundant components. By means of the redundancy, the system will stop working only if every redundant component fails. [10] underlines the important role of redundancy in both reliability and cost optimization. The approach considers the development of RAP (Redundant Allocation Problem) to find the best redundancy strategy to improve the system conditions. [11] presents a fuzzy multi-objective method to undertake the RAP development and to make the model more flexible and suitable to human decisions. Moreover, the partially redundant structure is a significant system to consider. It is also known as k-out-of-n configuration: it is constituted of n components and k of them ($k < n$) necessarily have to be operative so as not to compromise the functioning state of the system. That is, if $k + 1$ components fail the whole system will fail. For these kinds of systems, the literature presents different mathematical programming models in which functions such as costs and R&A are considered as objectives or constraints. For example, [12] observes how a k-out-of-n configuration permits to satisfy safety objectives based on the increasing of the system reliability level. [13] develops a method to quantify the unavailability for a k-out-of-n reactor protection system belonging to a nuclear power plant. This method also permits to investigate the more dangerous situations related to the entire system. Again, regarding the selection of maintenance policies, [14] highlights the cost minimization as one of the main objectives to pursue. About industrial systems with reparable components, it is known that the more interesting parameter used to drive the maintenance management is the stationary availability (SA), whose improvement is defined as a strategic objective in [15]. The aim of the present research is to investigate this parameter for a k-out-of-n system by suggesting a mathematical expression for the exact value of $A_{S(n,k)}$.

2 Proposed stationary availability equation

The availability $A_S(t)$, a fundamental parameter for reparable systems, represents the probability that, at time t, the considered system is available, i.e.

functioning, without keeping in consideration the occurrence of possible failures before t. The SA is defined, in general, as the ratio of functioning time and the total time of a general failure-reparation cycle and calculated in a wide time horizon by $A_S = \mu_S/(\mu_S + \lambda_S)$, where λ and μ are respectively the (constant) failure and repair rates of the system. We propose the following equation to calculate the exact SA $A_{S(n,k)}$ of a k-out-of-n system:

$$A_{S(n,k)} = \frac{\sum_{i=k}^{n} \binom{n}{i} \mu^i \lambda^{n-i}}{\sum_{i=k}^{n} \binom{n}{i} \mu^i \lambda^{n-i} + \binom{n}{k-1} \mu^{k-1} \lambda^{n-k+1}}, \tag{1}$$

λ and μ being the failure and repair rates of the components and $k \leq n$. Equation (1) is based on the following hypotheses: (i) all components are stochastically independent and identical from the reliability point of view, and (ii) there are not constraints about the availability of maintenance crews.

The equation is obtained by partition all the possible states (functioning or failing, mutually excluding) for system S at a generic time instant. The first component of the denominator of (1) corresponds to all functioning states while the second is the only possible failing state of the system. The ratio: functioning states over possible states gives the SA. Equation (1) may be validated by the Fundamental Theorem of Markov Chains (FTMC) [16].

3 Mathematical model

The problem considers the design of a k-out-of-n system, S, a fundamental sub-system of a generic process plant. It is supposed S is constituted by n components, identical from the reliability point of view, and k components have to work to guarantee the sub-system working. The problem is to determine the optimal number n of components maximizing the objective function $R - C$, where R expresses the trend of revenue associated to the products manufactured by sub-system S and C is the costs associated to the use of redundant components. R can be expressed as $R = A_{S(n,k)} Q_{max} r$, where Q_{max} = higher value of sub-system productivity achievable when $A_{S(n,k)} = 1$, calculating $A_{S(n,k)}$ by (1) and r = unitary revenue. C is the incremental cost due to the use of redundant components, $C = (n - k)c_u$, with c_u the unitary cost of one additional operating component, and is different from zero when the $k + 1$ component is activated. Note that C considers purchase of components and maintenance costs. There are no constraints about the availability of maintenance crews. A_S can be expressed by $A_{S(n,k)} = \sum_{m=k}^{n} A_{S(m,k)} x_m$, x_m being Boolean decisional variables. Only one, precisely the variable associated to the optimal configuration, takes a value of 1: $\sum_{m=k}^{n} x_m = 1$.

4 Numerical example

Table 1 gives the assumed input data for the model considered in Section 3.

Table 1: Input data

r	Q_{max}	c_u	λ	μ	k
0.01	1000	unitary	0.0002	0.002	2

The range of possible values for n is $[k, 5]$. Under the considered hypotheses, the availability and objective function values for the system configurations, calculated by 1, are given in Table 2.

Table 2: Availability and objective function values

n	$A_{S(n,k)}$	$R - C$
2	0.8333	8.3333
3	0.9774	8.7744
4	0.9959	7.9591
5	0.9997	6.9969

It is clear that the maximum occurs at $n = 3$. Thus, three is the number of redundant components for the k-out-of-n optimal sub-system S. These results are also shown through the graphical solution in Figure 1.

5 Conclusions

The proposed formula is a novel approach in calculating the exact value of the SA for k-out-of-n systems. It is validated through the FTMC, that is the classical approach used to calculate the SA for the considered reliability configuration. A fundamental advantage provided by the proposed equation is a lower computational effort over the FTMC approach. Future developments could regard the formulation of a multi-objective mathematical model aimed at obtaining the Pareto front to simultaneously optimise various objective

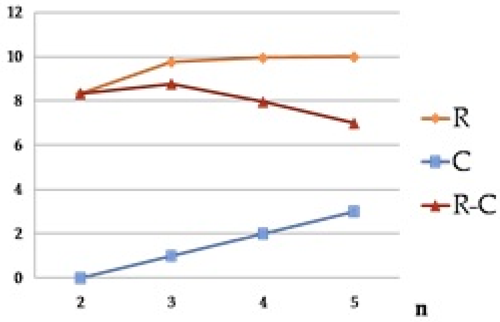


Figure 1: Graphical solution

functions and to provide the analyst with the possibility to hypothesize various design scenarios. The final choice, corresponding to one of the Pareto solutions, may be carried out by means of a MCDM approach.

References

[1] Akhavein A. and Fotuhi Firuzabad M. A heuristic-based approach for reliability importance assessment of energy producers *Energy Policy*, 39(3):1562–1568, 2011.

[2] Vasili M., Tang S.H., Ismail N., Vasili M.R. A Maintenance optimization models: a review and analysis. Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management, Kuala Lumpur, Malaysia, January 22–24, 2011.

[3] Yssaad B., Abene A. Rational Reliability Centered Maintenance Optimization for power distribution systems *International Journal of Electrical Power and Energy Systems*, 73:350–360, 2015.

[4] Moubray J. Reliability-centered maintenance Butterw.-Heinem., 1991.

[5] EN 60812 Standard. Analysis techniques for system reliability – Procedure for failure mode and effects analysis (FMEA), May 2006.

[6] Garg H. Performance and behavior analysis of repairable industrial systems using Vague Lambda–Tau methodology *Applied Soft Computing*, 22:323–338, 2014.

[7] Sharma R.K. and Kumar S. Performance modeling in critical engineering systems using RAM analysis *Reliability Engineering and System Safety*, 93(6):913–919, 2008.

[8] Lu J.-M., Wu X.-Y., Liu Y. and Lundteigen M.A. An approach for analyzing the reliability of industrial systems using soft-computing based technique *Expert Systems with Applications*, 41(2):489–501, 2014.

[9] Billinton R. and Allan R.N., Reliability evaluation of engineering systems: concepts and techniques. 2nd ed. Plenum, New York, 1992.

[10] Chambari A., Rahmati S. H.A., Najafi A.A. and Karimi A. A bi-objective model to optimize reliability and cost of system with a choice of redundancy strategies *Computers and Industrial Engineering*, 63(1):109–119, 2012.

[11] Garg H. and Sharma S.P. Multi-objective reliability-redundancy allocation problem using particle swarm optimization *Computers and Industrial Engineering*, 64(1):247–255, 2013.

[12] Lu L. and Lewis G. Configuration determination for k-out-of-n partially redundant systems *Reliability Engineering and System Safety*, 93(11):1594–1604, 2008.

[13] Kang H.G. and Kim H.E. Unavailability and spurious operation probability of k-out-of-n reactor protection systems in consideration of CCF *Annals of Nuclear Energy*, 49:102–108, 2012.

[14] Zhang Y. and Pham H., A Cost Model of an Opportunistic Maintenance Policy on k-out-of-n Surveillance Systems Considering Two Stochastic Processes. Proc. 20th ISSAT, 2014 - Seattle WA, U.S.A.

[15] Ahmed Q., Khan F. and Ahmed S. Improving safety and availability of complex systems using a risk-based failure assessment approach *Journal of Loss Prevention in the Process Industries*, 32:218–229, 2014.

[16] Häggström O, Finite Markov Chains and Algorithmic Applications, Cambridge University Press, 2002.

A firm model for the bank

R. Cervelló-Royo[(a)] [*], J.-C. Cortés[(b)],
R.M. Shoucri[(c)] , R.-J. Villanueva[(b)]

(*a*) Economics and Social Sciences Department,
Universitat Politècnica de València

(*b*) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València

(*c*) Department of Mathematics and Computer Science,
Royal Military College of Canada, Kingston, Ontario.

November 30, 2016

1 Introduction

A common assumption in the theory of the firm is that firms tend to maximize profit and minimize costs. Baumol [1, 2] developed the idea that a firm may decide to maximize revenue instead of maximizing profit, an idea that has been investigated since then by several researchers [3]. Whether a firm chooses to maximize profit or revenue depends often on the type of management, private owned firms planify differently than labour-managed firm (see Ward [4]). Moro [3] published a study in which he analyzed revenue maximization versus profit maximization, in his study he assumed that the price P is normalized to one. In this study the problem of revenue maximization and profit maximization is reviewed again by considering that the revenue $R(Q) = P(Q)Q(K, L)$, where $P(Q)$ is the price function and $Q = Q(K, L)$ is the production function (K is the capital and L the labour). The mathematical formalism for the firm developed is applied to data collected in the

[*] e-mail: rocerro@esp.upv.es

year 2013 for nine banks in Spain, where Q is the total equity of the bank and the coefficient $P(Q)$ reflects the ability of a bank to transform equity into revenue. K is the capital investment of bank, and L is the number of employees of the bank and reflects the work force. The data for the nine banks studied are given in Table 1.

	Q	K	L	prof. year	Ri	pay$D$
A	73871000	736746000	169460	8943000	45612000	18185000
B	30763000	346117000	104416	4595000	23775000	9893000
C	644154	341342	1207	28239	405325	201281
D	1362901	50532810	1866	131977	871944	483836
E	1610211	22632657	3747	102591	1156705	609287
F	2583011	40714676	4509	254404	324893	147999
G	5297370	65777852	9466	526309	3166233	1565586
H	5472536	98878760	8905	558824	3717540	1921101
I	8447984	102289399	14431	780347	5059068	2236515

Table 1: Values of Q (total equity), capital K (credit investment) and labour L (number of employees), prof. year (profit per year), Ri (interest return on loans) and pay$D$ (payment on deposit) for the nine Spanish banks.

2 Mathematical model

The revenue $R(Q)$ of the firm is represented by

$$R(Q) = P(Q)Q(K, L), \tag{1}$$

where P is the price and Q is the firm production function (output), K is capital and L is labour. In what follows we shall need to calculate the partial derivatives of R with respect to Q, K or L. We have

$$R' = P + P'Q, \quad R_K = R'Q_K, \quad R_L = R'Q_L, \tag{2}$$

where $R' = \mathrm{d}R/\mathrm{d}Q$ is the marginal revenue also denoted by MR and $P' = \mathrm{d}P/\mathrm{d}Q$, $R_K = \partial R/\partial K$ and $R_L = \partial R/\partial L$ are the partial derivatives with respect to K and L, respectively. The second derivatives can be expressed in the following form

$$R_{KK} = R''Q_K^2 + R'Q_{KK}, \tag{3}$$

$$R_{LL} = R''Q_L^2 + R'Q_{LL}, \tag{4}$$

$$R_{KL} = R''Q_KQ_L + R'Q_{KL}, \tag{5}$$

where

$$R'' = 2P' + P''Q. \tag{6}$$

We use the notation

$$R'' = \frac{\mathrm{d}^2R}{\mathrm{d}Q^2}, \quad P'' = \frac{\mathrm{d}^2P}{\mathrm{d}Q^2}.$$

The following assumptions are made on the production function $Q(K, L)$:

H1 : It is a homogeneous function

$$Q(K, L) = Lf(k), \tag{7}$$

where $f(k) = Q(k, 1)$ being $k = K/L$.

H2 : It is strictly quasi-concave, first derivative $f'(k) > 0$ and second derivative $f''(k) < 0$. It is assumed that $f(0) = 0$ and $\lim_{k\to+\infty} f(k) = +\infty$.

H3 : One satisfies the following representation

$$Q_K = f'(k), \quad \text{where } \lim_{k\to 0} f'(k) = +\infty, \quad \lim_{k\to+\infty} f'(k) = 0. \tag{8}$$

H4 : One satisfies the following representation

$$Q_L = f(k) - kf'(k), \quad \text{where } \lim_{k\to 0} f'(k) = 0, \quad \lim_{k\to+\infty} f'(k) = +\infty. \tag{9}$$

H5 : The following relations can easily be derived:

$$Q_{KK} = \frac{f''(k)}{L}, \quad Q_{LL} = k^2\frac{f''(k)}{L}, \quad Q_{KL} = -k\frac{f''(k)}{L}. \tag{10}$$

The balance sheet of the bank can be written in the form Revenue = Profit + Cost, i.e.,

$$R = \Pi + \text{Cost}, \tag{11}$$

where $R = P_1(Q)Q + P_2$ reflects the way the total equity Q is transformed into revenue, P_1 and P_2 are coefficients, the profit per year is $\Pi(K, L)$, and the cost is given by

$$\text{Cost} = wL + rK + B, \tag{12}$$

where wL is the labour cost, rK is the cost on capital, $B(Q)$ is the base expenses. From equations (2), (11) and (12), we have

$$(R-B)_K = (R-B)'Q_K = \frac{\partial \Pi}{\partial K} + r = \frac{\partial \Pi}{\partial Q} Q_K + r = \Pi' Q_K + r, \tag{13}$$

$$(R-B)_L = (R-B)'Q_L = \frac{\partial \Pi}{\partial L} + w = \frac{\partial \Pi}{\partial Q} Q_L + r = \Pi' Q_K + w, \tag{14}$$

where $(R-B)' = \frac{\partial(R-B)}{\partial Q}$. By multiplying expression (13) by K, and expression (14) by L and adding (we assume that Q is a homogeneous function), we get

$$\begin{array}{rcccl} (R-B)'Q & = & \Pi'Q + wL + rK, & & \\ (R' - \Pi' - B')Q & = & wL + rK & = & R - \Pi - B. \end{array} \tag{15}$$

From equation (15), we can derive the differential equation

$$\frac{\mathrm{d}(R - \Pi - B)}{R - \Pi - B} = \frac{\mathrm{d}Q}{Q} \Rightarrow R - \Pi - B = cqQ, \tag{16}$$

where cq is a constant. We can write the previous equations in the form

$$wL + rK = cqQ, \tag{17}$$

$$R = P_1 Q + P_2 = \Pi + B + cqQ. \tag{18}$$

In these two equations the coefficients r or cq and P_1 and P_2 need to be estimated in order to assess the performance of the banks, the coefficient w can be calculated from knowledge of staff cost and number of employees which is given in Table 1. Maximum profit implies that $\mathrm{d}\Pi/\mathrm{d}Q = 0$, and maximum revenue implies that $\mathrm{d}R/\mathrm{d}Q = 0$.

3 Applications and Results

From the data of the nine banks showed in Table 1. By taking Q as the total equity, K as the credit investment, Figure 1 shows the relation between the production $f(k) = Q/L$ and K/L, as well as the relation between Q/K and K/L. One can compare with similar curves published in [3].

For the calculation of cq and r we have used the relation between staff cost = staffc. = wL, and Q = total equity, the relation between staffc./k and Q/k is shown in Figure 2 ($k = K/L$) (staff cost = staffc = wL, and Q = total equity). Least squares fit of a second degree curve gives staffc/k = $-4.633\times10^{-6}\times(Q/k)^2+0.1953\times Q/k-31.93$. By comparing with expression (17) written in the form $wL/k = cq \times Q/k - r \times L$, we have taken $cq \approx 0.2$ and $r \times L \approx 4.633 \times 10^{-6} \times (Q/k)^2 + 31.93$ (or $r = (cq \times Q - wL)/K$).

Figure 3 shows the relation between the coefficient r (cost on capital) and K/L (left), and the relation between the investment cost $cq \times Q = w \times L + r \times K$ and K/L. Notice that the investment cost increases when K/L decreases, when the work force L becomes large.

Figure 4 shows two relations between cost of labour $w \times L$ and cost on capital $r \times K$, both seem to follow the same trend (a similar relation exists between L and K).

Figure 5 shows variation of Q (total equity) and $Ri - \text{pay}D$ (interest return on loans minus payment on deposits) with K/L, except for two banks (bank (A) and bank (B)) that show higher values. We have also looked then at the ratio $Q/(Ri - \text{pay}D)$ (Figure 6, right) the variation of which is small except for one bank (bank (F)).

Interesting to note from Figure 6 (left) that the ratio prof. year/$(Ri - \text{pay}D)$ is smaller than one except for one bank (bank (F)). This observation is more evident in Figure 7 (right) where the ratio prof. year/$(Ri - \text{pay}D)$/prof. year appears negative except for one bank (bank (F)). The net profit per year profyear can be expressed as follows

Figure 5 (left) and Figure 6 show relations involving the ratio $(Ri - \text{pay}D)$/prof. year or its inverse and $1 - (Ri - \text{pay}D)$/prof. year. The five relations shown suggest a distinctive pattern of the data along two lines, indicating that the ratio $(Ri - \text{pay}D)$/prof. year is constant for some banks, while the same ratio varies linearly with the variables used sown as ordinates for other banks. The ratio (prof. year $- (Ri - \text{pay}D)$)/prof.year is negative

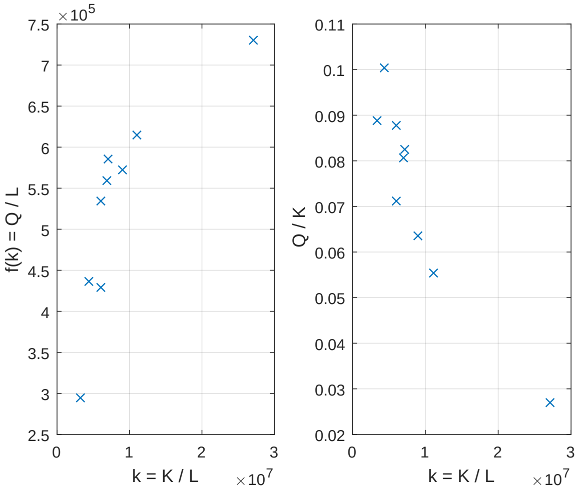


Figure 1: Left: Relation between $f(k) = Q/L$ and $k = K/L$. Right: Relation between Q/K and $k = K/L$; Q is the total equity, K is the credit investment, L is the number of employees.

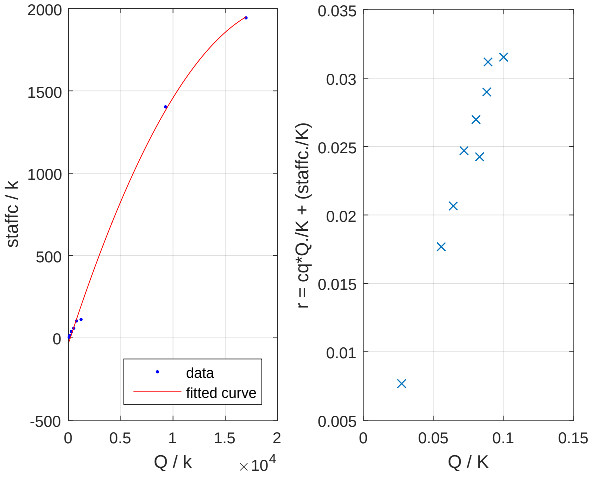


Figure 2: Left: Relation between staff cost$/k = wL/k$ and Q/k. Right: Relation between the coefficient $r = cq \times Q/K + \text{staffc}/K$ and Q/K for $cq = 0.2$; Q is the total equity, K is the credit investment, staffc = staff cost.

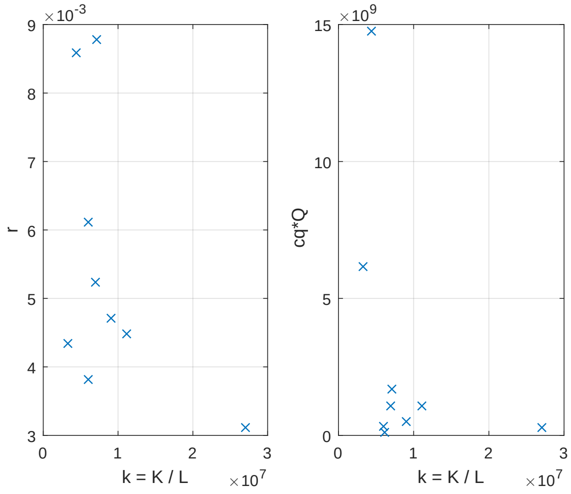


Figure 3: Left: Relation between the coefficient r and K/L. Right: Relation between investment cost $cq \times Q = w \times L + r \times K$ and K/L, Q is the total equity, K is the credit investment and L is the number of employees.

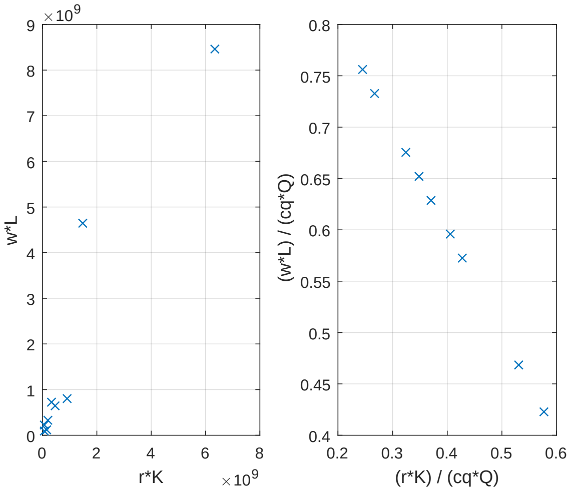


Figure 4: Left: Relations between cost of labour wL and cost on capital rK. Right: Relation $(w \times L)/(\text{cq} \times Q) + (r \times K)/(\text{cq} \times Q) = 1$.

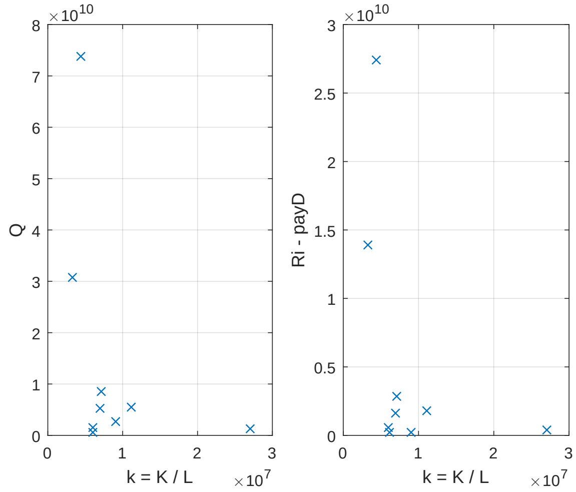


Figure 5: Left: Relations between K/L with Q. Right: Relation between $Q/(Ri - \text{pay}D)$ where Q is the total equity, Ri is the interest return on loans, K is the credit investment and L is the number of employees.

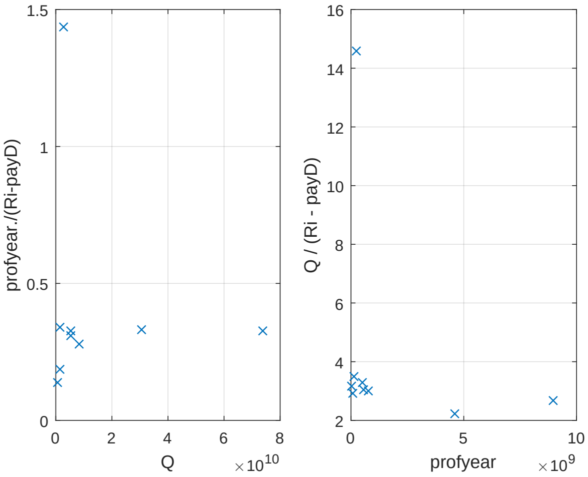


Figure 6: Left: Relation between Q and porf. year$/(Ri - \text{pay}D)$. Right: Relation between prof. year and $Q/(Ri - \text{pay}D)$ being Ri the payment on loan and pay$D$ the payment on deposit.

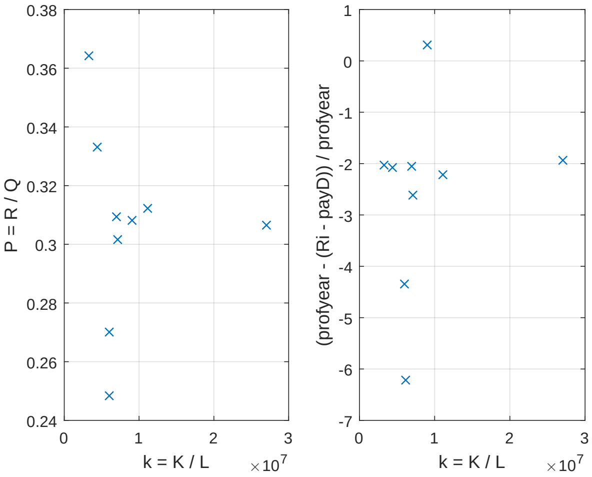


Figure 7: Left: Relation between $k = K/L$ and $P = R/Q$. Right: Relation between $k = K/L$ and (prof. year$-(Ri-\text{pay}D)$)/prof. year being Ri interest return on loans and pay$D$ the payment on deposit.

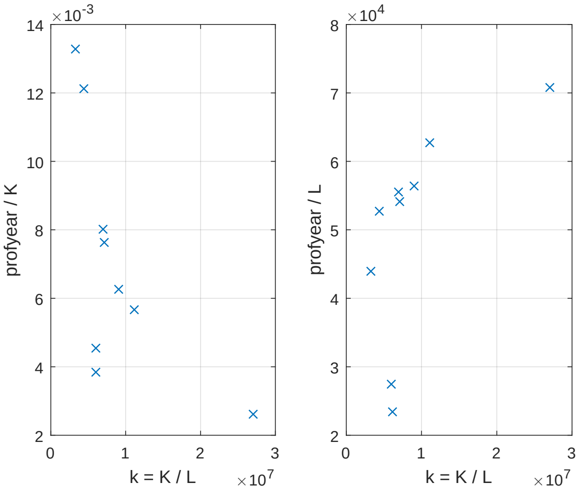


Figure 8: Left: Relation between K/L and prof. year$/K$. Right: porf. year$/L$ being K the credit investment and L the number of employees.

except for one bank (bank (f)). The net profit per year, prof. year, can be expressed as follows

$$\text{prof. year} = \text{prof. due to investment} + Ri - \text{pay}D. \qquad (19)$$

Equation (19) indicates that with the exception of one bank, all other banks are losing in the profit due to investment (negative) and they are relying on their income from Ri is the interest return on loans (a dangerous policy!). Figure 7 (left) shows the variation of the coefficient $P = R/Q$ (see (1)) with $k = K/L$. The results of Figure 7 seems to suggest that there are two trends in the data that need further analysis. Finally Figure 8 shows the variations of prof. year$/K$ and prof. year$/L$ with K/L, with a decrease in these ratios when the L increases (K/L decreases).

4 Conclusions

This study has presented a mathematical model that can be used for the evaluation of the performance of the banks. An important observation in this study is that most of the banks considered are losing in their investment and are relying on the interest return on loans for their income.

References

[1] W.J. Baumol, Business Behavior, Value, and Growth. New York, Macmillan, 1959.

[2] W.J. Baumol, On the theory of the expansion of the firm, American Economic Review, 52: 1078–1087, 1962.

[3] B. Moro, The theory of the revenue maximizing firm, Journal of service Science and Management, 1: 172–192, 2008.

[4] Ward B., The firm in Illyria: market syndicalism, American Economic Review, 48: 566–589, 1958.

A Model of Biological Control of Plant Virus Propagation with Delays

M. Jackson$^\flat$ and B. Chen-Charpentier† *

($\flat$)($\dagger$) Department of Mathematics, University of Texas at Arlington,
Arlington, TX 76019-0408.

November 30, 2016

1 Introduction

Plants play a vital role in almost every ecosystem on the planet. Sometimes plants may become infected with a virus. These infections can be devastating to not only the plants themselves but also the ecosystem that depends on them. Also, plant virus infections can have a negative impact on the crops necessary for human survival. For example, the cassava plant, which is a staple in many underdeveloped African countries, is susceptible to the cassava mosaic virus. This virus has ravaged plants in Kenya, Uganda and Tanzania.Viruses need a method of transportation to move from one plant to another. Typically an insect vector this is the mode of transportation. In fact, insects are responsible for 70 percent of all plant virus transmissions.The insect must come in contact with an infected plant, usually by feeding on it, obtain the virus, and transmit the virus to another healthy plant.

Mathematical models can be used to understand the dynamics of a particular situation. Ordinary differential equations (ODEs) have been used to model plants infected with viruses. In [**?**], the authors develop a model to combat plant viruses by continuously removing infected plants and replacing them with healthy plants. In [**?**] a system of ODEs is considered that explicitly models the interaction between plants and insects. Although ODE

[*] e-mail: bmchen@uta.edu

models can help one understand the interaction between plants, viruses, and vectors, they do not consider the time it takes for a virus to spread within a plant or insect. By introducing a delay to the system, we can account for this biological fact. For example, the authors in [**?**] introduced a delay to the model in [**?**] to account for the incubation period of the plants and noticed a change in the dynamics of the model. In [**?**], the authors modified the model in [**?**] by including multiple delays to account for the incubation periods of the virus in both the plants and insects. They too noticed changes in the dynamics in the system, specifically changes in the solutions. From a mathematical perspective, delays change the solution, may change the stability of steady solutions and may introduce discontinuities in the derivatives [**?**].

There are many different ways to combat the disease. One could try to breed a plant that is resistant to the virus and replant infected plants with the resistant ones. Another way is the use of pesticides. However, there are some drawbacks to the two methods. To breed a plant that is resistant to the virus, such a plant must first be discovered. For pesticides, too much can be harmful to the plants and to the environment as well. Another method we is to introduce a predator to the environment to feed upon the insects. This alternative can be more environmentally friendly compared to the pesticides. Predators may not be present naturally or their number may not be enough to control the vectors. We will explore the effects of introducing a predator to the system or increasing its number.

In this paper, we first discuss our modeling assumptions. Then we construct a system of ordinary differential equations modeling the interaction between the plant hosts, the insect vectors, and the predators of the vectors. Afterwards, a stability analysis is performed on the system. A couple of delays accounting for the time it takes for a plant and vector to become infected by the virus are introduced to the system. We use a programming software biftool to numerically approximate eigenvalues and bifurcation points of the delay differential equations (DDE's). Afterwards, we introduce a predator at a constant rate to the model and we analyze the dynamics. Results and conclusions are presented .

2 Mathematical model

In this paper, we extend the model in [**?**] to include a predator We consider 6 populations: susceptible plants $S(t)$, infected plants $I(t)$, recovered plants $R(t)$, susceptible insect vectors $X(t)$, infected insect vectors $Y(t)$, and predators $P(t)$. Each variable describes it's respective population at time t. Susceptible plants do not have the disease but could contract the disease if infected with the virus. The infected plants have the virus but cannot directly transmit the virus to susceptible plants. Additionally, since the infected plants can die from the viral infection their death rate is higher than that of plants that do not have the virus. We also assume that as soon as a plant dies either from the infection or from a natural death, it is immediately replaced with a new susceptible plant by a farm worker. Thus it is reasonable to assume that the plant population remains fixed and the total plant population will be denoted by K. This assumption has the modeling advantage that $K = S(t)+I(t)+R(t)$ can be used to eliminate the recovered population from the system of equations. The susceptible insects do not have the virus but can obtain the virus if they come in contact with a infected plant. Infected insects can transmit the virus to susceptible plants upon contact. We assume no vertical transmission of the virus with neither plants nor vectors. Moreover, we assume that the virus does not harm the vector and thus the vector does not defend against the virus and it retains the virus for the rest of its life. We assume that the predators consume both infected and healthy insects at the same rate. The predators use this energy from feeding on the vectors to grow the predator population. We also assume that even if a predator consumes an infected insect, it will not become infected with the virus. We will also include competition between predators for the insects. Moreover, predators can feed on the infected insects and susceptible insects at different rates, but since we are assuming that the vectors are asymptomatic in the calculations we will use the same rate. The interaction between vector and plant as well as that of predator and vector are assumed to have a limitation of the form of predator-prey Holling type 2.

The parameters of the model are: K, total plant host population, β, infection rate of plants due to vectors, β_1, infection rate of vectors due to plants, α, saturation constant of plants due to vectors, α_1, saturation constant of vectors due to plants, μ, natural death rate of plants, m, natural death rate of vectors, γ, recovery rate of plants, Λ, d, death rate of infected plants due to the disease, c_1, contact rate between predators and healthy insects, c_2, contact rate between predators and infected insects, δ, natural death rate of predators, ϵ, competition constant between predators, α_3, saturation of predators due to insects and α_4, conversion rate of predators due to insects.

After an infected vector bites a susceptible plant, it takes time for the virus to enter the plant cells, to replicate and to spread in the plant, and since it also takes time for the virus to infect a susceptible insect after it bites an infected plant, we introduce delays to the system. In particular, we consider two discrete delay,

τ_1, which is time it takes a plant to become infected after contagion and τ_2, the time it takes a vector to become infected after contagion. τ_1 is much larger than τ_2 since the virus needs to penetrate the plant cells, replicate and spread mover the plant. In the vector the virus does not replicate and usually stays only around the jaws of the insect.

The model with the two discrete delays is

$$\begin{aligned}
\frac{dS}{dt} &= \mu(K-S) + dI - \frac{\beta Y(t-\tau_1)}{1+\alpha Y(t-\tau_1)} S(t-\tau_1) \\
\frac{dI}{dt} &= \frac{\beta Y(t-\tau_1)}{1+\alpha Y(t-\tau_1)} S - (d+\mu+\gamma)I \\
\frac{dX}{dt} &= \Lambda - \frac{\beta_1 I(t-\tau_2)}{1+\alpha_1 I(t-\tau_2)} X(t-\tau_2) - \frac{c_1 X}{1+\alpha_3 X} P - mX \\
\frac{dY}{dt} &= \frac{\beta_1 I(t-\tau_2)}{1+\alpha_1 I(t-\tau_2)} X(t-\tau_2) - \frac{c_2 Y}{1+\alpha_3 Y} P - mY \\
\frac{dP}{dt} &= \frac{\alpha_4 c_1 X}{1+\alpha_3 X} P + \frac{\alpha_4 c_2 Y}{1+\alpha_3 Y} P - \delta P - \epsilon P^2
\end{aligned}$$

The system of equations with no delays, which is a system of ODE's, has one of equilibrium point that is easy to analyze:

$$S^* = K, I^* = 0, R^* = 0, X^* = \frac{\Lambda}{m}, Y^* = 0, P^* = 0.$$

This point is usually referred as the disease-free equilibrium. There are other equilibrium points but due to the complexity of their formulas it is not possible to determine whether they make physical sense. The local stability of the equilibrium points of the system of ODE's is established by linearizing the system about the equilibrium point and determining whether the eigenvalues of the Jacobian matrix have positive real parts. For the disease-free equilibrium we found that it is stable when $\frac{m^2\omega}{K\Lambda} > 1$ (the real part of all the eigenvalues is negative) and unstable when $\frac{m^2\omega}{K\Lambda} < 1$.

The other equilibrium points and their stability needs to be studied numerically for given values of the parameters and results are presented later.

The system of delay differential equations has the same equilibrium points as the system of ODE's. We can also construct a Jacobian matrix for our system of delay equations and determine a characteristic equation using the disease free equilibria. However, it is no longer a polynomial equation and it is very difficult to determine the eigenvalues from the equation. It is also troublesome to determine the endemic equilibria analytically because of the number of parameters in our system. This, in turn, is problematic when performing a stability analysis on the endemic equilibria. Therefore, we run numerical simulations for particular values to see how the system behaves. We use them to determine the equilibria values and we investigate the stability of these equilibria using numerical simulations. The study of the stability of the equilibrium points of the delay differential equation system has to be done numerically. We used the code *dde-biftool* [**?**]. This program will approximate the eigenvalues of a system of delay differential equations. It also allows the variation of some parameters and will approximate the location of bifurcation points.

3 Introducing a Predator at Constant Rate

The previous numerical solutions showed that the predator is not very effective in combating the disease if the initial population is small. We would like to see what happens if we introduce a certain amount of the predator every day. We will now define a new parameter, Λ_p. Λ_p is the amount of predator that a farm worker will artificially insert into the system each day. It is a constant rate. The system of differential equations stays the same except for the equation modeling the predator.

$$\frac{dP}{dt} = \Lambda_p + \frac{\alpha_4 c_1 X}{1+\alpha_3 X}P + \frac{\alpha_4 c_2 Y}{1+\alpha_3 Y}P - \delta P - \epsilon P^2$$

We notice that if the predators are not introduced at a high enough rate the disease will persist.

4 Conclusion

We presented two plant virus propagation models, one with no delays and the other with two delays. Even though we used a general model, we applied it to specific examples to give additional validation and to show the type of results that can be expected if more data was available. Showing that a model can be used for predictions may encourage plant and insect researchers to gather the data necessary for the model. With the data to be collected, the parameters can be modified to fit a particular situation.

The delay model can be useful for agriculture researchers and workers. They can utilize the model to help predict the behavior of the epidemic. For example, suppose an agriculture worker notices a plant endemic occurs, but after a period of time they notice that fewer plants are becoming infected. The delay model will let the worker know that there will be an increase in the number of infections at a later time, because of the oscillatory behavior of the model. This will force workers to not become relaxed in trying to combat the virus through vector control means.

References

[1] K. Engelborghs, et al., DDE-BIFTOOL: a Matlab package for bifurcation analysis of delay differential equations. Leuven, Belgium, Department of Computer Sciences, Free University, 2001.

[2] M. Jackson, B. Chen, Modeling plant virus propagation with delays, Journal of Computational and Applied Mathematics. 309 (2016) 611–621.

[3] X. Meng and Z. Li, The dynamics of plant disease models with continuous and impulsive cultural control strategies, Journal of Theoretical Biology, vol. 266, no. 1, pp. 29?40, 2010.

[4] M.A. Novak, Virus Dynamics: Mathematical Models of Immunology and Virology. New York, Oxford University Press, 2000.

[5] A.S. Perelson, D.E. Kirschner, R. De Boer, Dynamics of HIV Infection of CD4$^+$ T-cells, Math. Biosci. 114 (1993) 81–125.

[6] R. Shi, H. Zhao, S. Tang, Global Dynamic Analysis of a Vector-Borne Plant Disease Model, Advances in Difference Equations. 59 (2014).

[7] T. Zhang, et al. Dynamical Analysis of Delayed Plant Disease Models with Continuous or Impulsive Cultural. Control Strategies. Abstract and Applied Analysis (2012).

Retinal Diseases Characterization using Fractal Analysis

A. Colomer[♭][*], V. Naranjo[♭], and T. Janvier[†]

(♭) Instituto de Investigación e Innovación en Bioingeniería (I3B),
Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain.
(†) University of Orléans, I3MTO Laboratory,
Rue de Chartres, BP 6759, 45067 Orléans Cedex 2, France.

November 30, 2016

1 Introduction

Diabetic retinopathy (DR) and age-related macular degeneration (AMD) are two of the most common pathologies in the current society that provoke retinal damage and can be directly related to blindness and vision impairment.

Different lesions are manifested in the fundus images; these lesions are the basis to identify the pathologies and their stage of proliferation. Exudates and drusen are the bright lesions that characterize DR and AMD. In the literature, the most common procedure is to segment these lesions using different techniques [1, 2, 3] but in this work the characterization of the healthy and the pathological retina is studied applying texture analysis techniques, in particular, fractal descriptors.

Fractals have been recognized as an effective descriptor of complex structures in biology and medicine [4, 5, 6, 7]. Fractal objects are characterized by a high degree of complexity but also by self-similarity represented by a repetition of patterns or statistical properties over different scales. Based on this property of self-similarity it is possible to describe an image by means

[*] e-mail: adcogra@i3b.upv.es

of a fractal dimension (FD) parameter. The characterization of gray level images is described using the potential usefulness of the fractional Brownian motion (fBm) model [8].

The main objective of this work is to demonstrate that the fBm model suits well for the characterization of pathological texture in retinal images and as novelty we propose how it is possible to detect lesions as exudates or drusen by means of the fractal dimension.

2 Methods

2.1 Fractals

The minimum number of independent variables to describe an object is defined by the Euclidean dimension (E). A point is a 0-dimensional object; a line is 1-dimensional while a plane is 2-dimensional. However, this approach does not extract information about the "roughness" of the object, in other words, Euclidean dimension of a line will always remain constant (E=2) whatever it is straight or crooked.

Two centuries ago different mathematicians such as Koch, Sierpinski and Hausdorff established a geometric definition of fractals demonstrating that shapes and objects have fractional dimension. Fractal object can be described by a part of itself as a constant pattern repeating at different scales. This property is known as "self-similarity" (Figure 1).

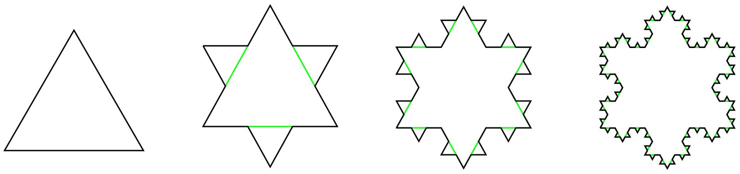


Figure 1: The first four iterations of the Koch snowflake.

The Hausdorff dimension (D_H) is an extension of E, where the dimension can be non-integer, and it is defined as:

$$D_H = \frac{ln(N)}{ln(1/r)} \tag{1}$$

where N is the number of the object's internal homotheties and r the common ratio of theses homotheties. In other words, D_H is the ratio between the number of elementary patterns N included in the object and their reduction factor r.

Mandelbrot [9] defined the fractal dimension for characterizing fractal patterns or sets by quantifying their complexity as a ratio of the change in detail to the change in scale. He also formalized the relation between the length of an object and its roughness.

$$L_\delta = K \cdot \delta(1 - D) \tag{2}$$

where L is the length of the object, δ the measuring scale, K a constant and D the object fractal dimension (FD).

2.2 Fractional Brownian motion model

In the work presented by Mandelbrot and Van Ness [10], fractional Brownian motion is defined by its stochastic representation based on the long-range dependence and self-similar behaviours. This representation is governed by a single parameter called the Hurst exponent (H) [11] which is linked to the fractal dimension by $H = E + 1 - D$, where E is the Euclidean dimension.

The spectral representation of fBm, $B_H(t)$, is given by [12]:

$$B_H(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{(\mathrm{i}\omega)^{H+1/2}}(e^{\mathrm{i}t\omega} - 1)dB(\omega) \tag{3}$$

It is a Gaussian, continuous, centered and non-stationary second-order process. Applying the initial condition $B_H(0) = 0$, its covariance function (ρ) is defined as:

$$\rho(s,t) = E[B_H(s)B_H(t)] = \frac{a^2 V_H}{2}(|t|^{2H} + |s|^{2H} - (t-s)^{2H}) \tag{4}$$

for $0 < s <= t$ where $E[\cdot]$ is the mathematical expectation, a is a constant and V_H a function of H defined as [13]:

$$V_H = \Gamma(1 - 2H)\frac{cos(\pi H)}{\pi H} \tag{5}$$

where Γ is the gamma function.

The self-similarity property of the fBm can be expressed as:

$$B_H(km) \equiv k^H B_H(m); \qquad \forall k \quad and \quad m > 0 \tag{6}$$

As fBm is a non-stationary process, it is more convenient to study its incremental process called fractional Gaussian noise (fGn) and is defined as:

$$G_m(k) = B_H(k+1) - B_H(k) \tag{7}$$

The corresponding autocovariance function $\gamma(\cdot)$ is given by:

$$\gamma(\tau) = E[G_m(k)G_m(k+\tau)] = \frac{\sigma_m^2}{2|m|^{2H}}(|\tau+m|^{2H} + 2|\tau|^{2H} + |\tau-m|^{2H}) \tag{8}$$

where $\sigma_m^2 = a^2 V_H |m|^{2H}$ is the variance of G_m.

The Power Spectra Density (PSD) of the fGn can be calculated from (8), and when m tends to infinity, a normalized spectrum of the fBm increments can be defined as [14]:

$$PSD_{G_m}(f) \propto |f|^{1-2H} \tag{9}$$

In this work the fGn-based Spectral Estimator (GSE) is used in order to compute the parameter H from the PSD of the fBm increments. Equation (9) shows that the PSD follows a law in $|f|^{1-2H}$. In a log-log scale, the PSD function of the fBm increments is a line of slope $1 - 2H$. The H parameter can be estimated by linear regression.

2.3 Application to retinal images

In fundus images, the green component of the RGB-representation (Figure 2a) shows the best contrast, the red channel is often saturated and has low contrast, and the blue channel is very noisy and suffers poor dynamic range. For these reasons, the green component is commonly used to segment the lesions [15, 16, 1] and it is used in this work (Figure 2b).

Blood vessels cover a high percentage of the fundus image and are considered as noise or artefacts that hamper the classification of pathologies based on background textures. For this reason it is necessary a pre-processing step in order to remove the contribution of this structure to the fractal analysis. Retinal vessels are detected using the algorithm proposed by Morales et al. [17]. This method is based on mathematical morphology and curvature evaluation for the detection of retinal vascular tree (Figure 2c).

A possible procedure to avoid blood vessels is to consider these structures as missing pixels and trying to restore them using the background. This technique is known as image inpainting and different kind of methods exist in the literature. In this work, diffusion-based inpainting category by means of the simplest isotropic diffusion model is used through the algorithm proposed in [18]. Figure 2d shows a retinal inpainted image.

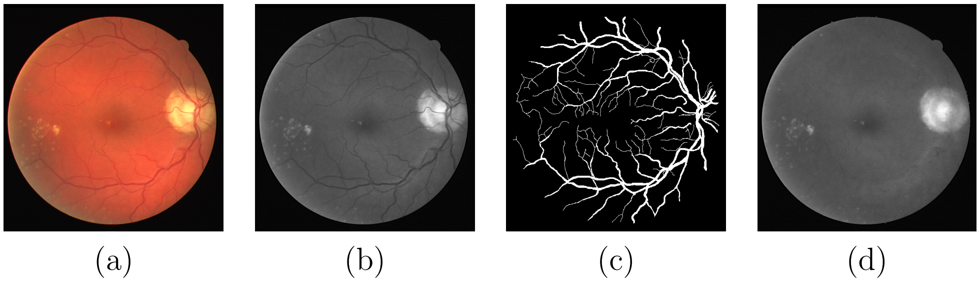
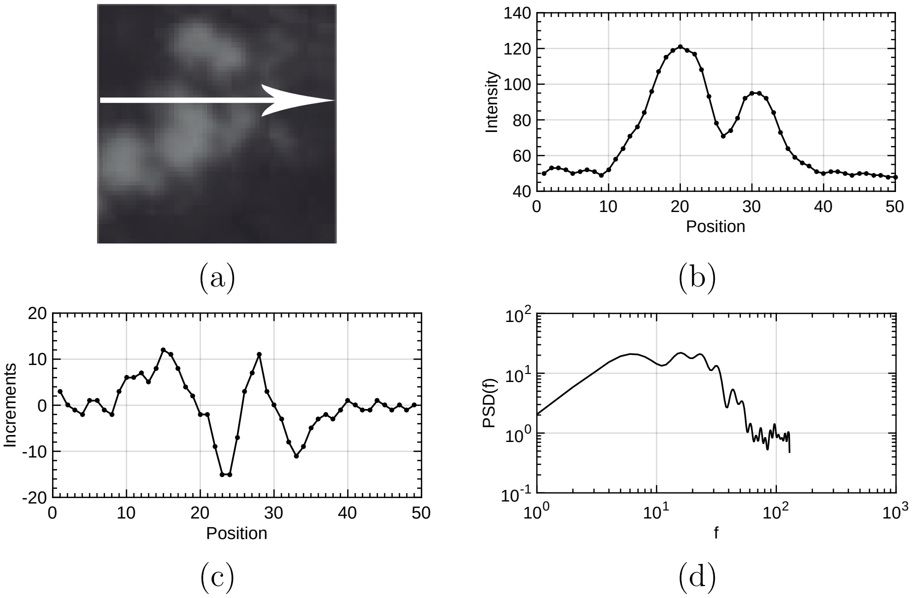


Figure 2: Image pre-processing steps. (a) original image, (b) green channel extracted from color fundus image, (c) vessel mask computed using the algorithm proposed in [17] and (d) the retinal image inpainted by the algorithm introduced by [18].

The lesions induced by macular degeneration or diabetic retinopathy present different size according to the stage of the disease. In most cases lesions represent less than one percent of the total number of pixels that compose the retinal image. For this reason, a texture analysis based on patch-looping is necessary, in other words, the image is divided in patches of the same size without overlapping and the fractal dimension is calculated for each block or patch. Note that patches containing optic disk are not considered in the process. Patches should also be completely contained within the field of view of the retinal image.

After the pre-processing stage, the PSD of the fGn is computed for all patches. Figure 5a shows an example of a line extracted from a retina background ROI of an image; we can notice the non-stationarity of the signal (Figure 5b). As recalled above, the increments of fBm are easier to study due to their stationarity (Figure 5c). In Figure 5d it is possible to observe the averaged periodogram of the increments of the lines extracted from the ROI presented.

Finally, in order to extract the H parameter from each block, a linear regression in log-log scale for each PSD curve is performed. After the H parameter estimation, a statistical analysis is carried out in order to prove the existence of significant differences between healthy and pathological regions



(a) (b)

(c) (d)

Figure 3: (a) A pathological ROI extracted from a retinal image, (b) the intensity of each pixel extracted from the ROI according to the arrow, (c) the increments of the line presented and (d) the periodogram of the increments of all lines of the ROI.

attending to the fractal dimension. Figure 4 shows the flow chart of the pre-processing and fractal analysis stages.

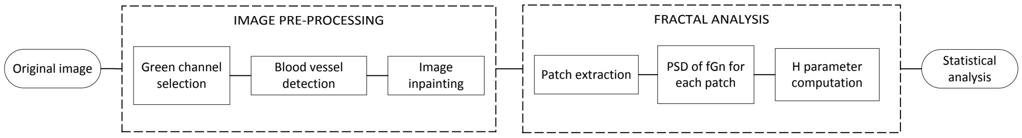


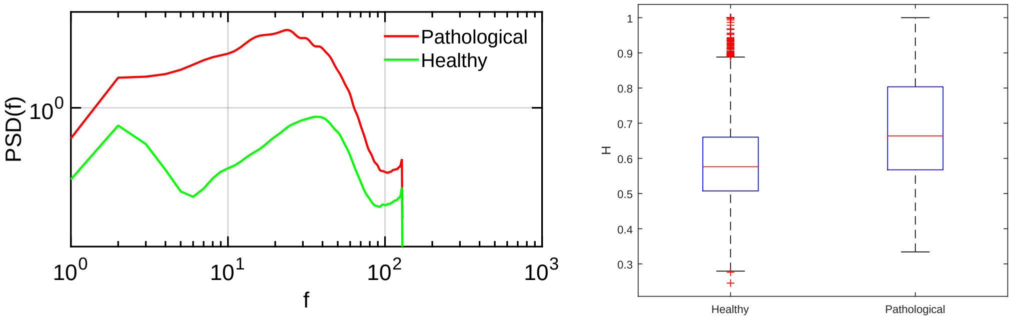
Figure 4: Flow chart of the pre-processing and fractal analysis stages.

3 Preliminary results

In order to discriminate between the pathological and the healthy retina using fractal analysis, the public database E-OPTHA_ex is used [19]. This database contains 47 color fundus images with exudates and 35 without lesions. The resolution of the 82 images is 2544 x 1696 pixels.

The methodology explained in the previous section is applied only in the 35 images with exudates. From these images, 4774 healthy patches and 492 patches with lesions are extracted. Figure 5a shows the average PSD of the fGn computed from each labelled patch.

The performance of fractal analysis to discriminate the two populations is evaluated with p-value statistical test by means of the students t-test [20]. To perform the statistical analysis, a random selection of 492 healthy patches is carried out in order to balance the groups. Figure 5b shows the box comparison of H parameter. The p-value obtained from the t-test is $p = 1.9429e - 29$.



(a)

(b)

Figure 5: (a) Average PSD comparison between the two populations, pathological (in red) and healthy (in green) and (b) Hurst exponents ± SD estimated using GSE for the healthy and pathological groups.

Acknowledgements

This work was supported by the Ministerio de Economía y Competitividad of Spain, Project ACRIMA (TIN2013-46751-R). The work of Adrián Colomer has been supported by the Spanish Government under the FPI Grant BES-2014-067889.

References

[1] T. Walter, J.-C. Klein, P. Massin, and A. Erginay, “A contribution of image processing to the diagnosis of diabetic retinopathy - detection of

exudates in color fundus images of the human retina.," *IEEE Trans. Med. Imaging*, vol. 21, no. 10, pp. 1236–1243, 2002.

[2] D. Welfer, J. Scharcanski, and D. R. Marinho, "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images.," *Computerized Medical Imaging and Graphics*, vol. 34, no. 3, pp. 228–235, 2010.

[3] M. Ghafourian and H. Pourreza, "Localization of hard exudates in retinal fundus image by mathematical morphology operations," in *2nd International eConference on Computer and Knowledge Engineering.*, pp. 185–189, 2012.

[4] R. Jennane, W. J. Ohley, S. Majumdar, and G. Lemineur, "Fractal analysis of bone x-ray tomographic microscopy projections," *IEEE Trans. Med. Imaging*, vol. 20, no. 5, pp. 443–449, 2001.

[5] G. Landini, "Fractals in microscopy," *Journal of Microscopy*, vol. 241, no. 1, pp. 1–8, 2011.

[6] G. Dougherty and G. M. Henebry, "Fractal signature and lacunarity in the measurement of the texture of trabecular bone in clinical ct images.," *Medical Engineering and Physics*, vol. 23, no. 6, pp. 369–380, 2001.

[7] O. M. Bruno, R. de Oliveira Plotze, M. Falvo, and M. de Castro, "Fractal dimension applied to plant identification," *Inf. Sci.*, vol. 178, pp. 2722–2733, June 2008.

[8] K. Harrar, L. Hamami, E. Lespessailles, and R. Jennane, "Piecewise whittle estimator for trabecular bone radiograph characterization," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 657–666, 2013.

[9] B. B. Mandelbrot, *The fractal geometry of nature.* W.H. Freeman, 1 ed., Aug. 1982.

[10] B. B. Mandelbrot and J. W. Van Ness, "Fractional brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.

[11] H. E. Hurst, R. P. Black, and Y. M. Simaika, *Long-term storage : an experimental study / by H.E. Hurst, R.P. Black, Y.M. Simaika.* Constable London, 1965.

[12] I. S. Reed, P. C. Lee, and T. K. Truong, "Spectral representation of fractional brownian motion in n dimensions and its properties," *IEEE Transactions on Information Theory*, vol. 41, pp. 1439–1451, Sep 1995.

[13] R. J. Barton and H. V. Poor, "Signal detection in fractional gaussian noise," *IEEE Transactions on Information Theory*, vol. 34, pp. 943–959, Sep 1988.

[14] P. Flandrin, "Wavelet analysis and synthesis of fractional brownian motion," *IEEE Transactions on Information Theory*, vol. 38, pp. 910–917, March 1992.

[15] X. Zhang, G. Thibault, E. Decencire, B. Marcotegui, and et al., "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Medical Image Analysis*, vol. 18, no. 7, pp. 1026–1043, 2014.

[16] M. Abràmoff, J. Folk, D. Han, and et al., "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA Ophthalmology*, vol. 131, no. 3, pp. 351–357, 2013.

[17] S. Morales, V. Naranjo, J. Angulo, J. J. Fuertes, and M. Alcañiz, "Segmentation and analysis of retinal vascular tree from fundus images processing.," in *BIOSIGNALS*, pp. 321–324, SciTePress, 2012.

[18] J. D'Errico, "Inpainting nans," 2004. `http://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint-nans`. Last accessed on 12th January 2016.

[19] E. Decencire, G. Cazuguel, X. Zhang, G. Thibault, *et al.*, "Teleophta: Machine learning and image processing methods for teleophthalmology," {*IRBM*}, vol. 34, no. 2, pp. 196 – 203, 2013.

[20] R. Mankiewicz, *The Story of Mathematics.* Cassell, 2000.

A family of parametric iterative methods for solving nonlinear models: dynamics and applications [*]

A. Cordero$^\flat$ [†] Lucía Guasp$^\flat$, and Juan R. Torregrosa$^\flat$

($\flat$) Instituto de Matemáticas Multidisciplinar,
Universitat Politècnica de València,
Camino de Vera, s/n, 46022-Valencia, Spain

November 30, 2016

1 Introduction

Calculating roots of a scalar equation $f(x) = 0$, where $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is a function defined in an open interval I, ranks among the most significant problems in the theory and practice not only of applied mathematics, but also of many branches of engineering sciences, physics, computer science, finance, chemistry, to mention only some fields. These problems lead to a rich blend of mathematics, numerical analysis and computing science.

In general, for solving these equations iterative methods must be used. In the last decades, a lot of schemes have been designed, proving their usefulness in practical problems of the different sciences. Many of them are designed almost ad-hoc, for solving specific types of problems, like derivative-free schemes, for those problems that do not allow to calculate the derivative of the nonlinear equation to be solved, like multi-point schemes for increassing the order of convergence of the classical methods, like schemes with

[*] This research was partially supported by Ministerio de Economía y Competitividad MTM2014-52016-C2-2-P and Generalitat Valenciana PROMETEO/2016/089.

[†] e-mail: acordero@mat.upv.es, luguaal@ade.upv.es, jrtorre@mat.upv.es

memory for improving the stability properties, etc. A good overview can be found in [1], and the references therein.

In the literature, iterative methods are analyzed under different points of view. A research area that is getting strength nowadays consists of applying discrete dynamics techniques to the associated fixed point operator of iterative methods. The dynamical behavior of such operators when applied on the simplest function (a low degree polynomial) gives us relevant information about its stability and performance. This study is focused on the asymptotic behavior of fixed points, as well as in its associated basins of attraction. Indeed, in case of families of iterative schemes, the analysis of critical points (where the derivative of the rational function is null), different from the roots of the polynomial, not only allows to select those members of the class with better properties of stability, but also to classify iterative methods of the same order in terms of their dynamics.

In the last years, the use of tools from complex dynamics has allowed the researchers in this area of numerical analysis to deep in the understanding of the stability of iterative schemes (see, for example, [2–6]). The analysis, in these terms, of the rational function R associated to the iterative procedure applied on quadratic polynomials, gives us valuable information about its role on the convergence's dependence on initial estimations, the size and shape of convergence regions and even on a possible convergence to fixed points that are not solution of the problems to be solved or to attracting cycles. Moreover, if a family of parametric schemes is studied, the most stable elements of the class can be chosen, by means of an appropriate use of the parameter plane.

In this paper, we begin a dynamical analysis of the family of parametric iterative methods designed by Kou et al. in [7] and whose iterative expression is

$$\begin{aligned} y_k &= x_k - \frac{2}{3}\frac{f(x_k)}{f'(x_k)}, \\ x_{k+1} &= x_k - \left(1 - \frac{3}{4}\frac{t_k - 1}{\gamma t_k + 1 - \gamma}\right)\frac{f(x_k)}{f'(x_k)}, \quad k = 0, 1, 2, \ldots, \end{aligned} \tag{1}$$

where $t_k = \dfrac{f'(y_k)}{f'(x_k)}$ and γ is a free parameter. The authors proved, for any value of parameter, the third-order of convergence of the elements of the family and fourth-order when $\gamma = 3/2$.

In order to analyze the dynamical behavior of family (1) on quadratic polynomials we choose a generic one $p(z) = (z - a)(z - b)$. If we apply (1) on $p(z)$, the following rational operator is obtained

$$T_{p,\gamma,a,b}(z) = z + \frac{(a-z)(b-z)(3a^2+3b^2+q_1(z))}{(a+b-2z)(3a^2+3b^2+q_2(z))},$$

where $q_1(z) = b(-15+4\gamma)z + (15-4\gamma)z^2 + a(b(9-4\gamma) + (-15+4\gamma)z)$ and $q_2(z) = 4b(-3+\gamma)z - 4(-3+\gamma)z^2 + a(b(6-4\gamma) + 4(-3+\gamma)z)$, depending on parameters γ, a and b.

By means of the conjugacy map $h\left(z\right) = \dfrac{z-a}{z-b}$, (a Möbius transformation), with the following properties:

$$\text{i)}\;\; h\left(\infty\right) = 1, \qquad \text{ii)}\;\; h\left(a\right) = 0, \qquad \text{iii)}\;\; h\left(b\right) = \infty,$$

operator $T_{p,\gamma,a,b}(z)$ on quadratic polynomials is conjugated to operator $O_\gamma\left(z\right)$,

$$O_\gamma\left(z\right) = \left(h \circ T_{p,\gamma,a,b} \circ h^{-1}\right)\left(z\right) = -z^3 \frac{6-4\gamma+3z}{-3-6z+4\gamma z}. \qquad (2)$$

We analyze the fixed and critical points of operator $O_\gamma\left(z\right)$. Some results about the stability of the fixed point are obtained and the behavior of the independent free critical point, used as initial guess, give us an interesting parameter plane. From this parameter plane we can extract important information about the stability of the different members of the family. Stable and pathological behaviors are obtained depending on parameter. We choose values of γ which give us stable iterative schemes and other ones with chaotic numerical deportment. The dynamical planes of all these cases are studied.

We check the numerical behavior of both types of methods (stable and unstable) on an applied chemical problem: when the flow within a round-section pipe is analyzed, different models are used (see, for example, [8]); these models show experimental relationship among the different variables in the flow transport in a pipe, such as Reynolds number Re with the longitude, inner diameter and rugosity of the pipe ε_r and its friction factor f_f.

Colebrook-White equation is one of the more precise and wide-rank ways to calculate the friction factor f_f associated to a pipe, but is is an implicit function that must be solved in an iterative way,

$$\frac{1}{\sqrt{f_f}} = -2.0\log_{10}\left(\frac{\varepsilon_r}{3.7065} + \frac{2.5226}{Re\sqrt{f_f}}\right).$$

In our study a particular case is shown, corresponding to the values of Reynolds number $Re = 4 \cdot 10^3$ and rugosity factor $\varepsilon_r = 1 \cdot 10^{-4}$ (in this case, the friction factor is $f_f \approx 0.0401$). The behavior of Newton's method and also the corresponding to different elements of (1) are shown. The numerical results confirm the dynamical behavior analyzed on quadratic polynomials, in spite of the equation that we need to solve is not a polynomial equation.

References

[1] M. Petković, B. Neta, L.D. Petković, J. Džunić, Multipoint Methods for Solving Nonlinear Equations, Academic Press, Amsterdam, 2013.

[2] S. Amat, S. Busquier, C. Bermúdez, S. Plaza, On two families of high order Newton type methods, Appl. Math. Lett. 25: (2012) 2209–2217.

[3] A. Cordero, A. Ferrero, J.R. Torregrosa, Study of the dynamics of third-order iterative methods on quadratic polynomials, Math. Comput. Simulations 119 (2016) 57–68

[4] A. Cordero, J. García-Maimó, J.R. Torregrosa, M.P. Vassileva, P. Vindel, Chaos in King's iterative family, Appl. Math. Lett. 26 (2013) 842–848.

[5] A. Cordero, J.R. Torregrosa and P. Vindel, Dynamics of a family of Chebyshev-Halley type method, Appl. Math. Comput. 219 (2013) 8568–8583.

[6] B. Neta, C. Chun, M. Scott, Basins of attraction for optimal eighth order methods to find simple roots of nonlinear equation, App. Math. Comput. 227: (2014) 567–592.

[7] J. Kou, Y. Li, X. Wang, Fourth-order iterative methods free from second derivative, Appl. Math. Comput. 184 (2007) 880–985.

[8] F. White, Fluid Mechanics, McGraw-Hill, Boston, 2003.

Approximating the first probability distribution function of the solution stochastic process to second-order random linear time-dependent differential equations with regular points

J.-C. Cortés[♭] [*]; A. Navarro-Quiles[♭],
J.V. Romero[♭] aand M.-D.Roselló[♭].

(♭) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Spain.

November 30, 2016

1 Introduction

The aim of this contribution is to provide a full probabilistic description, through the approximation of the first probability density function (1-p.d.f.), $f_1(x, t)$, of the solution stochastic process (s.p.), $X(t)$, to second-order random linear differential equations

$$X''(t) + p(t; A)X'(t) + q(t; A)X(t) = 0, \quad t > t_0 \in \mathbb{R}, \tag{1}$$

with initial conditions

$$X(t_0) = Y_0, \quad X'(t_0) = Y_1. \tag{2}$$

In the initial value problem (IVP) (1)–(2), A, Y_0 and Y_1 are assumed to be absolutely continuous real random variables (r.v.'s) defined on a common complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

[*] e-mail: jccortes@imm.upv.es

The computation of the 1-p.d.f. of the solution s.p. is advantageous since from it all one-dimensional statistical moments of the solution can be computed,

$$\mathbb{E}\left[(X(t))^k\right] = \int_{-\infty}^{+\infty} (x(t))^k f_1(x,t)\,\mathrm{d}x, \quad k = 0, 1, 2, \ldots$$

Hence, the mean $\mu_X(t) = \mathbb{E}\left[X(t)\right]$ and the variance, $\sigma_X^2(t) = \mathbb{V}\left[X(t)\right] = \mathbb{E}\left[(X(t))^2\right] - (\mu_X(t))^2$, are easily obtained as particular cases. In addition, $f_1(x,t)$ allows us to compute the probability that the solution lies in specific sets of interest,

$$\mathbb{P}\left[\{\omega \in \Omega : a \leq X(t)(\omega) \leq b\}\right], \quad -\infty \leq a < b \leq +\infty.$$

The Random Variable Transformation (RVT) technique will be used to compute approximations of $f_1(x,t)$. In its multi-dimensional version, this result can be stated as follows

Theorem 1 (Multidimensional RVT method) *[1, p.25]. Let us consider* $\mathbf{X} = (X_1, \ldots, X_n)^\top$ *and* $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ *two* n*-dimensional absolutely continuous random vectors defined on a probability space* $(\Omega, \mathfrak{F}, \mathbb{P})$*. Let* $\mathbf{r} : \mathbb{R}^n \to \mathbb{R}^n$ *be a one-to-one deterministic transformation of* $\mathbf{X}$ *into* $\mathbf{Z}$*, i.e.,* $\mathbf{Z} = \mathbf{r}(\mathbf{X})$*. Assume that* $\mathbf{r}$ *is continuous in* $\mathbf{X}$ *and has continuous partial derivatives with respect to each* X_i*,* $1 \leq i \leq n$*. Then, if* $f_{\mathbf{X}}(\mathbf{x})$ *denotes the joint probability density function of random vector* $\mathbf{X}$*, and* $\mathbf{s} = \mathbf{r}^{-1} = (s_1(z_1, \ldots, z_n), \ldots, s_n(z_1, \ldots, z_n))^\top$ *represents the inverse mapping of* $\mathbf{r} = (r_1(x_1, \ldots, x_n), \ldots, r_n(x_1, \ldots, x_n))^\top$*, the joint probability density function of random vector* $\mathbf{Z}$ *is given by*

$$f_{\mathbf{Z}}(\mathbf{s}) = f_{\mathbf{X}}\left(\mathbf{h}(\mathbf{z})\right)|J|, \tag{3}$$

where $|J|$*, which is assumed to be different from zero, is the absolute value of the Jacobian defined by the determinant*

$$J = \det\left(\frac{\partial \mathbf{s}^\top}{\partial \mathbf{z}}\right) = \det\begin{pmatrix} \frac{\partial s_1(z_1,\ldots,z_n)}{\partial z_1} & \cdots & \frac{\partial s_n(z_1,\ldots,z_n)}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_1(z_1,\ldots,z_n)}{\partial z_n} & \cdots & \frac{\partial s_n(z_1,\ldots,z_n)}{\partial z_n} \end{pmatrix}. \tag{4}$$

The following result, usually referred to as Poincaire's theorem, will be key throughout the analysis

Theorem 2 *Let us consider the initial value problem*

$$\left.\begin{array}{rcl}\mathbf{W}'(t) & = & \mathbf{f}(t,\mathbf{W}(t);\lambda), \quad t > t_0, \\ \mathbf{W}(t_0) & = & \mathbf{W}_0,\end{array}\right\} \tag{5}$$

where $\mathbf{f} :]t_0, +\infty[\times \mathbb{R}^n \times \mathbb{R} \longrightarrow \mathbb{R}^n$, $\mathbf{f} = [f_1, \ldots, f_n]^\top$, $\mathbf{W}(t) \equiv \mathbf{W} = [W_1, \ldots, W_n]^\top$ *and* $\mathbf{W}_0 = [W_{0,1}, \ldots, W_{0,n}]^\top$. *If* $\mathbf{f}(t, \mathbf{W}; \lambda)$ *admits a convergent power series expansion about* $(t_0, \mathbf{W}_0; \lambda_0)$*, i.e.,*

$$f_i(t,\mathbf{W};\lambda) = \sum_{j\geq 0}\sum_{\mathbf{k}\geq 0}\sum_{l\geq 0} c^i_{j,\mathbf{k},l}(t-t_0)^j(W_1 - W_{0,1})^{k_1}\cdots(W_n - W_{0,n})^{k_n}(\lambda-\lambda_0)^l,$$

$$\lambda_0 \in \mathbb{R},\ 1 \leq i \leq n, \quad \mathbf{k} = k_1, \ldots, k_n. \tag{6}$$

Then, for every $(t_0, \mathbf{W}_0)$*, the solution of the IVP* (5) *can also be represented as the following convergent power series*

$$\mathbf{W}(t) = \sum_{l\geq 0} \mathbf{L}_l(t)(\lambda - \lambda_0)^l. \tag{7}$$

The coefficients $\mathbf{L}_l(t)$ *are solutions of certain coupled unhomogeneous linear system of linear differential equations.*

2 Computing the approximations

We first consider the well-known representation

$$X(t) = Y_0 S_1(t;A) + Y_1 S_2(t;A), \tag{8}$$

$$S_1(t;A) = \sum_{n\geq 0} C_n(A)(t-t_0)^n, \quad S_2(t;A) = \sum_{n\geq 0} D_n(A)(t-t_0)^n, \tag{9}$$

of the solution s.p. of IVP (1)–(2) being $S_1(t;A)$ and $S_2(t;A)$ two linearly independent solutions. Then, we take its truncation

$$X_N(t) = Y_0 S_1^N(t;A) + Y_1 S_2^N(t;A), \tag{10}$$

$$S_1^N(t;A) = \sum_{n=0}^{N} C_n(A)(t-t_0)^n, \quad S_2^N(t;A) = \sum_{n=0}^{N} D_n(A)(t-t_0)^n, \tag{11}$$

being N a positive integer previously fixed, to construct the following approximation

$$f_1^N(x,t) = \iint_{\mathcal{D}_{A,Y_1}} f_{A,Y_0,Y_1}\left(z_1, \frac{x - y_1 S_2^N(t;a)}{S_1^N(t;a)}, y_1\right) \times \left|\frac{1}{S_1^N(t;a)}\right| \, \mathrm{d}y_1 \, \mathrm{d}a, \tag{12}$$

to the 1-p.d.f. of the truncated s.p. $X_N(t)$ to IVP (1)–(2).

Finally, we will introduce suitable hypotheses to legitimate that

$$\lim_{N\to+\infty} f_1^N(x,t) = f_1(x,t), \quad \text{for each } (x,t) \in \mathbb{R} \times [t_0,+\infty[\text{ fixed}, \tag{13}$$

being

$$f_1(x,t) = \iint_{\mathcal{D}_{A,Y_1}} f_{A,Y_0,Y_1}\left(z_1, \frac{x - y_1 S_2(t;a)}{S_1(t;a)}, y_1\right) \times \left|\frac{1}{S_1(t;a)}\right| \, \mathrm{d}y_1 \, \mathrm{d}a, \tag{14}$$

where $S_1(t,a)$ and $S_2(t;a)$ are defined in (9).

3 An illustrative example

In this example we will display the behaviour of second-order random differential equation

$$X''(t) - AtX(t) = 0, \tag{15}$$

with initial conditions

$$X(0) = Y_0, \quad X'(0) = Y_1. \tag{16}$$

We will consider that A, Y_0 and Y_1 are mutually independent r.v.'s. In a first step, by simplicity it is assumed that each one is uniformly distributed on the interval $]0,1[$, i.e. A, Y_0, $Y_1 \sim U(]0,1[)$. Then, we obtain the 1-p.d.f. of $X_N(t)$, $f_1^N(x,t)$. Figure 1 shows $f_1^N(x,t)$ on the time instant $t=3$ for different values of N (truncation order). On the left, for $N = 1,\ldots,3$, on the right for $N = 4,\ldots,8$. We can observe that when the truncation increase $f_1^N(x,3)$ converges to the exact 1-p.d.f, $f_1(x,3)$, of the solution s.p. of the IVP (15)–(16).

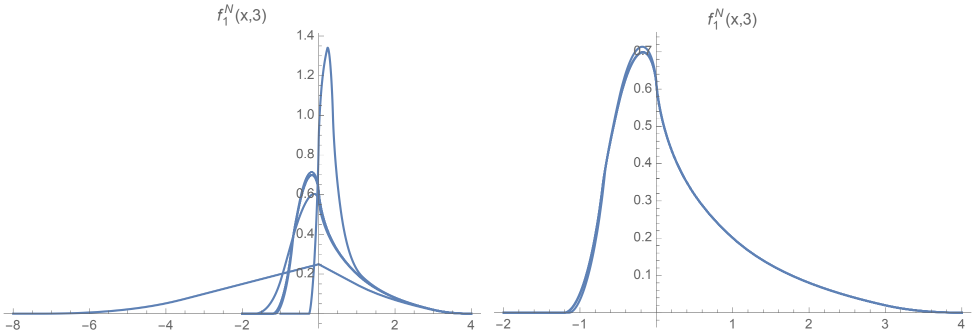


Figure 1: Plot of $f_1^N(x, 3)$ given by (12) for different values of N: $N = 1, \dots, 3$ (left), $N = 4, \dots, 8$ (right).

Acknowledgements

This work has been partially supported by the Ministerio de Economía y Competitividad grants MTM2013-41765-P. Ana Navarro Quiles acknowledges the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

References

[1] T.T. Soong, Random Differential Equations in Science and Engineering. New York, Academic Press, 1973.

[2] F. Verhulst, Nonlinear Differential Equations and Dynamical Systems. Berlin, Springer-Verlag, 1996.

A new efficient and accurate spline algorithm for the matrix exponential computation [*]

E. Defez[♭] [†] J. Ibáñez[§], J. Sastre[‡], J. Peinado[§] and P. Alonso[♯]

(♭) Instituto de Matemática Multidisciplinar,

(§) Instituto de Instrumentación para Imagen Molecular,

(‡) Instituto de Telecomunicaciones y Aplicaciones Multimedia,

(♯) Grupo Interdisciplinar de Computación y Comunicaciones,

Universitat Politécnica de Valencia, Camino de Vera s/n, 46022, Valencia, España,

November 30, 2016

1 Introduction

Matrix exponential computation has received remarkable attention in the last decades due to its usefulness in the solution of systems of linear differential equations. Moreover, in many cases, the resolution of these systems involve large or perturbated matrices, see [1]. So, the use of not only accurate, but also efficient methods becomes *absolutely necessary.* In this work, an accurate and efficient method based on matrix splines, see [2], for computing matrix exponential is given.

[*] **Acknowledgments:**This work has been supported by Spanish *Ministerio de Economía y Competitividad* and the *European Regional Development Fund (ERDF)* grant TIN2014-59294-P.

[†] e-mail:edefez@imm.upv.es

2 Taylor-Spline method

The matrix exponential can be defined for $A \in \mathbb{C}^{n\times n}$ by $e^A = \sum_{i=0}^{\infty} \frac{A^i}{k!}$, and let

$$T_m(A) = \sum_{i=0}^{m} \frac{A^i}{i!} \tag{1}$$

be the Taylor approximation of order m of e^A. $T_m(A)$ can be computed efficiently from by the Paterson-Stockmeyer's method [3], by using the following expression:

$$\begin{aligned} T_m(A) = & \left(\left(\cdots\left(\frac{A^q}{m!} + \frac{A^{q-1}}{(m-1)!} + \cdots + \frac{A}{(m-q+1)!} + \frac{I}{(m-q)!}\right)\right.\right. \\ & \times\ A^q + \frac{A^{q-1}}{(m-q-1)!} + \frac{A^{q-2}}{(m-q-2)!} + \cdots + \frac{A}{(m-2q+1)!} + \frac{I}{(m-2q)!}\Bigg) \\ & \times\ A^q + \frac{A^{q-1}}{(m-2q-1)!} + \frac{A^{q-2}}{(m-2q-2)!} + \cdots + \frac{A}{(m-3q+1)!} + \frac{I}{(m-3q)!}\Bigg) \\ & \cdots \\ & \times\ A^q + \frac{A^{q-1}}{(q-1)!} + \frac{A^{q-2}}{(q-2)!} + \cdots + A + I. \end{aligned} \tag{2}$$

The optimal values of m [4, p. 6455][5, p. 74] are:

$$\mathbb{M} = \{2, 4, 6, 9, 12, 16, 20, 25, 30, 26, 42, 49, \ldots\}, \tag{3}$$

If we denote the elements of $\mathbb{M}$ as m_1, m_2, $m_3, \ldots, m_k, \ldots$, respectively, the optimal values for m_k of q are $\lceil\sqrt{m_k}\rceil$ or $\lfloor\sqrt{m_k}\rfloor$. Both values divide to m_k and have the same cost. Table 1 shows the chosen values of q.

m_k	2	4	6	9	12	16	20	25	30	36	42	49
q_k	2	2	3	3	4	4	5	5	6	6	7	7

Table 1: Values of $q_k = \lceil\sqrt{m_k}\rceil$ for several values of m_k.

Taking into account Table 1, the cost of evaluating $T_{m_k}(A)$ in terms of matrix products, denoted by Π_{m_k}, for $k = 1, 2, \ldots$, is $\Pi_{m_k} = k$. The problem of applying algorithms based on the Taylor expression (1) is that these approximation is accurate only near the origin, hence the norm of matrix A must be reduced by using techniques based on the scaling an squaring method. The idea is to use the identity $e^A = \left(e^{A/2^s}\right)^{2^s}$ where s is a positive integer, and to apply the following approximation: $e^A \cong \left(T_m(A/2^s)\right)^{2^s}$.

2.1 Backward error analysis

The backward error, ΔA, of computing e^A by means $T_m(A)$ verifies $e^{A+\Delta A} = T_m(A)$. If we assume that $\|\log(T_m(A)) - I\| < 1$, where log is the principal logarithm, then $\Delta A = \log\left(T_m(A)\right) - A$. If we develop this expression in Taylor series, we obtain $\Delta A = \sum\limits_{i=m+1}^{\infty} p_i A^i$, and if we apply the Theorem 1.1 from [6], then the relative backward error e_b verifies that

$$e_b = \frac{\|\Delta A\|}{\|A\|} = \frac{\left\|\sum\limits_{i=m+1}^{\infty} p_i A^i\right\|}{\|A\|} \leqslant \frac{\sum\limits_{i=m+1}^{\infty} |p_i| \left\|A^i\right\|}{\|A\|} \leqslant \sum_{i=m}^{\infty} |p_{i+1}| \left\|A^i\right\|$$

$$\leqslant \sum_{i=m}^{\infty} |p_{i+1}| \left(\left\|A^i\right\|^{1/i}\right)^i \leqslant \sum_{i=m}^{\infty} |p_{i+1}| \alpha_m^i, \alpha_m = \max\left\{||A^i||^{1/i} : i \geqslant m \text{ and } p_{i+1} \neq 0\right\}$$

and, by Theorem 2 from [7],

$$\alpha_m = \max\left\{||A^i||^{1/i} :\ m \leq i \leq 2m-1 \text{ and } p_{i+1} \neq 0\right\}. \qquad (4)$$

Let be $\Theta_m^{(0)} = \max\left\{\theta \geqslant 0 : \sum\limits_{i=m}^{\infty} |p_{i+1}| \theta^i \leqslant u\right\}$, where $u = 2^{-53}$ is the unit roundoff in the double precision floating-point. For computing $\Theta_m^{(0)}$ the symbolic programs *Mathematica* and *Matlab* can be used. Values of $\Theta_{m_k}^{(0)}$, $m_k \in \{2, 4, 6, 9, 12, 16, 20, 25, 30\}$, appear in Table 2.

2.2 Forward error analysis

Applying as before Theorem 1.1 from [6] and Theorem 2 from [7], the forward relative error can be computed as follows:

$$e_f = \left\|e^{-A}\left(e^A - T_m(A)\right)\right\| = \left\|I - e^{-A}T_m(A)\right\| = \left\|\sum\limits_{i=m+1}^{\infty} |q_i| A^i\right\| \leq$$

$$\leq \sum\limits_{i=m+1}^{\infty} |q_i| \left\|A^i\right\| \leq \sum\limits_{i=m+1}^{\infty} |q_i| \left(\left\|A^i\right\|^{1/i}\right)^i \leq \sum\limits_{i=m+1}^{\infty} |q_i| \bar{\alpha}_m^i,$$

where

$$\bar{\alpha}_m = \max\left\{||A^i||^{1/i} : m+1 \leq \ i \leq 2m \text{ and } q_i \neq 0\right\}. \qquad (5)$$

Let be $\Theta_m^{(1)} = \max\left\{\theta \geqslant 0 : \sum\limits_{i=m+1}^{\infty} |q_i|\theta^i \leqslant u\right\}$. As as before, the symbolic programs *Mathematica* and *Matlab* can be used for computing $\Theta_m^{(1)}$. Values of $\Theta_{m_k}^{(1)}$, $m_k \in \{2,4,6,9,12,16,20,25,30\}$, appear in Table 2.

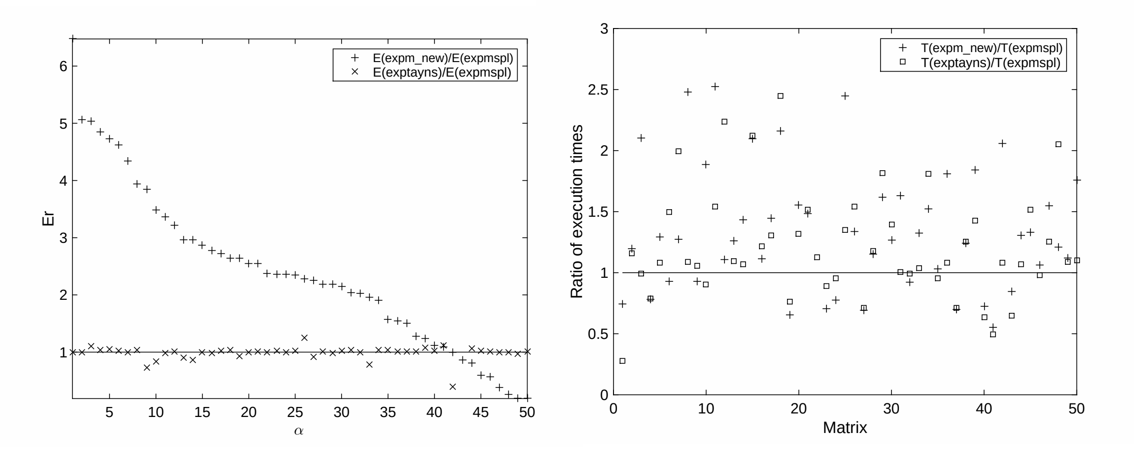
2.3 Determination of the values m and s

Let be $\Theta_m = \max\left\{\Theta_m^{(0)}, \Theta_m^{(1)}\right\}$ ($\Theta_m = \Theta_m^{(1)}$ for $m \leq 16$ and $\Theta_m = \Theta_m^{(0)}$ in other cases). We will consider $m_M \geqslant 20$. There are two possibilities:

- $\alpha_m \leq \Theta_m$, for some $m \leq m_M$. Int this case, $e_f < u$ or $e_b < u$:
  - If $m \leq 16$, $\bar{\alpha}_m \leq \alpha_m \leq \Theta_m = \Theta_m^{(1)}$, hence $e_f < u$.
  - If $m \geq 20$, $\alpha_m \leq \Theta_m = \Theta_m^{(0)}$, hence $e_b < u$.
- $\alpha_{m_M} > \Theta_{m_M} = \Theta_{m_M}^{(0)}$. In this case, we select the first positive integer s hat verifies $2^{-s}\alpha_{m_M} \leq \Theta_{m_M}$, that is $s = \left\lceil \log_2\left(\frac{\alpha_{m_M}}{\Theta_{m_M}}\right)\right\rceil$, and then the relative backward error for computing $e^{2^{-s}A}$ is lower than u, i.e. $\frac{\|\Delta 2^{-s}A\|}{\|2^{-s}A\|} \leq u$.

For computing α_m or $\bar{\alpha}_m$ we are going to compute $||A^m||^{1/m}$ instead of using Formulaes (4) or (5). Because this could affect to the error made in computing e^A, then we going to compute $\bar{T}_m(A) = \sum\limits_{i=0}^{m-1} \frac{A^i}{i!} + A_m$, where $A_m \in \mathbb{C}^{n\times n}$ is an unknown matrix that must be calculated, instead of $T_m(A) = \sum\limits_{i=0}^{m} \frac{A^i}{i!}$. To compute the matrix A_m we require that $\bar{T}_m(x) = \sum\limits_{i=0}^{m-1} \frac{A^i}{i!}x^i + A_m x^m$ is the solution at $x = 1$ of the ordinary differential matrix equation

$$Y'(x) = AY(x), Y(0) = I, x \in [0,1] \qquad (6)$$



(a) Ratio of relative errors.

(b) Ratio of relative execution times.

Figure 1: Test with fifty 128 × 128 non-diagonalizable real matrices

Thus, A_m is obtained by solving the equation: $(mI - A)\,A_m = \dfrac{A^m}{(m-1)!}$. Applying the Scheme (2) to $\bar{T}_m(A)$, we obtain:

$$
\begin{aligned}
\bar{T}_m(A) = {} & \left(\left(\cdots\left(\bar{A}_m + \frac{A^{q-1}}{(m-1)!} + \cdots + \frac{A}{(m-q+1)!} + \frac{I}{(m-q)!}\right)\right.\right. \\
\times {} & A^q + \frac{A^{q-1}}{(m-q-1)!} + \frac{A^{q-2}}{(m-q-2)!} + \cdots + \frac{A}{(m-2q+1)!} + \left.\frac{I}{(m-2q)!}\right) \\
\times {} & A^q + \frac{A^{q-1}}{(m-2q-1)!} + \frac{A^{q-2}}{(m-2q-2)!} + \cdots + \frac{A}{(m-3q+1)!} + \left.\frac{I}{(m-3q)!}\right) \\
& \cdots \\
\times {} & A^q + \frac{A^{q-1}}{(q-1)!} + \frac{A^{q-2}}{(q-2)!} + \cdots + A + I, \qquad (7)
\end{aligned}
$$

where $\bar{A}_m$ is the solution of the equation $(mI - A)\,\bar{A}_m = \dfrac{A^q}{(m-1)!}$.

3 Algorithm. Numerical experiments

Algorithm 1 computes the matrix exponential based on the above method. For computing the Taylor approximation we use Expression (7), if the condition number of $\bar{A}$ is lower than 100, else we use Expression (2).

m_k	$\Theta_{m_k}^{(0)}$	$\Theta_{m_k}^{(1)}$
2	2.675298260329713e-8	8.733457635286420e-6
4	3.397168839977002e-4	1.678018844321752e-3
6	9.065656407595296e-3	1.773082199654024e-2
9	8.957760203223343e-2	1.137689245787824e-1
12	2.996158913811581e-1	3.280542018037261e-1
16	7.802874256626574e-1	7.912740176600239e-1
20	1.438252596804337	1.415070447561532
25	2.428582524442827	2.353642766989427
30	3.539666348743690	3.411877172556770

Table 2: Values of $\Theta_{m_k}^{(0)}$ and $\Theta_{m_k}^{(1)}$.

Algorithm 1 E=expmspl(A,m_M)

Given a matrix $A \in \mathbb{C}^{n\times n}$ and m_M, this algorithm computes $E = e^A$ by a Taylor-spline approximation of order m_k lower or equal to m_M of e^A.

```
$B = mI - A$
[$m_k$,$s$,$pA$]=select_m_s($A$,$m_M$) or [$m_k$,$s$,$pA$]=select_m_s_w($A$,$m_M$)
if cond($B$) $< 100$ then                ▷ cond($B$) computes the condition number of $B$
    Compute $E = \bar{T}_{m_M}(pA, m_k, s)$ from (7)
else
    Compute $E = T_{m_M}(pA, m_k, s)$ from (2)
end if
for $i = 1 : s$ do
    $E = E^2$
end for
```

We show some numerical experiments with the proposed algorithm. MATLAB 8.4 (R2014b) implementation was tested on an Intel Core 2 Duo processor at 3.00 GHz with 4 GB main memory with fifty 128 × 128 non-diagonalizable real matrices, choose randomly, with eigenvalues whose algebraic multiplicity vary between from 1 to 10. The 1-norms of that matrices vary between 1 and 74.24. Algorithm accuracy was tested by computing the relative error E $= \dfrac{\|e^A - \tilde{Y}\|_1}{\|e^A\|_1}$, where $\tilde{Y}$ is the computed solution and e^A the exact solution. We compared our algorithm (`expmspl`) with the implemented based on the Taylor method (`exptayns`, see [7]) and Padé series (`expm_new`, see [6]).

Figure 1.a shows the ratio of relative errors E(`expm_new`)/E(`expmspl`) and E(`exptayns`)/E(`expmspl`), and Figure 1.b shows the ratio of execution times T(`expm_new`)/T(`expmspl`) and T(`exptayns`)/T(`expmspl`).

Numerical experiments show that in general the new algorithm has a higher accuracy than `expm_new` and `exptayns` functions in the majority of the tests, with a lower execution time than the implementation based on Padé series (`expm_new`) and similar execution times that the another Taylor implementation (`exptayns`).

References

[1] P. Bader, S. Blanes, M. Seydaoglu, The scaling, splitting, and squaring method for the exponential of perturbed matrices, SIAM Journal on Matrix Analysis and Applications 36 (2) (2015) 594–614.

[2] E. Defez, M. Tung, J. J. Ibáñez, J. Sastre, Approximating and computing nonlinear matrix differential models, Mathematical and Computer Modelling 55 (7).

[3] M. S. Paterson, L. J. Stockmeyer, On the number of nonscalar multiplications necessary to evaluate polynomials, SIAM J. Comput. 2 (1) (1973) 60–66.

[4] J. Sastre, J. J. Ibáñez, E. Defez, P. A. Ruiz, Efficient orthogonal matrix polynomial based method for computing matrix exponential, Appl. Math. Comput. 217 (2011) 6451–6463.

[5] N. J. Higham, Functions of Matrices: Theory and Computation, SIAM, Philadelphia, PA, USA, 2008.

[6] A. H. Al-Mohy, N. J. Higham, A new scaling and squaring algorithm for the matrix exponential, SIAM J. Matrix Anal. Appl. 31 (3) (2009) 970–989.

[7] P. Ruiz, J. Sastre, J. Ibáñez, E. Defez, High perfomance computing of the matrix exponential, J. Comput. Appl. Math. 291 (2016) 370–379.

Computational performance of analytical methods for the acoustic modelling of automotive exhaust devices incorporating monoliths

F.D. Denia[♭],[*] J. Martínez-Casas[♭], J. Carballeira[♭], E. Nadal[♭],
and F.J. Fuenmayor[♭]

(♭) Centro de Investigación en Ingeniería Mecánica, Universitat Politècnica de València
Camino de Vera s/n, 46022 Valencia, Spain.

November 30, 2016

1 Introduction

The acoustic modelling of automotive exhaust devices, such as catalytic converters (CC) and diesel particulate filters (DPF), usually requires the use of multidimensional methods [1-4]. The presence of higher order modes and three dimensional waves in the inlet/outlet cavities, as well as sound propagation within the monolith capillary ducts, can be considered through the finite element method [4, 5], although this approach is traditionally thought to be very time consuming [6]. With a view to overcome this limitation and to reduce the computational effort of the FEM, alternative modelling techniques are proposed in the current work to speed up transmission loss calculations in exhaust devices incorporating monoliths. These approaches are based on the mode matching method [6-11] and the point collocation technique [12-14].

As shown in recent studies [4, 15], the sound attenuation of an exhaust device incorporating a monolith (e.g., a catalytic converter) can be properly predicted if the monolith is replaced by a plane wave four-pole matrix providing a relationship between the acoustic fields at both sides of the capillary region (see Figure 1).

Therefore, the presence of higher order modes in the cylindrical inlet/outlet regions is combined with one dimensional wave propagation within the capillary ducts of the central monolith. The mode matching method and point collocation technique are applied to the continuity conditions of the acoustic fields at all the interfaces to couple the solutions of the wave equation in the corresponding exhaust device

[*] e-mail: fdenia@mcm.upv.es

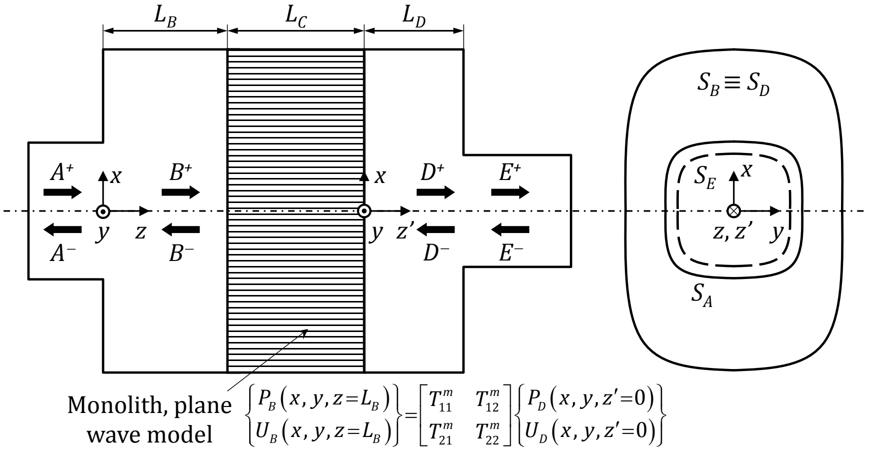

L_B
L_C
L_D
A^+
A^-
B^+
B^-
D^+
D^-
E^+
E^-
x
y
z
x
y
z'
$S_B \equiv S_D$
S_E
x
z, z'
y
S_A
Monolith, plane wave model
$\begin{Bmatrix} P_B(x,y,z=L_B) \\ U_B(x,y,z=L_B) \end{Bmatrix} = \begin{bmatrix} T_{11}^m & T_{12}^m \\ T_{21}^m & T_{22}^m \end{bmatrix} \begin{Bmatrix} P_D(x,y,z'=0) \\ U_D(x,y,z'=0) \end{Bmatrix}$

Figure 1: Scheme of an automotive exhaust device (CC/DPF) incorporating a monolith. The latter is replaced by a transfer matrix to model its acoustic behaviour.

subcomponents [7-9]. For rigid circular ducts, Bessel functions are considered as transversal pressure modes [16].
The computational efficiency and accuracy of the results associated with the analytical modelling techniques presented here are assessed, including the effect of the number of modes and collocation points. All the analytical approaches proposed in this work provide accurate predictions of the device attenuation performance and outperform the computational expenditure of a FE computation. Some differences are found, however, among the various analytical schemes in terms of computational speed and solution accuracy. From the results presented here, the most efficient technique for the particular configurations under study is the mode matching method.

2 Overview of the mathematical approach

In all the rigid ducts involved (A, B, D and E), the acoustic pressure and velocity fields can be written in terms of a series expansion. For example, the solution of the wave equation in region A is given by [16]

$$P_A(x,y,z) = \sum_{n=1}^{\infty} \left(A_n^+ e^{-jk_{A,n}z} + A_n^- e^{jk_{A,n}z} \right) \psi_{A,n}(x,y) \tag{1}$$

$$U_A(x,y,z) = \frac{1}{\rho_0 \omega} \sum_{n=1}^{\infty} k_{A,n} \left(A_n^+ e^{-jk_{A,n}z} - A_n^- e^{jk_{A,n}z} \right) \psi_{A,n}(x,y) \tag{2}$$

In the particular case of circular ducts, the transversal pressure modes in Eqs. (1) and (2) are given by suitable Bessel functions $\psi_{A,n}(x,y) = \psi_{A,n}(r) = J_0(\alpha_n r/R_1)$ satisfying the rigid wall boundary condition [9, 16].

The complete acoustic field of the system requires the computation of the wave amplitudes $A_n^{\pm}$, $B_n^{\pm}$, $D_n^{\pm}$ and $E_n^{\pm}$ in all the ducts. The continuity conditions of the acoustic fields at the interfaces between ducts A and B (inlet expansion), as well as ducts D and E (outlet contraction) are taken into account when applying both the mode matching method and the point collocation technique. Further details related to the computation of the corresponding equations can be found elsewhere [7-9]. Regarding the capillary ducts, the acoustic coupling between both sides of the monolith (at the interfaces $S_B \equiv S_D$, see Figure 1) can be expressed as [1, 3, 4, 15]

$$P_B(x, y, z = L_B) = T_{11}^m P_D(x, y, z' = 0) + T_{12}^m U_D(x, y, z' = 0) \; on \; S_B \equiv S_D \tag{3}$$

$$U_B(x, y, z = L_B) = T_{21}^m P_D(x, y, z' = 0) + T_{22}^m U_D(x, y, z' = 0) \; on \; S_B \equiv S_D \tag{4}$$

2.1 Mode matching method. Equations associated with the monolith

The equations and integrals associated with the inlet expansion and outlet contraction are omitted here for the sake of brevity. It is worth noting that advantage can be taken from the orthogonality properties of the rigid duct transversal modes, thus reducing the computational effort of the mode matching approach. Computation details of the corresponding integrals can be found in references [7-10]. Here, the application of the mode matching method focuses on the monolith. Thus, Eqs. (3) and (4) are multiplied by the transversal mode $\psi_{B,s}(x, y) = \psi_{D,s}(x, y)$, with $s = 1, 2, \ldots, N_m$ (a suitable series truncation is considered). Integrating over $S_B \equiv S_D$, taking advantage of the orthogonality relations [7] and removing common factors, the following equations are derived for $s = 1, 2, \ldots, N_m$

$$B_s^+ e^{-jk_{B,s}L_B} + B_s^- e^{jk_{B,s}L_B} = T_{11}^m (D_s^+ + D_s^-) + T_{12}^m \frac{k_{D,s}}{\rho_0 \omega} (D_s^+ - D_s^-) \tag{5}$$

$$\frac{k_{B,s}}{\rho_0 \omega} (B_s^+ e^{-jk_{B,s}L_B} - B_s^- e^{jk_{B,s}L_B}) = T_{21}^m (D_s^+ + D_s^-) + T_{22}^m \frac{k_{D,s}}{\rho_0 \omega} (D_s^+ - D_s^-) \tag{6}$$

It is worth noting here that very simple algebraic expressions have been obtained, where neither integrations nor modal summations appear, relating directly wave coefficients with equal modal number. These equations do not depend on the transversal cross section, provided that its geometry is axially uniform. As shown in the results of the work, the computational performance of this approach will deliver excellent results when compared with alternative techniques such as the point collocation technique described in the next section.

2.2 Point collocation technique. Equations associated with the monolith

Only the equations for the monolith are presented here, the expressions for the inlet expansion and outlet contraction being omitted. The reader is referred to works [12-

14] for additional information. Therefore, the four-pole relation given by Eqs. (3) and (4) is prescribed at N_m points over $S_B \equiv S_D$, establishing a pointwise connection between both sides of the monolith. The particular coupling expressions between the acoustic field in region B at coordinates $(x_p, y_p, z = L_B)$ and the acoustic field in region D at coordinates $(x_p, y_p, z' = 0)$ are given by

$$\begin{aligned}&\sum_{n=1}^{N_m}(B_n^+ e^{-jk_{B,n}L_B} + B_n^- e^{jk_{B,n}L_B})\psi_{B,n}(x_p, y_p)\\ &= T_{11}^m \sum_{n=1}^{N_m}(D_n^+ + D_n^-)\psi_{D,n}(x_p, y_p) + \frac{T_{12}^m}{\rho_0\omega}\sum_{n=1}^{N_m} k_{D,n}(D_n^+ - D_n^-)\psi_{D,n}(x_p, y_p)\end{aligned} \tag{7}$$

$$\begin{aligned}&\frac{1}{\rho_0\omega}\sum_{n=1}^{N_m} k_{B,n}(B_n^+ e^{-jk_{B,n}L_B} - B_n^- e^{jk_{B,n}L_B})\psi_{B,n}(x_p, y_p)\\ &= T_{21}^m \sum_{n=1}^{N_m}(D_n^+ + D_n^-)\psi_{D,n}(x_p, y_p) + \frac{T_{22}^m}{\rho_0\omega}\sum_{n=1}^{N_m} k_{D,n}(D_n^+ - D_n^-)\psi_{D,n}(x_p, y_p)\end{aligned} \tag{8}$$

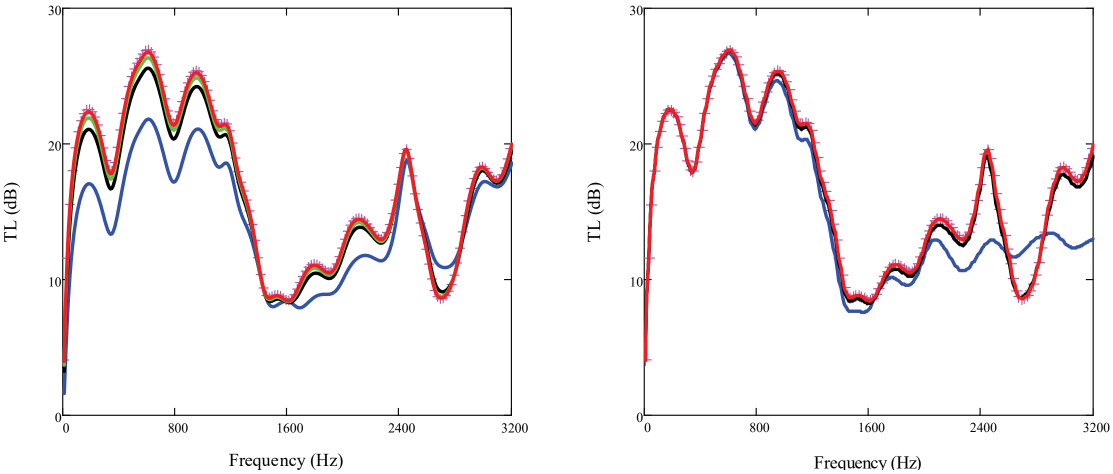
Note that, compared to Eqs. (5) and (6), the algebraic expressions (7) and (8) involve all the wave amplitudes, and a full set of equations is obtained with no direct relation between wave coefficients of equal modal number. As shown later, this will have an impact on the computational performance of the point collocation technique.

3 Results

An axisymmetric configuration is considered for validation and computation purposes. The inlet and outlet ducts are defined by the radii $R_1 = R_3 = 0.0268$ m, while the central chambers are characterized by the radial dimension $R_2 = 0.15$ m and the lengths $L_B = L_D = 0.1$ m (see Figure 1 for details). Cold flow hypotheses are retained through the computations [5, 14, 16], the properties for the air being $c_0 = 340$ m/s and $\rho_0 = 1.225$ kg/m^3. Regarding the monolith acoustic model, the following values are assumed: length $L_C = 0.2$ m, resistivity $R = 1000$ rayl/m, porosity $\phi = 0.8$, geometrical factor $\alpha_g = 1.07$, dynamic viscosity $\mu = 1.783 \cdot 10^{-5}$ Pa s, thermal conductivity $\kappa = 0.02534$ W/(m K) and specific heat $C_p = 1005$ J/(kg K). The detailed model for the four-pole matrix computation can be found in the bibliography [1, 4, 17].

3.1 Validation

First, the proposed analytical techniques based on point collocation and mode matching are validated by comparison with a reference numerical solution computed through FEM [4]. The numerical calculations have been carried out with a refined finite element mesh consisting of axisymmetric 8-node quadratic quadrilateral elements, whose approximate size is 0.0015 m. This provides 70 quadratic elements per wavelength for the maximum frequency $f_{max} = 3200$ Hz considered in the simulations, thus guaranteeing an accurate reference. For the point collocation technique, four solutions have been obtained with: (1) $N_m = 4$ points in the chambers and $N_a = N_e = 1$ point in the inlet/outlet ducts; (2) $N_m = 10$ and $N_a = N_e = 2$; (3) $N_m = 40$ and $N_a = N_e = 8$; and finally (4) $N_m = 75$ and $N_a = N_e = 14$. Regarding the mode matching method, three solutions are considered, with the following number of modes: (1) $N_m = 2$ and $N_a = N_e = 1$; (2) $N_m = 4$ and $N_a = N_e = 1$; and (3) $N_m = 10$ and $N_a = N_e = 2$. More information about the influence of N_m on the results will be provided in section 3.2. Figures 2(a) and 2(b) show the analytical results and the reference FE solution. As it can be seen, a suitable convergence appears as the number of collocation points/modes increases. The agreement for the highest numbers is excellent, with undistinguishable curves over all the frequency range, thus validating the proposed analytical techniques from a practical point of view.

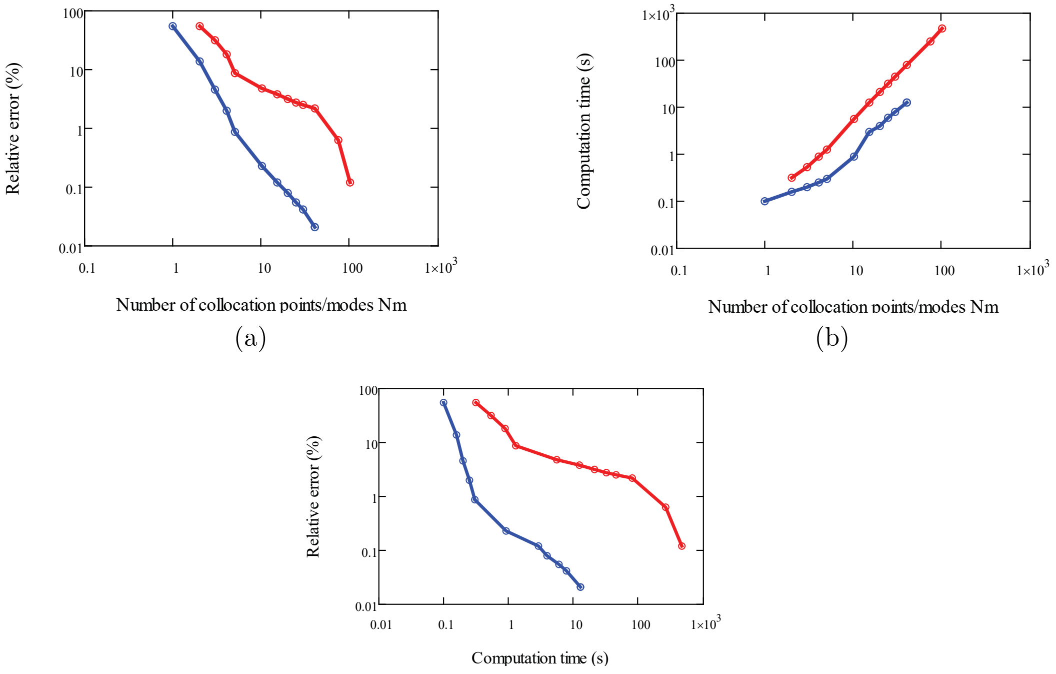


(a) (b)

Figure 2: (a) TL of a catalytic converter with monolith, point collocation: +++, FEM; —, PC, $N_m = 4$; —, PC, $N_m = 10$; —, PC, $N_m = 40$; —, PC, $N_m = 75$. (b) TL of a catalytic converter with monolith, mode matching: +++, FEM; —, MM, $N_m = 2$; —, MM, $N_m = 4$; —, MM, $N_m = 10$.

3.2 Comparison of the proposed techniques

The proposed analytical techniques are compared in this section in terms of their computational performance. From the previous results (see Figure 2), it is inferred that the solution converges as the number of collocation points/modes increases. This aspect is further confirmed by the results presented in Figure 3(a), where the TL relative error [4] is shown. For the mode matching technique, the decreasing error exhibits a more uniform trend (note that log-log scale is used), while the point collocation technique presents a less uniform behaviour. The computation time is depicted in Figure 3(b) and increases for higher N_m, as expected. Finally, the curves shown in Figure 3(c) are more useful to draw some conclusions since, for a given relative error, it is clear that the mode matching technique delivers a lower computation time. Therefore, the latter should be the preferred simulation tool, at least for the specific problem under consideration. One of the reasons is related to the orthogonality properties of the transversal modes for ducts with axially uniform cross section.



(a)

(b)

(c)

Figure 3: Computational performance of the proposed tools. (a) Relative error (%) versus number of collocation point/modes. (b) Computation time (s) versus number of collocation point/modes. (c) Relative error (%) versus computation time (s): —o—, point collocation; —o—, mode matching.

4 Conclusions

Two analytical models based on point collocation and mode matching have been presented in this work to assess the acoustic behaviour of automotive exhaust devices incorporating monoliths. Higher order modes have been included in the modelling methodology in order to be confident of accurate predictions in the high frequency range, while the monolith has been replaced by a four-pole transfer matrix to provide more realistic results. After validation by benchmarking against the finite element method, both analytical techniques have been compared in terms of accuracy and speed. The approaches proposed in this work have been shown to provide accurate predictions of the device attenuation performance and to outperform the computational expenditure of a FE computation. Some differences are found, however, between point collocation technique and mode matching method in terms of computational speed and solution accuracy. From the results presented here, the most efficient technique for the particular configurations under study is the mode matching method since, for a given error, it is shown to be significantly faster than point collocation. The reasons may be attributed, among others, to orthogonality properties of the transversal modes.

5 Acknowledgements

The authors gratefully acknowledge the financial support of Ministerio de Economía y Competitividad and the European Regional Development Fund (project TRA2013-45596-C2-1-R), as well as Generalitat Valenciana (project Prometeo/2016/007 of Conselleria d'Educació, Investigació, Cultura i Esport).

References

[1] A. Selamet, V. Easwaran, J. M. Novak, and R. A. Kach. Wave attenuation in catalytic converters: reactive versus dissipative effects. *Journal of the Acoustical Society of America*, 103 (2), 935–943, 1998.

[2] S. Allam, and M. Åbom. Sound propagation in an array of narrow porous channels with application to diesel particulate filters. *Journal of Sound and Vibration*, 291 (3-5), 882–901, 2006.

[3] C. Jiang, T. W. Wu, M. B. Xu, and C. Y. R. Cheng. BEM modeling of mufflers with diesel particulate filters and catalytic converters. *Noise Control Engineering Journal*, 58 (3), 243–250, 2010.

[4] F. D. Denia, J. Martínez-Casas, L. Baeza, and F. J. Fuenmayor. Acoustic modelling of exhaust devices with nonconforming finite element meshes and transfer matrices. *Applied Acoustics*, 73 (8), 713–722, 2012.

[5] F. D. Denia, E. M. Sánchez-Orgaz, J. Martínez-Casas, and R. Kirby. Finite element based acoustic analysis of dissipative silencers with high temperature and thermal-induced heterogeneity. *Finite Elements in Analysis and Design*, 101, 46–57, 2015.

[6] R. Kirby. A comparison between analytic and numerical methods for modelling automotive dissipative silencers with mean flow. *Journal of Sound and Vibration*, 325 (3), 565–82, 2009.

[7] A. Selamet, and Z. L. Ji. Acoustic attenuation performance of circular expansion chambers with offset inlet/outlet: I. Analytical approach. *Journal of Sound and Vibration*, 213 (4), 601–617, 1998.

[8] R. Kirby, and F. D. Denia. Analytic mode matching for a circular dissipative silencer containing mean flow and a perforated pipe. *Journal of the Acoustical Society of America*, 122 (6), 3471–3482, 2007.

[9] F. D. Denia, A. Selamet, M. J. Martínez, and F. J. Fuenmayor. Sound attenuation of a circular multi-chamber hybrid muffler. *Noise Control Engineering Journal*, 56 (5), 356–364,2008.

[10] F. D. Denia, A. G. Antebas, A. Selamet, and A. M. Pedrosa. Acoustic characteristics of circular dissipative reversing chamber mufflers. *Noise Control Engineering Journal*, 59 (3), 234–246, 2011.

[11] Z. Fang, Z. L. Ji, and C.Y. Liu. Acoustic attenuation analysis of silencers with multi-chamber by using coupling method based on subdomain division technique. *Applied Acoustics*, 116, 152–163, 2017.

[12] R. Kirby. Transmission loss predictions for dissipative silencers of arbitrary cross section in the presence of mean flow. *Journal of the Acoustical Society of America*, 114 (1), 200–209, 2003.

[13] L. Yang, Z. L. Ji, and T. W. Wu. Transmission loss prediction of silencers by using combined boundary element method and point collocation approach. *Engineering Analysis with Boundary Elements*, 61, 265–273, 2015.

[14] F. D. Denia, E. M. Sánchez-Orgaz, L. Baeza, and R. Kirby. Point collocation scheme in silencers with temperature gradient and mean flow. *Journal of Computational and Applied Mathematics*, 291, 127–141, 2016.

[15] F. D. Denia, A. G. Antebas, R. Kirby, and F. J. Fuenmayor. Multidimensional acoustic modelling of catalytic converters. *The Sixteenth International Congress on Sound and Vibration*, Kraków, 2009.

[16] M.L. Munjal. Acoustics of Ducts and Mufflers, Wiley, NewYork, 2014.

[17] J.F. Allard and N. Atalla. Propagation of Sound in Porous Media, Wiley, NewYork, 2009.

Starting points for Newton's method under a center Lipschitz condition for the second derivative

J. A. Ezquerro[♭] [*], M. A. Hernández-Verón[♭], and Á. A. Magreñán[†]

(♭) Universidad de La Rioja, Departamento de Matemáticas y Computación,
Edificio CCT. Calle Madre de Dios, 53, 26006 Logroño, Spain,

(†) Universidad Internacional de La Rioja, Escuela de Ingeniería,
Avenida Gran Vía Rey Juan Carlos I, 41, 26002 Logroño, Spain.

November 30, 2016

1 Introduction

By using mathematical modelling, many problems from computational sciences and other disciplines can be brought in the form of the equation $F(x) = 0$, where F is a nonlinear operator defined on a nonempty open convex subset Ω of a Banach space X with values in a Banach space Y. As the solutions of these equations can rarely be found in closed form, we usually look for numerical approximations of these solutions. That is why the solution methods for these equations are iterative. For this, starting from one initial approximation x_0 of a solution x^* of the equation $F(x) = 0$, a sequence $\{x_n\}$ of approximations is constructed such that $\|x_{n+1} - x^*\| < \|x_n - x^*\|$, $n \geq 0$, that leds to the sequence $\{x_n\}$ converges to the solution x^*.

The study about convergence matter of iterative procedures is usually centered on two types: semilocal and local convergence analysis. The semilocal convergence matter is, based on the information around an initial point, to

[*] e-mail:jezquer@unirioja.es

give criteria ensuring the convergence of iterative procedure; while the local one is, based on the information around a solution, to find estimates of the radii of convergence balls.

It is well-known that Newton's method,

$$x_{n+1} = x_n - [F'(x_n)]^{-1} F(x_n), \quad n \geq 0, \quad \text{with } x_0 \text{ given},$$

is the one of the most used iterative methods to approximate the solution x^* of $F(x) = 0$. We analyse the semilocal convergence of the method from conditions on the starting point x_0 and the operator F, along with a condition that connects the previous conditions. The first semilocal convergence result for Newton's method in Banach spaces was given by Kantorovich [4] under the following conditions:

(C1) There exists $\Gamma_0 = [F'(x_0)]^{-1} \in \mathcal{L}(Y, X)$, for some $x_0 \in \Omega$, with $\|\Gamma_0\| \leq \beta$ and $\|\Gamma_0 F(x_0)\| \leq \eta$, where $\mathcal{L}(Y, X)$ is the set of bounded linear operators from Y to X.

(C2) $\|F''(x)\| \leq K$ for $x \in \Omega$.

(C3) $K\beta\eta \leq \frac{1}{2}$.

A few years later, Ortega observes that condition (C2) implies that F' is Lipschitz continuous in Ω and presents in [5] a variant of the result given by Kantorovich where (C2) is replaced by condition:

$$\|F'(x) - F'(y)\| \leq L\|x - y\| \quad \text{for} \quad x \in \Omega. \tag{1}$$

A little later, in [2], conditon (C2) is relaxed by using this condition "centered" in the starting point x_0:

$$\|F'(x) - F'(x_0)\| \leq L_0\|x - x_0\| \quad \text{for} \quad x \in \Omega,$$

what is known as center Lipschitz condition for F'.

The use of the previous condition instead of condition (C2) leads to condition (C3) is replaced by another more restrictive: $L_0\beta\eta \leq \frac{14-4\sqrt{6}}{25} = 0.1680816\ldots$ This implies that the domain of starting points for Newton's method is smaller, since the point x_0 satisfies the conditions of semilocal convergence or there is no possibility of choosing another starting point. As a consequene, the domain of starting points for Newton's method consists of a single point, x_0, or is an empty set and Newton's method is never convergent.

To avoid the previous problem that presents the last condition, we use in [1] a center condition on an auxiliary point and obtain a domain of starting points which is not reduced to a point or to the empty set, since a nonempty set of possible starting points is found.

In this work, we propose to use a center condition for the second derivative of the operator F, which lead to a modification in the domain of starting points, allowing us to use the technique of the majorizing sequences [4, 5] from a scalar function. This technique allows relaxing the conditions imposed to the starting point of Newton's method.

2 Background

Huang proposes in [3] an alternative, that does not consist of relaxing the condition on the operator involved, and imposes a condition that leads to a modification, not a restriction, of the the domain of starting points. In particular, Huang proposes that F'' is Lipschitz continuous in Ω. But, if we pay attention to the proof of Huang, we see that it is not necessary that F'' is Lipschitz continuous in the entire domain Ω, since it is enough that F'' is center Lipschitz continuous at x_0. So, Huang's result on the semilocal convergence of Newton's method can be then proved under the following conditions:

(B1) There exists $\Gamma_0 = [F'(x_0)]^{-1} \in \mathcal{L}(Y, X)$, for some $x_0 \in \Omega$, with $\|\Gamma_0\| \leq \beta$ and $\|\Gamma_0 F(x_0)\| \leq \eta$; moreover, $\|F''(x_0)\| \leq M_0$.

(B2) $\|F''(x) - F''(x_0)\| \leq L\|x - x_0\|$ for $x \in \Omega$.

(B3) $6M_0^3\beta^3\eta + 9L^2\beta^2\eta^2 + 18LM_0\beta^2\eta - 3M_0^2\beta^2 - 8L\beta \leq 0$.

We observe that Huang changes the Lipschitz condition on the operator F' given in (1) for the Lipschitz condition on the operator F'' given in (B2). Obviously, condition (B2) limits the number of operator equations that can be solved by applying Newton's method.

In this work, following the idea of Huang modified and mentioned above, we propose to use condition (B2), but centered at a different point, $\widetilde{x} \in \Omega$, from the starting point x_0 of Newton's method, so that we modify condition (B2) in the following way:

$$\|F''(x) - F''(\widetilde{x})\| \leq \widetilde{L}\|x - \widetilde{x}\|, \quad \text{for} \quad x \in \Omega,$$

once the point $\widetilde{x} \in \Omega$ is fixed. This modification leads to a modification in the domain of starting points of Newton's method.

3 Semilocal convergence

As a consequence of the above mentioned, the semilocal convergence of Newton's method is now established in the following result.

Theorem. Let $F : \Omega \subseteq X \longrightarrow Y$ be a twice continuously differentiable Fréchet operator defined on a nonempty open convex domain Ω of a Banach space X with values in a Banach space Y. Suppose that the following conditions are satisfied:

(i) There exists $\widetilde{x} \in \Omega$ such that $\|x_0 - \widetilde{x}\| = \gamma$, where $x_0 \in \Omega$, and $\|F''(\widetilde{x})\| \leq \delta$.

(ii) There exists the operator $\Gamma_0 = [F'(x_0)]^{-1} \in \mathcal{L}(Y, X)$, with $\|\Gamma_0\| \leq \beta$ and $\|\Gamma_0 F(x_0)\| \leq \eta$.

(iii) $\|F''(x) - F''(\widetilde{x})\| \leq \widetilde{L}\|x - \widetilde{x}\|$ for $x \in \Omega$.

(iv) $\psi(\alpha) \leq 0$, where α is a positive real root of $\psi'(t) = 0$ and

$$\psi(t) = \frac{\widetilde{L}}{6}t^3 + \frac{1}{2}\left(\delta + \gamma\widetilde{L}\right)t^2 - \frac{t}{\beta} + \frac{\eta}{\beta}.$$

If $B(x_0, t^*) \subset \Omega$, where t^* is the smallest positive zero of polynomial ψ, then Newton's sequence, starting at x_0, converges to a solution x^* of the equation $F(x) = 0$ and $x_n, x^* \in \overline{B(x_0, t^*)}$, for all $n \geq 1$. In addition, $\|x^* - x_n\| \leq t^* - t_n$ for $n \geq 0$, where $\{t_n\}$ is defined as follows:

$$t_0 = 0, \qquad t_{n+1} = t_n - \frac{\psi(t_n)}{\psi'(t_n)}, \quad n \geq 0.$$

Observe that condition (B3) is replaced in this case by condition (iv). To prove the last result, we use the technique of majorizing sequences from polynomial ψ.

References

[1] J. A. Ezquerro and M. A. Hernández-Verón, On the domain of starting points of Newton's method under center Lipschitz conditions, *Mediterr. J. Math.*, 13(4):2287–2300,2016.

[2] J. M. Gutiérrez and M. A. Hernández, Newton's method under weak Kantorovich conditions, *IMA J. Numer. Anal.*, 20:521–532, 2000.

[3] Huang Zhengda, A note on the Kantorovich theorem for Newton method, *J. Comput. Appl. Math.*, 47:211–217, 1993.

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

[5] J. M. Ortega, The Newton-Kantorovich theorem, *Amer. Math. Monthly*, 75:658–660, 1968.

A comparison of machine learning techniques for the centerline segregation prediction in continuous cast steel slabs

P.J. García Nieto[♭] [*], E. García-Gonzalo[♭], J.C. Álvarez Antón[†],
V.M. González Suárez[†], R. Mayo Bayón[†], F. Mateos Martín[†]

(♭) Department of Mathematics, University of Oviedo,
Faculty of Sciences, 33007 Oviedo, Spain,

(†) Department of Electrical Engineering, University of Oviedo,
Campus de Viesques, 33204 Gijón, Spain

November 30, 2016

Abstract

Centerline segregation in steel cast products is an internal defect that can be very harmful when slabs are rolled in heavy plate mills. The aim of this study was to obtain a predictive model able to perform an early detection of central segregation severity in continuous cast steel slabs. This study presents a novel hybrid algorithm, based on support vector machines (SVMs) in combination with the particle swarm optimization (PSO) technique, for predicting the centerline segregation from operation input parameters determined experimentally in continuous cast steel slabs. Additionally, a multiple linear regression (MLR), a multilayer perceptron network (MLP) and a multivariate adaptive regression splines (MARS) approach, this last method also in combination with the particle swarm optimization (PSO) technique, were fitted to the experimental data with comparison purposes. Thus, some models for predicting segregation are obtained with success. Indeed, regression with optimal hyperparameters was performed

[*] e-mail: lato@orion.ciencias.uniovi.es

and coefficients of determination equal to 0.98 for continuity factor estimation and 0.97 for average width were obtained when this hybrid PSO–SVM–based model with the RBF kernel function was applied to the experimental dataset, respectively. The agreement between experimental data and the model confirmed the good performance of the latter. Finally, conclusions of this innovative research work are exposed.

1 Introduction

One of the most unpredictable defects of the steel slabs is the centerline segregation, which has a negative effect on further processing of the slabs and hence on the possible uses of the final product [1]. Specifically, this research work studies one type of macro-segregation, the central or centerline segregation, in a continuous cast steel slabs. It appears as a line of impurities in the central line of a transversal section of the slab. In this central area, cracks could appear too which can be very harmful when slabs are rolled into thick plates [1].

In this sense, the objective of this study is to evaluate the application of the support vector machines (SVMs) approach in combination with the evolutionary optimization technique known as Particle Swarm Optimization (PSO) as well as the multivariate adaptive regression splines (MARS) approach also in combination with the PSO technique and an artificial neural network known as Multilayer Perceptron (MLP) to identify central segregation in continuous cast steel slabs, comparing the results obtained.

2 Materials and methods

2.1 Experimental dataset

The experimental dataset used for the analyses was collected using a database from the continuous casting process of the LDA steelmaking belonging to the company Arcelor-Mittal located in Avilés (Northern Spain). Output variables are two indexes given by the tool used to evaluate segregation from sulfur prints: *Continuity factor* (C factor) and *Average width*. Specifically, C factor is a measure of the continuity of the segregated band and Average width is the average width of the spots forming the centerline segregation [1].

With respect to the input variables, we have selected the main input variables controlled in the casting process: (a) variables related to the analysis of steel in the tundish, that is to say, the composition of the steel (solute): Manganese (Mn); Sulfur (S); Carbon (C); Aluminum (Al); Silicon (Si); Phosphorus (P); and (b) variables related to the cooling conditions of the slab: specific flow (Specific_Flow (m^3/s)); average speed (m/s) (Ave_Speed); overtemperature in the tundish (superheating) (^{o}C); temperature in segment 8 and segment 17 (^{o}C) (Temp_Seg8 and Temp_Seg17); mold oscillation frequency (Freq_Oscillation); percentage of negative strip (Ratio_Strip). In order to determine the C factor, since big spots of segregation are more dangerous than small spots, this factor must take into account this question computing the standard deviation $\sigma\left(S_i\right)$ and the mean size $\bar{S}_i$ of the continuous areas of segregation, and the standard deviation $\sigma\left(NS_i\right)$ and the mean size $\overline{NS}_i$ of areas without segregation, respectively. Given these considerations, its expression is as follows [1]:

$$C\_Factor = \frac{\sum\limits_{i=1}^{n} S_i}{\sum\limits_{i=1}^{n} S_i + \sum\limits_{i=1}^{n} NS_i} \times \left[\frac{\sigma\left(S_i\right)}{\bar{S}_i} + \frac{\sigma\left(NS_i\right)}{\overline{NS}_i} \right]$$

Finally, the width factor is calculated as the distance between the upper and lower line of the segregation spots and the Average Width as the mean of these widths.

2.2 Support vector machine (SVM) method

The SVMs were originally developed for classification, and were later generalized to solve regression problems [2]. This last method is called *support vector regression* (SVR). The model produced by SVR depends on a subset of the training data, because the cost function for building the model tries to ignore any training data that are close (within a threshold ε) to the model prediction. When the regression SVM is applied to non–linear separable data, it is necessary to use the *kernel trick*. The reason that this kernel trick is useful is that there are many regression problems that are not linearly regressable in the space of the inputs $\mathbf{x}$, which might be in a higher dimensionality feature space given a suitable mapping $\mathbf{x} \rightarrow \psi\left(\mathbf{x}\right)$ [2].

2.3 Multivariate adaptive regression splines (MARS)

Multivariate adaptive regression splines (MARS) is a multivariate nonparametric classification/regression technique [3]. MARS model does not require any a priori assumptions about the underlying functional relationship between dependent and independent variables. The MARS model of a dependent variable $\mathbf{y}$ with M basis functions (terms) can be written as [3]:

$$\hat{\mathbf{y}} = \hat{f}_M(\mathbf{x}) = c_0 + \sum_{m=1}^{M} c_m B_m(\mathbf{x})$$

where $\hat{\mathbf{y}}$ is the dependent variable predicted by the MARS model, c_0 is a constant, $B_m(\mathbf{x})$ is the m-th basis function, which may be a single spline basis functions, and c_m is the coefficient of the m-th basis functions.

2.4 The particle swarm optimization (PSO) algorithm

The algorithm Particle Swarm Optimization (PSO) is an evolutionary optimization algorithm classified inside the group of swarm intelligence (SI) based bio-inspired algorithms [4], where a population of N_p particles or proposed solutions evolves with each iteration, moving towards the optimal solution of the problem.

2.5 Neural network: multilayer perceptron

The multilayer perceptron (MLP) consists of an input layer and an output layer and one or more hidden layers of nonlinearly-activating nodes [5]. It is a modification of the standard linear perceptron in that it uses three or more layers of neurons (nodes) with nonlinear activation functions.

3 Analysis of results and discussion

The total number of predicting variables used to build the hybrid PSO–SVM–based model, hybrid PSO–MARS–based model and MLP approach was 13. The total number of dependent variables (output variables) used to build these models was two: *Continuity factor* (C_Factor) and the *Average Width* of the spots (Ave_Width) forming the centerline segregation. Indeed, we have built three different models (specifically, the PSO–SVM–based model, the

Table 1: Coefficients of determination (R^2) and correlation coefficients (r) for the hybrid PSO–SVM–based model with RBF kernel, the hybrid PSO–MARS–based model, and multilayer perceptron fitted in this study for the C Factor.

Model	R^2/r
RBF–SVM	0.98/0.99
MARS	0.86/0.93
Multilayer perceptron	0.66/0.82

Table 2: Coefficients of determination (R^2) and correlation coefficients (r) for the hybrid PSO–SVM–based model with RBF kernel, the hybrid PSO–MARS–based model, and multilayer perceptron fitted in this study for the Average Width.

Model	R^2/r
RBF–SVM	0.97/0.98
MARS	0.80/0.89
Multilayer perceptron	0.63/0.79

PSO–MARS–based model and MLP approach) taking as dependent variables C_Factor and Ave_Width, respectively.

Table 1 shows the determination and correlation coefficients for the PSO–SVM–based model for the RBF kernel, the PSO–MARS–based model and MLP model fitted here for the C Factor prediction. Similarly, Table 2 shows the determination and correlation coefficients for the PSO–SVM–based model for the RBF kernel, the PSO–MARS–based model and MLP model fitted here for the Average Width variable. Fig. 1 shows the comparison between the C Factor values observed and predicted using the PSO–SVM–based model with RBF kernel.

Similarly, Fig. 2 shows the comparison between the Average Width values observed and predicted using the PSO–SVM–based model with RBF kernel.

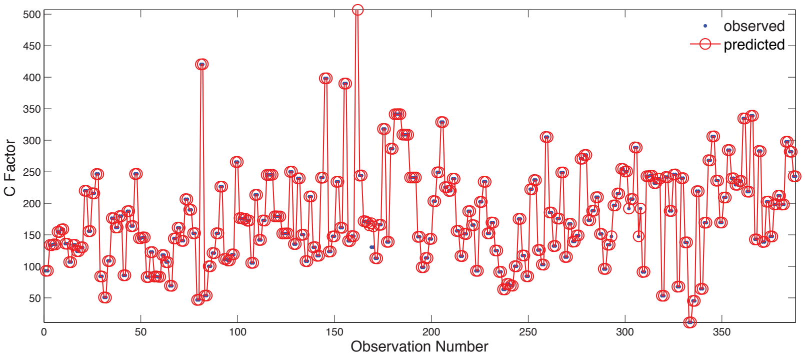


Figure 1: Comparison between the C Factor values observed and predicted by the PSO–SVM model with RBF kernel ($R^2 = 0.98$).

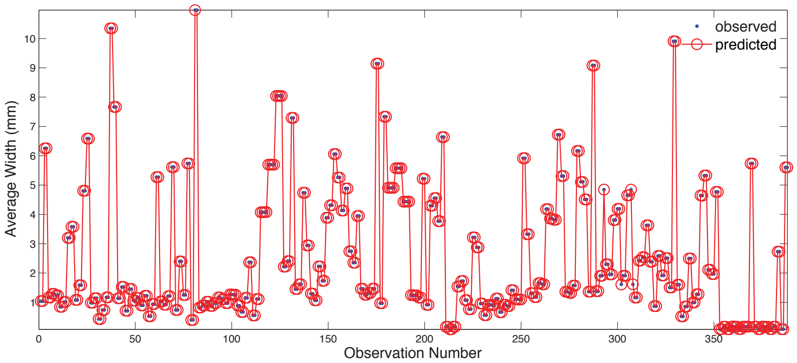


Figure 2: Comparison between the Average Width values observed and predicted using PSO–SVM model with RBF kernel ($R^2 = 0.97$).

4 Conclusions

Based on the experimental and numerical results, the main findings of this research work can be summarized as follows: segregation is a very common and serious problem in steel production. The diagnostic techniques commonly used based on the traditional methods are expensive from both the material and human points of view. Consequently, the development of alternative diagnostic techniques is necessary. Finally, the new hybrid PSO–SVM–based method with a RBF kernel function used in this work is the best choice to evaluate the segregation.

Acknowledgements

Authors wish to acknowledge the computational support provided by the Department of Mathematics at University of Oviedo. Additionally, we would like to thank Anthony Ashworth for his revision of English grammar and spelling of the manuscript.

References

[1] A.M. Díaz, L.F. Sancho, J.A. Sirgo, A.M. López, Application of techniques of dimension reduction to predict the steel quality at the end of the secondary steelmaking, in: IEEE Industry Applications Conference, 40th IAS Annual General Meeting, Hong Kong, October 2–6th 2005, pp. 537–542.

[2] I. Steinwart, A. Christmann, Support Vector Machines. New York, Springer, 2008.

[3] J.H. Friedman. Multivariate adaptive regression splines, *Annals of Statistics*, 19:1–141, 1991.

[4] R.C. Eberhart, Y. Shi, J. Kennedy, Swarm Intelligence. San Francisco, Morgan Kaufmann, 2001.

[5] S. Haykin, Neural Networks. A comprehensive foundation. New York, Prentice Hall, 1999.

[6] D. Freedman, R. Pisani, R. Purves, Statistics. New York, W.W. Norton & Company, 2007.

Improved railway train-track interaction model in curves in the high-frequency domain

J. Giner-Navarro, J. Martínez-Casas, L. Baeza, F. D. Denia and J. Carballeira

Centro de Investigación en Ingeniería Mecánica, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain. E-mail: juagina1@etsid.upv.es

EXTENDED ABSTRACT

1. Introduction

The interaction between a railway vehicle and the track is a very complex problem due to the vibrational coupling of both sub-systems through the forces appearing in the contact area. These contact forces depend on the surface imperfections, such as rail roughness and wheel out-of-roundness. Unwanted phenomena such as damage of the rolling surfaces in the form of high levels of noise and vibration [1], corrugation [2], wheelset axle fatigue [3] and stress damage are related to the large dynamic oscillation of the contact forces.

Although railways are generally considered an environmentally friendly mean of transportation, wheel-rail noise generation is one of their few environmental drawbacks. Curve squeal noise, the most annoying type of noise which generally appears when the train negotiates a sharp curve, is generated above 5 kHz according to the literature [1]. In order to get a better understanding of the phenomena, finite element (FE) wheel models have been introduced in railways research to take the flexibility of the wheelset into account, thus extending the frequency range; only very recently, further works have considered the inertial effects due to wheelset rotation running on a tangent [3] and curved track [4]. Additionally, a FE cyclic track model with a refined mesh in the contact area has been developed in [5] to extend the frequency range of validity of the track models commonly used (Timoshenko beam, [6]).

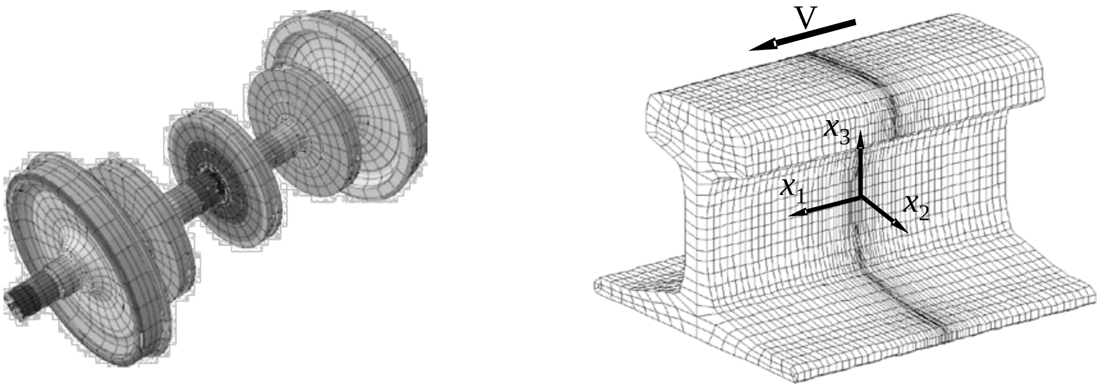
In this paper, some simulations for a curved track are carried out in the time domain using both the aforementioned wheelset and the track models and including the dynamics of the carbody and the bogie frames through two alternative strategies. Then, the results are compared with the simulations of a model using a Timoshenko beam, in order to verify if the proposed high-frequency interaction model can predict squeal noise by the occurrence of peaks in the tangential forces.

2. Overview of the mathematical approach

An Eulerian-modal approach is adopted for the FE wheelset model (see Fig. 1(a)), in which $\mathbf{\Phi}(\mathbf{u})$ is the mode shape function matrix of the free-boundary wheelset. This matrix does not depend on time since the rotation of the solid does not change the mode shape functions in fixed coordinates due to the axial symmetry of the wheelset. The modal properties are computed from a FE technique, resulting the following modal equation of motion

$$\ddot{\mathbf{q}} + \left(2\Omega\,\tilde{\mathbf{V}} + 2\,\tilde{\mathbf{P}}\right)\dot{\mathbf{q}} + \left(\Omega^2\left(\tilde{\mathbf{A}} - \tilde{\mathbf{C}}\right) + 2\Omega\,\tilde{\mathbf{S}} + \tilde{\mathbf{R}} - \tilde{\mathbf{B}} + \tilde{\mathbf{D}}\right)\mathbf{q} = \Omega^2\,\tilde{\mathbf{c}} - 2\Omega\,\tilde{\mathbf{U}} - \tilde{\mathbf{H}} + \tilde{\mathbf{N}} - \tilde{\mathbf{G}} + \mathbf{Q}_c + \mathbf{Q}_s\,, \qquad (1)$$

where Ω is the angular velocity of the wheelset; the matrices $\tilde{\mathbf{V}}$, $\tilde{\mathbf{P}}$, $\tilde{\mathbf{A}}$, $\tilde{\mathbf{C}}$, $\tilde{\mathbf{S}}$, $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{B}}$ account for inertial effects associated with the deformed configuration originated by Coriolis, centrifugal and tangential acceleration of the wheelset, as produced by the track frame motion and by the rotation of the wheelset around its axis; the vectors $\tilde{\mathbf{c}}$, $\tilde{\mathbf{U}}$, $\tilde{\mathbf{H}}$, $\tilde{\mathbf{N}}$ and $\tilde{\mathbf{G}}$ account for inertial effects not depending on wheelset deformation, which are also originated by Coriolis, centrifugal and tangential acceleration experienced by the wheelset and track frame; the diagonal matrix $\tilde{\mathbf{D}}$ is the modal stiffness matrix that contains the square of the undamped natural frequencies of the free-boundary wheelset; finally, $\mathbf{Q}_c$ and $\mathbf{Q}_s$ are the vectors of the generalized forces acting on the flexible wheelset resulting respectively from wheel-rail contact forces and from the forces applied by the primary suspension, respectively. A complete description of the wheelset formulation can be found in [4].



(a) (b)

Fig. 1. Finite element mesh of the flexible wheelset (a) and rail (b)

In this work a Timoshenko flexible beam [6] is used as the track model, which considers a simply supported infinite beam subjected to moving loads. Since Timoshenko beam is only valid up to 1.5 kHz for lateral rail vibration and up to 2 kHz for vertical vibration [1], Martínez-Casas *et al.* [5] replaced it by a tridimensional FE track model based on the Moving Element Method (MEM) initially proposed by Koh *et al.* [7]. This technique fixes the wheel-rail contact forces in a spatial point of the mesh through an Eulerian coordinate system attached to the moving vehicle, instead of a fixed coordinate system (see Fig. 1(b)). A new class of finite elements associated with the moving coordinate system are defined, hence, the mesh is moving with this mobile frame and consequently the material of the rail 'flows' into this mesh. This allows to refine only the contact area. This relative motion requires considering the material derivative for the formulation of the rail dynamics. Again, an Eulerian-modal approach is adopted and the following modal equation of motion is derived:

$$\tilde{\mathbf{M}}\ddot{\mathbf{q}} + \left(-2V\,\tilde{\mathbf{C}} + \tilde{\mathbf{C}}_w\right)\dot{\mathbf{q}} + \left(\tilde{\mathbf{K}} - V^2\tilde{\mathbf{A}} + \tilde{\mathbf{K}}_w + V\,\tilde{\mathbf{KC}}_w\right)\mathbf{q} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{F}, \qquad (2)$$

where V is the vehicle speed, $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{K}}$ are the standard mass and stiffness matrices from the FE technique, and $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{A}}$ are associated with the inertial force due to the convective velocity and the convective acceleration, respectively (the track formulation is detailed in [5]). $\tilde{\mathbf{C}}_w$, $\tilde{\mathbf{K}}_w$ and $\tilde{\mathbf{KC}}_w$ have been added for the inclusion of a viscoelastic Winkler foundation under the rail in case the discrete supports (sleepers) are modelled with this approach.

Once the formulation of the wheelset and track models have been defined, two techniques are separately developed to include the dynamics of the carbody and the bogie frames. Firstly, both elements are introduced as rigid bodies and their formulation is based on a general multibody approach from Shabana [8]; the model is capable to negotiate the entrance of the curve and set the wheelset in the full curve after the transition one (more details can be found in [9]). Secondly, a new approach has been developed in this paper. It starts from initial steady conditions of the wheel-rail contact in a curved track given by the multibody dynamic simulation ADAMS/RAIL software, from which the vertical and tangential dynamics are described by small displacements from the steady values (dashed magnitudes in Eqs. (3), (5) and (6)). Following this approach, the wheel-rail displacement is computed taking into account the initial position of the wheelset in the contact point *j*, $\Delta\bar{\mathbf{r}}_{c,j}^{w}$, and rail, $\Delta\bar{\mathbf{r}}_{c,j}^{r}$. This corresponds to the pseudo-static state that is previously calculated by means of a static equilibrium from the creep forces given by ADAMS/RAIL [10].

$$\Delta\mathbf{r}_{c,j}^{wr} = \left(\Delta\mathbf{r}_{c,j}^{r} - \Delta\bar{\mathbf{r}}_{c,j}^{r}\right) - \left(\Delta\mathbf{r}_{c,j}^{w} - \Delta\bar{\mathbf{r}}_{c,j}^{w}\right). \tag{3}$$

The incremental elastic penetration is computed by projecting the relative wheel-rail displacements in the contact point along the direction normal to the contact plane:

$$\Delta\delta = \Delta\mathbf{r}_{c,j}^{wr} \cdot \mathbf{x}_3. \tag{4}$$

Now, the tangential forces are defined from the creepages, calculated at each instant by means of an incremental estimation once computed the wheel-rail material velocity $\Delta\dot{\mathbf{r}}$, the local reference frame $\mathbf{x}_\tau$ and the pseudo-static creepages $\bar{\xi}_\tau$ given by ADAMS/RAIL:

$$\xi_\tau = \frac{\Delta\dot{\mathbf{r}}}{V} \cdot \mathbf{x}_\tau + \bar{\xi}_\tau, \ \tau = 1,2. \tag{5}$$

The normal force is defined by a non-linear simplified normal relationship [10]:

$$N = \bar{N} + \Delta N = \begin{cases} K_H\left(\bar{\delta} + \Delta\delta\right)^{3/2} & \text{if } \left(\bar{\delta} + \Delta\delta\right) > 0, \\ 0 & \text{if } \left(\bar{\delta} + \Delta\delta\right) \leq 0, \end{cases} \tag{6}$$

where $\bar{\delta}$ refers to the contact interpenetration and K_H is a constant coefficient calculated from the steady magnitudes.

3. Results

A set of simulations has been carried out using the complete high-frequency wheelset-track interaction model described previously. A high-speed ETR 500 carbody has been considered in ADAMS/RAIL for the estimation of the pseudo-static solution. The wheelset has been meshed using 20-node quadratic hexahedron, and 300 vibration modes have been considered for the truncation of the modal approach, covering a frequency range up to 8.0 kHz. Regarding the track, a UIC60 rail profile has been extruded and meshed along 42 m (and refined only around the contact point in the centre of the rail) using again 20-node quadratic hexahedron; 400

vibration modes have been selected to cover a range up to 8.5 kHz. Continuously supported track has been modelled adopting a Winkler foundation under the rail.

Curve radii are given by R^r = 120 m and 500 m, while the friction coefficient values μ = 0.20, 0.32, 0.40, 0.60 have been selected as inputs of the simulations. The vehicle speed V is set to obtain zero non-compensated acceleration when the vehicle negotiates the curve. This parameter study pretends to figure out if squeal can occur for a constant friction coefficient only through a wheel modal coupling mechanism.

For a case of R^r = 500 m and μ = 0.40, the temporal evolution of the tangential contact forces is presented in Fig. 2. The zoomed view shows a response following a harmonic behaviour with a marked principal frequency. Applying the Fourier transformation, the frequency spectrum of the tangential contact forces is presented in Fig. 3, in which a main peak around 2755 Hz corresponding to the previous zoomed view is highlighted. Secondary peaks around 4065, 4127 and 5513 Hz also appear in the spectrum. The frequency content of these peaks corresponds to the high-frequency range, justifying the need of including flexibility on both wheelset and track models.

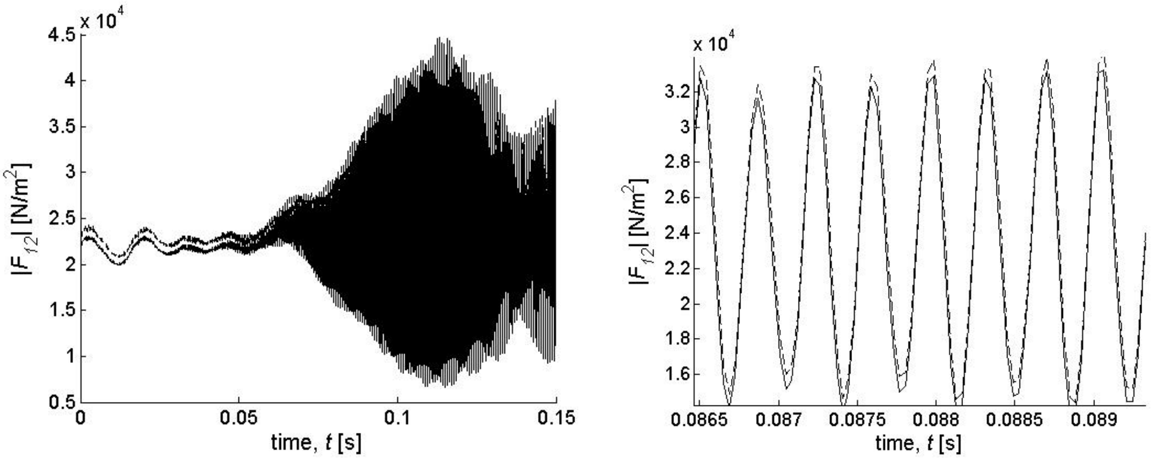


Fig. 2. Temporal evolution of the tangential contact forces. The right figure shows a zoomed view around 0.088 s.

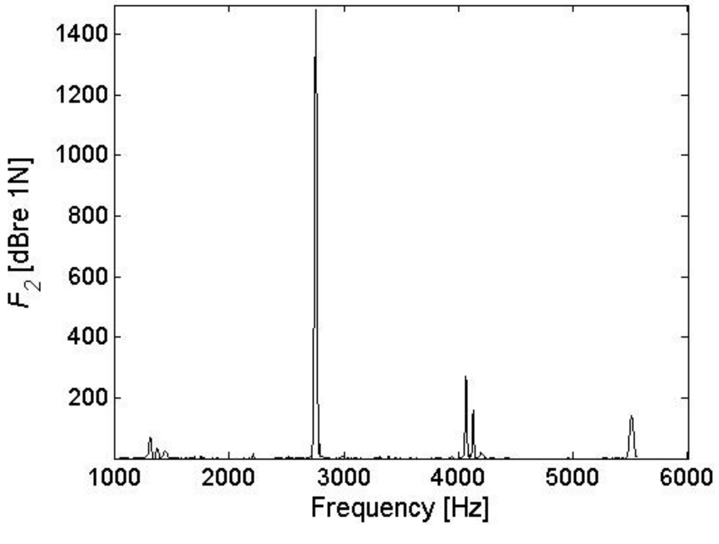


Fig. 3. Frequency spectrum of the tangential contact forces.

Table 1 gathers the corresponding frequencies of the previous peaks and the closest wheel modes associated with these frequencies. Fig. 4 shows the deformed configuration of the wheelset corresponding to the two closest eigenfrequencies around 2755 Hz (main peak) with one nodal circles, and three and four nodal diameters, respectively. The additional combinations of curve radii and friction coefficients give simulations which will be included in a future work.

Frequency [Hz]	Closest wheel modes
2755	(3,1,a), (4,1,a)
4065	Axle mode
4127	(6,0,a)
5513	(7,0,a)

Table 1. Main frequency component in simulations.

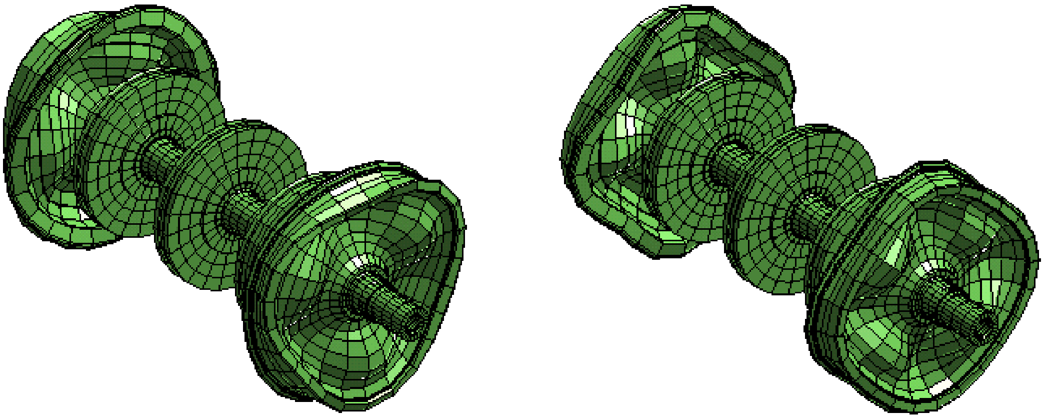


Fig. 4. Deformed configuration of the wheelset corresponding to wheel coupling modes: (3,1,a)–2747 Hz (left) and (4,1,a)–2778 Hz (right).

4. Conclusions

A new approach has been adopted for the estimation of the contact forces in curve conditions. The pseudo-static state of the wheel-rail system is given by the multibody dynamic simulation ADAMS/RAIL software. From the pre-calculated pseudo-static wheel and rail displacements, the vertical and tangential dynamics are described by small displacements. The normal force is defined by a non-linear simplified normal relationship from the elastic penetration deviation calculated in each step of the simulation and the tangential forces are estimated from the creepages derived by using the same approach.

A new model for a curved and continuously supported track based on a Winkler bedding has been developed through the Moving Element Method by adopting an Euler-modal approach to make the model more efficient. The flexibility introduced in the model widens the frequency range above the range of validity of the Timoshenko beam (up to 1.5 kHz), crucial to detect any high-frequency phenomena associated with the rail curve.

For a radius curve of 500 m and a constant friction coefficient of 0.40, the frequency spectrum of the temporal evolution of the tangential contact forces shows remarkable peaks in the high-frequency range associated with wheel coupling modes. These peaks can be associated with the squeal phenomenon, confirming that the wheel modal coupling mechanism may be sufficient to induce squeal even with constant friction coefficient.

5. Acknowledgements

The authors gratefully acknowledge the financial support of Ministerio de Economía y Competitividad and the European Regional Development Fund (project TRA2013-45596-C2-1-R), Generalitat Valenciana (project Prometeo/2016/007 and GV/2016/011 of Conselleria d'Educació, Investigació, Cultura i Esport).

6. References

[1] D. J. Thompson, *Railway Noise and Vibration: Mechanisms, Modelling and Means of Control*, Elsevier, 2009.

[2] Paloma Vila, Luis Baeza, José Martínez-Casas, Javier Carballeira, *Rail corrugation growth accounting for the flexibility and rotation of the wheelset and the non-Hertzian and non-steady state effects at contact patch*, Vehicle System Dynamics 52 (Supplement 1) (2014) 92-108.

[3] J. Martínez-Casas, L. Mazzola, L. Baeza and S. Bruni, *Numerical estimation of stresses in railway axles using a train-track interaction model*, International Journal of Fatigue 47 (2013) 18-30.

[4] J. Martínez-Casas, E. Di Gialleonardo, S. Bruni and L. Baeza, *A comprehensive model of the railway wheelset-track interaction in curves*, Journal of Sound and Vibration 333 (2014) 4152-4169.

[5] J. Martínez-Casas, J. Giner-Navarro, F. D. Denia, P. Vila and L. Baeza, *Improved railway wheelset-track interaction model in the high–frequency domain*, 309 (2017), 642-653.

[6] S. Timoshenko, D. H. Young, W. Weaver Jr., *Vibration Problems in Engineering* (4th edn), John Wiley: New York, 1974.

[7] C. G. Koh, J. S. Y. Ong, D. K. H. Chua and J. Feng, *Moving Element Method for train-track dynamics*, International Journal for Numerical Methods in Engineering 56 (2003) 1549-1567.

[8] A. A. Shabana, *Dynamics of multibody systems*. Cambridge University Press, 2013.

[9] J. Martínez-Casas, L. Baeza, E. Di Gialleonardo and S. Bruni, *Dynamic model of the track-railway vehicle interaction on curves*, 24th International Symposium on Dynamics of Vehicles on Roads and Tracks, Graz (Austria), 2015.

[10] L. Chen, Y. Zhang, W. Ren, G. Qin, *Dynamic Analysis of Mechanical Systems and Application Guide ADAMS*. Beijing: Tsinghua University Publishing Company, 2000.

A General Framework for Nonlinear Approximations with Applications to Image Restoration

Vicente F. Candela$^{\flat}$ [*], A. Falcó† [†] and Pantaleón D. Romero† [‡]

($\flat$) Universitat de València,

Departamento de Matemática Aplicada, València Burjassot 46100, Valencia (Spain),

($\dagger$) Universidad CEU-Cardenal Herrera

Departamento de Matemática, Física y Ciencias Tecnológicas,

Alfara del Patriarca, 46115,Valencia (Spain).

November 30, 2016

1 Introduction

In this paper for a given a weakly-closed (non-convex) cone in Hilbert space we establish sufficient conditions for the existence of optimal nonlinear approximations to a closed subspace generated by this (non-convex) cone. Most nonlinear problems do not only have difficulties in order to implement good projection algorithms, but also the subsets where we project the functions do not have the geometric properties necessary for classic existence results (such as convexity, for instance). The theoretical results given in this note overcome some of these difficulties. We illustrate these results by applying them to a fractional model for image deconvolution. In particular, we explain the convergence of the computational algorithm and some examples are given.

[*] vicente.candela@uv.es

[†] afalco@uchceu.es

[‡] pantaleon.romero@uchceu.es

Let V be a Hilbert space; we denote by $(\cdot,\cdot)$ and $\|\cdot\|$ a general inner product on V and its associated norm. Let $\mathcal{C}$ be a nonempty subset in V such that

(A1) $\mathcal{C}$, is a cone, that is, if $v \in \mathcal{C}$ then $\lambda v \in \mathcal{C}$ for all $\lambda \geq 0$, and

(A2) $\mathcal{C}$ is weakly closed in V.

Clearly, every closed and convex cone in V satisfy (A1) and (A2).

Let us consider the non-convex cone

$$\mathcal{C} = \left\{ u \in L_2[0,1] : u(x) = \alpha x^{\beta} \text{ where } \alpha \geq 0 \text{ and } \beta \in [0,2] \right\}.$$

It is weakly closed in $L_2[0,1]$.

Let be the map $\sigma(\cdot|\mathcal{C}) : V \to \mathbb{R}$ defined by

$$\sigma(z|\mathcal{C}) = \max_{\substack{w\in\mathcal{C} \\ \|w\|=1}} |(z,w)|, \tag{1}$$

Proposition 1. *For each $z \in V$, there exists $v^* \in \mathcal{C}$ such that*

$$\|z - v^*\|^2 = \min_{v\in\mathcal{C}} \|z - v\|^2 = \|z\|^2 - \sigma(z|\mathcal{C})^2. \tag{2}$$

Moreover, $\sigma(z|\mathcal{C}) = \|v^*\|$, and

$$(z - v^*, v^*) = 0. \tag{3}$$

We will denote by $U(\mathcal{C}) = \overline{\operatorname{span}\mathcal{C}}^{\|\cdot\|}$ the closed linear subspace generated by $\mathcal{C}$. Now, we introduce the set

$$\mathcal{V}(z|\mathcal{C}) = \{w \in \mathcal{C} : \|w\| = 1 \;\text{ and }\; \sigma(z|\mathcal{C}) = |(z,w)|\}. \tag{4}$$

Then the projector $\Pi(\cdot|\mathcal{C})$ can be written as

$$\Pi(z|\mathcal{C}) = \sigma(z|\mathcal{C})\mathcal{V}(z|\mathcal{C}), \tag{5}$$

which means that for $v^* \in \Pi(z|\mathcal{C})$, there exists $w^* \in \mathcal{V}(z|\mathcal{C})$ such that $v^* = \sigma(z|\mathcal{C})w^*$.

Proposition 1 allows to construct a sequence $\{e_n\}_{n\geq 0} \subset V$ by means of the following iterative scheme. Let $z_0 = 0$, and, for each $n \geq 1$, take

$$\begin{aligned} e_{n-1} &= z - z_{n-1}, \text{ and update} && (6)\\ z_n &= \sum_{i=1}^{n} \sigma(e_{i-1}|\mathcal{C})w^{(i)}, \quad w^{(i)} \in \mathcal{V}(e_{i-1}|\mathcal{C}) && (7)\\ & && (8) \end{aligned}$$

Definition 1. *We define the $\mathcal{C}$-rank of an element $z \in V$, denoted by* $\operatorname{rank}(z|\mathcal{C})$*, as follows:*

$$\operatorname{rank}(z|\mathcal{C}) = \min\{n : \sigma(e_n|\mathcal{C}) = 0\}, \tag{9}$$

where by convention $\min(\emptyset) = \infty$*.*

Now, we state the main result of this paper:

Theorem 1. *For $z \in V$, the sequence $\{e_n\}_{n \geqslant 0}$ constructed in (6) satisfies that $\lim_{n\to\infty} e_n = e^*$ and $e^* \in U(\mathcal{C})^{\perp}$. Moreover,*

$$P_{U(\mathcal{C})}(z) = z - e^* = \sum_{i=1}^{\operatorname{rank}(z|\mathcal{C})} \sigma(e_{i-1}|\mathcal{C}) w^{(i)},$$

where $P_{U(\mathcal{C})}$ is the orthogonal projection over $U(\mathcal{C})$, and

$$\|e_n\|^2 = \|z\|^2 - \sum_{i=1}^{n} \sigma(e_{i-1}|\mathcal{C})^2 = \sum_{i=n+1}^{\operatorname{rank}(z|\mathcal{C})} \sigma(e_{i-1}|\mathcal{C})^2.$$

In consequence,

$$\|z - P_U(z)\|^2 = \|z\|^2 - \sum_{i=1}^{\operatorname{rank}(z|\mathcal{C})} \sigma(e_{i-1}|\mathcal{C})^2.$$

The above results provide a theoretical ground for a large class of nonlinear approximation problems. We will illustrate this with a particular example modelling blind deconvolution.

A problem arising frequently in image processing is that of recovering the original image from a degraded one. It is well known that an image $z_0(x, y)$ gets degraded due to different (natural or computational) causes, which can usually be mathematically formulated as follows:

$$z_1(x, y) = (K * z_0)(x, y) + n(x, y), \tag{10}$$

In [4], the authors develop a deconvolution model in the context of images degraded by weather and time conditions (in particular, artistic restoration of paintings). Under these circumstances, stochastic errors can be neglected

$n \approx 0$ (because of the focus distance or exposal time for the acquisition of the image) and the kernel can be e modelled by a Lèvy distribution.

Deconvolution problems consist of recovering the original image z_0 from the convolved, observed, one z_1. The problem should be solved in the context of Fourier transforms, due to the fundamental theorem of convolution:

$$\widehat{K * z_0}(\xi, \eta) = \widehat{K}(\xi, \eta)\widehat{z}_0(\xi, \eta) = \widehat{z}_1(\xi, \eta).$$

A naive way to deconvolve is thus to obtain $\widehat{z}_0$ by a simple division. The problem arises when $\widehat{z}_0$ is close to zero, then the direct deconvolution is unstable, not allowing the recovery of high frequences of $\widehat{z}_0$. In consequence, a regularizing term must be included in order to stabilize the problem.

In [3], the authors propose to regularize that equation by using a fractional power of the Laplacian. For each $1 \leq l \leq n$, the regularized equation appears as

$$(-\Delta)^{\beta_l} \upsilon_l + \psi \, \overline{k}_{\alpha_l}^{\beta_l} * \left(k_{\alpha_l}^{\beta_l} * \upsilon_l - \upsilon_l \right) = 0, \tag{11}$$

here $\overline{k}$ denotes the complex conjugate of k, and υ_l is obtained from

$$\widehat{\upsilon}_l(\xi, \eta) = \frac{\widehat{k}_{\alpha_l}^{\beta_l}(\xi, \eta)\widehat{\upsilon}_{l-1}(\xi, \eta)}{\epsilon(\xi^2 + \eta^2)^{\beta_l} + |\widehat{k}_{\alpha_l}^{\beta_l}(\xi, \eta)|^2}. \tag{12}$$

where ϵ and $\widehat{\upsilon}_l$ was previously computed applying the Fourier transform to the equation (11).

Since $\mathcal{C}$ is a weakly closed cone, it allows us to introduce the iterative fractional deconvolution algorithm.

1. Given $\upsilon_0 = z_1$ take $\widehat{\upsilon}_0$ and for each $l \geq 1$ proceed until convergence as follows.

2. Take $\widehat{\upsilon}_l$ given by (12) and then compute α_l, β_l such that

   $$\Pi(\log(|\widehat{\upsilon}_l(\xi, \eta)|) - \log(|\widehat{\upsilon}_{l-1}(\xi, \eta)|)|\mathcal{C}) = e_l,$$

   where $e_l(r) = \alpha_l r^{\beta_l}$, that is, we solve

   $$\min_{(\alpha_l, \beta_l) \in \mathbb{R}_+ \times [0,2]} \| \log(|\widehat{\upsilon}_l(\xi, \eta)|) - \log(|\widehat{\upsilon}_{l-1}(\xi, \eta)|) - e_l \|_{L_2[0,1]}$$

3. Put $\widehat{\upsilon}_l(\xi, \eta) := \widehat{\upsilon}_{l-1}(\xi, \eta) \cdot \exp(\alpha_l (\xi^2 + \eta^2)^{\beta_l})$ take $l = l + 1$ and goto 2.

References

[1] A. Falcó, W. Hackbusch. *On minimal subspaces in tensor representations*, Found. Comput. Math. Volume 12, Issue 6, 765-803 (2012).

[2] Candela V., Marquina A., Serna S., "A Local Spectral Inversion of a Linearized TV Model for Denoising an Deblurring",IEEE Transctions on Image Processing,2003, 12,7,pp.808-816

[3] P.D. Romero, V.F. Candela, *Blind deconvolution models regularized by fractional powers of the Laplacian*, J. Math Imaging Vis., 32, 181-191 (2008).

[4] P.D. Romero, V.F. Candela, *Mathematical models for restoration of Baroque paintings*, Lecture Notes in Computer Sciences, 4179, 24-34 (2006).

[5] P. D. Romero, V.F. Candela. *Modelos de deconvolución ciega fraccionaria. Aplicaciones a la restauración de obras pictóricas.* Servici de Publicacions de la Universitat de Valéncia, 2009, ISBN: 978-84-370-7562-4.

[6] T. Chan, C.K. Wong, *Total variation blind deconvolution*, IEEE Trans. Image Process, 7(3), 370-375 (1998).

[7] L. Rudin, S. Osher, *Total variation based image restoration with free local constraints*, Proc. IEEE Int. Conf. Image Process., 31-35 (1994).

Mathematical modelling to monitor the differences of the appraised housing values in Valencia Province.

Natividad Guadalajara[1] and Miguel Ángel Lopez[2]

[1] *Centro de Ingeniería Económica, Universitat Politècnica de Valencia, Spain*

[2] *Universitat Politècnica de Valencia, Spain*

Emails: nguadala@omp.upv.es, mangel.lopez@bde.es

1. Introduction

The aim of this paper is to perform a multivariate hedonic regression model of the price of multifamily houses in Valencia Province, and measure the differences in the value obtained from different appraisal companies. Data were obtained from certified appraisal companies.

Multivariate hedonic regression models applied to estimate housing prices have been broadly used by different authors, and most of this studies have been analyzed by Thanasi (2015).

Formerly other authors (Tabales et al., 2007; McGreal & Taltavull, 2012; Ugarte et al. 2015) have performed hedonic regression models using appraisal companies data, or offered prices, in order to measure the accuracy of the mortgage lending values in different provinces in Spain. Also Fernández et al. (2012), have produced neural network analysis to determine the incidence of location in housing prices in Valencia province. Alemán et al. (2008) developed a model of appraisals fraud detection in Mexico.

In order to optimize the number of factors with influence over the price of the house, the first step is to perform a factorial analysis to comply with the parsimonious principle. Most representative variables from each factor are selected in this stage.

Once the hedonic model is completed, the second step is to analyze if statistically significant differences appear, depending on which appraisal company has made the valuation. If differences are detected, we proceed to determine the amount of the difference and which are the companies that produce the differences.

The third step consists on developing a model, where each appraisal company is represented with a dummy variable; depending on the value of the coefficient of the dummy variable, it is possible to measure the magnitudes of the differences.

Software and variables applied: IBM SPSS software was used. We worked with two types of variables: continuous and categorical variables. Some variables have been discarded because data collected were not accurate, or because the variable was not statistically significant. An example of continuous variable is the housing area, in our study the area ranges from 28 m^2 to 639 m^2. A categorical variable is by example a lift, in this case values are 1 or 0 depending on the existence of a lift in the building or not.

One of the main problems detected to perform the model, was to comply with the regression principles. As a consequence of valuations were performed by different appraisers, and that the final aim of the study is to detect the differences between appraisers, bias and errors in the model estimates, appear due to the fact of the existence of differences in values generated by different appraisers.

2. Methodology

The regression model was estimated by Ordinary Least Square (OLS). This approach has the advantage of simplicity, and also the coefficients interpretation is very intuitive, although is difficult to fulfill the entire theoretical hypothesis.

2.1 Data and Setting

Data base used to perform the regression model and the subsequent tests and analysis, refers to 18.101 multi-family apartments located in Valencia province. All data were collected from Official appraisals of 36 Certified Appraisal Companies performed during 2014, according to legal information requirements and methodology[1].

Table 1 shows variables considered, in a first step, to perform the hedonic regression model. Dependent variable of the model is the total housing value. 10 out of 22 variables are dummies. The rest are continuous variables.

Table 1. Variables initially considered to perform the regression model

VARIABLE	MEAN	DEV STD	MÍN	MÁX	MODE	RANG	Q1	MEDIAN	Q3	SKEWNESS	KURTOSIS
VTOTAL	109.69	82.818	8.35	1.600.3	91.403	1.591.973	61.851	91.403	131.513	4	38
VUNI_AD	1.040,0	537,3	150	6.122	800	5972,42	674	916	1.263	1,6	4,4
EDAD_T	31,45	23,6	1	245	6	244	8	36	47	1,1	4,1
SUPA	103,35	32,7	28	639	90	611	82	100	120	1,8	13,9
NUBA	1,62	0,7	1	7	2	6	1	2	2	2	9,2
NUDO	2,93	1	1	9	3	8	2	3	3	0,5	2,4
IZVE	0,08	0,3	0	1	0	1	-	-	-	3,2	8
IPIS	0,11	0,3	0	1	0	1	-	-	-	2,4	4
IZDE	0,12	0,3	0	1	0	1	-	-	-	2,3	3,3
ICAL	0,22	0,4	0	1	0	1	-	-	-	1,3	-0,2
IAIR	0,3	0,5	0	1	0	1	-	-	1	0,8	-1,3
CALI	2,71	0,5	1	3	3	2	1	1	2	-1,4	0,8
CONS	2,63	0,7	1	4	3	3	2	2	3	-0,6	0,3
MCTC	0,7	4,9	0	210	0	210	-	-	-	23	715,7
MCTD	0,95	7,2	0	235	0	235	-	-	-	14	269
IASC	0,74	0,4	0	1	1	1	-	1	1	-1,1	-0,7
NPLA	2,35	2,3	0	52	0	52	1	2	4	2	15,4
TPLA	0,5	0,9	-1	1	1	2	-	1	1	-1,2	-0,6
NGAR	0,04	0,2	0	1	0	1	-	-	-	4,7	20,4
JERARQUIA	32,74	12,5	1	67	35	66	28	35	40	-0,2	-0,1
RENTM	9.609	2.456	5.24	16.025	12.041	10.776	7.860	10.391	12.041	0,00	-0,80
HABT	290.21	360.159	188	786.18	786.18	786.001	18.570	43.320	786.189	0,60	-1,60

2.2 Modelling

[1] Ministerial Order: "*Orden ECO/805/2003, sobre normas de valoración de bienes inmuebles y de determinados derechos para ciertas finalidades financieras*".

An initial model was undertaken without any variable transformation, applying OLS and taken the total house value as dependent variable and the rest of the variables as independents. Results are in general coherent to the expected values for each variable.

Consequently, variable signs show the logical relationships to the total value, in general terms. For example, the increase of the house age (EDAD_T) should lead to a decrease in the house value (coeff = -316).

However, in four variables, coefficients don't show the expected sign: porches (MCTC), sports area (IZDE), number of bedrooms (NUDO), and the average city income (RENTM).

Furthermore, two variables display coefficients not statistically significant at 5% level: terrace surface, and the type of floor (penthouse, basement, others).

There could be several reasons to explain those coefficients. After some data checking, was detected that in some cases, data wear not properly collected or reported.

Although model results are acceptable, the model setting requires enhancing and simplifying it.

In order to reduce the model dimension and comply with the parsimony principle, we developed the exploratory factorial analysis. Others (Stadelmann, 2010), have performed with the same objective a Bayesian analysis, pointing out the great importance of location factors.

Factorial analysis disclosures 8 components: **neighborhood** (AREA), **house structure** (PROGRAMA), **condo facilities** (URBANIZACION), **age**, **climate**, **extras**, **house quality**, and **penthouse**. Malpezzi (2002) defined some relevant variables such as the number and type of rooms, area, house typology, heating and air conditioned, age, parking spaces, neighborhood, distance to city COB, and date of data collection.

2.3 Final multivariate hedonic regression model

After factorial analysis, variables with a higher influence on the dependent variable have been detected, and consequently, its results are considered to develop the hedonic model.

Others have settled a similar methodology: Lehner (2011), regarding to Singapore, performs an exhaustive analysis of similar studies from South-East Asia, and Reddy (2015) shows the advantage of not being necessary to develop different models for every sub-area, achieving good results even if the estimators are not optimal.

Like some other works (Hu et al., 2013; Goodman et al., 2003), continuous variables have been transformed into logarithmic variables. This way, the coefficients value is equal to the value of the elasticity with respect to the dependent variable (the total house value), and is enhanced residual normality. Regressors are calculated applying OLS.

Modell specification is:

$$(1)\ LN(y) = \alpha + \sum_1^N \beta_i\, LN\,(X_i) + \gamma\, D_i + \varepsilon$$

Where: y = Total property value

X_i= Continuous Variables

D_i= Dummy Variables.

ε = error

Consequently, the original model once retransformed is:

$$(2) \quad Y = [\, e^{\propto} * \; \prod_i^j x_i^{\beta}] * \; \prod_j^n e^{\gamma_j}] * \eta$$ [2]

Model setting and variables are selected according to the results obtained in the first regression model, the results from factorial analysis, and the empirical knowledge of the variables behavior.

The optimized model is compound of the variables:

Predictives:(Constante),LN_EDAD_T^2;LN_SUPA^2;IPIS;IAIR;LN_CALI_T;LN_CONS_T;IASC; LN_JERARQUIA.
Dependent Variable: LN_VTOTAL

Finally, outlier values are detected and eliminated. According to (Fotheringham et al. 2002), we have considered outlier value if studentized residual >=3. Although there are other options (Olewuezi, 2011), one of the most popular is the "Outlier Labeling Rule" , also known as Tukey's method (boxplot) .

As 111 outlier values were detected, filtered data base contains 17.990 properties.

3. Results

The final model contains the variables outlined above, plus the dummy variables of each appraisal company (table 2).

ANOVA is significant at 95% level (P-VALUE < 0,05), and Adjusted R squared is 0,798. Normality and homoscedasticity of errors is not complain. No multicollinearity problems are detected.

Table 2						
	Heteroscedasticity-Consistent Regression Results					
Variables	**Coeff**	**SE(HC)**	**t**	**P>|t|**	**Levels sig.**	
Hedonic Variables						
(Constante)	8,8659	0,0416	213,0495	0	***	
LN_EDAD_T^2	-0,0697	0,0014	-49,4615	0	***	
LN_SUPA^2	0,4601	0,0037	123,9843	0	***	
IPIS	0,1536	0,0069	22,3763	0	***	
IAIR	0,0679	0,0046	14,8717	0	***	

[2] The error term ε, requires a specific transformation (η) , Duan "Smearing" (Duan 1983)

LN_CALI_T	-0,037	0,0105	-3,538	0	***	
LN_CONS_T	-0,0959	0,0073	-13,0778	0	***	
IASC	0,2842	0,006	47,5427	0	***	
Variable with socioeconomic and location characteristics						
LN_JERARQUIA	-0,457	0,0044	-104,8688	0	***	
Dummies Variables of each Certified Appraisal Company (CAC)						**OVERVALUATION**
S1	0,201	0,023	8,85	0	***	71%
S2	0,538	0,008	68,705	0	***	66%
S3	0,393	0,035	11,255	0	***	59%
S4	0,411	0,715	0,575	0,565		58%
S5	0,509	0,125	4,072	0	***	52%
S6	0,419	0,179	2,346	0,019	**	51%
S7	0,454	0,01	43,581	0	***	51%
S8	0,049	0,03	1,628	0,104		48%
S9	0,466	0,01	49,186	0	***	47%
S10	0,191	0,008	23,669	0	***	40%
S11	0,197	0,015	13,027	0	***	38%
S12	0,119	0,014	8,651	0	***	31%
S13	0,25	0,31	0,805	0,421		28%
S14	0,409	0,025	16,265	0	***	27%
S16	0,165	0,008	21,18	0	***	23%
S17	0,382	0,016	23,59	0	***	22%
S18	0,323	0,012	26,929	0	***	22%
S19	0,336	0,096	3,487	0,001	***	22%
S20	0,239	0,01	23,649	0	***	22%
S22	0,162	0,034	4,735	0	***	21%
S23	0,054	0,039	1,38	0,168		21%
S24	0,271	0,063	4,31	0	***	18%
S25	0,152	0,014	11,029	0	***	18%
S26	0,121	0,013	9,016	0	***	16%
S27	-0,104	0,034	-3,061	0,002	***	14%
S28	0,199	0,011	17,449	0	***	13%
S29	0,208	0,012	17,279	0	***	13%
S30	0,191	0,012	15,788	0	***	6%
S31	0,195	0,034	5,765	0	***	5%
S33	0,128	0,049	2,598	0,009	***	-10%

The model does not comply with the assumption of Homoscedasticity at 95% level. This is the reason to the perform Heteroscedasticity-Consistent Regression (HCR) test (Hayes & Cai, 2007).

Only coefficients of 4 companies are no significant at 95% level of confidence. (HCR test produce similar results to the test if homoscedasticity hypothesis is fully complied, companies S8 and S13 coefficients are not significant in both cases, but S23 and S4 are significant under homoscedasticity hypothesis, and instead S24 is not)

4. Conclusions

The aim of this paper is to determine whether the certified appraisal companies show significant differences regarding the total housing values estimates in the Valencia province.

Valuation differences range from -10% al +71% (the S15 has being taken as control company in order to measure the overvaluation).

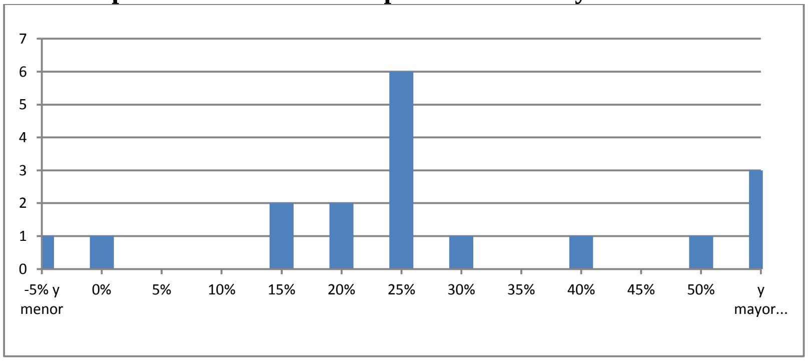
Graph 1 represents the histogram of the distributions of the companies ordered by overvaluation level. Clearly a bias in higher values can be stated, but it is important to outline that the control company could have negative bias in valuation produced by bad reported information.

Two critical types of misreporting have to be considered: Firstly, in some reports house value is divided into several components artificially (considering terraces, basement, porches or others such as single properties not linked to the others elements). The bias is negative.

Secondly, in other reports, parking spaces, storeroom, or other elements different to the house, are considered jointly with it. In this case the bias is positive.

If the misreporting is present in the control company, all the results are biased. Fortunately, is easy to trim the results if bias is known in the control company.

Graph 1: % number of companies ordered by % overvalued



Nevertheless, in order to avoid overvalue identification mistakes referable to error estimates of the model, companies with less than 60 valuations are not considered. Likewise, those companies with not significant coefficients are not considered.

Consequently are not reflected: S8 and 23 because of lack of significance (at 95%) and S31, S3, S22, S33, S19, S2, S5, S6, S4 , S13 y S24, due to have valued less than 60 properties (last two both because of lack of significance).

The conclusions are:

1) 3 companies display overvaluations higher than 50%, and no outlier were detected.
2) In 9 companies differences ranged from 20% to 50%, and outlier bias is always positive (The company with a higher level of overvaluation, shows the higher outlier bias and ratio).
3) 4 companies show overvaluation levels ranged from 10% y el 20%, low outlier ratio and negative bias.
4) The remaining two, including the control company, show undervaluation below 0%, high outlier ratio (1%-2%), but negative.

5. References

Alemán, J., Gutiérrez, J. & Gómez, G., 2008. Modelo de Detección de Fraude por Sobrevaluación del Valor de la Vivienda. *Sociedad Hipotecaria Federal, S.N.C.*, pp.1–28.

Duan, N., 1983. Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, 78(383), pp.605–610.

Fernández, L., Llorca, A., Valero, S. & Botti, VJ., 2012. Incidencia de la localización en el precio de la vivienda a través de un modelo de red neuronal artificial. Una aplicación a la ciudad de Valencia. *Catastro. Revista del Centro de Gestión Catastral y Cooperación Tributaria*, abril, pp.7–26.

Hu, G., Wang, J. & Feng, W., 2013 Multivariate Regression Modeling for Home Value Estimates with Evaluation using Maximum Information Coefficient. p.2013.

Goodman, A.C., Thibodeau, T.G. & Allen C., 2003. Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, 12(3), pp.181–201.

Hayes, A.F. & Cai, L., 2007. Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, 39(4), pp.709–722.

Lehner, M., 2011. Modelling housing prices in Singapore applying spatial hedonic regression. *ETH Zürich*, (July), p.108.

Malpezzi, S., 2002. Hedonic Pricing Models: A Selective and Applied Review. *Housing Economics and Public Policy*, 2002 pp.67–89.

McGreal, S. & Taltavull, P., 2012. An analysis of factors influencing accuracy in the valuation of residential properties in Spain. *Journal of Property Research*, 29(1), pp.1–24.

Olewuezi, N.P., 2011. Note on the Comparison of Some Outlier Labeling Techniques. *Journal of Mathematics and Statistics*, 7(4), pp.353–355.

Reddy, B.Y.S., 2015. Residential Property Value Estimation via Linear Mixed Model Methods. *Journal of property tax assessment & Administration*, 12(2), pp.73–94.

Stadelmann, D., 2010. Which factors capitalize into house prices? A Bayesian averaging approach. *Journal of Housing Economics*, 19(3), pp.180–204.

Tabales, J.M.N., Villamandos, N.C. & Torre, G.M.V. de la, 2007. Aproximación a la valoración inmobiliaria mediante la metodología de precios hedónicos (mph). *Conocimiento; innovación y emprendedores: camino al futuro Universidad de La Rioja.*, p.190.

Thanasi, M., 2014. Hedonic Pricing Model – Literature Review. Paper presented at the 2d Scientific Conference, organized by FEUT, Tirana, Albania.

Ugarte, M.D., Goicoa, T. & Militino, A.F., 2015. Searching for housing submarkets using mixtures of linear models. In *Spatial and Spatiotemporal Econometrics*. pp. 259–276.

Epidemiological approach to forecast water demand consumption through SAX

C. Navarrete-López[♭][*], B. M. Brentan[†], M. Herrera[◇], E. Luvizotto Jr.[†],
J. Izquierdo[‡], and R. Pérez-García[‡]

(♭) Faculty of Environmental Engineering,
Universidad Santo Tomás, Bogotá (Colombia),

(†) LHC – Faculty of Civil Engineering,
Universidade Estadual de Campinas, Campinas (Brazil),

(◇) EDEn – Dept. of Architecture and Civil Eng.
University of Bath, Bath (UK),

(‡) Fluing – Institute for Multidisciplinary Mathematics,
Universitat Politècnica de València, Valencia (Spain).

November 30, 2016

1 Introduction

Water demand forecasting is paramount for water utilities both for project and operation processes. Several methods proposed in the literature use classical approaches such as multivariate linear regression, auto-regressive integrated moving average and its variants, or machine learning and artificial intelligence methods, such as artificial neural networks, support vector regression and, more recently, various hybrid models [1].

Epidemiological data analysis approaches (EDAAs) can be understood as whole-system approaches that focus on empirical research and provide a multidisciplinary framework to better study and understand customer water

[*] e-mail: claudianavarrete@usantotomas.edu.co

demand behaviour together with new capabilities to analyse various risks and vulnerabilities related to water distribution. The classical approaches on epidemiological studies are associated with health-related states or events in specified populations, and applications on control of the different problems that arise in such a context. However, recent advances in Energy on Buildings [2] and also in Hydraulics [3] point to epidemiology as a promising data analysis tool-set with several concepts that can be adapted to other engineering applications.

This paper proposes a novel EDAA for the technical management of urban water distribution systems (WDSs). The approach is specifically developed for the analysis of urban water demand. This subject is one of the main topics in water supply, and several approaches have been proposed in the literature [1, 4]. EDAAs intend to ease simultaneous analyses of water demand in various areas of a WDS. The main advantage is to study the water network balance in multiple discrete areas of supply, and to provide a way of comparison of water use and costumer behaviour depending on the characteristics of each zone. The current proposal ultimately handles these objectives through a Symbolic ApproXimation (SAX) [5] of the individual values of each of the targeting time series.

2 Epidemiology-based forecast model

Epidemiology-based forecast models traditionally involve the assessment of relations between time series, i.e. how time series change, and applying usual forecasting methods. It is also typical to work with multiple time series models to establish associations between exposure to a threat and health outcome [6]. It is a usual requirement to work with transfer functions and intervention models that limit the complexity of the desired model performance [7]. This work adapts the epidemiology-based forecast framework to analysing water demand in the different parts in which a WDS is usually divided.

2.1 Epidemiology-based forecast for water demand

In water demand forecasting, an EDAA has the benefit of providing insight into the impact on water consumption linked to the effects of such disparate events as valve manoeuvres or extreme weather conditions, among others. Under the epidemiology-based forecast model, it is possible to disaggregate those previously mentioned effects for the individual district metered areas (DMAs) in which a WDS is usually divided [8]. A patient-level data analysis may propose a different predictive model for each DMA, and then, the study may be completed through a suitable correlation analysis among the different DMAs depending on their inter-connectivity level [9].

2.2 Epidemiology-based forecast and SAX

This paper introduces SAX to deal with epidemiology-based forecast models. Its use is suitable, as SAX is a powerful tool to compute distances between time series and to find out how they are related to each other [5]. SAX follows a two-step process: (1) Piecewise Aggregate Approximation (PAA), which divides the time series dataset into equally spaced segments and computes the average of each segment. (2) Conversion of a PAA sequence into a series of letters. SAX has been primarily developed to reduce the dimensionality of a numerical series into a short chain of characters. However, it is also useful for estimating time series distances and correlations based on pattern matching on the symbolic strings. Furthermore, SAX provides a suitable framework to foster the study of potential event effects straightforwardly by Suffix Tree Analysis [10] on the alphabetic composition of symbolic strings.

3 Analysis of urban water demand

The current work proposes the implementation of SAX on large water consumption time series for related DMAs in utility WDSs. The practical approach is to use SAX to find out patterns, or "words", for each time series, computing their distances to study a potentially different water use depending on the DMA, and to ultimately scrutinize water demand similarities. The process puts into practice the SAX epidemiology-based forecast model introduced in Section 2.

3.1 Case-study description

The presented methodology is applied to 4 DMAs from Franca, a Brazilian city with approximately 315,000 inhabitants. The choice of these 4 DMAs for our analyses takes into account their spatial distribution and hydraulic properties. Considering the number of demand nodes, 3 of the DMAs are of similar size: `SA/3` with 2168 connections, `Leporace`, 2506 connections and `SA/ZA`, 2728 connections; these 3 DMAs are responsible to supply urban districts with small businesses, a typical configuration of Brazil's residential areas. The fourth DMA included in the case-study is called `ETA/ZA`, with 10439 connexions. This DMA supplies a prison, thus it is a sector of particular importance for Franca's water management. The time series for each of the 4 DMAs under study are composed of 4,000 water demand consumption data metered in litres per second. As the data has been collected every hour, it makes approximately a total length of approximately 5 months.

3.2 Results

The 4 time series corresponding to each DMA of the case-study are coded using SAX. The most suitable number of segments forming the PAA partition is 400 with an alphabet of 4 letters {`a,b,c,d`} (corresponding to various levels of water demand ranging from `a`, lower, to `d`, higher). Figure 1 shows how the PAA tuning process considers both dimensionality reduction and the sum of squares (SS) of the distance between the PAA averages and the original time series values.

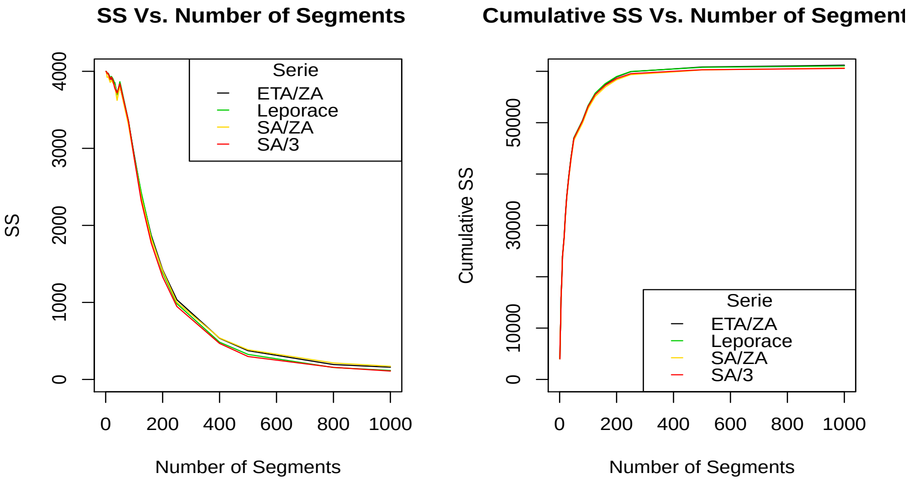


Figure 1: Number of segments for the PAA configuration

Computing the distance (MINDIST) between two SAX words requires looking up the distances between each pair of symbols, squaring them, summing them, taking the square root and multiplying by the square root of the compression rate. Table 1 shows the MINDIST values between the 4 DMAs. The DMAs `Leporace`, `SA/ZA` and `SA/3` are closer in SAX distance computed by differences at their corresponding codifications in a 400-letter word per each time series. Furthermore, SAX coded time series corresponding to `SA/ZA` can be obtained in combination with the other 3 DMAs.

Table 1: MINDIST for the 4 DMAs of the case-study

	ETA/ZA	Leporace	SA/ZA	SA/3
ETA/ZA	0	–	–	–
Leporace	4.77	0	–	–
SA/ZA	5.22	0.00	0	–
SA/3	4.27	2.13	2.13	0

The longest SAX pattern, `aaddcdbaaddddbaacdddcaa`, is found by using suffix trees on the SAX words. This is the longest sub-string common to all the series. This pattern is in the same position in the case of `ETA/ZA` and `Leporace`. Considering that each letter represents 10 water consumption registers, this pattern is showing the same tendency for 230 hours of the original time series (that is more than 9 consecutive days of similar level of water demand). Several shorter patterns have also been found in the four DMAs and each one comes up at least twice within each series.

4 Conclusions

The epidemiological data analysis approach is a new framework in energy related topics. The proposal of this work is to tailor it as a novel approach for Hydraulic Engineering as well. This has been partially done by adapting an epidemiology-based forecasting process to water demand prediction. EDAAs need further investigation in water distribution systems as a promising framework for related tasks to assess water network resilience.

The second proposal of this work is to use SAX for epidemiology-based forecast modelling. SAX performs well as a support for finding motifs and surprising patterns from time series of water demand and for approaching visualization analytics. SAX is suitable to work with long time series and it is an accurate framework for computing similarities. It can be further expanded to also use quantiles or local regressions instead of the average.

References

[1] Brentan B, Luvizotto Jr E, Herrera M, Izquierdo J, and Pérez-García R. Hybrid regression model for near-real time urban water demand forecasting. *Computational and Applied Mathematics*, 309(1):532-541, 2017.

[2] Hamilton IG, Summerfield AJ, Lowe R, Ruyssevelt P, Elwell CA, and Oreszczyn T. Energy epidemiology: a new approach to end-use energy demand research. *Building Research & Information*, 41(4):482-97, 2013.

[3] Bardet JP and Little R. Epidemiology of urban water distribution systems. *Water Resources Research*, 50(8):6447-65, 2014.

[4] Herrera M, Torgo L, Izquierdo J, and Pérez-García R. Predictive models for forecasting hourly urban water demand. *Hydrology*, 387(1-2):141-150, 2010.

[5] Lin J, Keogh E, Wei L, and Lonardi S. Experiencing SAX: a novel symbolic representation of time series *Data Mining and Knowledge Discovery*, 15(2):107-144, 2007.

[6] Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, and Armstrong B. Time series regression studies in environmental epidemiology *International Journal of Epidemiology*, 42:1187-1195, 2013.

[7] Helfenstein U. The Use of Transfer Function Models, Intervention Analysis and Related Time Series Methods in Epidemiology *International Journal of Epidemiology*, 20(3):808-815, 1991.

[8] Herrera M, Improving water network management by efficient division into supply clusters. Valencia, PhD Thesis - Universitat Politècnica de València, Spain, 2011.

[9] Herrera M, Abraham E, and Stoianov I. A graph-theoretic framework for assessing the resilience of sectorised water distribution networks. *Water Resources Management*, 30(5):1685-99, 2016

[10] Chen G, Puglisi SJ, Smyth WF. Fast and practical algorithms for computing all the runs in a string. In Combinatorial Pattern Matching (pp. 307-315). Springer Berlin Heidelberg, 2007.

Optimal sector selection for a gradual transition from intermittent to continuous water supply

[(1)]A. E. ILAYA-AYZA, [(2)]E. CAMPBELL, [(3)]J. IZQUIERDO and [(4)]R. PÉREZ-GARCÍA

[(1)]Facultad Nacional de Ingeniería, Universidad Técnica de Oruro, Ciudad universitaria s/n, Oruro, Bolivia.

[(2)]Berliner Wasserbetriebe. Neue Jüdenstraße 1, Berlin, Germany

[(1),(2),(3),(4)]FluIng-Instituto de Matemática Multidisciplinar (IMM) - Universitat Politècnica de València, Camino de Vera SN, pc: 46015, Valencia, Spain.

1. Introduction

Despite using intermittent supply should be the last measure to take in a water scarcity scenario, since it damages the system infrastructure [1-3] and involves health risks [4, 5], it still remains the main form of water access for millions of people around the world. To avoid the associated drawbacks, it is possible to guarantee continuous supply with positive and continuous pressure across the network [5, 6].

In the literature, two clearly differing tendencies that face intermittent water supply may be identified. The first aims to achieve continuous supply by reducing water losses, improving the infrastructure, and incorporating new supply sources [7-9]; the second considers the assumption of intermittent water supply as a fact and, based on this paradigm, design and operation methods are sought to minimize the negative impacts caused by this type of supply [10, 11].

The first approach to solve problems associated to intermittent supply systems seeks a transition to continuous supply. In this regard, we perform the following classification: direct transition, involving large investments with results in the short/mid-term; and gradual transition, limited by the available water company resources, and is the result of a set of steps seeking sustainable long-term continuous supply.

A difficulty for the transition to continuous supply is economic scarcity. Under this condition, the water company does not have the possibility to make the large investments needed for a direct transition. Thus, profitable and planned long-term strategies should be sought. Therefore, the gradual transition is a good option.

An intermittent water supply network has sectors with hourly differentiated water supply. These sectors compete in the gradual transition process configured in stages, to have continuous supply until the entire network reach this state. For the selection of sectors, it is necessary to improve the network previously, also gradually. In this paper, the sector selection process is analysed.

Thus, in each stage some sectors continue with intermittent supply and others already have continuous supply. For the selection of the sectors that have continuous supply the benefit of the greatest number of system users should be considered, ensuring equity of supply in sectors that are still with intermittent supply and convenience for the actions of water company operation and maintenance.

2. Methodology

The problem of optimal sector selection start to have continuous supply in each of the stages of the network upgrade is addressed by using multi-criteria analysis and genetic algorithms. Using binary variables, which define sector intermittency, qualitative criteria for non-objective variables, and the non-linear relationship between flow and head loss in the mathematical model of the network water turn the optimization process into a complex task.

The selection of sectors that change to continuous supply must guarantee the supply equity of sectors that still work with intermittent supply. We propose a sector evaluation based on the equity index (I_{eq}), which uses the uniformity (CU) of all nodes (n_e) of sector n, the supplied and demanded flow rate (s_k), the theoretical maximum flow (Q_{maxt}), and the required maximum flow (Q_{maxr}) of each sector.

$$CU = 1 - \frac{\sum \left| s_k - s_{average} \right|}{s_{average} \cdot n_e},$$

$$I_{eq} = \frac{Q_{maxt}}{Q_{maxr}} \cdot CU \ .$$

The network topology is included in the optimization through a reorganization of the network as a directed graph, with adjacency matrix A, to establish an order in the selection and ensure a more efficient transition to continuous water supply. By multiplying the transposed of A with a selection vector S, we obtain a vector G_e that defines the sector possibility to be selected.

$$A^T \times S = G_e.$$

The transition process involves a pattern supply change in the whole network, from an intermittent pattern to a continuous pattern.

First, a vector H represents every sector water delivery schedule, since sectors belong to an intermittent supply system. The product of H_n by the average volume during a supply period of time (V_{sn}) of sector n gives the supply vector B_n.

$$H_n = \left(h_{1n},\, h_{2n}, \ldots,\, h_{mn} \right), \text{ with } h_{ij} = \begin{cases} 1 & \text{hour with supply} \\ 0 & \text{hour without supply} \end{cases} \quad ,$$

$$B_n = Vs_n \cdot \left(h_{1n}, h_{2n}, \ldots, h_{mn} \right) = \left(b_{1n}, b_{2n}, \ldots, b_{mn} \right).$$

In continuous supply, the daily volume has to be multiplied by every demand factor per hour,

$$K = \left(k_1,\, k_2, \ldots,\, k_m \right),$$

which gives the consumption pattern of each sector

$$A_n = \frac{Vd_n}{m} \cdot \left(k_1, k_2, \ldots, k_m \right) = \left(a_{1n}, a_{2n}, \ldots, a_{mn} \right).$$

We propose two vectors to define every sector status. First, Y defines the intermittency status.

$$Y = (y_1, y_2, \ldots, y_n),\ y_j = \begin{cases} 1 & \text{sector with intermittent supply} \\ 0 & \text{sector with continuous supply} \end{cases},$$

Then, *X* is complementary to vector *Y*, and is related to a continuous supply.

$$X = (x_1, x_2, \ldots, x_n),\ x_j = \begin{cases} 1 & \text{sector with continuous supply} \\ 0 & \text{sector with intermittent supply} \end{cases}.$$

Based on these values, we define a transition curve using vector *T*, whose maximum value represents the transition process peak flow Q_t.

$$t_m = \sum_{j=1}^{n} \left(a_{mj} \cdot x_j + b_{mj} \cdot y_j \right),\ T = (t_1,\ t_2,\ \ldots,\ t_m),\ \text{being}\ Q_t = \max(t_1, t_2, \ldots, t_m).$$

To define the objective function and the various constraints, the criteria and weights used are: number of consumers (wc_j), pressure (wp_j), distance from source of supply (wd_j), equity in the supply sector (we_j), and operation difficulty (wo_j). The perception of water company experts on sector operating conditions is introduced through surveys based on the AHP methodology [12, 13]. Also, the network topology is included in the optimization process with vectors G_e.

$$\text{Maximize} \sum_{j=1}^{n} \left(wc_j \cdot x_j + wp_j \cdot x_j + wd_j \cdot x_j + we_j \cdot x_j + wo_j \cdot x_j \right)$$

$$\text{Subject to}:\ x_1 + y_1 = 1,\ x_2 + y_2 = 1,\ \ldots,\ x_n + y_n = 1,$$

$$t_1 \le Q_t,\ t_2 \le Q_t,\ \ldots,\ t_m \le Q_t,$$

$$Pc_1 \ge P_{ref},\ Pc_2 \ge P_{ref},\ \ldots,\ Pc_m \ge P_{ref},$$

$$x_1 \le g_{e1},\ x_2 \le g_{e2},\ \ldots,\ x_n \le g_{en},$$

The minimum pressure calculated in the network (P_c) must be higher than the assumed reference pressure (P_{ref}).

3. Example of implementation

We perform the transition process in one of the water supply sub-systems of Oruro City (Bolivia). Part of Oruro's network works with intermittent supply. The selected sub-system has 15 sectors that are fed by a solely source.

The network will be gradually improved in three stages. In this process, pipes will be substituted and consequently hydraulic conditions will be improved.

Stage	Replaced pipe	Investment (Bs)
First	P-12	562303.90
Second	P-17	581264.92
Third	P-11	532875.95
	P-2	
	P-13	
Total		1676444.78

We take experts' advice about the importance of each criterion for the sector-selection process.

Network improvements produce different hydraulic scenarios that must be evaluated.

Stage	Number of sectors with intermittent water supply	Number of sectors with continuous water supply
First	15	0
Second	2	13
Third	0	15

One of the most important restrictions to select a sector is the node pressure, which needs to be higher to be viable. The reference pressure (P_{ref}) for every selection-process stage is 10 m.

In the first stage, no sector turns to continuous supply, since the optimization problem restrictions are not met. Figure 1 shows the group of sectors that can work with a continuous supply after the second stage. In the third stage, the remaining intermittent sectors are selected.

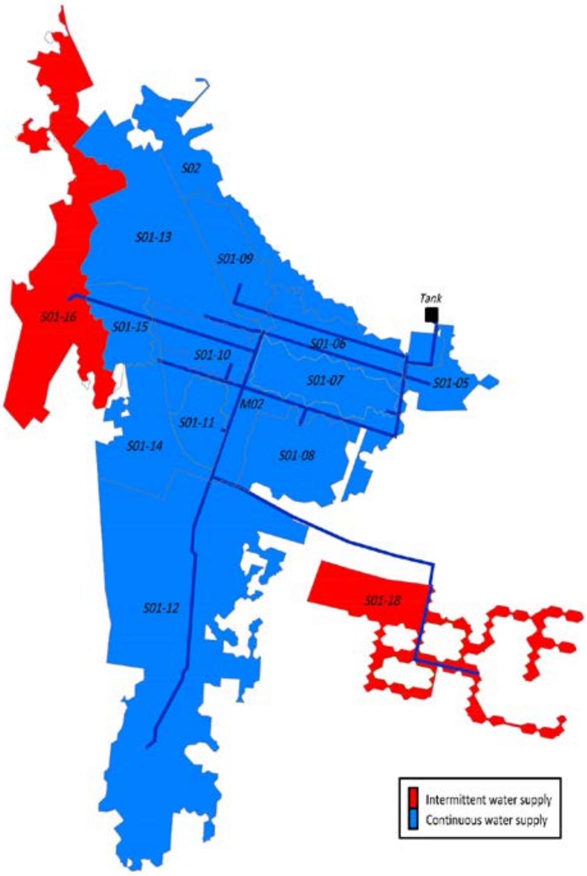


Figure 1. Second stage sectors selection.

Finally, the whole sub-system works with continuous supply, and the incoming pressures for every sector are higher than 20 m.

4. Conclusions

This study makes it possible a transition from intermittent to continuous supply based on an optimal sector selection in every network-improvement stage.

To improve drinkable water services in systems with economic limitations, a gradual transition is a recommendable option. By including the network topology in the optimizing process, we obtain more efficient solutions.

Including water-company experts' opinion in the optimizing process is very important, because intermittent-supply systems management is usually empirical.

Dedication

In memoriam Rafael Pérez-García.

References

[1] S. V. Dahasahasra, «A model for transforming an intermittent into a 24x7 water supply system,» *Geospatial today,* pp. 34-39, 2007.

[2] F. Faure y M. Pandit, «Intermittent Water Distribution,» 2010. [En línea]. Available: http://www.sswm.info/category/implementation-tools/water-distribution/hardware/water-distribution-networks/intermittent-water.

[3] B. Charalambous, «The Effects of Intermittent Supply on Water Distribution Networks,» de *Water Loss 2012*, Manila, Philippines, 2012.

[4] S. Tokajian y F. Hashwa, «Water quality problems associated with intermittent water supply,» *Water Sci Technol.*, vol. 47, nº 3, pp. 229-234, 2003.

[5] E. Kumpel y K. L. Nelson, «Mechanisms affecting water quality in an intermittent piped water supply,» *Environmental science & technology,* vol. 48, nº 5, pp. 2766-2775, 2014.

[6] E. E. Geldreich, Microbial quality of water supply in distribution systems, Boca Raton, Florida: CRC Lewis Publishers, 1996.

[7] A. C. McIntosh, Asian water supplies. Reaching the Urban Poor, Asian Development Bank, 2003.

[8] R. Franceys y A. Jalakam, «The Karnataka Urban Water Sector Improvement Project. 24x7 Water Supply is Achievable,» 2010. [En línea]. Available: http://www.wsp.org/sites/wsp.org/files/publications/WSP_Karnataka-water-supply.pdf.

[9] R. Mrunalini, «Water supply scheme to provide 24 x 7 water case study: science city area (zone-i), Ahmedabad,» *Sarjan SOCET Journal of Engineering & Technology,* pp. 12-22, 2015.

[10] K. Vairavamoorthy, E. Akinpelu, Z. Lin y M. Ali, «Design of sustainable water distribution systems in developing countries,» de *ASCE conference*, Orlando, Florida, 2001.

[11] M.-J. Soltanjalili, O. B. Haddad y M. A. Mariño, «Operating Water Distribution Networks during Water Shortage Conditions Using Hedging and Intermittent Water Supply Concepts,» *Journal of Water Resources Planning and Management,* vol. 139, nº 6, pp. 644-659, 2013.

[12] T. L. Saaty, «A scaling method for priorities in hierarchical structures,» *Journal of mathematical psychology,* vol. 15, nº 3, pp. 234-281, 1977.

[13] T. L. Saaty y L. Vargas, Models, Methods, Concepts & Applications of the Analytic Hierarchy Process, New York: Springer, 2012.

FLEET-ASSIGNMENT ON TIME-SPACE NETWORKS WITH STABLE MARRIAGES AND COLLEGE ADMISSION ALGORITHMS

J. Alberto Conejero[♭], Cristina Jordán[†,*], Esther Sanabria-Codesal[‡]

(♭) Instituto Universitario de Matemática Pura y Aplicada,

Universitat Politècnica de València, Camino de Vera, s/n, 46022, Valencia,

(†) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Camino de Vera, s/n, 46022, Valencia

Departamento de Matemática Aplicada,

Universitat Politècnica de València, Camino de Vera, s/n, 46022, Valencia

November 30, 2016

1 Introduction

Fleet assignment is one of the core problems in the management and scheduling of transport carriers. These problems arise from restrictions imposed by the acceptance of bookings from the customers and from the necessity of the carrier to assign vehicles and crew for attending the services. They started to be considered in air transportation in the eighties and were modeled in terms of graph theory and integer linear programming, see for instance [1, 2, 9].

However, with the coming of driverless cars, the rearrangement of vehicles among the depots can be simplified since no staff must be responsible of these tasks. The authors have considered the solution to this problem in [3], where they show a heuristic algorithm that permits to minimize the number of cars that have to be subcontracted from an external provider in order to attend a list of reservations from customers. Its efficacy is shown in comparison with the solution given by the integer linear programming method. These results can be of interest for other autonomous systems that need to be controlled.

The *stable marriage problem* is the problem of finding a stable matching between two equally sized sets of elements. Every element of each set orders its preferences of elements of the other set. A matching consists on making pairs of elements of both sets. A matching is said to be stable if there is no pair of elements of both sets that can join into a new pair by increasing their preferences respect to the elements of the other set.

[*] e-mail: cjordan@mat.upv.es

A version of this problem is given when both sets have a different number of elements and some of the elements of one set can accept several elements from the other set, this is know as the *college admission problem*. Both problems were introduced by Gale and Shapley in [8]. These ideas translated into the problem of stable allocations and the practice of market design lead to Roth and Shapley to win the Nobel prize in Economics in 2012 ([10]).

We show how to use the ideas from the stable marriage and college admission solution algorithms to give a solution to the self-organization of a fleet of vehicles that have to attend a list of bookings. Our results are of particular interest because they permit to easily reschedule the vehicles if cancelations of the bookings along the time are permitted.

2 Statement of the problem

We consider the logistics of a car-rental company of driverless cars with several depots and an initial number of cars at each depot. If required, the cars can move on their own from a depot to another depot. Taking this into account, we can study the problem of how to organize the company fleet of vehicles for attending a list of reservations given in advance. The problem will be tied to the following considerations:

(C1) A number of reservations is given. We consider that the satisfaction of all customer petitions is mandatory. So that, all reservations must be accepted.

(C2) If some cars are needed at a depot at the initial time, then we bring there from an external provider.

(C3) If there are not enough cars available at a depot i at some time t (not the initial), then we first try to bring cars from another depot i'. This is only done if we do not leave unattended bookings at this depot before time t.

(C4) If after these rearrangement of vehicles, there are not enough cars available at any depot to attend a booking, then we subcontract them from an external provider.

(C5) There is no limit to the parking places at each depot.

This problem can be compared with the ones described in [6, 7] and with the one of the management of an electric car-rental service [4].

In general these problems are solved by integer linear programming (ILP). In the paper [3] we present an heuristic algorithm that permits to obtain an optimal solution to the problem in terms of reducing the need of cars from an external provider.

Now, we define our problem as follows. The assumptions that we consider are:

- There are n depots $v_1, v_2, \ldots, v_n$ at m different times $t_1, t_2, \ldots, t_m$.
- The initial number of cars at every depot is given.

- A list of bookings to be accepted is also given.
- If needed, the cars can be moved among the depots (driverless cars).

And our goal is to minimize the number of additional cars needed to serve all the bookings, taking into account that we want to accept all the bookings.

3 Modelling

First, we are going to model it by using Graph Theory. For the sake of clarity we show in the Figure 1 the case of two depots with p_1 cars at depot v_1 and p_2 cars at depot v_2, where $v_{i,j}$ denotes the depot v_i at the time t_j and consider a reservation of k cars from depot 1 at time t_2 to depot 2 at time t_3, i. e., of $v_{1,2}$ to $v_{1,3}$.

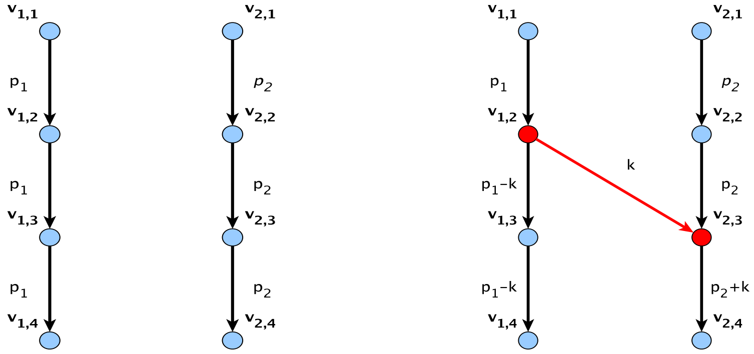


Figure 1: Reservation of k cars from $v_{1,2}$ to $v_{1,3}$.

If we accept all reservations, we can have a negative number of cars somewhere. Then we should move cars from others depots where there are free cars for the next times. Our goal is to find the optimal strategy for doing it. In order to model this situation we will use a time-space network $N = (V, E)$, (see [5]), where the nodes $v_{i,j}$ represent the depot v_i at the time t_j and the arcs represent changes of cars between nodes $(s, v_{i,1})$, $(p, v_{i,j})$ or $(v_{i,j}, v_{i',j'})$, $1 \leq i \leq n$, where p is the external provider and s is the source of network, as we show in the Figure 2.

Once we have set the network N, we can consider flows on them.

We recall that a **flow** c is a function $c : E \to \mathbb{N}$. For every node $u \in V$ with positive incoming and outgoing degree, we have that the conservation law of the flow holds:

$$\sum_{v \in V,\, (v,u) \in E} c(v,u) \;=\; \sum_{v \in V,\, (u,v) \in E} c(u,v). \tag{1}$$

That is, the sum of the flows that enter into u is the same as the sum of flows that depart from u. In our model a flow will represent how cars are moved between the depots through the time.

Bookings are given by a 5-tuple $r = (i_p, t_p, i_d, t_d, w)$, where $1 \leq i_p, i_d \leq n$, $1 \leq t_p < t_d \leq m$, and $n \in \mathbb{N}$, being i_p the pick up depot, t_p the pick up time, i_d the drop off city, t_d the drop off time, and w the number of cars to be reserved.

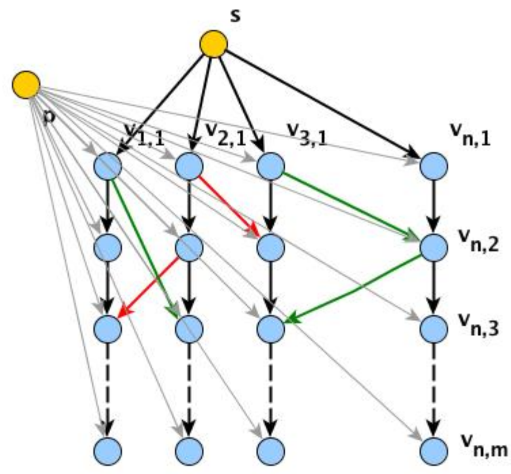


Figure 2: A time-expanded network with n depots at m different times.

Each booking $r = (i_p, t_p, i_d, t_d, w)$ is converted into a new edge from node v_{i_p,t_p} to node v_{i_d,t_d} with weight w on it.

4 Formulation as an Integer Linear Problem (ILP)

We briefly report how this problem can be optimized as an Integer Linear Problem (ILP). In order to solve the problem let us consider the following function to be minimized:

$$Z(x) = \sum_{\substack{1\le i,i'\le n\\ 1\le j<j'\le m}} w(v_{i,j}, v_{i',j'})x(v_{i,j}, v_{i',j'}) + \sum_{\substack{1\le i\le n\\ 1\le j\le m}} w(p, v_{i,j})x(p, v_{i,j}) \tag{2}$$

The weights $w(\cdot)$ are assigned in order to give prevalence to move cars from another depot before taking them from an external provider p ($\lambda_1 < \lambda_2$).

The unknowns to be found are denoted by $x(\cdot) \in \mathbb{Z}_+$

$$w(v_{i,j}, v_{i',j'}) = \begin{cases} 0 & \text{if } i = i',\ j' = j+1, \\ \lambda_1 & \text{if } i \ne i',\ j < j', \end{cases} \tag{3}$$

$$w(p, v_{i,j}) = \lambda_2 \ \text{ for } 1 \le i \le n,\ 1 \le j \le m. \tag{4}$$

The function $Z(x)$ has to be minimized tied to the following restrictions that stand for the conservation law at every vertex $v_{i,j}$. In particular, for $j = 1$ and $1 \le i \le n$ we have

$$k_i + x(p, v_{i,1}) = \sum_{1\le i'\le n} x(v_{i,1}, v_{i',2}), \tag{5}$$

and for $2 \le j \le m$ and $1 \le i \le n$ we have

$$\sum_{1\leq i'\leq n} x(v_{i',j-1},v_{i,j})+x(p,v_{i,j}) = \sum_{1\leq i'\leq n} x(v_{i,j},v_{i',j+1}). \tag{6}$$

Some other restrictions can be easily included in this way with inequalities such as: a maximum number of bookings per day despite of overbooking, a maximum number of cars at each depot (electric cars) or by including variable deadhead times and crew cost times.

As we have previously pointed out, by using the inherent properties of the problem we design an ad-hoc heuristical algorithm that substantially improves the runtime. We refer the reader to [3] for its description and a comparison with the approach as an optimization of an ILP problem.

5 Stable marriages

A different approach based in stable allocations and the practice of market design [10]. The opportunity of using this approach is based on some practical aspects of the problem. Since bookings can be canceled, long term predictions are not usually suitable for designing the rearrangement of vehicles. The theory of stable allocation is based in the seminal work of Gale and Shapley [8] on matching assignments. First, we recall what is understood as a stable marriage.

Definition 1 (Stable marriage) *Let us consider two sets of applicants M and W. Let us assume that they have the same size and every applicant of M has set a total order of the elements of W based on his preferences. Conversely, we also assume that every applicant of W has set a total order of the elements of M based on his preferences.*

An assignment of pairs of applicants will be unstable if there are two applicants $m\in M$ and $w\in W$ that are matched but the following conditions hold:

There is some $m'\in M$ and $w'\in W$ such that m prefers w' respect to w, and w prefers m' isntead of m.

If this condition never holds, then we say that the marriage is stable.

Let us illustrate this situation with two sets M, of three men, and W, of three women. In Figure 3, by (m_{i1},m_{i2},m_{i3}) and (w_{i1},w_{i2},w_{i3}) we denote the order of preferences of each one the three women w_i, $i=1,2,3$ and of the three men m_i, $i=1,2,3$ respectively. The couple (w_1,m_1) is stable while the one (w_2,m_2) is unstable since woman w_2 prefers man m_3 to man m_2, and man m_3 prefers woman w_2 to woman w_3.

In [8], Gale and Shapley give a constructive proof of the following result, based on the "deferred acceptance procedure", that they introduced there:

Theorem 1 *There always exists a stable set of marriages.*

In fact, they also provide further insight into these assignments:

Theorem 2 *The assignment is as least as well off under the assignment given by the deferred acceptance procedure as the one obtained under any other stable assignment.*

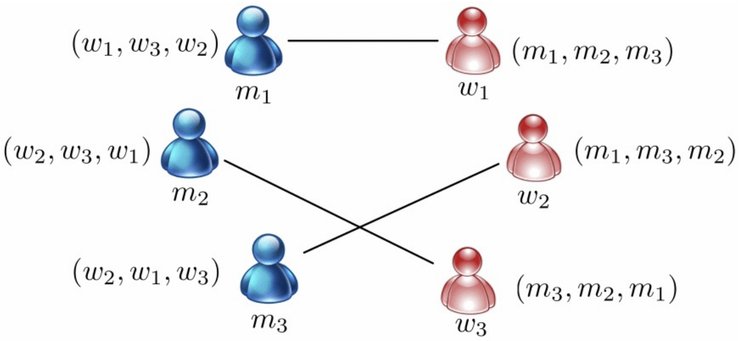


Figure 3: Assignment by using stable marriage.

Based on these results and by applying the algorithm based on the deferred acceptance procedure by Gale and Shapley we can give a solution to the fleet-assignment problem considered before, provided that we consider the following additional assumptions:

- Every depot $v_{i,j+1}$ has a quota of $q_{i,j+1}$ that are unattended.
- The cars waiting at certain depots at time t_j are the ones that can be used to fill the quotas at time t_{j+1}.
- One has to take out cars that need some maintenance.
- If a car is about to run out off battery, the booking is split in two new bookings, where at some intermediate depot, the passenger must change of vehicle.
- We give preference to the depots respect to the cars.

The details of this procedure will appear in a forthcoming paper. We briefly outline the key issues of that:

- Every car has to order all the depots according to its preference. One computes the average of benefit among all the bookings at some depot. Cars would prefer to attend depots with a higher average of expected benefit for the service.
- Every depot has to order the cars. Only cars with enough battery for the longest booking at some depot are admitted. They are rank according to a different criteria. The criterion is based in proximity (If a car is too far away, then it can be omitted).
- Once a car arrives, it is assigned to FIFO rule (first in, first out).

Acknowledgements

This work is supported by Vicerrectorado de Estudios, Calidad y Acreditación, PIMEB08: “Modelos de la Teoría de Grafos aplicados a problemas de competiciones

de programación".

This work is comprised within the context of extending the contribution of graph theory methods and algorithms to improve programming skills for the solution of engineering problems appearing in different fields.

References

[1] J. Abara. Applying integer linear programming to the fleet assignment problem. *Interfaces* 19(4): 20–28, 1989.

[2] L. Bodin, B. Golden, A. Assad, and M. Ball. Routing and scheduling of vehicles and crews: the state of the art. *Comput. Oper. Res.* 10: 63-212, 1983.

[3] J.A. Conejero, C. Jordan, and E. Sanabria-Codesal. An algorithm for self-organization of driverless vehicles of a car-rental service. *Nonlinear Dynamics* 84(1): 107-114, 2016.

[4] J.A. Conejero, C. Jordan, and E. Sanabria-Codesal. An iterative algorithm for the management of an electric-car-rental service, *J. Appl. Math.* 2014, Article ID 483734, 11 pp, 2014.

[5] J.R. Evans and E. Minieka. Optimizaction algorithms for networks and graphs *Marcel Dekker, Inc.*, 1992.

[6] A. Fink and T. Reiners. Modeling and solving the short-term car rental logistics problem, *Transportation Research: Part E* 42: 272–292, 2006.

[7] A. Hertz, D. Schindl, and N. Zufferey. A solution method for a car fleet management problem with maintenance constraints, *J. Heuristics* 15: 425–450, 2009.

[8] D. Gale and L.S. Shapley. College admissions and the stability of marriage. *Amer. Math. Monthly* 69: 9-14, 1962

[9] M. Lohatepanont and C. Barnhart. Airline schedule planning: integrated models and algorithms for schedule design and fleet assignment. *Transp. Sci.* 38: 19-32, 2004.

[10] Ll. Shapley and A. Roth. Stable allocations and the practice of market design. *Nobel Prize in Economics*, 2012.

Numerically stable and quadratic convergent method for solving absolute value equation

Taher Lotfi [*], Katayoun Mahdiani, and Nahid Zainali

Department of Applied Mathematics, Hamedan Branch, Islamic Azad University, Hamedan, Iran

November 30, 2016

1 Introduction

We consider the following absolute value equation (AVE)

$$G(x) = Ax - |x| - b = 0, \tag{1}$$

where $A \in R^{n\times n}$, $b \in R^n$, and $|.|$ denotes the absolute value. To solve (1), Mangasarian applies a generalized Newton's method for solving it provided that the singular values of A are not less than one (see Lemma 6 in [6]). Although, the generalized Newton's method is linear convergent, a quadratically convergent method under the same condition has been developed [1]. When the singular values of A exceed 1, the AVE (1) has a unique solution [6]. Hence it seems that under this limitation, such an iterative method converge globally [9]. On the other hand, Prokopyev proves that checking whether the AVE (1) has unique or multiple solutions is an NP-complete problem [7]. Therefore, it is not generally possible to construct a polynomial algorithm for solvability of AVE.

We develop an iterative method to solve the NP-complete AVE (1). Indeed, we focus on modifying the generalized method introduced in [6] in such a way that it has convergence order two, and it is a numerically stable method.

[*] lotfi@iauh.ac.ir, lotfitaher@yahoo.com

2 Main results

In this section, reconsidering the generalized Newton's method [6], we modify it in such a way that it has convergence order two with general conditions compared with the provided conditions by Managasarian in [6]. To this end, let the generalized Jacobian of (1) be given by

$$J_G(x) = A - T_z(x), \tag{2}$$

where $T_z(x)$=diag(sign(x)). Let x^0 be a suitable starting vector to the exact solution, say x^*, of (1). Then, we propose the following modified Newton-Mangasarian method

$$(A - T_z(x^k))\Delta x^k = -Ax^k + |x^k| + b, \tag{3}$$

$$x^{k+1} = x^k + \Delta x^k, \quad k = 0, 1, 2, \dots. \tag{4}$$

It should be noted that we solve the linear system (3), and then, update the value x^{k+1} from (4). Therefore, we reduce the numerical solution of solving a nonlinear system of equations to the numerical solution of a linear systems of equations. For more details, one can consult [2-5]. We will prove that this method is numerically stable and has convergence order two.

Remark 2.1 *In practice, we have to use floating point arithmetic with finite digits accuracy. Consequently, rounding errors occur, and we compute $\hat{x}^k$ instead of x^k [3]. Actually, the computed $\hat{x}^k$ satisfies*

$$(A - T_z(\hat{x}^k) + E_1^k)\Delta\hat{x}^k = -Ax^k + |x^k| + b + E_2^k, \tag{5}$$

$$\hat{x}^{k+1} = \hat{x}^k + \Delta\hat{x}^k + E_3^k, \quad k = 0, 1, 2, \dots. \tag{6}$$

where E_1^k, E_2^k, and E_3^k are the errors that are made in computing $G(\hat{x}^k) = A\hat{x}^k - |\hat{x}^k| - b$, forming $J_G(\hat{x}^k)$ and solving the linear system (3), and updating (4), respectively.

To prove the quadratic convergence order of the method (3)-(4), we need the following lemma:

Lemma 2.2 *Let D be an open convex set in R^n, and let J_G be Lipschitz continuous at x in the neighborhood D. Then, for any $y = x + \Delta x \in D$,*

$$\|G(y) - G(x) - J_G(x)\Delta x\| \leq \frac{L_J}{2}\|\Delta x\|^2 \tag{7}$$

where L_J is the Lipschitz constant for J_G at x.

Now, we can establish the quadratic convergence of the proposed method (3)-(4). Let $N_r(x^*) = \{x \in R^n : \|x - x^*\| < r\}$, and $r_k = \|x^k - x^*\|$.

Theorem 2.3 *Suppose that x^* is a solution of the AVE (1), i.e, $G(x^*) = 0$. In addition, suppose that the assumptions of Lemma (2.2) hold, and $\|J_G(x^k)^{-1})\| \leq M$, M is a constant, for any $x^k \in N_r(x^*) \subset D$. Then, the sequence $\{x^k\}$, $k > 0$, generated by (3)-(4) satisfies*

$$\|x^{k+1} - x^*\| \leq \frac{L_J}{2}\|x^k - x^*\|^2. \tag{8}$$

Theorem 2.4 *Let the assumptions and the assertion of the Theorem 2.3 hold. Moreover, let $r_0 = \|x^* - x^0\| < 2/L_J$, and $N_{r_0}(x^*) \subset D$. Then, (3)-(4) generates the sequence $\{x^k\}$ such that $x^k \in N_{r_0}(x^*)$, and $x^k \to x^*$ as $k \to \infty$.*

Theorem 2.5 *If the conditions of the Theorem 2.4 hold, then the solution x^* is unique in $N_{2/L_J}(x^*)$.*

2.1 Numerical stability analysis

Now we concern with studying numerical stability of (3)-(4). For this purpose, similar to Wozniakowski [3], we need to consider and study $G(x) = G(x; d) = 0$ which means that G depends on an input data d. The condition number of $G(x; d)$ is defined by

$$\text{cond}(G; d) = \|G'_x(x^*; d)^{-1}G'_d(x^*; d)\|\frac{\|d\|}{\|x^*\|}, \tag{9}$$

where G'_x and G'_d stand for the Frechet derivatives with respect to x and d. Let $G(x; A) = Ax - |x| - b$, where the data vector is supposed to be the given matrix A. For the sake of simplicity, we do not consider the vector b as a part of the data vector. Hence, $G'_x(x; A) = A - T_z(x)$, and $G'_A(x; A) = \text{diag}(x)$ where $\text{diag}(x)$ denotes an $n \times n$ diagonal matrix whose elements are $d_{i,i} = x_i$ for $i = 1, \ldots n$. Then,

$$\text{cond}(G; A) = \|(A - T_z(x^*))^{-1}\|\|A\|.$$

Hence

Lemma 2.6 *Let the matrix A be the data vector in (1). Then,*

$$cond(G; A) = \|(A - T_z(x^*))^{-1}\|\|A\|. \tag{10}$$

As can be seen from (10), if $T_z(x^*) = 0$, then $G(x; A) = Ax - b = 0$, and we obtain the classic condition number of the matrix A, say $K(A) = \|A^{-1}\| \|A\|$.

Now, we suppose that x^k generated by (3)-(4) is close enough to x^*. It is crucial that the numerical accuracy of x^{k+1} highly depends on the condition number (10). For a moment, we stop and focus on the $G(x^k)$ in (3). Based on our assumption, if k is large, then the value of $G(x^k)$ tends to zero. So, in this case, we have a homogenous linear system and no matter how ill-conditioned it is. This the reason why we only pay attention to the condition number $\mathrm{cond}(G; A)$, and not to the condition number $\mathrm{cond}(G'; A)$;

As Wilkinson says [8], we are not generally able to solve the equation $G(x; A)$=0 exactly because we have to compute in the finite digits of the floating point arithmetic. Let $\mathrm{eps} = 5 \times 10^{-t}$ denotes the machine precision. Let

$$\mathrm{fl}(G(x^k; A)) = (I + \Delta G_\epsilon^k) G(x^k + x_\epsilon^k; A + A_\epsilon) = G(x^k) + \delta G(x^k), \qquad (11)$$

where $\|\Delta G_\epsilon^k\| \sim \mathrm{eps}$, $\|x_\epsilon^k\| \sim \mathrm{eps}$, $\|A_\epsilon\| \sim \mathrm{eps}\|A\|$, and

$$\delta G(x^k) = \Delta G_\epsilon^k G(x^k) + (A - T_z(x^k)) x_\epsilon^k + \mathrm{diag}(x^k) \mathrm{D}(A_\epsilon) + O(\mathrm{eps}^2), \qquad (12)$$

where D is a vector whose elements are given by $d_i = \max A_\epsilon(i, j)$, $j = 1, \dots n$.

Similarly, let

$$\mathrm{fl}(G'(x^k; A)) = G'(x^k) + \delta G'(x^k), \quad \delta G'(x^k) = O(\mathrm{eps}). \qquad (13)$$

Hence, we can assume that the numerical solution of (3) satisfies

$$(G'(x^k) + \delta G'(x^k) + E^k)) \Delta \tilde{x}^k = G(x^k) + \delta G(x^k), \qquad (14)$$

with $E_k = O(\mathrm{eps})$. For more details on solving (13) consult [3]. Consequently, the next improvement, x^{k+1}, is computed by

$$x^{k+1} = (I + \xi^k)(x^k + \Delta \tilde{x}^k), \qquad (15)$$

where ξ^k is a diagonal matrix with $\|\xi^k\| \sim \mathrm{eps}$. Now, we are ready to state the numerical stable features of (14)-(15).

Theorem 2.7 *Let (11)-(13) hold. Then the method (14)-(15) is numerically stable.*

References

[1] Caccetta, L., Qu, B., Zhou, G., A globally and quadratically convergent method for absolute value equations. *Comput. Optim. Appl.* 48, 45–58 (2011)

[2] Cordero, A., Lotfi, T., Mahdiani, K., Torregrosa, J.R., A stable family with high order of convergence for solving nonlinear equations. *Appl. Math. Comput.* 254, 240-251 (2015)

[3] Higham, N.J., Accuracy and stability of numerical algorithms. SIAM, Philadelphia (1996)

[4] Hueso, J.L., Martinez, E., Torregrosa, J.R., Modified Newtons method for systems of nonlinear equations with singular Jacobian. *J. Comput. Appl. Math.* 224, 77-83 (2009)

[5] Lotfi, T., Bakhtiari, P., Cordero, A., Mahdiani K., Torregros J.R., Some new efficient multipoint iterative methods for solving nonlinear systems of equations. *Int. J. Comput. Math.* 5 **92**(9), 1921-1934 (2015)

[6] Mangasarian, O.L., A generalized Newton method for absolute value equations. *Optim. Lett.* **3**(1), 101-108 (2009)

[7] Prokopyev, O.A.: On equivalent reformulations for absolute value equations. *Comput. Optim. Appl.* **44**(3), 363-372 (2009)

[8] Wilkinson, J.H., Rounding errors in algebraic processes. Prentice-Hall, Englewood Cliffs (1964). Reprinted by Dover Publications, New York, 1994

[9] Zhang, C.,Wei, Q.J.: Global and finite convergence of a generalized Newton method for absolute value equations.*J. Optim. Theory Appl.* 143, 391-403 (2009)

Improved convergence analysis of the Secant method using restricted convergence domains

I. K. Argyros[♭], Á. A. Magreñán[†*]
and J. A. Sicilia[†].

(♭) Cameron University, Department of Mathematics Sciences
Lawton, OK 73505, USA.

(†) Universidad Internacional de La Rioja, Escuela de Ingeniería
Avenida Gran Vía Rey Juan Carlos I, 41, 26002 Logroño, Spain.

November 30, 2016

1 Introduction

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

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

where, F is a Fréchet-differentiable operator defined on a nonempty subset $\mathcal{D}$ of a Banach space $\mathcal{X}$ with values in a Banach space $\mathcal{Y}$. Several problems from Applied Sciences including Engineering can be expressed in a form like equation (1) using mathematical modelling [1, 2, 3, 4, 5, 6]. The solutions of these equations can be found in closed form only in special cases. That is why the most solution methods for these equations are iterative.

In this paper we consider the convergence of the Secant method defined by

$$x_{n+1} = x_n - \mathcal{A}_n^{-1} F(x_n),\ \mathcal{A}_n = \delta F(x_n, x_{n-1}) \ \text{ for each } n = 1, 2, \ldots, \tag{2}$$

[*] e-mail:alberto.magrenan@unir.net

where x_{-1}, x_0 are initial points. Here $\mathcal{A}_n \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ is an approximation of the Fréchet-derivative F' of F and $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ stands for the space of bounded linear operators from $\mathcal{X}$ into $\mathcal{Y}$. There is a plethora of sufficient convergence criteria for the Secant method (2) under Lipschitz-type conditions (see [2]–[6]). It is interesting to notice that although we use very general majorizing sequences for $\{x_n\}$ our technique leads in the semilocal case to: weaker sufficient convergence criteria; more precise estimates on the distances $\|x_n - x_{n-1}\|$, $\|x_n - x^*\|$ and an at least as precise information on the location of the solution x^* in many interesting special cases such as Newton's method or the Secant method.

2 Mathematical backgrund

We shall study the Secant method for triplets $(\mathcal{F}, x_{-1}, x_0)$ belonging to the class $\mathcal{K} = \mathcal{K}(\nu, c, k, k_0, k_1, k_2)$ defined as follows.

Let there be parameters $c \geq 0$, $\nu \geq 0$,$k > 0$, $k_0 > 0$, $k_1 > 0$ and $k_2 \geq 0$. Define the scalar sequence $\{\alpha_n\}$ by

$$\begin{cases} \alpha_{-1} = 0,\ \alpha_0 = c, \alpha_1 = c + \nu, \\ \alpha_{n+2} = \alpha_{n+1} + \dfrac{[k_1(\alpha_{n+1} - \alpha_n) + k_2(\alpha_n - \alpha_{n-1})](\alpha_{n+1} - \alpha_n)}{1 - (k_0(\alpha_{n+1} - c) + k\alpha_n)} \quad \text{for each } n = 0, 1, 2, \ldots \end{cases} \tag{3}$$

Remark 1 **(a)** *Let us introduce the notation*

$$c^N = \alpha_{N-1} - \alpha_{N-2},\ \nu^N = \alpha_N - \alpha_{N-1}$$

for some integer $N \geq 1$. Notice that $c^1 = \alpha_0 - \alpha_{-1} = c$ and $\nu^1 = \alpha_1 - \alpha_0 = \nu$. The results in the Lemmas in [6] can be weakened even further as follows. Consider the convergence criteria (C_*^N) for $N > 1$: (C^1) (see [6] for the definition of C^1) with c, ν replaced by c^N, ν^N, respectively

$$\alpha_{-1} < \alpha_0 < \alpha_1 < \ldots < \alpha_N < \alpha_{N+1},$$

$$k_0(\alpha_{N+1} - c^N) + k\alpha_N < 1.$$

Then, the preceding results hold with $c, \nu, \delta_1, \delta_2, b_1^1, b_2^1$ replaced, respectively by $c^N, \nu^N, \delta_N, \delta_{N+1}, b_1^N, b_2^N$.

(b) *Notice that if*

$$k_0(\alpha_{N+1} - c^N) + k\alpha_N < 1 \textit{ holds for each } n = 0, 1, 2, \ldots, \tag{4}$$

then, it follows from (3) *that sequence* $\{\alpha_n\}$ *is increasing, bounded from above by* $\frac{1+k_0c}{k_0+k}$ *and as such it converges to its unique least upper bound* α^*. *Criterion* (4) *is the weakest of all the preceding convergence criteria for sequence* $\{\alpha_n\}$. *Clearly all the preceding criteria imply* (4). *Finally, define the criteria for* $N \geq 1$

$$(I^N) = \begin{cases} (C_*^N) \\ (4) \textit{ if criteria } (C_*^N) \textit{ fail.} \end{cases} \tag{5}$$

Definition 1 *Let* ν, c, k, k_0, k_1, k_2 *be constants satisfying the hypotheses* (I^N) *for some fixed integer* $N \geq 1$. *A triplet* $(\mathcal{F}, x_{-1}, x_0)$ *belongs to the class* $\mathcal{K} = \mathcal{K}(\nu, c, k, k_0, k_1, k_2)$, *if:*

$(\mathcal{D}_1)$ $\mathcal{F}$ *is a nonlinear operator defined on a convex subset* D *of a Banach space* x^* *with values in a Banach space* $\mathcal{Y}$.

$(\mathcal{D}_2)$ x_{-1} *and* x_0 *are two points belonging to the interior* D^0 *of* D *and satisfying the inequality*

$$\|x_0 - x_{-1}\| \leq c,$$

for some constant $c \geq 0$.

$(\mathcal{D}_3)$ $\mathcal{F}$ *is Fréchet-differentiable on* D^0 *and there exists an operator* $\delta\mathcal{F} : \mathcal{D}^0 \times D^0 \to \L(X, Y)$ *such that* $\delta\mathcal{F}(x, y)(x-y) = F(x) - F(y)$ *for each* $x \neq y$, $\delta\mathcal{F}(x, x) = F'(x)$, $x \in D^0$, $F'(x_0)^{-1}$, $\mathcal{A}^{-1} = \delta\mathcal{F}(x_0, x_{-1})^{-1} \in \L(Y, X)$ *for all* $x, y \in D$ *then, the following hold*

$$\|\mathcal{A}^{-1}\mathcal{F}(x_0)\| \leq \nu,$$

$$\|F'(x_0)^{-1}(\delta\mathcal{F}(x, y) - F'(x_0))\| \leq k_0\|x - x_0\| + k\|y - x_0\|$$

and for each $x, y, z \in D_0 := U(x_0, \frac{1}{k_0+k}) \cap D$

$$\|F'(x_0)^{-1}(\delta\mathcal{F}(x, y) - F'(z))\| \leq k_1\|x - z\| + k_2\|y - z\|$$

for some constants $k > 0$, $k_0 > 0$, $k_1 > 0$, $k_2 \geq 0$ *and* $\nu \geq 0$.

$(\mathcal{D}_4)$

$$\overline{U}(x_0, \alpha^* - c) \subseteq D \text{ or } U(x_0, \frac{1}{k_0 + k}) \subset D.$$

where α^* is given by

$$\alpha^* = \frac{1}{L_0(k_0 + k)}.$$

3 Semilocal convergence analysis

Theorem 1 *If $(\mathcal{F}, x_{-1}, x_0) \in \mathcal{K}(\nu, c, k, k_0, k_1, k_2)$ then, the sequence $\{x_n\}$ $(n \geq -1)$ generated by the Secant method is well defined, remains in $\overline{U}(x_0, \alpha_0^*)$ for each $n = 0, 1, 2, \ldots$ and converges to a unique solution $x^* \in \overline{U}(x_0, \alpha^* - c)$ of* (1)*. Moreover, the following assertions hold for each $n = 0, 1, 2, \ldots$*

$$\|x_n - x_{n-1}\| \leq \alpha_n - \alpha_{n-1} \tag{6}$$

and

$$\|x^* - x_n\| \leq \alpha^* - \alpha_n, \tag{7}$$

where sequence $\{\alpha_n\}$ $(n \geq 0)$ is given in (3)*. Furthermore, if there exists R such that*

$$\overline{U}(x_0, R) \subseteq D,\ R \geq \alpha^* - c \text{ and } k_0(\alpha^* - \alpha_0) + kR < 1, \tag{8}$$

then, the solution x^* is unique in $\overline{U}(x_0, R)$.

References

[1] Argyros, I.K., Hilout, S., Weaker conditions for the convergence of Newton's method, Journal of Complexity, 28 (2012), 364–387.

[2] Argyros, I.K., Cho, Y.J., Hilout, S., Numerical method for equations and its applications. CRC Press/Taylor and Francis, New York, 2012.

[3] Argyros, I.K., González, D., Magreñán, Á.A., A semilocal convergence for a uniparametric family of efficient Secant-like methods, Journal of Function Spaces, vol. 2014, Article ID 467980, 10 pages, 2014. doi:10.1155/2014/467980.

[4] Argyros, I.K., Magreñán, Á.A., Relaxed Secant-type methods, Nonlinear studies, 21 (3) (2014), 485-503.

[5] Argyros, I.K., Magreñán, Á.A., A unified convergence analysis for Secant-type methods, Bulletin of the Korean Mathematical Society, 52 (3) (2015), 865-880.

[6] Argyros, I.K., Magreñán, Á.A., Expanding the applicability of the secant method under weaker conditions, Applied Mathematics and Computation, 266 (2015), 1000-1012.

Preconditioners for nonsymmetric linear systems with low-rank skew-symmetric part

J. Marín [*], J. Cerdán[*] , J. Mas[*] and D. Guerrero[*]

November 30, 2016

Keywords: Iterative methods, nearly symmetric matrices, preconditioning, low-rank update.

1 Introduction

In this paper we study the iterative solution of nonsingular, nonsymmetric linear systems

$$Ax = b \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, the coefficient matrix is large, sparse and its skew-symmetric part $\frac{1}{2}(A - A^T)$ has low rank or can be approximated by a low-rank matrix. Consider $A = H + K$ where H and K are the symmetric and skew-symmetric parts of A, respectively. It is supposed that the skew-symmetric matrix can be written as $K = FCF^T + E$ where is a full rank rectangular matrix $F \in \mathbb{R}^{n\times s}$, $C \in \mathbb{R}^{s\times s}$ is a skew-symmetric matrix, with $s \ll n$ and $\parallel E \parallel \ll 1$. Systems like this appear, for example, in discretization of PDEs with certain Neumann boundary conditions, in discretization of integral equations, as well as path following methods.

Different strategies have been proposed to solve (1) when the skew-symmetric part K has exactly rank $s \ll n$. In [1] it is presented the progressive GMRES method which shows than an orthogonal Krylov subspace basis can be generated with short recursion formulas for this kind of matrices. As pointed out in [5], although the method is mathematically equivalent to full GMRES [6], in practice

[*]Institut de Matemàtica Multidisciplinar, Universitat Politècnica de València, Camì de Vera, 14, Valencia, Spain, 46022. Email: jmarinma@imm.upv.es, jcerdan@imm.upv.es, jmasm@imm.upv.es, danguefl@doctor.upv.es. Supported by Spanish Grants MTM2014-58159-P and MTM2015-68805-REDT.

it may suffer from instabilities due to the loss of orthogonality between the vectors of the generated Krylov subspace basis. In the same paper, the authors propose a Schur complement method (SCM) that also permits the application of short-term formulas. The method obtains an approximate solution by applying the MINRES method $s+1$ times. The authors also suggest that SCM can be successfully applied as a preconditioner for GMRES.

In this paper we propose a method based on the framework proposed in [4]. Assuming that the matrix $H + FCF^T$ is nonsingular, our approach computes an incomplete LU factorization (ILU) of the matrix

$$\mathbf{A} = \begin{pmatrix} H & F \\ F^T & -C^{-1} \end{pmatrix} \tag{2}$$

as for instance, with the Balanced Incomplete Factorization (BIF) algorithm [2, 3] or ILUT [8]. Interestingly, the matrix in (2) is similar to the one used in [5] to develop SCM method, but in this work this matrix is used to compute an explicit preconditioner for system (1). It can be viewed as a low-rank update of an incomplete factorization of the symmetric part H.

The paper is organized as follows. In Section 2 we describe the proposed preconditioner and in Section 3 the results of the numerical experiments for different problems are presented. Finally, we give our conclusions.

2 Updated Preconditioned Method

Our preconditioner **M** is obtained by computing an incomplete LU of the matrix **A** in (2). Assuming that we have calculated an incomplete LU factorization of the symmetric part H, that is, $H \approx L_H D_H L_H^T$, one has

$$\mathbf{M} = \begin{pmatrix} L_H & 0 \\ F^T L_H^{-T} D_H^{-1} & I \end{pmatrix} \begin{pmatrix} D_H & 0 \\ 0 & R \end{pmatrix} \begin{pmatrix} L_H^T & D_H^{-1} L_H^{-1} F \\ 0 & I \end{pmatrix}$$

with $R = -(C^{-1} + F^T L_H^{-T} D_H^{-1} L_H^{-1} F)$. Note that

$$\left[\begin{array}{cc} I & O \end{array}\right] \mathbf{M}^{-1} \left[\begin{array}{c} I \\ O \end{array}\right]$$

approximates A^{-1} of (1). Thus, our preconditioner consists in the application of this operator. The preconditioning step is done by solving linear systems of the form

$$\mathbf{M} \left[\begin{array}{c} s \\ s_1 \end{array}\right] = \left[\begin{array}{c} r \\ 0 \end{array}\right]$$

The preconditioned vector s is obtained in three steps:

1. Solve $L_H D_H r_1 = r$.
2. Solve $Rs_1 = -F^T L_H^{-T} r_1$.
3. Solve $L_H^T s = r_1 - D_H^{-1} L_H^{-1} F s_1$.

The computation and application of the preconditioner is inexpensive provided that $s \ll n$. Note that step 2 implies the solution of a $s \times s$ linear system which can be done with a direct method.

3 Numerical experiments

In this section we compare our proposed preconditioner, referred to as $Upd.\ Prec.$, with the SCM method and also an incomplete LU factorization of the symmetric part H. The iterative methods used are the full GMRES, restarted GMRES(m) and BISCGSTAB. The experiments have been performed with Matlab. The iterative methods were run until the relative initial residual was reduced to 10^{-8}, allowing a maximum number of 1000 iterations. We present the results obtained for two different examples in which the skew-symmetric part has low rank and the can be represented exactly as FCF^T.

The first example was used in [5] to show the performance of SCM method. Consider the block-diagonal matrix

$$A = \begin{bmatrix} \Lambda_- & & \\ & \Lambda_+ & \\ & & Z \end{bmatrix},$$

where $\Lambda_- = diag(\lambda_1, \ldots, \lambda_p)$, $\Lambda_+ = diag(\lambda_{p+1}, \ldots, \lambda_{n-s})$ with $\lambda_1, \ldots, \lambda_p$ uniformly spaced in $[-\beta, -\alpha]$ and $\lambda_{p+1}, \ldots, \lambda_{n-s}$ uniformly spaced in $[\alpha, \beta]$ for some positive constants $\alpha < \beta$, $p \ll n$ and $2 \leq s \ll n$. $Z = tridiag(-\gamma, 1, \gamma) \in \mathbb{R}^{s \times s}$ with $\gamma > 0$. The matrix A is indefinite with eigenvalues

- $\lambda_1, \ldots, \lambda_p \in [-\beta, -\alpha]$,
- $\lambda_{p+1}, \ldots, \lambda_{n-s} \in [\alpha, \beta]$,
- s complex eigenvalues of the skew-symmetric part of Z.

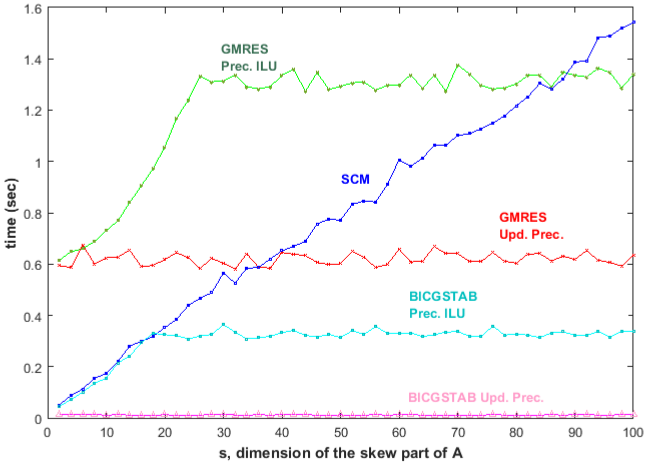


Figure 1: CPU solution time for the first example with the different methods tested preconditioning all methods with an ILU of the symmetric part of A, when increasing the rank s of the skew part of A, from 2 to 100.

We study how to solve the system (1) with b equal to $1/\sqrt{n}$ in all its components, $n = 10^5$, $\alpha = 1/8$, $\beta = 1$, $\gamma = 1$. Figure 1 compares the CPU time of the different methods tested.

For all the values of the rank s it can be observed that using BICGSTAB preconditioned with the updated preconditioned method performs the best. In the case of full GMRES, it starts to be competitive compared with SCM for values of s greater than 40. Note that the solution time of the SCM increases linearly with the rank of the skew symmetric part, while its remains almost constant for the other methods.

The second example corresponds to the 2-dimensional Bratu problem. It consists on finding the solution $u(x, y)$ of the nonlinear boundary problem

$$-\Delta u - \lambda \exp(u) = 0 \ \text{ in } \ \Omega, \quad \text{with } \ u = 0 \ \text{ on } \ \partial\Omega \tag{3}$$

depending on the parameter λ, Δ is the Laplacian, Ω the unit square and $\partial\Omega$ its boundary. We discretize this problem using the five-point finite differences as in [1, 5], in a grid of 500×500 points. After this, we obtain a system with coefficient matrix of order $n = 2.5 \times 10^5$ with skew-symmetric part of exactly rank equal to 2. Table 1 shows the results for the tested methods. The non-preconditioned BICGSTAB and restarted GMRES(m) were also tested.

It can be observed that BIGSTAB preconditioning with our technique has the

Method	Time (s)	Iter
GMRES(100)	†	
BICGSTAB	26.6754	827
GMRES(100) Prec. ILU	45,1028	123
GMRES(100) Upd. Prec.	46,2829	131
BICGSTAB Prec. ILU	13,1653	194
BICGSTAB Upd. Prec.	11,2569	156
SCM	38,2014	255

Table 1: CPU solution time and iterations for the Bratu problem

edge over the SCM method and also works better than the ILU preconditioner. Compared with the preconditioned restarted GMRES(100) both preconditioners performed similarly.

4 Conclusions

We have proposed a preconditioner for almost symmetric matrices that shows good performance for the problems tested compared with other methods used in the bibliography. In a future work we will present results for problems for which the skew-symmetric part must be approximated with a low-rank matrix.

References

[1] B. Beckermann, L. Reichel, The Arnoldi process and GMRES for nearly symmetric matrices, SIAM J. Matrix Anal. Appl., 30(1):102-120, February 2008.

[2] R. Bru, J. Marín, J. Mas, M. Tuma, Balanced incomplete factorization, SIAM J. Sci. Comput., 30(5):2302-2318, 2008.

[3] R. Bru, J. Marín, J. Mas, M. Tuma, Improved balanced incomplete factorization, SIAM J. Matrix Anal. Appl., 31(5):2431-2452, 2010.

[4] J. Cerdán, J. Marín, J. Mas, Low-rank updates of balanced incomplete factorization preconditioners, To appear in Numerical Algorithms.

[5] M. Embree, J. Sifuentes, K. Soodhalter, D. Szyld, F. Xue, Short-term recurrence Krylov subspace methods for nearly hermitian matrices, SIAM. J. Matrix Anal. and Appl., 33-2:480-500, 2012.

[6] Y. Saad, M. H. Schulz, GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM Journal on Scientific and Statistical Computing, 7:856-869, 1986.

[7] H. A. van der Vorst, Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear systems, SIAM Journal on Scientific and Statistical Computing, 12:631-644, 1992.

[8] Y. Saad, ILUT: A dual threshold incomplete LU factorization, Numer. Linear Algebra Appl., 1 (1994), pp. 387-402.

Nonstandard finite difference numerical schemes for delay differential models

M.A. García [*], F.Rodríguez, M.A. Castro, and J.A. Martín

Department of Applied Mathematics, University of Alicante,

Apdo. 99, 03080 Alicante, Spain.

November 30, 2016

1 Introduction

Delay differential equations (DDE) are basic modelling tools in scientific and technical problems where time lags or hereditary characteristics are present [1]. For non-delay, ordinary or partial, differential problems, nonstandard finite difference (NSFD) numerical schemes [2] have found wide applications in the last decades, as they can result in exact numerical solutions for particular equations, and in other cases they may provide schemes that compete in accuracy with standard methods while being dynamically consistent with the original differential problems.

The construction of NSFD schemes for delay differential models has not been much explored. In a recent work [3], a NSFD method was proposed for the linear delay problem

$$\begin{aligned} x'(t) &= \alpha x(t) + \beta x(t-\tau), & t > 0, \quad (1)\\ x(t) &= f(t), & -\tau \le t \le 0, \quad (2) \end{aligned}$$

which was exact only in the initial time interval $0 \le t \le \tau$. Beyond this first interval, a nonstandard method was proposed, and some dynamical properties satisfied by this scheme were partly proved or suggested.

[*] e-mail: miguel.garcia@ua.es

In the present work, exact schemes, valid for the whole range of time values, are constructed for the general linear problem (1)-(2), which are also valid for the restricted linear pure delay problem

$$\begin{aligned} x'(t) &= \beta x(t-\tau), \qquad t>0, && (3)\\ x(t) &= f(t), \qquad -\tau \le t \le 0. && (4) \end{aligned}$$

The new exact schemes are based on simplified expressions for the explicit solutions of the corresponding delay problems given in [4] and [5]. These exact schemes are used to derive a family of nonstandard methods, which can provide simpler computational properties and high order of accuracy.

It can be proved that the new nonstandard schemes are dynamically consistent with the exact problems, in terms of asymptotic stability, oscillation behaviour, and positivity preserving properties.

Numerical examples illustrating computational, accuracy, and dynamical behaviours of the methods are provided.

2 An exact numerical scheme

Based on the explicit expressions for the solutions of the problems (1)-(2) and (3)-(4) given in [4] and [5], it can be shown that the exact solution of problem (1)-(2) can be written in the form, for $(m-1)\tau < t \le m\tau$,

$$\begin{aligned} x(t) &= f(0)\sum_{k=0}^{m-1}\frac{\beta^k(t-k\tau)^k}{k!}e^{\alpha(t-k\tau)}\\ &+ \sum_{k=0}^{m-2}\frac{\beta^{k+1}}{k!}\int_{-\tau}^{0}(t-(k+1)\tau-s)^k e^{\alpha(t-(k+1)\tau-s)}f(s)ds\\ &+ \frac{\beta^m}{(m-1)!}\int_{-\tau}^{t-m\tau}(t-m\tau-s)^{m-1}e^{\alpha(t-m\tau-s)}f(s)ds, \qquad (5) \end{aligned}$$

which is also valid, for $\alpha = 0$, for the particular case of the pure delay model (3)-(4).

Writing $Nh=\tau$, $t_n \equiv nh$, and $x_n \equiv x(t_n)$, for $(m-1)\tau < nh \le m\tau$, and $m>1$, the following exact scheme can be derived,

$$x_{n+1} = e^{\alpha h}\sum_{k=0}^{m-1}\frac{\beta^k h^k}{k!}x_{n-kN} \qquad (6)$$

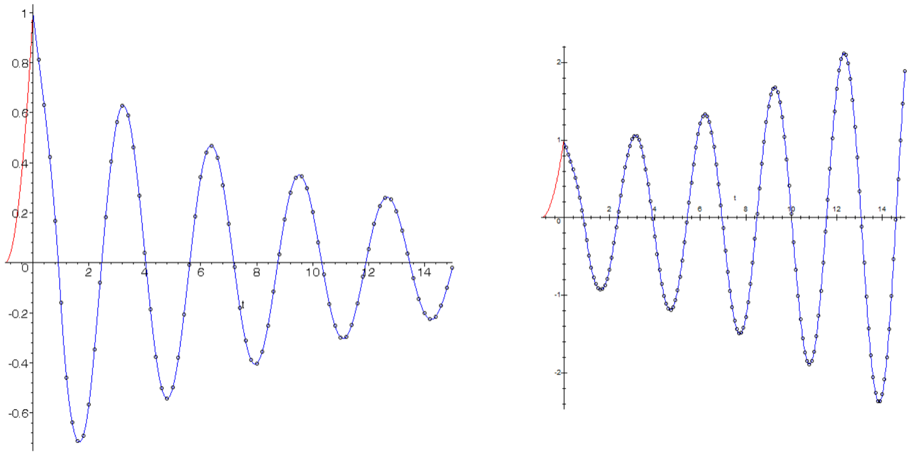


Figure 1: Exact solutions (lines) and numerical solutions provided by the exact scheme (points) for the problem (1)-(2) with initial function $f(t) = (t+1)^2$ and parameters $\tau = 1$, $\alpha = -1$, and $\beta = -2$ (left, $N = 5$) or $\beta = -2.5$ (right, $N = 10$).

$$+ \quad e^{\alpha h}\beta^m \sum_{k=0}^{m-1} \frac{h^k}{k!(m-1-k)!} \int_{t_n - m\tau}^{t_n - m\tau + h} (t_n - m\tau - s)^{m-1-k} e^{\alpha(t_n - m\tau - s)} f(s) ds.$$

Examples of numerical computations for problems with different asymptotic dynamics are shown in Figure 1.

3 Nonstandard methods

Although (6) provides an exact solution, which is the ideal case, the second term in this expression requires the computation of definite integrals, that in general have to be numerically aproximated, and also the number of the integral to be computed grows with increasing time intervals. Thus, it could be preferred to have an approximate method that avoids this problem, provided that sufficient accuracy and adequate dynamical properties are guaranteed.

To this end, an aproximate, nonstandard method is proposed to compute the numerical solution from the $M+1$ interval on, where previous values are assumed to be computed either by the exact method or by any numerical method of sufficient high accuracy.

Fix $M \geq 1$, and compute the numerical solution in the first M intervals either with the exact method (6) or with any other numerical method of

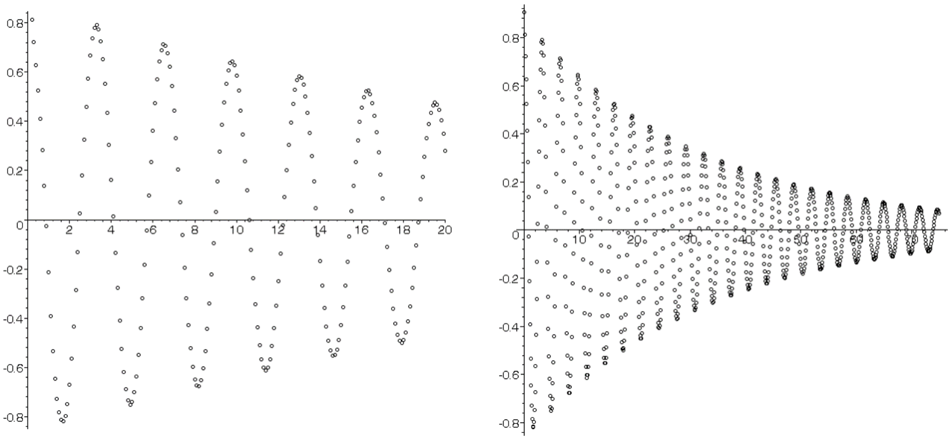


Figure 2: Numerical solution provided by the nonstandard scheme, with $M = 1$ and $N = 10$, for the problem (1)-(2) with initial function $f(t) = (t+1)^2$ and parameters $\tau = 1$, $\alpha = -1$, and $\beta = -2.2$, for $t \in [0, 20]$ (left), and $t \in [0, 75]$ (right).

order at least $O(h^{M+1})$. Then, for $m \geq M+1$ and $(m-1)\tau < nh \leq m\tau$, the expression

$$x_{n+1} = e^{\alpha h} \sum_{k=0}^{M} \frac{\beta^k h^k}{k!} x_{n-kN} \tag{7}$$

defines a nonstandard numerical scheme of uniform accuracy $O(h^{M+1})$.

It can be shown that these new schemes are dynamically consistent with the original delay problems, so that the numerical solutions obtained with these schemes, for any choice of M, are asymptotically stable whenever the exact solution is stable. An example illustrating this property is shown in Figure 2, where the values of the parameters have been selected close to the limits that define the region of asymptotic stability.

Another interesting dynamic property of delay models is the behaviour in terms of oscillation and non-oscillation of the solutions. It can also be proved that whenever every solution of (1)-(2) oscillates, then the numerical solution obtained with the nonstandard method defined by (7) also oscillates.

An example illustrating this consistency with the oscillation properties of the exact solution is presented in Figure 3, where the values of the parameters have also been selected close to the limits that define the region where all the solutions oscillate.

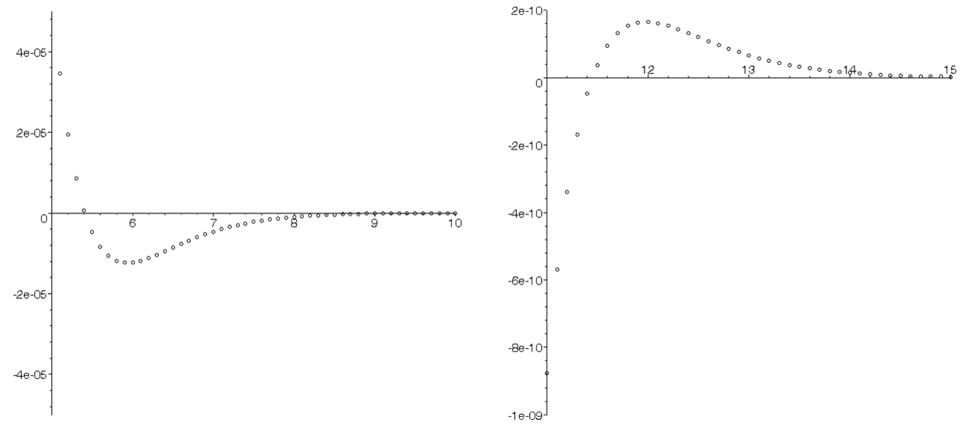


Figure 3: Numerical solution provided by the nonstandard scheme, with $M = 1$ and $N = 10$, for the problem (1)-(2) with initial function $f(t) = (t+1)^2$ and parameters $\tau = 1$, $\alpha = -1$, and $\beta = -0.15$, for $t \in [0, 10]$ (left), and $t \in [11, 15]$ (right).

Finally, a desirable property of any numerical method is the preservation of positive solutions, as this positivity is usually required in real applications of the models. Here again it can be proved that in the conditions where the exact solution preserves positivity, for any positive initial function $f(t)$, the numerical solutions provided by the new standard methods also guarantee that positivity is preserved.

References

[1] V. Kolmanovskii and A. Myshkis, Introduction to the Theory and Applications of Functional Differential Equations. Dordrecht, Kluwer Academic Publishers, 1999.

[2] R.E. Mickens, Nonstandard Finite Difference Models of Differential Equations. Singapore, World Scientific, 1994.

[3] S.M. Garba, A.B. Gumel, A.S. Hassan, J.M.-S. Lubuma. Switching from exact scheme to nonstandard finite difference scheme for linear delay differential equation, *Appl. Math. Comput.,* 258:388–403, 2015.

[4] J.A. Martín, F. Rodríguez, R. Company, Analytic solution of mixed problems for the generalized diffusion equation with delay, *Math. Comput. Modelling,* 40:361-369, 2004.

[5] E. Reyes, F. Rodríguez, J.A. Martín, Analytic-numerical solutions of diffusion mathematical models with delays, *Comput. Math. Appl.,* 56:743-753, 2008.

Pricing commodity options in jump-diffusion models[*]

L. Gómez-Valle[†], Z. Habibilashkary[‡], and J. Martínez-Rodríguez[§]

Facultad de Ciencias Económicas y Empresariales, Universidad de Valladolid,

Avenida del Valle Esgueva, 6, 47011-Valladolid, Spain

November 30, 2016

1 Introduction

In recent years, in the markets, there have been a steady growth in the number and type of derivatives. Moreover, the valuation of commodity derivatives differs considerably from the valuation of common stock derivatives, especially because of the particularities of the underlying commodity. In energy markets, as natural gas or electricity, we should explore the seasonal behaviour of the commodity price and take this fact into account in the pricing models.

The aim of this paper is to propose a new technique for estimating the functions of the risk-neutral stochastic processes of the seasonal pricing model, by means, of market data. This fact allows us to obtain the market prices of risk and price derivatives accurately when a closed-form solution is

[*] L. Gómez-Valle and J. Martínez-Rodríguez were supported in part by the GIR Optimización Dinámica, Finanzas Matemáticas y Utilidad Recursiva of the University of Valladolid and the projects: VA191U13 of the Consejería de Educación JCyL and MTM2014-56022-C2-2-P of the Spanish Ministerio de Economía y Competitividad and European FEDER Funds.

[†] e-mail: lourdes@eco.uva.es

[‡] e-mail: ziba.habibilashkary@alumnos.uva.es

[§] e-mail: julia@eco.uva.es

not known. Moreover, we analyse the impact of the seasonality when pricing natural gas futures options. NYMEX (New York Mercantile Exchange Market) data is used in the empirical applications.

2 The valuation model

The following section presents a two-factor jump-diffusion model with seasonality. We assume that the two factors of the model are: the dynamics of the spot price, S, and the instantaneous convenience yield, δ. Let define $(\Omega, \mathcal{F}, \mathcal{P})$ as a probability space with a filtration $\mathcal{F}$ that satisfies the usual conditions, [5].

This research describes the behaviour of the log-price process in terms of two types of components. The first one is a known predictable deterministic function of time $f(t)$ and the second one, X, is stochastic. Both of them verify that

$$\ln S(t) = f(t)X(t), \quad t \in [0, \infty).$$

In particular, we assume that X and δ follow a jump-diffusion and a diffusion stochastic process, respectively. Moreover, we assume that:

$$\begin{aligned} dX &= \left(\mu_X - \sigma_X \theta^{W_X} + \lambda^{\mathcal{Q}} E_Y^{\mathcal{Q}}[J]\right) dt + \sigma_S dW_X^{\mathcal{Q}} + J d\tilde{N}^{\mathcal{Q}}, && (1) \\ d\delta &= \left(\mu_\delta - \sigma_\delta \theta^{W_\delta}\right) dt + \sigma_\delta dW_\delta^{\mathcal{Q}}, && (2) \end{aligned}$$

where $\mu_X(X, \delta)$ and $\mu_\delta(X, \delta)$ are the drifts and $\sigma_X(X, \delta)$ and $\sigma_\delta(X, \delta)$ are the volatilities. The jump amplitude J is a function of X, δ and Y, which is a normal random variable, $Y \rightsquigarrow N(0, \sigma_Y^2)$. Moreover, $W_X^{\mathcal{Q}}(t)$ and $W_\delta^{\mathcal{Q}}(t)$ are the Wiener processes under $\mathcal{Q}$-measure and $Cov(W_X^{\mathcal{Q}}, W_\delta^{\mathcal{Q}}) = \rho t$.

The market prices of risk of the Wiener processes are $\theta^{W_X}(X, \delta)$ and $\theta^{W_\delta}(X, \delta)$, and $\tilde{N}^{\mathcal{Q}}$ represents a compensated Poisson process with intensity $\lambda^{\mathcal{Q}}(X, \delta)$. As usual in the literature, for simplicity and tractability, we assume that, under $\mathcal{Q}$-measure, the jump size distribution of the jump-diffusion process is known and equal to the distribution under $\mathcal{P}$-measure. This means that all risk premium related to the jump are absorbed by the change in the intensity of the jump λ^Q, see [3].

By means of Itô's Lemma and provided that the function f satisfies the appropriate regularity conditions [4], the process followed by the risk-neutral spot price can be expressed as the solution to the following stochastic differential equation

$$\begin{aligned} dS &= S\left(f' + f(\mu_X - \sigma_X\theta^{W_X}) + \frac{1}{2}\sigma_X^2 f^2 + \lambda^Q\left(e^{\frac{f^2\sigma_Y^2}{2}} - 1\right)\right)dt \\ &+ \sigma_X f^2 S dW_X^Q + S(e^{fJ} - 1)d\widetilde{N}^{\mathcal{Q}}. \end{aligned} \quad (3)$$

In case we do not consider seasonality in the model then, $\ln S = X$ and $f \equiv 1$ and $f' \equiv 0$.

A commodity futures price at time t with maturity at time T, $t \leq T$, can be expressed as $F(t, S, \delta; T)$ and at maturity it verifies that $F(T, S, \delta; T) = S$. We also assume that there exists a replicating portfolio for the futures price and then, the futures price can be expressed by

$$F(t, S, \delta; T) = E^{\mathcal{Q}}[S(T)|S(t) = S, \delta(t) = \delta], \quad (4)$$

where $E^{\mathcal{Q}}$ denotes the conditional expectation under the $\mathcal{Q}-$measure.

Let $V(t, S, \delta, T_2; T_1)$ be the price of an European call option that matures on T_1 on a futures contract that expires at T_2, $T_1 \leq T_2$, and K is the strike price. Then, analogously to (4), an European commodity futures option is priced as the expected discounted payoff under the $\mathcal{Q}-$measure, see [6],

$$\begin{aligned} &V(t, S, \delta, T_2; T_1) = \\ &\quad E^{\mathcal{Q}}[\mathrm{e}^{\,-\int_t^{T_1} r(u)\,du} \max\left(F(T_1, X, \delta; T_2) - K, 0\right)|S(t) = S, \delta(t) = \delta], \end{aligned} \quad (5)$$

where r denotes the instantaneous risk-free interest rate, which is assumed constant. Moreover, $\tau_1 = T_1 - t$ and $\tau_2 = T_2 - T_1$ are the maturity of the option contract and futures contract, respectively.

3 Exact results and approximations

This section proposes some results for estimating the functions of the risk-neutral stochastic factors of a seasonal commodity model directly from market data.

Theorem 1 *Let $F(t, S, \delta; T)$ be the price of the future (4), with $\ln S(t) = f(t)X(t)$, and X and δ follow the stochastic processes given by (1) and (2), respectively, then:*

$$\frac{\partial F}{\partial T}(t, S, \delta; T) = F(t, S, \delta; T)\left(f' + f(\mu_X - \sigma_X \theta^{W_X}) + \frac{1}{2}\sigma_X^2 f^2 + \lambda^{\mathcal{Q}}\left(e^{\frac{f^2\sigma_Y^2}{2}} - 1\right)\right)(T)$$

$$\frac{\partial (SF)}{\partial T}(t, S, \delta; T) = F(t, S, \delta; T)\left(2\frac{\partial F}{\partial T} + S(f^4\sigma_X^2 + \lambda^{\mathcal{Q}}\left(e^{f^2\sigma_Y^2} + 1 - 2\left(e^{\frac{f^2\sigma_Y^2}{2}}\right)\right)\right)\right)(T)$$

$$\frac{\partial (\delta F)}{\partial T}(t, S, \delta; T) = \left(\delta\frac{\partial F}{\partial T} + S(\mu_\delta - \sigma_\delta \theta^{W_\delta}) + S f^2 \rho\, \sigma_X \sigma_\delta\right)(T)$$

We proved these results by means of (4) with a similar reasoning to that in [2].

4 Commodity Derivatives Pricing

In this section, we price some natural gas derivatives using the model in Section 2 and results in Section 3.

Natural gas spot and futures prices are obtained from the E.I.A. (Energy Information Administration of the US Department of Energy) and Quandl platform. The sample period goes from January 2004 to December 2014.

The deterministic seasonal function, $f(t)$, is a fit of the monthly average of the historical data with a fourth order Fourier approximation.

As far as the convenience yield is concerned, it is not observable in the markets but we approximate it as in [1] with T-Bill rates obtained from the Federal Reserve h.15 database.

In order to price natural gas futures and options we obtain the conditional expectations in (4) and (5), respectively. We get these expectations by means of Monte Carlo approach and (1) and (2). We estimate all the necessary functions with Theorem 1 and a Kernel method. The derivatives are approximated by means of a fourth-order approximation.

So as to make comparisons, we also consider a model without seasonality, that is $\ln S = X$. In this case, we obtain a similar result to Theorem 1 and we also use a Kernel method to estimate the whole functions.

Table 1 shows the RMSE (Root Mean Square Error) of the futures with different maturities (F1 means one month, F2 two months and so on), for the out-of sample (January-July 2015) with the SM (seasonal model) and

Futures	F1	F6	F9	F12	F18	F24	F30	F36	F42	F44
NSM	0.1607	0.2445	0.1446	0.1797	0.1435	0.2490	0.2346	0.3624	0.3088	0.3850
SM	0.1232	0.1473	0.0920	0.1577	0.0949	0.1847	0.1099	0.2453	0.1674	0.2595

Table 1: RMSE for the out-of-sample, January-July 2015, for NSM and SM.

the NSM (non seasonal model). Note that for the whole maturities, the SM provides lower errors than the NSM. Moreover, the higher the maturity, the higher the differences.

In Table 2, we show the ratios between the option prices obtained with the NSM and the SM, for several strike prices and maturities. We assume that the maturity of the option is equal to that of the underlying futures contract. In this table, we show that for short maturities, that is three months, the SM underprices with respect to the NSM. However, for maturities higher or equal to 6 months, the SM overprices the natural gas options. This fact should be taking into account by practitioners in the markets.

Strike \ Maturity	3 months	6 months	9 months	12 months
90%	1.20	0.75	0.50	0.37
100%	1.39	0.78	0.49	0.36
110%	1.70	0.81	0.48	0.34

Table 2: Ratios between the NSM and SM option prices.

References

[1] Gibson, R. and E.S. Schwartz, Stochastic convenience yield and the pricing of oil contingent claims, *The Journal of Finance*, 45 (3), 959–976, 1990.

[2] Gómez-Valle, L., Z. Habibilashkary and J. Martínez-Rodríguez, A new technique to estimate the risk-neutral processes jump-diffusion commodity futures models, *Journal of Computational and Applied Mathematics*, 309, 435–441, 2017.

[3] S.K. Nawalkha, N. Beliaeva and G. Soto, Dynamic Term Structure Modeling: The Fixed Income Valuation Course, John Wiley & Sons, Inc, 2007.

[4] Øksendal, B.,and A. Sulem, Applied Stochastic Control of Jump Diffusions, Berlin, Heidelberg, Springer-Verlag, 2007.

[5] Runggaldier, W.J., Jump-diffusion models, in: S.T. Rachev (Ed.), Handbook of Heavy Tayled Distributions in Finance. Universitat Karisruhe, Karisruhe, North Holland, Germany, 169–209, 2003.

[6] X. Yan, Valuation of commodity derivatives in a new multi-factor model, *Rev. Derivatives Res*, 5, 251-271, (2002).

Updating preconditioners for least squares problems [*]

J. Marín$^\flat$ [†], J. Mas$^\flat$, and K. Hayami[†]

($\flat$) Instituto de matemática Multidisciplinar,

Universitat Politècnica de València,

(†) National Institute of Informatics,

SOKENDAI, (The Graduate University for Advanced Studies), Tokio

November 30, 2016

1 Introduction

We are interested in computing the least squares (LS) solution of the overdetermined linear system

$$Ax = b, \tag{1}$$

where A is a large and sparse $m \times n$ matrix, $m > n$, using a preconditioned iterative method. We assume that A has full rank n. As it is well known the LS solution is given by the vector x that minimizes $\|b - Ax\|_2$, and can be obtained by solving the normal equations corresponding to (1) given by

$$A^T Ax = A^T b. \tag{2}$$

This solution can be obtained using iterative methods. To improve the convergence of the iterative method a preconditioner can be applied, we will focus on Incomplete Cholesky (IC) preconditioners (see [1, 3, 6]).

[*] Partially supported by Spanish Grants MTM2014-58159-P and MTM2015-68805-REDT.

[†] e-mail: jmarinma@imm.upv.es, jmasm@imm.upv.es, hayami@nii.ac.jp

Suppose that the system (1) is modified, i.e., updated or downdated. These situation can arise in some applications from statistics, optimization and signal processing, since it is necessary to solve a sequence of modified least squares problems, see [2, Chapter 3] for more information.

If some new relations between the unknowns are considered and these relations are given as the system of r linear equations $Bx = c$, the new coefficient matrix $\left[\begin{smallmatrix}A\\B\end{smallmatrix}\right]$ has also full rank, and the corresponding normal equations are $(A^TA + B^TB)x = A^Tb + B^Tc$, that is, the new normal equations are the result of a low-rank update of the initial ones. If we put $f = A^Tb + B^Tc$, the components in x of the solution of the bordered linear system

$$\begin{bmatrix} A^TA & B^T \\ B & -I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}.$$

give the solution of the normal equations of the updated system.

On the other hand, if some linear equations are removed from the initial linear system (1), we can assume, without loss of generality, that they are the last equations so that, the normal equations of the original system are $(\tilde{A}^T\tilde{A} + C^TC)x = \tilde{A}^T\tilde{b} + C^Td$ and we want to remove the information contained in C and d to get $\tilde{A}^T\tilde{A}x = (A^TA - C^TC)x = \tilde{A}^T\tilde{b}$. This system is the result of a low-rank modification of the initial normal equations, and it has the same solution as component x in the solution of the augmented linear system

$$\begin{bmatrix} A^TA & C^T \\ C & I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \tilde{A}^T\tilde{b} \\ 0 \end{bmatrix}$$

We proposed in [4] a strategy to update an incomplete LU preconditioner when the coefficient matrix of a square system of linear equations is updated by an small rank matrix using the bordering technique. In the next section we adapt this technique for the two problems mentioned above.

2 Preconditioner update

We briefly describe this bordering technique in the symmetric case. Assume one has computed an incomplete Cholesky factorization of the symmetric and positive definite $n \times n$ matrix M $(= A^TA)$, and then the matrix is updated by a low-rank matrix that can be written as P^TP, where P is a $r \times n$ matrix with $r \ll n$, i.e, the updated matrix is $M \pm P^T P$. Clearly one has that,

$$M \pm P^T P = \left[\begin{array}{cc} I & O \end{array} \right] \left[\begin{array}{cc} M & P^T \\ P & \mp I \end{array} \right] \left[\begin{array}{c} I \\ P \end{array} \right] \tag{3}$$

$$(M \pm P^T P)^{-1} = \left[\begin{array}{cc} I & O \end{array} \right] \left[\begin{array}{cc} M & P^T \\ P & \mp I \end{array} \right]^{-1} \left[\begin{array}{c} I \\ O \end{array} \right]. \tag{4}$$

The preconditioner update technique consists in computing a preconditioner for the augmented linear operator in (3) that is used to approximate the inverse linear operator in (4) by direct preconditioning, i.e., solving the corresponding upper and lower triangular systems. Therefore we avoid the computation of a new preconditioner for the updated matrix $M \pm P^T P$ from scratch because we can reuse the previously computed IC of $M \approx R_M^T R_M$, then one gets a block LDL^T square root free Cholesky factorization of the augmented matrix in (3)

$$\left[\begin{array}{cc} M & P^T \\ P & \pm I \end{array} \right] = \left[\begin{array}{cc} R_M^T & 0 \\ R_{12}^T & I \end{array} \right] \left[\begin{array}{cc} I & 0 \\ 0 & \pm S \end{array} \right] \left[\begin{array}{cc} R_M & R_{12} \\ 0 & I \end{array} \right], \tag{5}$$

where $R_{12} = R_M^{-T} P^T$, $S = I \mp R_{12}^T R_{12}$. To maintain these factors sparse some dropping strategy can be used when computing R_{12} and an incomplete factorization of the Schur complement S as well, but if r is small enough this block can be factorized exactly. Using the updated preconditioner in the augmented form (3), the preconditioned vector is obtained from the solution of

$$\left[\begin{array}{cc} R_A^T & 0 \\ R_{12}^T & I \end{array} \right] \left[\begin{array}{cc} I & 0 \\ 0 & \pm S \end{array} \right] \left[\begin{array}{cc} R_A & R_{12} \\ 0 & I \end{array} \right] \left[\begin{array}{c} s \\ s' \end{array} \right] = \left[\begin{array}{c} r \\ 0 \end{array} \right]. \tag{6}$$

3 Numerical experiments

In this section we study the numerical performance of the preconditioner update method proposed. We present results obtained with matrices arising in different areas of scientific computing. The performance of the updated preconditioner is compared with two other preconditioning strategies. The first one consists in reusing the initial preconditioner computed for the normal equations of the unmodified matrix, while the second strategy corresponds to the computation of a new preconditioner for the updated matrix from scratch. Also, the results for the non-preconditioned iterations are reported.

Matrix name	rows	cols	nnz	Application
TESTBIG	17613	31223	61639	Linear programming problem
DELTAX	68600	21961	247424	High fillin with partial pivoting
FOME13	48568	97840	285046	Linear programming problem
LP_OSA_30	4350	104375	604488	Linear programming problem
MESH_DEFORM	234023	9393	853829	Image mesh deformation problem
SLS	1748122	62729	6804304	Statistics

Table 1: Set of tested matrices, nnz is the number of nonzero entries.

The tested matrices can be downloaded from the University of Florida Sparse Matrix Collection [5] and are shown in Table 1.

The preconditioned CGLS method [2] was used for a relative initial residual decrease of 10^{-8}, allowing a maximum number of $2,000$ iterations. The right hand side vector was computed as $b = Ae$, where e is the vector of all ones. The initial approximation to the solution x was the vector of all zeros. The experiments where obtained with MATLAB. The function **ilut**() was used to compute a Cholesky factorization for the normal equations, and also to factorize the Schur complement matrix S.

Matrix	r	ρ	non-prec	non-upd.	recomp.	updated
			Its./T.	Its./T.	Its./T.*	Its./T.*
TESTBIG	300	[0.54,0.56]	69/0.1	42/0.1	45/0.3	42/0.1
	600		70/0.1	42/0.1	45/0.3	40/0.1
	3200		137/0.2	94/0.2	76/0.4	72/0.1
DELTAX	50	[0.82,0.84]	†	1424/6.2	1332/5.8	1151/5.4
	300		1407/3.6	996/4.4	886/5.1	809/3.5
	1200		767/2.0	715/3.1	515/3.8	650/2.7
FOME13	100	[0.79,0.82]	†	1499/16.1	1306/14.9	1285/13.9
	1000		1040/6.3	237/2.6	183/2.3	181/2.0
	2000		713/4.3	196/2.1	134/1.9	166/2.0
LP_	500	[0.36,0.37]	525/1.9	167/0.8	103/0.8	118/0.5
OSA_30	5000		611/2.3	304/1.4	77/0.6	276/1.3
	10000		582/2.3	374/1.9	81/0.7	331/2.0
MESH_	100	[0.16,0.26]	1107/12.8	23/0.3	16/1.4	16/0.2
DEFORM	2300		691/5.1	45/0.4	15/2.4	16/0.2
	11500		427/3.4	72/0.7	13/5.7	39/0.4
SLS	175	[0.02]	424/30.6	133/9.9	132/10.1	119/8.1
	3500		436/31.3	157/11.5	128/9.8	137/9.2
	17500		465/33.4	188/14.2	122/9.5	172/10.9

Table 2: Effect of the rank of the update for the updating case.

Matrix	r	ρ	non-prec	non-upd.	recomp.	updated
			Its./T.	Its./T.	Its./T.*	Its./T.*
TESTBIG	100	[0.55,0.64]	123/0.1	84/0.1	71/0.1	52/0.1
	600		116/0.1	86/0.1	72/0.1	52/0.1
	3200		110/0.1	89/0.1	69/0.1	52/0.1
DELTAX	50	[0.81,0.92]	1154/2.7	811/3.6	812/4.9	741/2.9
	1200		1942/5.0	1300/5.6	1191/6.5	853/3.7
	6800		†	†	†	853/3.7
FOME13	10	[0.79]	1040/4.4	234/1.6	234/2.1	247/1.6
	100		1097/5.2	290/2.2	268/2.4	247/1.6
	300		1282/5.7	301/2.3	†	247/1.6
LP_OSA_30	500	[0.36,0.42]	130/0.5	36/0.2	23/0.3	31/0.2
	5000		129/0.4	61/0.3	35/0.3	39/0.2
	10000		136/0.4	65/0.3	38/0.3	39/0.2
MESH_DEFORM	100	[0.16,0.32]	1134/6.6	29/0.2	16/1.1	17/0.2
	2300		1262/7.3	117/0.8	18/1.1	27/0.8
	4600		1336/7.7	162/1.1	21/1.2	39/1.5
SLS	175	[0.02]	1394/108.5	150/12.1	146/12.4	131/9.4
	3500		†	178/14.7	178/15.2	149/9.9
	17500		†	303/24.4	621/51.2	149/9.8

Table 3: Effect of the rank of the update for the case of downdating.

Tables 2 and 3 report the results for the cases of adding and removing equations, respectively. In these tables, r represents the rank of the update, i.e., the number of equations added or removed. The column ρ shows the minimum and maximum relative densities of the preconditioner with respect to the updated matrix. Normally, the minimum value corresponds to the non-updated preconditioner while the maximum was achieved for either, the recomputed or the updated preconditioner. The number of iterations and CPU solution time are indicated with Its. and T., respectively and T.* indicates total timing corresponding to the preconditioner recomputation or update and the iterative solution.

We start analyzing the results for the case of adding equations that are shown in Table 2. The equations added were obtained by selecting at random r rows of the original matrix, and ordering in reverse order their column entries to avoid duplicated rows. The size of the modification ranges from a very few rows compared with the size of the matrix to roughly 10% of its size, depending on the matrix. With respect to the preconditioner density, we observe that it was roughly equal for all the preconditioners.

From the number of iterations we see that the updated preconditioner performed better than the non-updated one, and closely to the case of recomputing the preconditioner for moderate values of r while it tends to degrade when the number of equations added increase. But, taking into account the overall time, our strategy performed better in most of the cases.

In Table 3 we observe that, if the modification consists on removing a block of equations, the situation changes a bit and in favor of our proposed algorithm. In this case the computation of a preconditioner for the new matrix becomes more unstable for an increasing number of equations removed.

We can conclude from our numerical experiments that the proposed algorithm is competitive and robust. Moreover, we also think that instead of reusing the preconditioner it is better to update or recompute a new preconditioner from scratch. This last strategy has two drawbacks: an increment on the set-up time that usually only pays off in the case of adding equations if the size of the update is quite large and that, when removing equations, the preconditioner computation may become unstable.

References

[1] M. Benzi and M. Tůma. A robust incomplete factorization preconditioner for positive definite matrices. *Numer. Linear Algebra Appl.*, 10(5-6):385–400, 2003.

[2] Å. Björck. *Numerical methods for Least Squares Problems.* SIAM, Philadelphia, 1996.

[3] R. Bru, J. Marín, J. Mas, and M. Tůma. Preconditioned iterative methods for solving linear least squares problems. *SIAM J. Sci. Comput.*, 36(4):A2002–A2022, 2014.

[4] J. Cerdán, J. Marín, and J. Mas. Low-rank updates of balanced incomplete factorization preconditioners. *Numer. Algorithms*, pages 1–34, 2016.

[5] T. A. Davis. *University of Florida Sparse Matrix Collection.* available online at http://www.cise.ufl.edu/∼davis/sparse/, NA Digest, vol. 94, issue 42, October 1994.

[6] Y. Saad. ILUT: a dual threshold incomplete *LU* factorization. *Numer. Linear Algebra Appl.*, 1(4):387–402, 1994.

A Greedy-one Rank algorithm in Mobile Robot Applications

N.Montes[♭] [*], A.Falco[♭] , L.Hilario[♭] , M.C.Mora[†], and F.Chinesta[‡]

(♭) University CEU Cardenal Herrera,

C/San Bartolomé 55, Alfara del Patriarca, Valencia,

(†) University Jaume I,

Avda. Vicent Sos Baynat s/n, Castellón, Valencia

(‡) Institute of High Performance Computing. Ecole Central de Nantes,

Rue de la Noe, Nantes, Fránce.

November 30, 2016

1 Introduction

The present paper shows, for the first time, as the so-called Greedy Rank-one Algorithm can be use to path planning mobile robots. The main idea of this method is to obtain a separated representation form of a parametrised solution of a particular Poison equation with a source function which represents a starting and goal positions (the parameters) derived from a harmonic potential field. The Greedy Rank-one Algorithm is a tensor numerical technique with three main advantages. The first one is the ability to bring together all the possible Poisson equation solutions for all start and goal combinations in a map, guaranteeing that the resulting potential field does not have deadlocks. The second one is that the constructed solutions expressed as a sum of uncoupled multiplied terms: the geometric map and the start and goal configurations. Therefore, the harmonic potential

[*] e-mail: nicolas.montes@uchceu.es

field can be reconstructed extremely fast, in a nearly negligible computational time, allowing real-time path planning. The third one is that only a few uncoupled parameters are required to reconstruct the potential field with a low discretisation error. Simulation results are shown to validate the abilities of this technique.

2 Potential flow theory

Path planning based on potential flow theory has been used in the literature during the last years, see[1] -[2], focused mainly in the resolution of the Laplace equation. First of all, lets us outline the mathematical model describing the flow of an inviscid incompressible fluid. Assuming a steady state irrotational flow in the Eulerian framework, the velocity V obeys the relation

$$\bigtriangledown \times V = 0 \tag{1}$$

As a consequence, velocity is the gradient of a scalar u named potential function, so that the potential u is harmonic (Solution of the Laplace equation)

$$\bigtriangledown^2 \times u = 0 \tag{2}$$

The resolution of the Poisson equation under these conditions produces a potential field from the Start point A (source) to the Goal point B (sink), without deadlocks, see [2].

2.1 Greedy one rank algorithm at a glance

Consider the solution of the Poisson equation

$$\Delta u(x,y) = f(x,y) \tag{3}$$

in a two-dimensional rectangular domain $\Omega_{\underline{X}} = \Omega_X \times \Omega_Y$ with Newman conditions $\left.\frac{\partial u}{\partial n}\right|_{\Gamma} = q$ where for our particular case q=0. For all suitable test functions u^* , the weighted residual forms reads

$$\int_{\Omega_{\underline{X}}} u^* \cdot (\Delta u - f) d\Omega_{\underline{X}} \tag{4}$$

The classical way of accounting for Neuman conditions is to integrate by parts the weighted residual form and implement the flux condition as a so-called natural boundary condition:

Our goal is to obtain an approximate solution in the separated form

$$u(\Omega_{\underline{X}}) = \sum_{i=1}^{N} X_i(x) \cdot Y_i(y) \tag{5}$$

We shall do so by computing each term of the expansion one at a time. Thus enriching the Greedy one approximation until a suitable convergence criterion is satisfied.

2.2 Source term definition

First of all, it is neccesary to assume that a constant source term f is in really a non-uniform source term $f(\Omega_{\underline{X}}, \Omega_{\underline{S}}, \Omega_{\underline{T}})$ where $\Omega_{\underline{X}} = \Omega_{\underline{x}} \times \Omega_{\underline{y}}$, $\Omega_{\underline{S}} = \Omega_{\underline{r}} \times \Omega_{\underline{s}}$,$\Omega_{\underline{T}} = \Omega_{\underline{r}} \times \Omega_{\underline{t}}$. In this definition, the start point S and the target point T are defined a Gaussian model with variance and median $\underline{S} = (r, s)$, $\underline{S} = (r, t)$, where r is the variance and s,t are the median value located in an specified point $\underline{X} = (x, y)$ in each separated space $\Omega_{\underline{S}}, \Omega_{\underline{T}}$. Therefore, the source term is defined as:

$$\begin{aligned} f(\underline{X}, \underline{S}) &= \sum_{j=1}^{F} \alpha_j^S \cdot F_j^S(\underline{X}) \cdot G_j^S(\underline{S}) \\ g(\underline{X}, \underline{T}) &= \sum_{j=1}^{F} \alpha_j^T \cdot F_j^T(\underline{X}) \cdot G_j^T(\underline{T}) \end{aligned} \tag{6}$$

Then, the Posion equation to solve is now in the form

$$\Delta u(x, y) = f(\underline{X}, \underline{S}) + g(\underline{X}, \underline{T}) \tag{7}$$

2.3 Greedy rank one algorithm definition

For all suitable test functions u^*, the weighted residual forms reads

$$\int\limits_{\Omega_{\underline{X},\underline{S},\underline{T}}} u^* \cdot (\Delta u - f)\, d\Omega_{\underline{X},\underline{S},\underline{T}} = 0 \tag{8}$$

where f is in the form

$$f = f(\underline{X}, \underline{S}) + g(\underline{X}, \underline{T}) \tag{9}$$

Now, equation reads to

$$\begin{aligned} \int_{\Omega_{\underline{X},\underline{S},\underline{T}}} \bigtriangledown u^* \cdot \bigtriangledown u \, d\Omega_{\underline{X},\underline{S},\underline{T}} = \int_{\Omega_{\underline{X},\underline{S},\underline{T}}} u^* \cdot f \, d\Omega_{\underline{X},\underline{S},\underline{T}} - \\ - \int_{\Omega_{\underline{X},\underline{S},\underline{T}}} u^*(x, y = \Gamma) \cdot q \, d\Omega_{\underline{X},\underline{S},\underline{T}} \end{aligned} \tag{10}$$

And the Greedy one formulation is now as

$$u(\underline{X}, \underline{S}, \underline{T}) = \sum_{i=1}^{N} R_i(\underline{X}) \cdot W_i(\underline{S}) \cdot K_i(\underline{T}) \tag{11}$$

To do so, the alternating directions to construct the separated representation is used to compute it in a 5x5m square environment. This environment is discretized using 50 nodes in each side, $D_x \times D_y = 50 \times 50$, that is 2500 nodes. In the source term r, the variance is selected to 1.2 and s,t, the median value is selected to 1. The computational costs of the reconstruction is 0.0101 Sec in a Mac with an Intel Core 2 Duo, 3.06 GHz and 4 GB of RAM memory. This computational cost is compared with a FEM simulation using linear approximation techniques and the computational costs is 4.7 Sec.

References

[1] A.Saudi, J.Sulaiman. Path Planing for mobile robots using 4EGSOR via Nine-Point Laplacian (4EGSOR9L) Iterative method. *International Journal of computer applications*, Volume(53):38-42, 2012. . Vol 53, No16 pp 38-42. 2012.

[2] D.Gingras, E.Dupuis, G.Payre, J.Lafontaine. Path Planning Based on Fluid mechanics for mobile robots used Unstructured Terrain models.*EEE International Conference on Robotics and Automation.*, Anchorage, Alaska, USA 2010.

VAACA: Virtual Adaptive Agent Cluster Architecture

F.J. Mora, F. Aznar, M. Sempere, J.A. Puchol
P. Arques, M. Pujol, R. Rizo

University of Alicante

e-mail: [mora,fidel,mireia,puchol,arques,mar,rizo] @dccia.ua.es

Acknowledgements. Ministerio de Economia y Competitividad (Spain), Project TIN2013-40982-R. Project co-financed with FEDER funds.

November 30, 2016

1 Introduction

Multi-agent systems are used today in many domains where it is needed to work with heterogeneous systems located along different corporations and continents. Over the past 15 years many frameworks for working with agents have been used, such as Zeus, JADE, Comtec ... These frameworks try to fulfill standards of various international organizations, such as FIPA [1]. These organizations are responsible for describing the reference models of agents: roles of main agents (AMS, DF, ACL), communication languages between agentes (KQML, ACL, IIOP), software integration for agents, mobility, security, ontologies, etc. In this way, developers take advantage of these agent frameworks to build complex systems. Among these advantages we can mention the ease of implementation, modularity, execution speed and reliability. However, as other distributed systems, these frameworks are prone to failure. Agents and resources may become unavailable due to failures in communications, in the machine that contains the agents, in some process and other hardware and software problems.

2 Fault-tolerance strategies of AMS

When an application-agent fails we can look for another agent in the yellow pages agent (DF) to replace it. However, when a platform coordinator agent fails, things are much more complicated, especially if it is the agent management system (AMS).

For example, in JADE (Java Agent Development Framework) is possible to clone or to move an agent between different containers [2] but it is not possible between different platforms. Container architecture does not contain the AMS, the DF and the ACC. These modules are only implemented in the Front End Container. This is the problem of JADE concerning fault-tolerance: if Front End container fails and specifically the AMS, the system will collapse, requiring a mechanism to recovery the AMS copy or to restart the Front End. These mechanisms are not implemented in JADE.

Another interesting alternative is to build a cluster of agents (VAC) that propose a decentralized AMS into a distributed AP. VAC provides fault-tolerance based on communication layers that are separated into different machines. The main advantage of this architecture is that it has more than one access point (centralized AMS) obtaining a higher performance and avoiding the dependence on an element of the architecture. In [4] we can to see the architecture AAA (Adaptive Agent Architecture), that uses a team of persistent brokers.

There are a lot more techniques for fault recovery: hot backups, object group replication, virtual synchrony, n-version voting, but all of these techniques require important changes in existing agent platforms.

3 Proposed architecture: VAACA - Virtual Adaptive Agent Cluster Architecture

We will to present an architecture that improves fault-tolerance in MAS. One the one hand, this architecture supports communication between platforms (many other platforms, such as JADE, do not support it). On the other hand, our architecture improve VAC architecture using standart protocols for communication between different clusters.

VAACA architecture is presented in figure 1. This consists of a cluster of agent platforms that can interact with each other. Each platform can be distributed in different hosts using containers. Moreover, a broker agent can be found in each platform. This agent is responsible for coordinating the information of all AMS agents of different platforms.

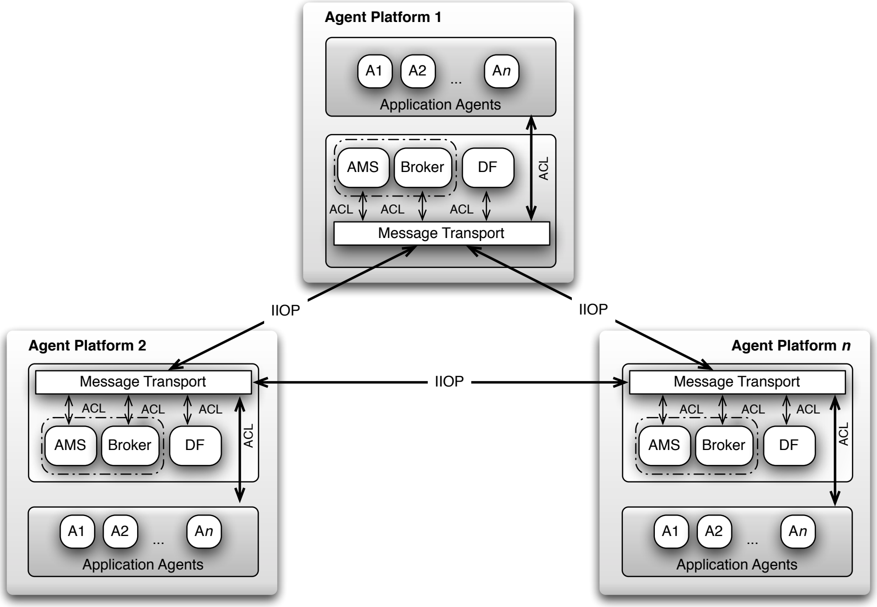


Figure 1: VAACA architecture. Brokers team will be synchronized and will enable the connection in case of the failure of the AMS.

Agent communication is one of the most important parts of this architecture. Different agents could be in different containers or even platforms so that communication methods could be quite different.

As in other multi-agent systems, agents can communicate directly with each other. This characteristic is used by brokers and it will be crucial when AMS fails. The VAACA brokers act as a team to achieve fault-tolerance. These agents share the knowledge about who is connected to whom in the team.

Broker agent behaviours are defined using logical characterization rather than using conventional algorithms, because we found easy to mantain the system specifying in a formal way both, goals and behaviours. More specifically, we use this characterization for implementing the synchronization between teams and broker agents, based on the AAA architecture [4]. Table 1 shows the description of the missions and main rules used in this architecture.

Mission Statement 1	**Mission Statement 2**	**Mission Statement 3**
Main goal of brokers is to register new agents and to keep the connection with the team until the agent is not longer registered.	Second objective is to keep the number of brokers, promoting agents to the category of broker when necessary.	Third objective of brokers is to synchronize and share the information and the status of registered agents and the team brokers.
Mission derived theorems		
Brokers have a commitment to accept agent registration.	When a broker realizes that the number of brokers in the team is less than necessary, this broker will ask to the team to confirm this hypothesys and to clarify the current situation.	When an agent registers with a broker, this broker has to inform to the team.
When a broker loses the connection with a registered agent, the broker will look for this agent.	When a broker team realizes that the number of brokers is less than required, each broker of the team has the objective of recruiting new members.	When an agent is not longer registered, the broker has to inform to the team.
When the team loses the connection with a specifyc agent, all the team have the commitment to find this agent.	When a broker recruits a new broker agent into the team, the broker has to inform to all the team.	When a broker loses the connection with a registered agent, this broker has to inform to the team.
When previously disconneted agents, try to reconnect to the team, the broker have the commitment to connect directly with these agents.		When a broker reconnects with a registered agent, this broker has to inform all the members.

Table 1: VAACA Logical Characterization. Behaviour is implemented by logical rules (missions and rules).

4 Discusion

In order to develop the proposed architecture it is necessary to define a platform that implements communication between agent platforms (IIOP), as can be the IPMS Project (Inter-Platform Mobility Service) of the Autonomous University of Barcelona [3]. It is also recommended a library for the logical characterization, such as the AAA agent library [4], which provides a facilitator agent that is used as a broker and as a matchmaker.

Moreover, we want to highlight that, depend on the application, we may need to add some elements to this infrastructure (for example, when this architecture is used in a robotic swarm). In this application, we found the problem of loss of coverage by a robot or a group of robots that belongs to a platform, as can be seen in figure 2. Initially we used a host with an access point (AP1). This host contained the AMS of the platform that centralizes the management of the swarm. When an agent lost the access point coverage, this agent was removed from the system and was no longer usable. To solve the problem of centralized management we use the proposed architecture (VAACA). In this case, apart from the logical architecture, we had to provide to all the platforms two network cards, setting up one of them in infrastructure mode to connect to a wireless Access Point and the other in adhoc mode for the communication between broker-agents.

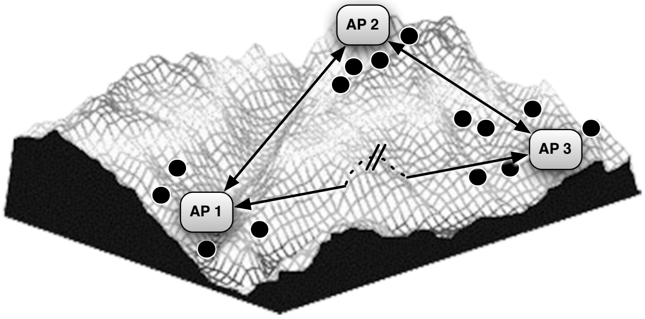


Figure 2: Coverage loss between the main agent platform (AP1) and the agent platform AP3.

In this paper we have presented VAACA architecture, which distributes agents accross platforms and containers, and implements a team of brokers to enable decentralized management using logical characterization. This architecture guides the user to work with agent mobility and with communication between platforms in order to improve fault-tolerance in MAS and to consolidate this paradigm to a broad range of real worl applications.

References

[1] Foundation for Intelligent Physical Agents. FIPA Agent Management Specification, ipa agent management sc00023k edn. (2004)

[2] Bellifemine, F., Poggi, A., Rimassa, G.: Developing multi-agent systems with jade. In: ATAL00: Proceedings of the 7th International Workshop on Intelligent Agents VII. Agent Theories

[3] Cucurull, J., Mart, R., Navarro-Arribas, G., Robles, S., Overeinder, B.J., Borrell, J.: Agent mobility architecture based on ieee-fipa standards. Computer Communications 32, 712729 (2009)

[4] Kumar, S., Cohen, P.R., Levesque, H.J.: The adaptive agent architecture: Achieving faulttolerance using persistent broker teams. In: In Proceedings of the Fourth International Conference on Multi-Agent Systems, pp. 159166. IEEE Computer Society (2000)

Markov process and random variable transformation technique to model a stroke disease

J.-C. Cortés[♭] [*], A. Navarro-Quiles[♭],
J.V. Romero[♭], and M.-D.Roselló[♭]

(♭) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Spain.

November 30, 2016

1 Introduction

In this paper we propose a general methodology to study the stroke disease using a Markov model, where some parameters on the transition matrix will be considered random variables (r.v.'s). In Markov models individuals can only remain in a particular state and total population is constant. We consider that the model advances by fixed time increments, called cycles and denoted by n. We will consider three states, Susceptible (S), Reliant (R) and Dead (D). In Figure 1, we represent the influence diagram associated to the Markov model where transitions among states are includes.

Taking into account [2], the Markov model is formulated as follows

$$\begin{pmatrix} S_{n+1} \\ R_{n+1} \\ D_{n+1} \end{pmatrix} = T \begin{pmatrix} S_n \\ R_n \\ D_n \end{pmatrix}, \quad (S_0, R_0, D_0) = (s_0, r_0, 0), \quad n = 0, 1, 2, \ldots \quad (1)$$

[*] e-mail: jccortes@imm.upv.es

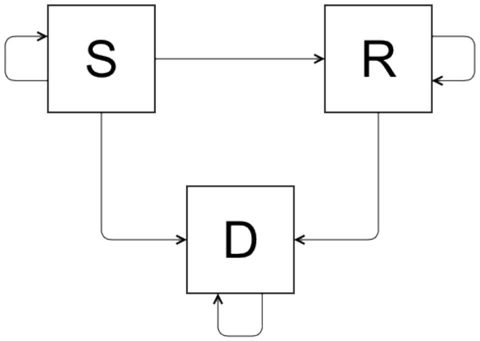


Figure 1: Influence diagram for the Markov model (1)–(2).

being

$$T = \begin{pmatrix} e^{-t_1 RR/1000} + e^{-(T_2+t_3(RR-1))/1000} - 1 & 0 & 0 \\ 1 - e^{-t_1 RR/1000} & 1-P & 0 \\ 1 - e^{-(T_2+t_3(RR-1))/1000} & P & 1 \end{pmatrix}. \qquad (2)$$

In (1)–(2), S_n, R_n and D_n are the proportion of susceptibles, reliants and deads in cycle n (as a population is constant then $S_n + R_n + D_n = 1$ for each n), $(s_0, r_0, 0)^{\mathsf{T}}$ is the initial cohort and t_1 and t_3 are the non–moral stroke and the stroke death rates, respectively. In addition we will assume that the relative risk, RR, the death rate due to any cause, T_2, and the probability of transition $R \to D$, P, are r.v.'s. Since P represent a probability of a particular transition it lies between 0 and 1. With regard T_2 and RR we know that they are positive parameters. Then, their respective domains are

$$\begin{array}{rcl} \mathcal{D}_{RR} & = & \{\, rr = RR(\omega), \omega \in \Omega : \, 0 \le rr_1 \le rr \le rr_2 \}\,, \\ \mathcal{D}_{T_2} & = & \{\, t_2 = T_2(\omega), \omega \in \Omega : \, 0 \le t_{2_1} \le \gamma \le t_{2_2} \}\,, \\ \mathcal{D}_{P} & = & \{\, p = P(\omega), \omega \in \Omega : \, 0 \le p_1 \le p \le p_2 \le 1 \}\,. \end{array} \qquad (3)$$

Hereinafter, we will denote their joint p.d.f. as $f_{RR,T_2,P}(rr, t_2, p)$.

2 Theoretical results

We will to determine the 1-p.d.f. and the statistical properties of each, susceptibles, reliants and deads, as well as we will to obtain the p.d.f. of the time until a given proportion of the population remains susceptible. We will apply random variable transformation (RVT) technique, [1] to obtain the full probabilistic description of the solution stochastic process to (1)–(2). Then, we will compute:

- The first probability density function (1-p.d.f.) of susceptible, reliant and dead subpopulations. For example, for susceptibles the 1-p.d.f. is given by

$$\begin{aligned} f_1(s,n) \;\; = \;\; & \int_{\mathcal{D}(R_n)} \int_{\mathcal{D}(P)} f_{RR,T_2,P} \left(-\frac{1000 \ln(1-m_1)}{t_1}, \right. \\ & \left. t_3 + \frac{1000 t_3 \ln(1-m_1)}{t_1} - 1000 \ln(1-m_2), p \right) \\ & \times \left| \frac{1000000}{t_1(1-m_1)(1-m_2)} \right| \left| \frac{\left(\frac{s}{s_0}\right)^{1/n} \left(-1 + \left(\frac{s}{s_0}\right)^{1/n} + p \right)}{ns \left(s - s_0 \left(1-p\right)^n\right)} \right| \mathrm{d}p \, \mathrm{d}r, \end{aligned}$$

  where

$$m_1 = \frac{\left(-1 + p + \left(\frac{s}{s_0}\right)^{1/n} \right) \left(r_0 \left(1-p\right)^n - r \right)}{\left(1-p\right)^n s_0 - s}$$

  and

$$m_2 = \frac{s - s\left(\frac{s}{s_0}\right)^{1/n} + \left(-s_0 + \left(r_0 + s_0\right) \left(\frac{s}{s_0}\right)^{1/n} + r_0 \left(-1+p\right) \right) \left(1-p\right)^n}{s - s_0 \left(1-p\right)^n}.$$

- Time until a given proportion of the population remains susceptible, reliant or dead.

With these 1-p.d.f.'s we can compute, for example, for susceptible subpopulation:

- The mean and the variance for each cycle n.
- Confidence intervals.

- The proportion of susceptibles that lies between a and b at a specific time period, say $\hat{n}$,

$$\mathbb{P}[a \leq S_{\hat{n}} \leq b] = \int_a^b f_1(s, \hat{n})\mathrm{d}s.$$

3 Graphical Examples

Based on [2], we assume that

- The initial condition is $(s_0, r_0, 0)^\intercal = (1, 0, 0)^\intercal$.
- RR, is a lognormal r.v. with parameters $(1.793, 0.143)$, i.e., $\ln(RR) \sim \mathrm{N}(1.793, 0.143)$.
- P is a beta r.v. with parameters $(80, 120)$, i.e., $P \sim \mathrm{Be}(80; 120)$.
- T_2 is a uniform r.v. on the interval $]21.27, 22.27[$, $T_2 \sim \mathrm{U}(]21.27, 22.27[)$.
- $t_1 = 1.11$ and $t_3 = 1.76$.
- RR, P and T_2 are pairwise independent r.v.'s.

In Figure 2, the 1-p.d.f.'s of susceptibles, reliants and deads have been plotted. We can observe that when time increases the percentage of susceptibles decreases. Besides, the percentage of reliants increases at the beginning, specifically from $n = 1$ to $n = 6$, and afterwards this percentage decreases towards zero. With regard to dead population, as is an absorbent state, all the population tends to this state. This is in agreement with results shown in Figure 2, where we can see that the percentage of deads increases over the time. Besides, the variability in both susceptible and dead subpopulations increases when times goes on. It is also interesting to observe that the 1-p.d.f. becomes sharper as standard deviation decreases.

Moreover, in Figure 3 it is shown the mean plus/minus the standard deviation functions of the three subpopulations. Notice that graphical representations shown in Figure 2 and Figure 3 are in agreement.

In addition, for example, we can obtain the proportion of reliant subpopulation which lies between $a = 0.010$ and $b = 0.015$ in the time period $\hat{n} = 5$:

$$\mathbb{P}[0.010 \leq R_5 \leq 0.015] = \int_{0.010}^{0.015} f_1(r, 5)\mathrm{d}r = 0.700602.$$

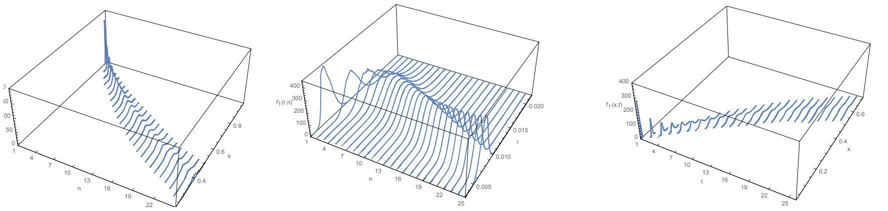


Figure 2: lot of the 1-p.d.f.'s of susceptibles (left), reliants (center) and deads (right) at the following values of $n \in \{1, 2, \ldots, 25\}$.

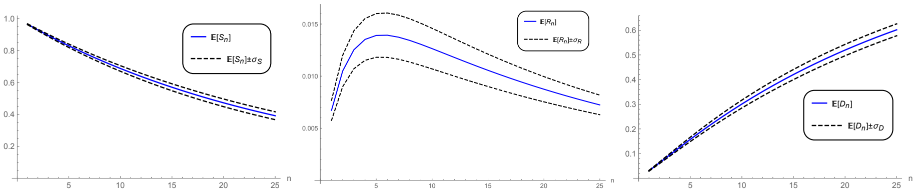


Figure 3: Expectation plus/minus standard deviation functions for susceptibles (left), reliants (center) and deads (right).

Acknowledgements

This work has been partially supported by the Ministerio de Economía y Competitividad grants MTM2013-41765-P. Ana Navarro Quiles acknowledges the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.

References

[1] T.T. Soong, Random Differential Equations in Science and Engineering. New York, Academic Press, 1973.

[2] J. Mar, F. Antoñanzas, R. Pradas and A. Arrospide. Los modelos de Markov probabilísticos en la evaluación económica de tecnologías sanitarias: una guía práctica *Gaceta Sanitaria*, 24 (3): 209–214 ,2010.

Modelling social systems: a new network point of view in labour markets

Miguel LLoret-Climent[♭], *Josue Antonio Nescolarde-Selva[†],
Higinio Mora-Mora[‡] and Maria Teresa Signes-Pont[♮]

(♭) Department of Applied Mathematics. University of Alicante. Alicante. Spain,
Carretera San Vicente del Raspeig s/n. 03690. San Vicente del Raspeig. Alicante. Spain,
(†) Department of Applied Mathematics. University of Alicante. Alicante. Spain,
Carretera San Vicente del Raspeig s/n. 03690. San Vicente del Raspeig. Alicante. Spain,
(‡) Department of Computer Technology and Computation. University of Alicante. Alicante. Spain,
Carretera San Vicente del Raspeig s/n. 03690. San Vicente del Raspeig. Alicante. Spain,
(♮) Department of Computer Technology and Computation. University of Alicante. Alicante. Spain,
Carretera San Vicente del Raspeig s/n. 03690. San Vicente del Raspeig. Alicante. Spain.

November 30, 2016

1 Introduction

Complex Systems is a new field of science studying how parts of a system give rise to the collective behaviours of the system, and how the system interacts with its environment. Graph theory is a fundamental tool in the study of social systems and economic issues, the input-output tables are precisely one of the main examples of it. We use the interpretation of labour market through networks to get a better understanding on its overall functioning. One benefit of the network perspective is that a large body of mathematics exists to help analyse many forms of networks models. If an economic system has a suitable model, then it becomes possible to utilize relevant mathematical

*e-mail: josue.selva@ua.es

tools, such as graph theory, to better understand the way the labour market works. This approach makes it possible to present and to understand the two most important relations in a labour market supply-demand and competition in a different way. Many problems in theoretical economics are mathematically formalized as dynamical systems. In this article, we apply concepts including structural functions, coverage and invariant sets to a social system modelling.

Many problems in theoretical economics are mathematically formalized as dynamic systems. Recently, an open approach to studying the dynamic behaviour of these models has appeared. In this paper, we present a mathematical model of the labour market based on a similarity with ecosystems, such as the relationships of predator-prey and competition [1] that can be equated with their analogues supply-demand and competition [2]. The model has exploited the knowledge and the progress made by the research group in modelling of complex systems [3-6].

Networks play an important role in a wide range of economic phenomena [7-11]. The diffusion of information across a network only requires a single contact between nodes, making network connectivity the crucial determinant of whether or not these simple contagions will spread [12]. Despite this fact, standard economic theory rarely considers economic networks explicitly in its analysis. A wide range of empirical studies of labour markets have shown that a significant fraction of all jobs are found through social networks. The role of informal social networks in labour markets has been emphasized initially by Granovetter [13]. He found that over 50% of jobs were found through personal contacts. In a recent paper, Jackson and Calvo-Armengo [14] introduced a network model of job information transmission. Network sampling is only useful if a researcher can produce accurate global network estimates. Recently, Smith [15], explored the practicality of making network inferences and examine networks with a skewed degree distribution surveying the limit that the number of social ties a respondent can list.

We interpret the labour market through networks. One benefit of the network perspective is that a large body of mathematics exists to help analyse many forms of network models. If an economic system has got a suitable model, then it becomes possible to use relevant mathematical tools, such as graph theory, to better understand the way the labour market works. This interpretation allows us to use the concepts of coverage and invariance alongside other related concepts. The latter will allow us to present the two most important relations in a labour market -supply-demand and competition- in a different way.

2 Basic concepts

Our basis for studying economic systems was the General Systems Theory of Ludwig von Bertalanffy [16] who defined a system as: a set of elements standing in interrelation among themselves and with the environment. Here, we present the labour market simply without reference to its properties. We define a labour market as a pair of companies and people (employed or unemployed) and by determining interactions between the elements. Real determination is causal determination and causal interactions may be of two classes: transactions, with material and monetary changes; and relations, which are indirect consequences of transactions such as competition and supply-demand relations.

Economics is a social science in which the phenomena are produced by the interaction of thousands of personal and business decisions in many social areas. It is appropriate to think that many of these phenomena are related, although these cause-effect relationships are not evident. In this sense, the application of network theory to analyze these phenomena may help to explain the operation of social activities or economic sectors. Labour markets are a good example for this. In this connection, it is crucial to establish correct analogies between theory and elements of the labour market to find explanations that help to understand how socio-economic reality works. Therefore, in the development of this research we have established analogies between network theory and the labour market that fit observable social behavior.

The structural function is a mathematical function that works in a qualitative way. A mathematical function usually assigns numerical values to numerical values whereas the structural function assigns to each set of agents another set of agents.

In our model, the state equations are represented by the structural functions associated with competition and supply-demand relationships. The structural function associated with competition assigns to the state variable of each competitor the set of all of its competitors. Thus, our first analogy of the model with the labour market is to consider any employed or unemployed people associated with the set of all the people that compete with them according to competition and supply-demand relationships.

Here, we present the labour market by adapting the concept of system-linkage [17], [3]. We only give a simple view of them and avoid analysing their features.The advantage of using the concepts of coverage and invariability to deal with these issues instead of analysing the labour market relationships from a classical point of view is that they enable us to use a mathematical function and in this manner we work with unions and intersections or create varied compositions of the structural function over different sets as is demonstrated by the results presented in this article. We are creating a Mathematical Formulation of the labour market which uses the concepts of coverage and invariability to obtain conclusions regarding the behaviour of these sets. The majority of the analytical techniques are based on a mathematical modelling process which does not always faithfully reflect the real model. On the other hand, the type of analysis involving coverage and invariability is accurately based on the real behaviour of the labour market. Relations with examples and analogies to the labour market have been described, which, without wishing to be exhaustive, illustrate certain behaviours of the real world that can be explained with the functions and proven properties. A more rigorous analysis may identify other situations in labour sectors of the economy that can be explained with the proposed theoretical model.

References

[1] Lloret. M., Amoros R., Gonzlez. L., and Nescolarde. J. A. Coverage and invariance for the biological control of pests in mediterranean greenhouses. *Ecological Modelling*, 292: 37-44, 2014.

[2] Xu. Ch., Li. P. Almost periodic solutions for a competition and cooperation model of two enterprises with time-varying delays and feedback controls. *Journal of Applied Mathematics and Computing*, DOI: 10.1007/s12190-015-0974-7, 2015.

[3] Lloret. M. and Esteve. P. F. A systemic theory of orbits in ecological networks. *Kybernetes*, 36(3/4): 469-475, 2007.

[4] Lloret-Climent. M. and Nescolarde-Selva. J. Data analysis using circular causality in networks. *Complexity*, 19(4): 15-19, 2014.

[5] Mora-Mora. H., Mora-Pascual. J., Signes-Pont. M. T. and Snchez-Romero J. L. Mathematical model of stored logic based computation. *Mathematical and Computer Modelling*, 52(7): 1243-1250, 2010.

[6] Signes. M. T., Garca. J. M., de Miguel. G. and Mora. H. Computational framework for behavioral modelling of neural subsystems. *Neurocomputing*, 72(7): 1656-1667, 2009.

[7] Laussel. D. and Resende. J. Dynamic price competition in aftermarkets with network effects. *Journal of Mathematical Economics*, 50: 106-118, 2014.

[8] Navarro. N. Price and quality decisions under network effects. *Journal of Mathematical Economics*, 48(5): 263-270, 2012.

[9] Steffi Yang, J.-H. Social network influence and market instability. *Journal of Mathematical Economics*, 45(3/4): 257-276, 2009.

[10] Hellerstein. J. K., McInerney. M., Neumark. D. Neighbors and Coworkers: The Importance of Residential Labor Market Networks. *Journal of Labor Economics*, 29(4): 659-695, 2011.

[11] Luo. J. The power-of-pull of economic sectors: A complex network analysis. *Complexity*, 18: 37-47, 2013.

[12] Centola D. Failure in Complex Social Networks. *The Journal of Mathematical Sociology*, 33(1): 64-68, 2008.

[13] Granovetter, M., Failure in Complex Social Networks. Chicago, University of Chicago Press, 1995.

[14] Jackson. M. and Calvo-Armengol. T. Networks in labor markets: Wage and employment dynamics and inequality. *The Journal of Economic Theory*, 132(1): 27-46, 2007.

[15] Smith. J. A. Global Network Inference from Ego Network Samples: Testing a Simulation Approach. *The Journal of Mathematical Sociology*, DOI:10.1080/0022250X.2014.994621, 2015.

[16] Bertalanffy L.V. General Systems Theory. Foundations, Development, Applications. New York, George Braziller Ed, 1968.

[17] Lloret. M., Villacampa. Y. and Uso-Domenech. J.L. System-linkage: structural functions and hierarchies. *Cybernetics and Systems*, 29: 29-39, 1998.

Definition, properties and applications of multiplex PageRank

F. Pedroche[♮] [*], M. Romance[♭], and R. Criado[♭]

(♮) Institut de Matemàtica Multidisciplinària,

Universitat Politècnica de València. Camí de Vera s/n, 46022. Valencia. Spain,

(♭) Department of Applied Mathematics (Rey Juan Carlos University) &

Center for Biomedical Technology (Technical University of Madrid)

Universidad Rey Juan Carlos, C/ Tulipan s/n. 28933.Móstoles. Madrid.

November 30, 2016

1 Introduction

In the literature devoted to complex systems, a new area is growing up around the concept of multilayer systems [2]. Multilayer systems are formed by several layers (graphs) having different number of nodes and with different types of links. For example, we can imagine the behaviour of some people on WhatsApp, Facebook and Linkedin to realize that not all the users are on all three networks and that their connections in each network are radically different. Generally speaking, multilayer networks can be considered as networks of networks and illustrative examples can be found in fields such as multimodal transportation networks, financial markets or biological systems in which their components (nodes) can interact in different ways and defining, as a consequence, different layers of interactions. The importance of the study of multilayer -or multiplex- networks is also enhanced by the fact that some authors agree [3] that some key traits of complex systems remain invisible

[*] e-mail: pedroche@imm.upv.es

when a multilayer network is considered as if it was a single (monoplex) network. Is thus of big interest to have analytical tools -similar to those existing in monoplex networks- to analyze the properties of multiplex networks.

2 Centrality measures

One of the key concepts in the study of networks is the idea of centrality. Centrality refers to a score that can be assigned to a node to describe its capability to perform some actions. For example, a node can be located in many short paths connecting a lot of nodes. This trait is measured by betweenness centrality. The betweenness centrality b_i of a node i is defined by

$$b_i = \sum_{j \neq k} \frac{n_{jk}(i)}{n_{jk}}$$

where n_{jk} is the number of shortest paths connecting j and k, and $n_{jk}(i)$ is the number of shortest paths that connect j and k and pass through i.

Another popular centrality measure is the local clustering coefficient, which measures to what extend a node is connected to friends which, in turn, are also friends. The local clustering coefficient of a node i can be defined by

$$c_i = \frac{\sum_{j,m} a_{ij} a_{jm} a_{mi}}{k_i(k_i - 1)}$$

where a_{ij} are the entries of the adjacency matrix associated to the graph ($a_{ij} = 1$ if there is a link from node i to node j, and 0 otherwise) and $k_i = \sum_j a_{ij}$ is the degree of node i. Note that the degree itself is a centrality measure.

3 PageRank

PageRank belongs to the class of centrality measures called spectral measures since they are derived from some properties related to eigenvalues and eigenvectors of an adjacency matrix or an ad-hoc defined matrix (e.g., the Google matrix). Since 1998 PageRank has been used in several fields ranging from website ranking to identification of leaders in social network sites [7], scientific literature retrieval [11], ranking in computational biology [1], ranking in water supply networks [5], tennis players ranking [10], ranking in protein interaction networks [6] and many more.

The PageRank vector π can be defined by

$$\pi^T = \alpha\pi^T(P_A + \mathbf{d}\mathbf{u}^T) + (1-\alpha)\mathbf{v}^T$$

with $\pi > 0$ and $\pi^T\mathbf{e} = 1$, where $\mathbf{e}$ is the (column) vector of all ones. In the above formula, P_A is a row-stochastic matrix obtained from the adjacency matrix, vectors $\mathbf{d}$ and $\mathbf{u}$ are accountable for the dangling nodes, α is a constant between 0 and 1, and $\mathbf{v}$ is the personalization vector (see [8] for further details).

It holds that $\pi^T G = \pi^T$, where the $n \times n$ matrix G is given by

$$G = \alpha(P_A + \mathbf{d}\mathbf{u}^T) + (1-\alpha)\mathbf{e}\mathbf{v}^T.$$

4 Biplex PageRank

We have recently shown in [8] that a formulation based on the $2n \times 2n$ matrix

$$M_A = \begin{pmatrix} \alpha P_A & (1-\alpha)I \\ \alpha I & (1-\alpha)\mathbf{e}\mathbf{v}^T \end{pmatrix}$$

where I is the $n \times n$ identity matrix, allows to define a vector similar to the PageRank vector. In fact, we define $\hat{\pi}_M$ such that:

$$\hat{\pi}_M^T = \hat{\pi}_M^T M_A$$

with $\hat{\pi}_M^T\mathbf{e} = 1$, and verifying $\hat{\pi}_M^T = [\pi_u^T \quad \pi_d^T] \in \mathbb{R}^{2n}$ with $\pi_u, \pi_d \in \mathbb{R}^n$, $\pi_u^T\mathbf{e} = \alpha$ and $\pi_d^T\mathbf{e} = 1 - \alpha$.

These ingredients allow us to define the two-layer approach PageRank of A as the column vector of n components given by

$$\hat{\pi}_A = \pi_u + \pi_d$$

It is easy to show that $\hat{\pi}_A$ coincides with the (usual) PageRank π in some cases (see [8]).

5 Multiplex PageRank

The above definition allows to extend the concept of PageRank to the case of multilayer systems in a very straightforward manner. It only suffices to add a matrix like M_A for every layer. For example, if we consider a multiplex composed by only two layers we define the matrix

$$M_2 = \frac{1}{2}\left(\begin{array}{cc|cc} \alpha P_{A_1} & I & (1-\alpha)I & 0 \\ I & \alpha P_{A_2} & 0 & (1-\alpha)I \\ \hline 2\alpha I & 0 & (1-\alpha)\mathbf{e}\mathbf{v}_1^T & (1-\alpha)\mathbf{e}\mathbf{v}_2^T \\ 0 & 2\alpha I & (1-\alpha)\mathbf{e}\mathbf{v}_1^T & (1-\alpha)\mathbf{e}\mathbf{v}_2^T \end{array}\right)$$

and compute the vector $\hat{\pi}_M^T = \hat{\pi}_M^T\, M_2$ with $\hat{\pi}_M^T\mathbf{e} = 2$, and

$$\hat{\pi}_M^T = [\pi_{u1}^T \quad \pi_{u2}^T \quad \pi_{d1}^T \quad \pi_{d2}^T]$$

with $\pi_{ui}, \pi_{di} \in \mathbb{R}^n$ for every $i = 1, 2$. Finally, we define the PageRank of the multiplex composed by two layers as

$$\hat{\pi}_2 = \frac{1}{2}\left(\pi_{u1} + \pi_{u2} + \pi_{d1} + \pi_{d2}\right) \in \mathbb{R}^n.$$

Some numerical results in [8] and [9] show that the introduced multiplex PageRank can be used to rank nodes within the framework of multiplex systems, giving results that differ from those obtained by considering (a combination) of the usual PageRank in each layer, or by computing the usual PageRank of the graph obtained as a projection of all the layers.

Acknowledgements

This work is supported by Spanish DGI grant MTM2014-59906.

References

[1] S. Allesina, and M. Pascual, *Googling food webs: can an eigenvector measure species importance for coextinctions?*, PLoS Computational Biology 5 (9), (2009).

[2] S. Boccaletti, G. Bianconi, R. Criado, C.I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang and M. Zanin, *The structure and dynamics of multilayer networks* Physics Reports **544**(1) (2014), 1-122.

[3] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gmez, and A. Arenas *Ranking in interconnected multilayer networks reveals versatile nodes* Nature Communications 6, Article number: 6868. (2015).

[4] E. García, F. Pedroche and M. Romance, *On the localization of the Personalized PageRank of Complex Networks*, Linear Algebra and its Applications **439**, 640 (2013).

[5] J. A. Gutiérrez-Pérez, M. Herrera, J. Izquierdo, and R. Pérez-García, *An approach based on ranking elements to form supply clusters in water supply networks as a support to vulnerability assessment*, in International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.), (2012).

[6] G. Iván, and V. Grolmusz, *When the Web meets the cell: using personalized PageRank for analyzing protein interaction networks.* Bioinformatics. 2011 Feb 1;27(3):405-7.

[7] F. Pedroche, F. Moreno, A. González, and A. Valencia, *Leadership groups on Social Network Sites based on Personalized PageRank* Mathematical and Computer Modelling **57**, (2013) 18911896.

[8] F. Pedroche, M. Romance, and R. Criado, *A biplex approach to PageRank centrality: From classic to multiplex networks*, Chaos: An Interdisciplinary Journal of Nonlinear Science . 26, 065301 (2016).

[9] F. Pedroche, M. Romance, and R. Criado, *Some rankings based on PageRank applied to the València Metro-Tram system*, Submitted to IJCSS.

[10] F. Radicchi, *Who is the best player ever? a complex network analysis of the history of professional tennis*, PLoS ONE 6 (2) (2011).

[11] X. Yin, J.X. Huang, and Z. Liet, *Mining and modeling linkage information from citation context for improving biomedical literature retrieval, Information Processing and Management* 47 (1) (2011) 5367.

Use of connected components of graphs in image processing

Cristina Pérez-Benito[♭], [*]J. Alberto Conejero[†], [†]
Cristina Jordán[‡], [‡]Samuel Morillas[†], [§]

(♭) Instituto de Biomecánica de València, Universitat Politècnica de València

(†) Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València

(‡) Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València

November 30, 2016

1 Introduction

Colour image denoising is a topic which has been extensively studied in the fields of computer vision and digital image processing. The denoising (or filtering) step is essential for almost every computer vision system because noise can significantly affect the visual quality of the images as well as the performance of most automatic image processing tasks. Among the different sources of noise in digital imaging, probably the most common one is the so-called thermal noise, which is mainly due to CCD sensor malfunction and specially intense with inappropriate illumination conditions. This kind of noise is modeled as additive white Gaussian noise [4].

Many methods for reducing Gaussian noise from color images have been proposed in the literature [3, 4], with the aim of smoothing image noise while keeping intact desired image features such as edges, texture and fine details. Recent non-linear methods exhibit an improved performance with respect to the earliest linear approaches, above all, from the edge and detail preservation point of view, since they try to detect and preserve these features. However, as the noise in the image is higher, many times image noise in homogeneous regions is confused with an image structure that should be preserved and so it is not properly reduced.

Graph theory has been used in image processing for image cut-based segmentation, edge detection pattern recognition in order to improve performance by overcoming this drawback. The underlying idea is the following one:

ì) Consider each pixel as a node,

[*] email: cripebe1@posgrado.upv.es

[†] email: aconejero@upv.es

[‡] email: cjordan@mat.upv.es

[§] email: smorillas@mat.upv.es

ii) Fix an arbitrary pixel,

iii) This pixel can be characterized depending on the features of its related pixels.

So as to, one can join it with pixels close to it with edges, and assign weights to these edges attending to certain characteristics of them.

Here, given a color image $\mathbf{F}$, which is represented in the RGB color space, we build a graph-based model for each pixel in $\mathbf{F}$. Let us consider a pixel of the image that does not belong to the frame of the image, namely $\mathbf{F}_0$. This pixel is defined by the triple (F_0^R, F_0^G, F_0^B) of its three RGB color components. We also consider the neighbors around this pixel in a window centered on it of size $N \times N$ where $N = 2n+1$ and $n = 1,2,\ldots$. The rest of the neighbor pixels in the window are denoted as $\mathbf{F}_i, i = 1,\ldots,N^2-1$, starting from the one on the top left corner of the window, following the clockwise order.

Given this pixel $\mathbf{F}_0$ we define a local weighted graph $G_{\mathbf{F}_0}$ where $V(G_{\mathbf{F}_0}) = \{\mathbf{F}_i, i = 0,\ldots,N^2-1\}$ and $L(G_{\mathbf{F}_0}) = \{(\mathbf{F}_i,\mathbf{F}_j), i \neq j, ||\mathbf{F}_i - \mathbf{F}_j||_2 < \mathscr{U}\}$. So, a link exists between pixel $\mathbf{F}_i$ and $\mathbf{F}_j, i \neq j$ if the euclidean distance between their color vectors is lower than a certain threshold $\mathscr{U}$. If the link exists, its weight is $w(\mathbf{F}_i,\mathbf{F}_j) = ||\mathbf{F}_i - \mathbf{F}_j||_2$.

Depending on the image and the value of $\mathscr{U}$ one can get a unique connected component or some of them. Following the aforementioned procedure, we assume the hypothesis that for *reasonable* values of $\mathscr{U}$, all the pixels in the same connected component belong to the same region of the image. As a consequence, if in a window of size $N \times N$ we have k connected components, we assume that there are k regions of the image.

An analysis of the sizes of the connected components give us an idea of the characteristics of the region. These features are able to properly distinguish flat image regions in front of edge and detail regions. An updated review on border detection on images can be found in [6].

Kruskal algorithm permits to compute maximum/minimum generating trees of each connected component of a weighted graph. Minimum generating trees are also useful for clustering similar pixels in an image. These trees have been of particular interest for smoothing colour images as it was seen in [2]. Depending on the size of the generating tree that contains the central pixel of a given $N \times N$ window, one can consider several intuitive situations. We list, as example, some of them:

- If the cardinal of the component of central pixel is one and two, and all the other vertex are in the same connected component, then we have that the central pixel can be probably affected by an impulsive noise.

- If the cardinal of the connected component of central pixel is between three or six, then we probably have two regions in the image.

- If the cardinal of the connected component of central pixel is between seven and nine, then we probably have a unique region in the image.

For illustrating these characteristics, we show some graphic representations of the image Lenna attending to the number of pixels of the connected components to

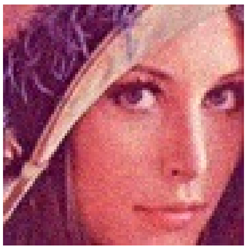


Figure 1: Lenna with an additive Gaussian noise with $s = 10$.

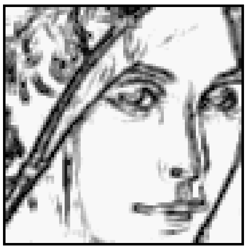


Figure 2: Lenna with an additive Gaussian noise with $s = 10$, processed with $\mathscr{U} = 37$.

which the central pixel belong. This is done by assigning to each pixel a grey level depending on the size of certain sub-generating tree. A black pixel will represent that a pixel is disconnected from the others, and a white pixel will represent that this pixel is connected with the rest of pixels of the window. The rest of values are assigned a grey level proportionally to the number of vertex in the connected component of that pixel in the window.

Then dark pixels correspond to edges in the image and to impulsive noise, and light pixels to flat regions. As an example we consider an image of Lenna corrupted with noise using the classical white additive Gaussian model of standard deviation $s = 10$, that represents the noise intensity, see Figure 1. In Figure 2, we have created an image from Figure 1, by assigning a level of grey to the number of pixels in the connected component of the central pixel as we have already indicated.

We point out, that all these results depend on how big was $\mathscr{U}$. There is no a clear way to set an optimal $\mathscr{U}$, since it depends on the characteristics of the image. One approach can be done based on mutually image information (MII). This measure gives a comparison. Further details on the use of MII can be found in [5].

The ideas explained in this We show how to use these ideas in order to define a border detection method on colour images as a linear combination of the AMF and FNRM filters. Our results permit us to improve the image processing respect to the results from [2] already mentioned.

We present some results regarding its robustness to Gaussian and impulsive noise. The results are compared respect to the peak signal-to-noise ratio (PSNR) and to the recent new metric Fuzzy Color Structural Similarity (FCSS), as it is presented in [1].

Acknowledgements

This work is supported by Vicerrectorado de Estudios, Calidad y Acreditación, PIMEB08: “Modelos de la Teoría de Grafos aplicados a problemas de competiciones de programación”.

This work is comprised within the context of extending the contribution of graph theory methods and algorithms to improve programming skills for the solution of engineering problems appearing in different fields.

References

[1] S.Grecova and S. Morillas. Perceptual similarity between color images using fuzzy metrics. *J. Vis. Commun. Image R.*, 34: 230-235, 2016.

[2] C. Jordan, S. Morillas, and E. Sanabria-Codesal. Colour image smoothing through a soft-switching mechanism using a graph model. *IET Image Processing*, 6(9): 1293-1298, 2012.

[3] Lukac, R., Plataniotis, K.N., A taxonomy of color image filtering and enhancement solutions, *Advances in Imaging and Electron Physics (Elsevier Acedemic Press)*, pp. 187-264, 2006.

[4] K.N. Plataniotis, A.N. Venetsanopoulos, Color Image processing and applications *Springer-Verlag, Berlin,* 2000.

[5] D.B. Russakoff, C. Tomasi, T. Rohlfing, and C.R. Maurer Jr. Image similarity using mutual information of regions. *Lecture Notes in Computer Science. Computer Vision - ECCV 2004*, 3023:596–607, 2004.

[6] N. Senthilkumaran and R. Rajesh. Edge detection techniques for image segmentation – A survey of soft computing approaches. *International Journal of Recent Trends in Engineering*, 1(2):250–254, 2009.

Numerical solutions of moving boundary problems for concrete carbonation preserving positivity and stability

M. A. Piqueras[†] [*], R. Company[†] , and L. Jódar[†]

([†]) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València,
Camino de Vera s/n, 46022 Valencia, Spain.

November 30, 2016

1 Introduction

This paper deals with the construction and computation of numerical solutions of a coupled mixed partial differential equation system arising in concrete carbonation problems. Apart from the stability and the consistency of the numerical solution, constructed by a finite difference scheme, qualitative properties of the numerical solution are established. We also confirm numerically the $\sqrt{t}$-law of propagation.

In a recent paper [1], the authors studied a one-dimensional free boundary problem modeling the carbonation process. The unknown CO_2 mass concentrations in air and water phases of pores are denoted by $U(t, x)$ and $V(t, x)$ respectively, depending on variables time t and space x. The space variable x is measured from the exposed boundary $x = 0$ to the unknown carbonation front $x = S(t)$. In the the system (2)-(9) it is assumed that κ_1 and κ_2 are positive diffusion constants ($\kappa_1 \gg \kappa_2$) and the functions $f(U, V)$ and $\psi(r)$ are defined as

$$f(U, V) = \beta(\gamma V - U), \quad \beta > 0, \quad \gamma > 0. \tag{1}$$

[*] e-mail: mipigar@cam.upv.es

The continuous model is described by

$$\frac{\partial U}{\partial t}-\frac{\partial}{\partial x}\left(\kappa_1\frac{\partial U}{\partial x}\right)=f(U,V),\ \ 0<t<T,\ \ 0<x<S(t), \tag{2}$$

$$\frac{\partial V}{\partial t}-\frac{\partial}{\partial x}\left(\kappa_2\frac{\partial V}{\partial x}\right)=-f(U,V),\ \ 0<t<T,\ \ 0<x<S(t), \tag{3}$$

together with the left boundary conditions

$$U(t,0)=G(t),\quad V(t,0)=H(t),\ \ 0\le t\le T. \tag{4}$$

The propagation front behaviour comes out from the Stefan-like conditions, involving function $\psi(r)$ linked to the chemical reactions:

$$S'(t)=\psi(U(t,S(t))),\ \ 0<t<T, \tag{5}$$

$$-\kappa_1\frac{\partial U}{\partial x}(t,S(t))=\psi(U(t,S(t)))+S'(t)U(t,S(t)),\ \ 0<t<T, \tag{6}$$

$$-\kappa_2\frac{\partial V}{\partial x}(t,S(t))=S'(t)V(t,S(t)),\ \ 0<t<T. \tag{7}$$

Function $\psi(r)$ is given by

$$\psi(r)=\alpha|r|^p,\ \ r\in\mathbb{R},\ \ \alpha>0,\ \ p\ge 1, \tag{8}$$

where p is the so called order of the chemical reaction.

The bounded initial conditions functions are described by

$$S(0)=S_0,\quad U(0,x)=U_0(x),\quad V(0,x)=V_0(x),\ \ 0<x<S_0. \tag{9}$$

2 Front-fixing transformation and discretization

Let us begin this Section by transforming the moving boundary problem (2)-(9) into another one with fixed boundary conditions. The Landau transformation, [2], suggests the substitution

$$L(t)=S^2(t),\ \ z(t,x)=\frac{x}{\sqrt{L(t)}},\quad 0\le t\le T,\ \ 0<x<\sqrt{L(t)}. \tag{10}$$

Using substitution (10), the problem (2)-(9) becomes

$$L(t)\frac{\partial W}{\partial t} - L'(t)\frac{z}{2}\frac{\partial W}{\partial z} - \kappa_1\frac{\partial^2 W}{\partial z^2} = L(t)\beta(\gamma Y - W),\;\; 0 < t < T,\;\; 0 < z < 1, \tag{11}$$

$$L(t)\frac{\partial Y}{\partial t} - L'(t)\frac{z}{2}\frac{\partial Y}{\partial z} - \kappa_2\frac{\partial^2 Y}{\partial z^2} = -L(t)\beta(\gamma Y - W),\;\; 0 < t < T,\;\; 0 < z < 1, \tag{12}$$

where

$$W(t,z) = U(t,x), \quad Y(t,z) = V(t,x). \tag{13}$$

In addition, the new boundary conditions take the form

$$W(t,0) = G(t), \quad Y(t,0) = H(t),\;\; 0 \le t \le T. \tag{14}$$

The Stefan-like conditions (5)-(7) are transformed into

$$L'(t) = 2\sqrt{L(t)}\alpha[W(t,1)]^p,\;\; 0 < t < T, \tag{15}$$

$$-2\kappa_1\frac{\partial W}{\partial z}(t,1) = L'(t)(1 + W(t,1)),\;\; 0 < t < T, \tag{16}$$

$$-2\kappa_2\frac{\partial Y}{\partial z}(t,1) = L'(t)Y(t,1),\;\; 0 < t < T, \tag{17}$$

and the initial conditions (9) become

$$L(0) = L_0;\;\; W(0,z) = W_0(z) = U_0(zS_0);\;\; Y(0,z) = Y_0(z) = V_0(zS_0),\;\; 0 < z < 1. \tag{18}$$

Let N and M be positive integers and let us consider the step sizes discretizations $k = \Delta t = T/N$, $h = \Delta z = 1/M$ and the mesh points (t^n, z_j), with $t^n = nk$, $z_j = jh$, $0 \le n \le N$, $0 \le j \le M$. Numerical approximations of the involved variables are denoted by: $w_j^n \approx W(t^n, z_j)$, $y_j^n \approx Y(t^n, z_j)$, $l^n \approx L(t^n)$, while we denote $G^n = G(t^n)$, $H^n = H(t^n)$. To transit from the time level n to $n+1$, one needs to obtain the values $\{w_M^n, y_M^n, l^{n+1}\}$ solving a nonlinear equation in w_M^n

$$F_n(w_M^n) = 0,\;\; 0 \le n \le N, \tag{19}$$

where $F_n : [0,\infty[\to \mathbb{R}$, is given by

$$F_n(\xi) = 2\alpha(l^n)^{\frac{1}{2}}\xi^{p+1} + 2\alpha(l^n)^{\frac{1}{2}}\xi^{p} + \frac{3\kappa_1}{h}\xi - \frac{\kappa_1}{h}(4w^n_{M-1} - w^n_{M-2}). \tag{20}$$

The solutions at the interior points $\{w_j^{n+1}, y_j^{n+1};\ 1 \le j \le M-1\}$ are calculated explicitly as follows

$$w_j^{n+1} = a^n_{1,j}w^n_{j-1} + b^n_{1,j}w^n_j + c^n_{1,j}w^n_{j+1} + k\beta\gamma y^n_j, \quad 0 \le n \le N-1,\ \ 1 \le j \le M-1, \tag{21}$$

$$y_j^{n+1} = a^n_{2,j}y^n_{j-1} + b^n_{2,j}y^n_j + c^n_{2,j}y^n_{j+1} + k\beta w^n_j, \quad 0 \le n \le N-1,\ \ 1 \le j \le M-1, \tag{22}$$

where

$$\begin{aligned} a^n_{i,j} &= \frac{\kappa_i k}{h^2 l^n} - \frac{z_j}{4h}\Delta^n, \quad c^n_{i,j} = \frac{\kappa_i k}{h^2 l^n} + \frac{z_j}{4h}\Delta^n, \quad i = 1,2, \\ b^n_{1,j} &= 1 - k\beta - \frac{2\kappa_1 k}{h^2 l^n}, \quad b^n_{2,j} = 1 - k\beta\gamma - \frac{2\kappa_2 k}{h^2 l^n}, \quad \Delta^n = \frac{l^{n+1}}{l^n} - 1. \end{aligned} \tag{23}$$

3 Positivity, stability and monotonicity of the numerical solution

Let us consider the following condition for the time step size k:

$$k < k_0 = \min\left\{\frac{h^2 l_0}{2\kappa_1 + h^2\beta l_0}, \frac{h^2 l_0}{2\kappa_2 + h^2\beta\gamma l_0}\right\}. \tag{24}$$

Theorem 1. *With previous notation, for small enough values of the step sizes h and k linked by the condition (24), the following conclusions hold true:*

i) Concentration solutions of the scheme (21)-(22) w^n_j and y^n_j are positive for $1 \le j \le M$, $1 \le n \le N$.

ii) The moving carbonation front is positive and increasing.

Definition Let us denote the vectors of CO_2 concentrations $w^n = [w^n_0, w^n_1, \ldots, w^n_M]^T$ and $y^n = [y^n_0, y^n_1, \ldots, y^n_M]^T$. We say that the numerical solution $\{w^n, y^n,\ \ 0 \le$

$n \leq N\}$ is $\|\cdot\|_\infty$-stable if there exist positive constants C_1 and C_2 independent of n, k and h, such that

$$\|w^n\|_\infty \leq C_1, \;\; \|y^n\|_\infty \leq C_2, \;\; 0 \leq n \leq N. \tag{25}$$

Theorem 2. *With previous notation, for small enough values of h and k satisfying the positivity step size condition (24), the numerical solution of scheme (21)-(22) is $\|\cdot\|_\infty$-stable.*

In the following example we show that when the positivity condition is satisfied, then we have both positivity and $\|\cdot\|_\infty$-stability.

Example 1. Taking step sizes $h = 0.05$ and $k = 0.005$, the stability of the solutions $U(t,x)$ and $V(t,x)$ is guaranteed as it is shown in Figure 1. Units in x-axe are taken in cm and y-axe in 10^{-6} g cm^{-3}.

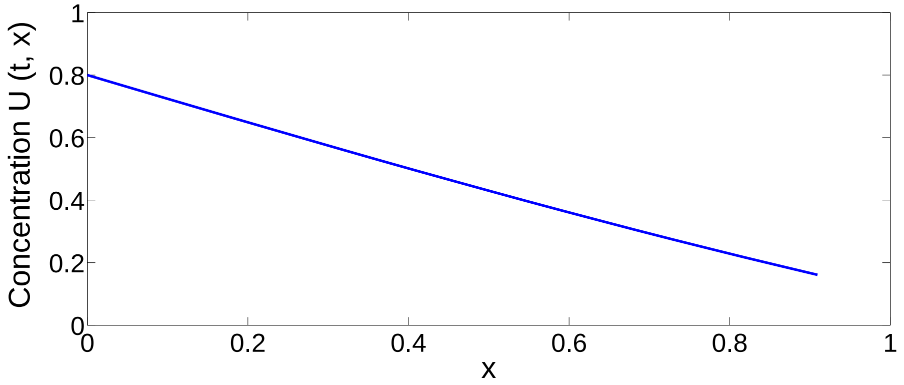


Figure 1: Numerical solution of $U(t,x)$ and $V(t,x)$ in Example 1 for $t = 13$ years, under stability condition.

Definition We say that the numerical scheme (21)-(22) is spatial monotone preserving, if assuming that the numerical solution is spatial monotone decreasing at time level n, $0 \leq n \leq N-1$, i. e.,:

$$w^n_{j+1} \leq w^n_j, \;\; y^n_{j+1} \leq y^n_j, \;\; 0 \leq j \leq M-1, \tag{26}$$

then, one satisfies

$$w^{n+1}_{j+1} \leq w^{n+1}_j, \;\; y^{n+1}_{j+1} \leq y^{n+1}_j, \;\; 0 \leq j \leq M-1. \tag{27}$$

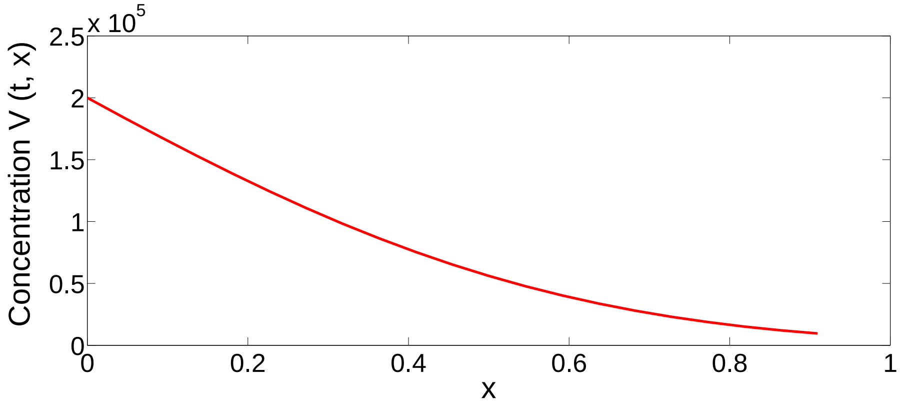


Figure 2: Numerical solution of $U(t,x)$ and $V(t,x)$ in Example 1 for $t = 13$ years, under stability condition.

Theorem 3. *Under the positivity condition (24), assuming that the concentrations at the exposed boundary are monotone non decreasing functions such that $G(t) = \gamma H(t)$, then the numerical scheme (21)-(22) is spatial monotone preserving.*

4 Numerical evidences of the $\sqrt{t}$-law of propagation

In this Section we confirm numerically, under appropriated positivity conditions, that the proposed numerical solution behaves as the theoretical solution suggested. In fact, in the next example, we match the numerical solution of the carbonation front as a function of the type $C\sqrt{t}$.

Example 2. Table 1 shows long time values of the carbonation front $S(t)$ on of time. These points $(t_i, S(t_i))$ have been fitted to a curve with two parameters of the type $S(t) = at^b$. The best fit is matched by $a = 0.2715$ and $b = 0.4568$, and the coefficient of determination $R^2 = 0.9999$.

t_i **(years)**	33.00	33.50	34.00	34.50	35.00
$S(t_i)$ **(cm)**	1.3407	1.3499	1.3591	1.3682	1.3772

Table 1: Carbonation depth for several times.

5 Consistency

Writing equations of the problem (11)-(17) in vector form as $\mathcal{L}(W,Y,L)=0$, and the finite difference scheme in a compact way as $\ell(w,y,l)=0$, the following result has been established:

Theorem 4. *With previous notation, the scheme $\ell(w,y,l)$ is consistent with the problem $\mathcal{L}(W,Y,L)$ and the local truncation error behaves as:*

$$T_j^n(W,Y,L)=\mathcal{O}(k)+\mathcal{O}(h^2). \tag{28}$$

References

[1] T. Aiki, and A. Muntean. A free-boundary problem for concrete carbonation: rigorous justification of the $\sqrt{t}$-law of propagation *Interfaces and Free Boundaries*, 15(2):167–180, 2013.

[2] H. G. Landau. Heat conduction in a melting solid *Quarterly of Applied Mathematics*, 8:81–95, 1950.

Symplectic integrators for time-dependent linear wave equations

P. Bader[♭] [*], S. Blanes[†], F. Casas[‡], N. Kopylov[†] and E. Ponsoda[†]

(♭) Dep. Math. Stat., La Trobe University, Australia.

(†) IMM, UPV, València, Spain.

(‡) IMAC, UJI, Castellón, Spain.

1 Introduction

We consider the numerical integration of the second order time-dependent linear matrix Hill's equation

$$x''(t) + M(t)x(t) = 0\,, \qquad x(t_0) = x_0, \quad x'(t_0) = x'_0, \tag{1}$$

where $t \in \mathbb{R}$, $x(t) \in \mathbb{C}^r$. Equation (1) can be written as

$$z'(t) = A(t)z(t)\,, \quad A(t) = \begin{pmatrix} 0 & I \\ -M(t) & 0 \end{pmatrix}, \tag{2}$$

with $z = (x, x')^T$, $z(0) = (x_0, x'_0)^T \in \mathbb{C}^{2r}$. For example, the linear time-dependent wave equation

$$\partial_t^2 u(x,t) = f(x,t)\partial_x^2 u(x,t) + g(x,t)u(x,t), \ x \in [-1,1], \ t \geq 0,$$

equipped with initial conditions $u(x,0) = u_0(x)$, and $u_t(x,0) = u'_0(x)$, after spatial discretization, is also given by (1).

The solution of (1) evolves through a linear evolution operator given by $z(t) = \Phi(t,0)z(0)$. In the usual case in which $M(t)$ is a real and symmetric $r \times r$ matrix valued function ($M^T = M$) then $\Phi(t,0)$ is a symplectic

[*] e-mail: p.bader@latrobe.edu.au

transformation. The eigenvalues of a symplectic matrix occur in reciprocal pairs, say $\{\lambda, \lambda^*, 1/\lambda, 1/\lambda^*\}$, where * denotes the complex conjugate. As a result, for stable systems, all of the eigenvalues must lie on the unit circle (see for example [4]). This is a very important property that is not preserved by standard numerical integrators. In this work we consider new families of explicit integrators tailored for this problem that, in the case M is real and symmetric, correspond to symplectic integrators.

Symplectic approximations for the autonomous case. The methods proposed in this work are based on the following symplectic approximations to the autonomous equations where the fundamental matrix solution for one time step can be written in closed form. From the exponential series, it follows that for real-valued matrices $C = -M$ [2, 11.3.3]

$$\Phi(h) = \exp\left(h\begin{pmatrix} 0 & I \\ C & 0 \end{pmatrix}\right) = \begin{pmatrix} \cosh h\sqrt{C} & \sqrt{C}^{-1}\sinh h\sqrt{C} \\ \sqrt{C}\sinh h\sqrt{C} & \cosh h\sqrt{C} \end{pmatrix}. \qquad (3)$$

We decompose the matrix exponential into a product of simple matrices such that in the case C is a symmetric matrix, the approximations preserve the symplectic structure by construction.

I) From [1] we have that if $h\rho(\sqrt{C}) < \pi$, where $\rho(A)$ is the spectral radius of the matrix A, then $\Phi(h)$ can be decomposed as follows

$$\Phi(h) = \begin{pmatrix} I & 0 \\ \sqrt{C}\tanh\left(\frac{h\sqrt{C}}{2}\right) & I \end{pmatrix} \begin{pmatrix} I & \frac{\sinh h\sqrt{C}}{\sqrt{C}} \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ \sqrt{C}\tanh\left(\frac{h\sqrt{C}}{2}\right) & I \end{pmatrix}. \qquad (4)$$

We denote by $\Upsilon^{[p]}$ an approximation when replacing the function matrices of C in (4) by the Taylor series expansion up to order p. If C is a symmetric matrix, by construction, $\Upsilon^{[p]}$ is a symplectic matrix $\forall p$. We will refer to $\Upsilon^{[p]}$ as a symplectic decomposition. This decomposition can be very useful for problems where matrix-matrix products and storage of matrices are computationally feasible.

II) Alternatively, we can consider different decompositions that are suitable for problems where only matrix–vector products are allowed. For instance, we can consider the following m-stage splitting method of order p

$$\Psi_m^{[p]} = \begin{pmatrix} I & 0 \\ hb_mC & I \end{pmatrix} \begin{pmatrix} I & ha_mI \\ 0 & I \end{pmatrix} \cdots \begin{pmatrix} I & 0 \\ hb_1C & I \end{pmatrix} \begin{pmatrix} I & a_1I \\ 0 & I \end{pmatrix},$$

where the coefficients $a_i, b_i,\ i = 1, \ldots, m$ have to be properly chosen (they must be obtained numerically by solving a set of non-linear polynomial equations). Here, the product $\Psi_m^{[p]} z_0$ is done with only m products of the matrix C on a vector of dimension r.

The non-autonomous case. We know that the solution of (2) can not be written in a closed form. Even the formal solution up to a given order $p > 2$ does not correspond to the exponential of matrix with the simple structure given in (3). For this reason we propose to generalize the symplectic methods for the autonomous case (4) in order to approximate the solution of the non-autonomous problem up to different orders. The computational cost of each method will depend on the application to a given particular problem.

The solution of (2) can formally be written as a single exponential using the Magnus expansion [3] where (under convergence conditions) the occurring integrals decrease in size as the oscillations become faster. The following examples illustrate such compositions which we refer to as *Magnus-decomposition* methods

$$\Upsilon_1^{[4]} = \begin{pmatrix} I & 0 \\ hD_2^{[4]} & I \end{pmatrix} \exp\left(h \begin{pmatrix} 0 & I \\ C_1^{[4]} & 0 \end{pmatrix} \right) \begin{pmatrix} I & 0 \\ hD_1^{[4]} & I \end{pmatrix}.$$

$$\Upsilon_2^{[6]} = \begin{pmatrix} I & 0 \\ hD_2^{[6]} & I \end{pmatrix} \exp\left(\frac{h}{2} \begin{pmatrix} 0 & I \\ C_2^{[6]} & 0 \end{pmatrix} \right) \exp\left(\frac{h}{2} \begin{pmatrix} 0 & I \\ C_1^{[6]} & 0 \end{pmatrix} \right) \begin{pmatrix} I & 0 \\ hD_1^{[6]} & I \end{pmatrix}.$$

Here, $\Upsilon_k^{[p]}$ denotes a method of order p that contains k exponentials, $C_i^{[p]}$ are linear combinations of $M(t)$ evaluated in a set of quadrature points of order p or higher, $D_i^{[p]}$ are linear combinations of $M(t)$ that additionally contain one product of such linear combinations. These schemes show a high performance when the solution is oscillatory since this property is retained by the exponentials.

When the problem is not highly oscillatory or for high dimensional problems where only matrix-vector products are allowed, we propose the following *Magnus-splitting* method,

$$\Psi_{11}^{[6]} = \begin{pmatrix} I & ha_{12}I \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ hC_{11} & I \end{pmatrix} \begin{pmatrix} I & ha_{11}I \\ 0 & I \end{pmatrix} \cdots \begin{pmatrix} I & 0 \\ hC_1 & I \end{pmatrix} \begin{pmatrix} I & ha_1 I \\ 0 & I \end{pmatrix}, \quad (5)$$

where

$$C_i = -(b_{i,1}M_1 + b_{i,2}M_2 + b_{i,3}M_3)$$

and $M_i = M(t_n + c_i h)$ with c_i being the nodes of the 6th-order Gauss-Legendre quadrature rule. The coefficients $a_i, b_{i,j}$ are properly chosen.

2 Magnus-based methods

The superiority of the proposed methods for oscillatory problems stems from the Magnus expansion [3]. For simplicity, assume that $A(t)$ has the dominant frequency ω, then, the higher order terms in the Magnus expansion can be regarded as an asymptotic expansion in $1/\omega$. The Magnus expansion expresses the solution to (2) as

$$\Phi(t_n, h) = \exp\left(\Omega(t_n, h)\right), \qquad \Omega(t_n, h) = \sum_{k=1}^{\infty} \Omega_k(t_n, h), \tag{6}$$

where the first terms of the Magnus series $\{\Omega_k\}$ are given by

$$\Omega_1(t_n, h) = \int_{t_n}^{t_n+h} A(t_1)\, d\tau_1\,,\ \ \Omega_2(t_n, h) = \frac{1}{2}\int_{t_n}^{t_n+h}\int_{t_n}^{\tau_1} [A(\tau_1), A(\tau_2)]\ d\tau_2\, d\tau_1, \ldots$$

where $[P, Q] = PQ - QP$ is the matrix commutator of P and Q. Here, Ω as well as any truncation of the series at order p, $\Omega^{[p]}$, belong to the symplectic Lie algebra, and symplecticity is preserved.

In order to obtain an approximation to Ω defined by (6) for a time step from t_n to $t_{n+1} = t_n + h$ in terms of the matrix $A(t)$ evaluated at the nodes of a quadrature rule, we consider the polynomial $\tilde{A}(t)$ of degree $s-1$ in t that interpolates $A(t)$ on the interval $[t_n, t_n+h]$ at the points $t_n + c_i h,\ i = 1, \ldots, s$, where c_i are the nodes of the Gauss – Legendre quadrature rule of order $2s$. The perturbed problem becomes

$$\frac{d\tilde{z}(t)}{dt} = \tilde{A}(t)\,\tilde{z}(t), \qquad \tilde{z}(t_n) = z(t_n), \qquad t \in [t_n, t_n + h],$$

where $z(t_n)$ is the exact solution of (2) at t_n. From a direct application of the Alekseev – Gröbner lemma [5], we have that

$$\tilde{z}(t_n + h) - z(t_n + h) = \mathcal{O}(h^{2s+1}).$$

Letting $t = t_n + \frac{h}{2} + \sigma$, we write the interpolation polynomial as

$$\tilde{A}(t) = \sum_{i=1}^{s} \mathcal{L}_i\Big(\frac{t - t_n}{h}\Big) A_i = \frac{1}{h}\sum_{i=1}^{s} \Big(\frac{\sigma}{h}\Big)^{i-1} \alpha_i, \qquad \sigma \in \left[-\frac{h}{2}, \frac{h}{2}\right],$$

with $A_i = A(t_n + c_i h) = \tilde{A}(t_n + c_i h)$ and the usual Lagrange polynomials $\mathcal{L}_i(t)$.

A 6th-order approximation $\Omega^{[6]} = \Omega + \mathcal{O}(h^7)$, is given by

$$\Omega^{[6]} = \alpha_1 + \frac{1}{12}\alpha_3 - \frac{1}{12}[12] + \frac{1}{240}[23] + \frac{1}{360}[113] - \frac{1}{240}[212] + \frac{1}{720}[1112], \quad (7)$$

where $[ij \dots kl]$ represents the nested commutator $[\alpha_i, [\alpha_j, [\dots, [\alpha_k, \alpha_l] \dots]]]$.

At this point, we can proceed in different ways and we present two successful strategies to build new methods.

2.1 Magnus-decomposition integrators

Firstly, we examine the structure of the Lie algebra generated by the α_i. We immediately notice that, $[ij] = 0$ for $i, j > 1$. Furthermore, the element [212] has the same sparse structure as α_2, α_3 and can be computed in the same exponential at a minor extra cost.

We distinguish the following types of exponentials that can occur as elements of the Lie algebra generated by $\alpha_i,\ i = 1, \dots, s$

$$E_1 = \exp\begin{pmatrix} D & B \\ C & -D^T \end{pmatrix}, \quad E_2 = \exp\begin{pmatrix} 0 & I \\ C & 0 \end{pmatrix}, \quad E_3 = \exp\begin{pmatrix} 0 & 0 \\ C & 0 \end{pmatrix}.$$

Now we propose a simplified composition given by

$$\begin{aligned} \Upsilon_1^{[4]} = \exp\big(x_3\alpha_2 + x_4\alpha_3 + x_5[212]\big)\ \exp\big(x_1\alpha_1 + x_2\alpha_3\big) \\ \times \exp\big(-x_3\alpha_2 + x_4\alpha_3 + x_5[212]\big) \end{aligned}$$

$$x_1 = 1, \quad x_2 = \frac{1}{20}, \quad x_3 = \frac{1}{12}, \quad x_4 = \frac{1}{60}, \quad x_5 = -\frac{1}{2880},$$

obtained from the scheme $\Upsilon_1^{[6]}$ in [1] where some costly terms used to force to get order six are removed. This scheme can be considered as an optimized 4th-order method because it uses a 6th-order quadrature rule and vanishes most of the leading error terms. The two-exponential scheme is also considered and in all cases the exponentials are replaced the symplectic decomposition at a sufficiently high order.

2.2 Magnus-splitting integrators

Consider the split

$$A(t) = B(t) + D \qquad B(t) = \begin{pmatrix} 0 & 0 \\ -M(t) & 0 \end{pmatrix}, \qquad D = \begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix},$$

which generates new graded Lie algebra

$$\delta_1 = hD, \qquad \beta_i = \frac{h^i}{(i-1)!} \frac{d^{i-1}\tilde{B}(s)}{ds^{i-1}}\Big|_{s=t+\frac{h}{2}}, \ i \geq 1$$

where $\tilde{B}(s)$ is the interpolating polynomial to $B(s)$ in the interval, and $\alpha_1 = \delta_1 + \beta_1 \ \alpha_i = \beta_i, \ i > 1$. It is easy to see that $[\beta_i, \beta_j] = 0$ as well as $[\delta_1, \delta_1, \delta_1, \beta_i] = [\beta_i, \beta_j, \beta_k, \delta_1] = 0$ for any value of i, j, k, and the formal solution (7) simplifies considerably to

$$\begin{aligned} \Omega^{[6]} &= \delta_1 + \beta_1 + \frac{1}{12}\beta_3 + \frac{1}{12}[\beta_2, \delta_1] + \frac{1}{360}\big(- [\delta_1, \beta_3, \delta_1] + [\beta_1, \delta_1, \beta_3]\big) \\ &\quad - \frac{1}{240}[\beta_2, \delta_1, \beta_2] + \frac{1}{720}\big([\delta_1, \beta_1, \delta_1, \beta_2] - [\beta_1, \delta_1, \beta_2, \delta_1]\big). \end{aligned}$$

The 11-stage method (5) satisfies all required order conditions and shows a high performance in practice.

References

[1] P. Bader, S. Blanes, E. Ponsoda and M. Seydaoğlu. Symplectic integrators for the matrix Hill equation. *J. Comp. Appl. Math.*, (In Press, http://dx.doi.org/10.1016/j.cam.2016.09.041).

[2] D. S. Bernstein. Matrix mathematics: theory, facts, and formulas, vol. 1. U.S.A. 2009.

[3] S. Blanes, F. Casas, J. A. Oteo, J. Ros. The Magnus expansion and some of its applications. *Physics Reports*, Volume (470): 151–238, 2009.

[4] A. J. Dragt. Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics, University of Maryland, 2015.

[5] A. Zanna. Collocation and relaxed collocation for the Fer and the Magnus expansions. *SIAM J. Numer. Anal.* Volume (36): pp. 1145–1182, 1999.

Mathematical modelling of bullying propagation in Spain

Elena De la Poza[1]*, Lucas Jódar**, Lucía Ramírez**

(*) Faculty of Business and Management.
Universitat Politècnica de València.
Building 7J, Camino de Vera s/n., Valencia, Spain.
(**) Instituto Universitario de Matemática Multidisciplinar,
Building 8G, 2nd floor, Universitat Politècnica de València, 46022 Valencia, Spain

1. Introduction

Bullying in schools can be defined as a category of aggressive behaviour in which there is an imbalance of power, and the aggressive event is repeated over time [1, 2]. Bullying occurs as a social process in nature, taking place in groups, in which the victim has scarce possibility of avoiding his/her tormentors, and the bully often gets support from other group members [3]. The attacks are mostly unprovoked and may be physical or verbal, direct or indirect. Sexual violence must be included in the category of physical violence when it occurs between scholars.

Researchers in the field emphasize the social character of bullying [4] describes bullying as violence in a group context in which the scholars reinforce each other's behaviour in their interaction. Studies on bullying have typically concentrated only on bully-victim relationship,[5,6]. Until recently bullying behaviour was interpreted as a function of certain characteristics of the bully and/or the victim, while the group was set aside, or forgotten.

Lagerspetz el al. [7] have pointed out that bullying among schoolchildren has a collective character and that it is based on social relationships in the group. Bullying often takes place in situations in which the victim is surrounded by several members of the group; even the ones belonging to the group but not in presence of the bullying action are usually aware of what is going on, due to the fact that bullying by definition happens repeatedly, over a period of time. Thus, there are different participant roles in the bullying process, [8, 9, 10].

This model may be regarded as a continuation of previous social population models dealing with electoral behaviour [11] and propagation of social addictions.
[12,13].

This paper focuses on modelling the propagation of bullying in the Spanish school population aged [12, 18] during the period July 2015-January 2020 identifying and quantifying its main drivers. Then, the study provides recommendations to reduce and prevent the growth of this social problem but also to mitigate a correlated problem such as the intimate partner violence among adults.

[1] E-mail: elpopla@esp.upv.es

2. Methods

2.1 Population of study

The population of study (S) embraces all Spanish school children aged in the interval [12,18] during the period of study July 2015-January 2020.

The first step of the work consists of splitting our population into five categories:

- Victims, (V): those school children who become targets for bullies due to a particular physical or psychological characteristic.
- Aggressors (bullies) (A): emotionally unbalanced schooled people with low self-esteem who demonstrate power exercising violence against peers.
- Defenders (D): those school children who take side with the victim direct or indirectly (complaining to school staff).
- Outsiders, (O): merely observing the events without taking part at them.
- Co-operators (C): those that reinforce bully's behaviour by taking part of the aggression (collective). The persistence of bullying is due to the collective character of the aggression. [8].

Then, from the S(0)=3,100,000 population aged [12,18] in July 2015, [15], we estimate the initial subpopulations at n=0 (July 2015) obtained from different reports as [9] and [10]:

- V(0)=288,300, victims are the 9.3% of S(0).
- D(0)=158,100, defenders represent the 5.1% of S(0).
- O(0)=1,525,200, outsiders are the 49.2% of S(0).
- C(0)=821,500, co-operators represent the 26.5% of S(0).
- A(0)=306,900, are the 9.9% of S(0).

Both V(n) and A(n) are cumulative categories, what implies, those victims in first semester (July 2015) will be also collected by V(9), number of victims in January 2020 and A(9), number of aggressors in January 2020.

Thus, V(n) is defined as the amount of students aged [12, 18] turning into victims at any time before the end of semester n. Then, A(n)= amount of students aged [12, 18] becoming aggressors at any time before the end of semester n.

Then, a discrete mathematical model is built considering the dynamic behaviour of these subpopulations by semester in the period of study [2015, 2020] by identifying the main drivers of the problem.

2.2. Transition Coefficients' Modelling

The primary causes that determine the growth of the problem are: legal environment, demographic (birth and death rates, emigration), economic stress (unemployment), emotional stress, alcohol and drug consumption, technology, and contagion by experiential learning, [14].

We assume that under the period of study the bullying legal environment doesn't change but also there is a deficit of anti-bullying policies implemented in Spanish high- school system.

Then, the transit coefficients are modelled and estimated:

Starting by the demographic factor, we compute the difference between the inputs to the system (school children aged 12), and the outputs (children becoming older than 18 years old and those who emigrate or pass away), [15]. We denote the demographic coefficient as β. So we estimate β value for each subpopulation. This coefficient is assumed constant for the period of study.
$\beta(V) = 1.804$; $\beta(D)= 989$; $\beta(O) = 9.543$; $\beta(C) = 5.140$; $\beta(A) = 1.921$.

About defenders (D): we identify three possible transits. Those defenders who can become victims, ofenders or remain as defenders.
$D \rightarrow V$ *$\alpha(DV) = 0.06$; estimated as the 10% of the population aged (12, 15) acting as defenders become victims of bullies. Thus, since 60% of S(n) are aged (12, 15),* $\alpha(DV) = 0.6 * 0.1$.
$D \rightarrow D$ We assume there are two kind of defensive behaviour, those who intermediate physically during the aggression and also those who set a query. Also, we assume scholars act as defenders due to their own principles in an 80% times, while 20% of the times they do based on friendship with the victim. $\alpha(DD) = 0.74$.
$D \rightarrow O$ $\alpha(DO) = 0.2$.

About outsiders (O): there are two possible transits:
$O \rightarrow D$ $\alpha(oD) = 0.081$, estimated as the result of three factors, the relation with the victim (5%), defender contagion (2.5%), and the outsider family advise (0.06%), indirect or by complaint defence.
$O \rightarrow C$ $\alpha(oc) = 0.06 - 1/2\ 0.0$ dn (previous), due to the economic stress of the family, where dn is the rate of unemployment drop per semester based on [16, 17].
$O \rightarrow O$ $\alpha(oo) = 0.839 + 1/2\ 0.0$ dn, due to the economic situation. This is measured by the rate of unemployment (dn).

About co-operators (C): there are two possible transits:
$C \rightarrow O$ $\alpha(CO) = 0.125$, estimated through three causes: relation to victim (5%), non violent partner contagion (6.9%), and the cooperator family advise (0.06%)
$C \rightarrow A$ $\alpha(CA) = 0.05 - 1/2\ 0.0$ dn. The reference rate 0.05 means that 3 out of 4 co-operators becoming one year old transit to aggressors, dn is the % of employ improvement per semester [16, 17].

2.3 Model

The model is expressed as follows:

$$V(n+1) = V(n) + \alpha(DV)\, D(n) + \alpha(oV)\, O(n) + \beta(V)$$

$$D(n+1) = \alpha(oD)\, O(n) + (1 - \alpha(DV) - \alpha(DO))\, D(n) + \beta(D)$$

$$O(n+1) = \alpha(DO)\, D(n) + \alpha(co)\, C(n) + (1 - \alpha(oV) - \alpha(oD) - \alpha(oC))\, O(n) + \beta(O)$$

$$C(n+1) = \alpha(OC)\, O(n) + \alpha(AC)\, A(n) +(1 - \alpha(CO) - \alpha(CA))\, C(n) + \beta(C)$$

$$A(n+1) = A(n) + \alpha(CA)\, C(n) + \beta(A)$$

Then, solving the difference system of equations of the model the subpopulations V(n),D(n), O(n),C(n) and A(n) are computed for each semester forecasting those at the end of the period of study, January 2020.

3. Results

Figure 2 shows the trend of subpopulations for the period of study. By assuming a constant legal environment, the V and A subpopulations grow, while outsiders reduce, and co-operators remain constant.

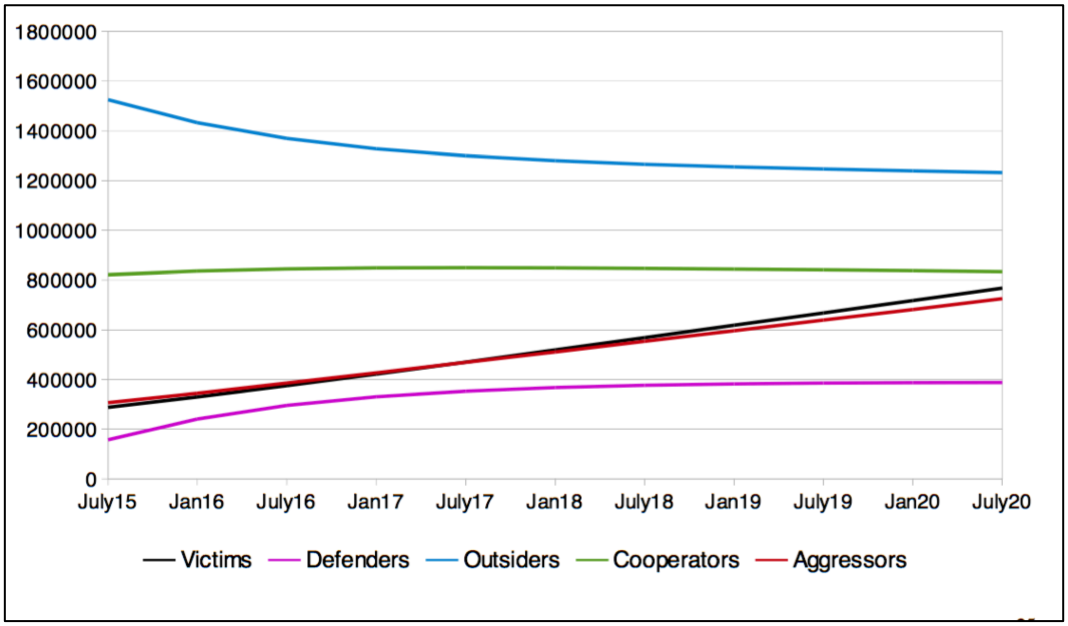


Figure 2. Forecast of the subpopulations.

Conclusions

Bullying is a social problem affecting Spanish school population. However the magnitude of this problem is uncertain and hidden. The physical and psychological damages to children are visible in the short term but also in the lon run; school absenteeis, loss self-esteem, mistrustfulness, depression or suicide. Also, bullies are potential parter and mobbing aggressors.

This work presents a model that allow us to forecast the number of victims and aggressors in the period of time 2015, 2020 but also let us know the key drivers of the problem. Thus, some recommendations to stop the problem are given:

- Student's Encouragement to report bullying incidents
- Implementation of anti-bullying measures at school (i.e. visual advertising in public school spaces).
- Protection of victims
- Development of solidarity among students
- Stimulation of accountability of both educators and centres to eradicate these practices
- Intense and massive campaigns advising families
- Legal support for teachers and centres to require and empower them against bullying

References

[1] D. Olweus, Bully/victim problems among school children: Basic facts and effects of a school-based intervention program. In Rubin K. Pepler eds. The development and Treatment of Childhoold Agression. Hillsdale, N.J. Erlbaum, 1991, pp.411-418.

[2] P.K. Smith. D. Thompson, Practical Approach to Bullying, David Fulton, London 1993.

[3] K. Bjorqvist, K. Ekman, K.M.J. Lagerpetz, Bullies and Victims: Their ego picture, ideal ego picture and normative ego picture. Scandinavian Journal of Psychology 23 (1982), pp.307-313.

[4] A. Pikas, Sa stppar vi mobbning, Prisma Ed., Stockholm 1975.

[5] C. Salmivalli, K. Peets, Bullies, Victims, and Bully-Victim Relationships published in Middle Childhood and Early Adolescence in Handbook of Peer Interactions, Relationships, and Groups edited by Kenneth H. Rubin, William M. Bukowski, Brett Laursen, New york, 2011.

[6] Smith, Peter K.; Bowers, Louise; Binney, Valerie; Cowie, Helen Duck, Steve, Relationships of children involved in bully/victim problems at school published in Learning about relationships. Understanding relationship processes series, Vol. 2., pp. 184-204. Thousand Oaks, CA, US: Sage Publications, Inc, xiii, 252 pp., 1993.

[7] K. M. J. Lagerspetz, K. Björkqvist, M. Berts, E. King, Group aggression among school children in three schools. Scandinavian Journal of Psychology, 1982, 23, 45–52.

[8] C. Salmivalli, K. Lagerspetz, K. Bjorvist, K. Osterman and A. Kaukiainen, Bullying as a Group Process: Participant Roles and Their Relations to Social Status Within the Group, Agressive Behaviour, 22 (1996), pp.1-15.

[9] Oñate, A., Piñuel, I, (2007), “Acoso y Violencia Escolar en España”, Informe Cisneros, Instituto de Innovación Educativa y Desarrollo Directivo, Madrid, Ed. IIEDDI.

[10] “Yo a eso no juego. Bullying y ciberbullying en la infancia" Fundación Save the Children, Coord: Ana Sastre, 2016, Ed. Save the Children España, Madrid.

[11] E. De la Poza, L. Jódar, and A. Pricop, Mathematical Modeling of the Propagation of Democratic Support of Extreme Ideologies in Spain: Causes, Effects, and Recommendations for Its Stop, Abstract and Applied Analysis, vol. 2013, Article ID 729814, 8 pages, 2013. doi:10.1155/2013/729814

[12] E. De la Poza Plaza, M.N., Guadalajara Olmeda, L.A. Jódar Sánchez, P. Merello Giménez, Modeling Spanish anxiolytic consumption: Economic, demographic and behavioral influences. Mathematical and Computer Modelling, 2013, 57(7):1619-1624.

[13] E. de la Poza, M. del Líbano, I. García, L. Jódar, P. Merello, Predicting workaholism in Spain: a discrete mathematical model, International Journal of Computer Mathematics, 2014, 91, 233-240, DOI: 10.1080/00207160.2013.783205

[14] Christakis & Fowler, Connected: The Surprising Power of Our Social Networks (2009)

[15] Spanish Statistics Institute, 2016.

[16] IMF, International Monetary Fund, 2016.

[17] OECD, Organisation for Economic Co-operation and Development, 2016.

Effects of rail cracking in different railway track typologies.

Beatriz Baydal Giner[1*], Carlos Miñana Albanell[1], Clara Zamorano Martín[1], Julia Irene Real Herraiz[1]

[1]Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

[2]Foundation for the Research and Engineering in Railways, 160 Serrano, 28002 Madrid, Spain

* Corresponding author. E-mail: beabaygi@cam.upv.es. Telephone: +34 96 387 70 00

July 18, 2016

1. Introduction

The need of improving railway infrastructures has arisen during last decades as a consequence of the increasing traffic loads. To do so, the first step should be to reduce the strain-stress state of the track induced by the rolling stock.

In this sense, track imperfections compel a potential danger, since usually produce dynamic overloads which deteriorate both vehicle and track infrastructure. Furthermore, cracks have strong influence on modal parameters (specifically of the natural frequencies), since strong variations of stiffness on the cracked zone take place.

Thus, in order to study the stress-strain response of different track typologies against cracks, the present investigation develops a three-dimensional Finite Elements (FE) model of three railway tracks in which cracked and uncraked rails are merged. The model is calibrated and validated with real data following the studies developed by Montalbán et al. [1] from an Experimental Modal Analysis (EMA).

2. FE model of the isolated rail

A FE model of 1.5 m long is done (Fig. 1). The rail is modeled using SOLID95 brick elements and the mesh size selected for this study was 0.005 m. The reference values to calibrate FEM models of the isolated rails have been obtained in [2].

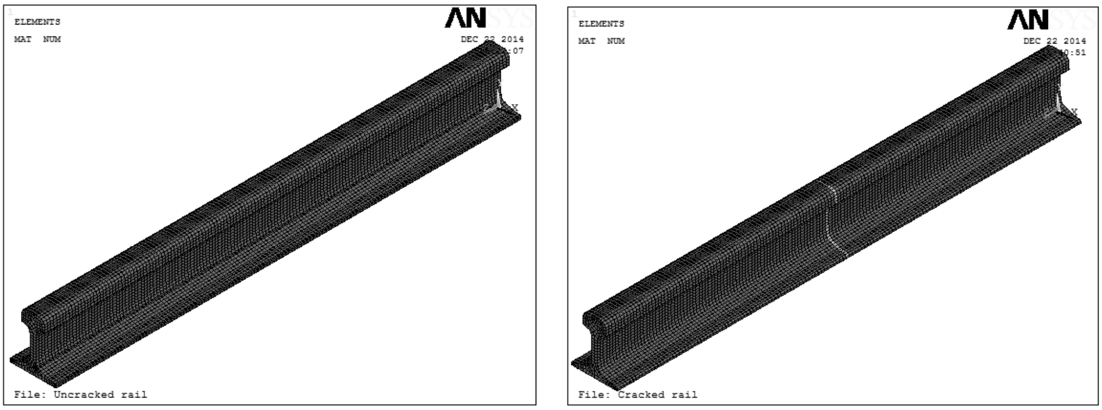


Fig. 1. FE model of the un-cracked rail (left) and cracked rail (right)

In order to simulate the effect of the crack, a stiffness discontinuity was induced according to Eq. (1) by reducing the Young's Modulus (E) keeping constant the inertia (I), since the crack introduces a discontinuity that results in a lower flexural stiffness caused by a drop of the inertia in the cracked section [3,4].

$$Flexural\ Stiffness = EI \tag{1}$$

All the calculations were done using the Modal Analysis tool provided by ANSYS software, which solves the equation of undamped free vibration Eq. (2), where *[M]* is the mass matrix, *[K]* is the stiffness matrix, *{u}* the displacement vector and *{ü}* the acceleration vector

$$[M]\{\ddot{u}\} + [K]\{u\} = 0 \tag{2}$$

3. FE model of the scenarios studied

Once the isolated model of the rail is done, three scenarios are studied: a ballasted track; a concrete slab Rheda 2000 track and an asphalt slab track Getrac (Fig. 2). The track model has been developed on the basis of the works compiled by the ORE committee D-117 [5] and the Recommendations for the projects of railway infrastructures of the Spanish Ministry of Development [6].

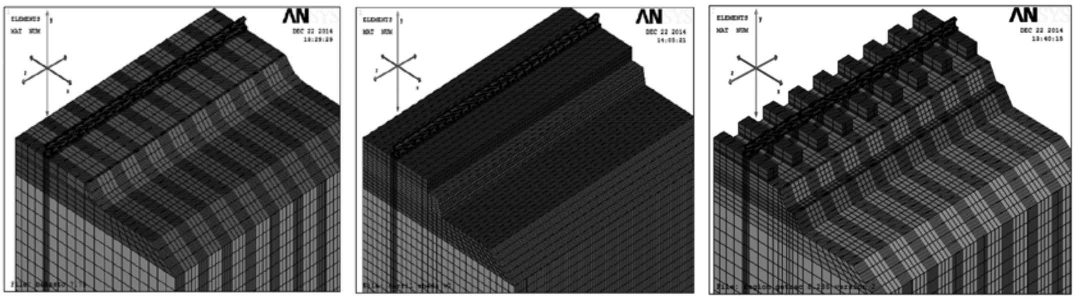


Fig. 1. FE model of the ballasted track (left), Rheda 2000 (centre) and Getrac (right)

The cross section of the track is included in the x-y plane with the dimensions of the different layers and elements specified in section 3. It must be highlighted that in the case of the Getrac asphalt slab track all the asphaltic layers have been considered as a single 0.03 m thick layer (this simplification and its effect on the model results has been proven to be insignificant).

The track length is determined by Pandolfo's Theory, which states that the vehicle load acting on a point of the rail affects up to 9 adjacent sleepers. In addition, the model has been reduced thanks to the symmetry about the x-y plane. Boundary conditions are defined setting equal to zero (Ux=Uy=Uz=0) the total displacements of the nodes located in the vertical planes of the model as well as in the horizontal plane in the base [41]. The different track models are shown in Figs. 7-9.

Track geometry has been already developed in the isolated rail models for a length comprising 9 sleepers. For railpads, sleepers and slabs a linear elastic behavior is assumed, since the contribution to the total displacements of the permanent deformation of the asphaltic layers is negligible compared with the settlements of the granular layers. Nevertheless, this simplification cannot be assumed for the granular layers. The ballast, subballast and the embankment are modeled with elastoplastic behavior according to the Drucker Prager's model [7].

The element used for the track model is again the SOLID95. The mesh size on the sleeper and the slab is 0.03m. The 3D model has been subjected to a static analysis in which the vertical load has been augmented to account for the dynamic effects according to Eissenmann's formulation.

After modeling these sections, the obtained results are compared with the experimental data obtained during a data gathering campaign with freight trains circulating at 100 km/h (Tables 1-3).

Ballasted track						
	Total settlement (mm)	Error %	Vertical displacement in the interface platform-subballast (mm)	Error %	Stresses in the interface platform-subballast (kN/m^2)	Error %
EXPERIMENT	-0,740	0.21	-0,544	-1,42	-76,902	0,48
MODEL	-0,738		-0,552		-76,531	

Table 1. Validation of the ballasted track FE model.

RHEDA 2000						
	Total settlement (mm)	Error %	Vertical displacement in the interface platform-subballast (mm)	Error %	Stresses in the interface platform-subballast (kN/m^2)	Error %
EXPERIMENT	-1,420	0.11	-0,051	-4,13	-15,800	-0.94
MODEL	-1,418		-0,053		-15,950	

Table 2. Validation of the Rheda 2000 FE model.

GETRAC						
	Total settlement (mm)	Error %	Vertical displacement in the interface platform-subballast (mm)	Error %	Stresses in the interface platform-subballast (kN/m^2)	Error %
EXPERIMENT	-1,560	-0,14	-0,301	1,54	-24,711	1,55
MODEL	-1,562		-0,308		-24,332	

Table 3. Validation of the GETRAC FE model.

This comparison ensures the model validation for the subsequent comparison with the cracked rails.

4. Simulations and results

Once both FE model of the rail and 3D FE track models are developed, it is the moment to merge the cracked rails in the track model. To do so, four different static analyses are performed for each track system: one for the non-cracked rail and three for the cracked rails with different crack depths. Therefore, the total number of scenarios calculated is 12 (Table 4).

		Total settlement (mm)	**%**	**Vertical displacement (mm)**	**%**	**Stresses (kN/m^2)**	**%**
Ballasted track	Uncracked rail	-0,7384		-0,5517		76,531	
	Cracked 2.2 cm	-0,7543	2,15	-0,5533	0,29	76,683	0.20
	Cracked 4.3 cm	-0,7697	4,24	-0,5554	0,67	76,897	0.48
	Cracked 9.8 cm	-0,8513	15,29	-0,5781	4,78	78,788	2.95
RHEDA 2000	Uncracked rail	-1,4184		-0,0531		15,950	
	Cracked 2.2 cm	-1,4389	1,44	-0,0531	0,09	16,106	0,98
	Cracked 4.3 cm	-1,4599	2,93	-0,0532	0,21	16,176	1,42
	Cracked 9.8 cm	-1,6089	13,43	-0,0539	1,53	16,752	5,03
GETRAC	Uncracked rail	-1,5621		0,3084		24,332	
	Cracked 2.2 cm	-1,5822	1,29	-0,3089	0,18	24,393	0,25
	Cracked 4.3 cm	-1,6027	2,60	-0,3096	0,39	24,459	0,52
	Cracked 9.8 cm	-1,7426	11,56	-0,3167	2,72	25,077	3,06

Table 2.Calculated results for the different track systems with non-cracked and cracked rails

According to Table 4, if stresses and displacements of the infrastructure with a cracked rail are compared with the uncracked rail, it is deduced that the cracks are only influent in the case of the deepest crack. In this case, the crack reaches the web of the rail, modifying the behavior of the whole track. Moreover, the influence of the cracked rail on the deflections is more pronounced in the upper layers since the vertical displacement of the rail top is greater than the vertical displacement in the interface platform-subballast in all the studied cases.

Focusing on the displacements of the rail top, the most susceptible track to cracking is the ballasted track (15%), followed by the Rheda 2000 system (13.5%) and the Getrac system (11.5%).

The displacements in the platform are more affected by rail cracking in the ballasted track, with a relative variation of a 5% for the deepest crack, being this relative variation irrelevant in the rest of the cases. The Getrac track experiences a variation of a 3% with regard to the initial situation when the crack depth is 9.8 cm. Finally, the concrete slab track displacements only increase by 1.5 %. This fact is caused by the low stiffness of the railpad in the concrete slab track; this element absorbs the most part of the displacements. Thus, these displacements become almost imperceptible in the elements located beneath the elastic railpad.

Looking at the stresses produced in the different scenarios, the most affected system is the concrete slab track. The stresses raise may cause the concrete cracking and the possible structure failure. The increment of the 5% of the subballast-platform stresses in the most unfavorable scenario for the Rheda 2000 track falls to a 3% in the case of the Getrac system. The best static behavior in terms of stresses corresponds to the ballasted track. The elastic systems, such as the ballasted track, are able to spread the forces originated from the traffic loads, creating a pressure bulb under point of load application. In contrast, slab track systems concentrate the stresses.

Settlement variations induced by rail cracks are very important since if these settlements are high, the impact loads originated when the train is passing on the track may damage the rolling stock and cause its sudden breakage. Moreover, the increase of the stresses and the displacements in the interface subballast-platform may cause a structural failure. Apart from the sudden fracture of the rail, rail cracks may trigger other important deterioration mechanisms: the rail top settlements cause a stress increment in the different support layers and therefore in the whole structure and the dynamic overloads will be higher in the cracked zones when the vehicle wheel passes over the crack. These overloads would begin a deterioration process in the rail and in the rolling stock. Finally, the dynamic and impact loads produced as a consequence of the crack affect the passenger comfort and cause increased levels of noise and vibration

5. Conclusions

The work presented analyzes influence of the rail vertical cracking on the stress-strain behavior of three different track types: ballasted tracks, concrete slab tracks and asphalt slab tracks. The followed procedure is divided in three stages.

First, different FEM models for the cracked and non-cracked rails were developed. These models were calibrated in a previous research using Experimental Modal Analysis results. The low errors obtained in the model calibration reveal that the FEM model is a useful tool to describe the cracked and non-cracked rail behavior.

The second part of the research focuses on the comparison of three different railway track systems: ballasted track, concrete slab track and asphalt slab track. The results from an extensive experimental campaign were used to describe the structural behavior of the studied tracks as well as to calibrate the FEM models of these tracks. On the basis of the results, the advantages and disadvantages of each track were identified. Moreover, the FEM models constructed allowed studying the static behavior of the tracks.

Finally, the 3D FEM track models including the cracked rails have revealed the stress-strain changes in the railway structures in different scenarios in which the rail crack had different depths. The results show that only the deepest cracks have strong influence on the track static behavior. The most influenced tracks by rail cracking were the concrete slab track in terms of stresses and the ballasted track in terms of displacements. The best performance was achieved by the asphalt slab track since both stresses and displacements induced by the rail cracks were lower than in the ballasted and the concrete slab track. The structural changes of the railway tracks are not severe but the renovation of the cracked rails is absolutely necessary since the rail may suddenly fail when the train loads are passing over the imperfection

References

[1] Montalbán, L., Zamorano, C., Palenzuela, C. and Real, J.I. (2014). “Analysis of the influence of cracked sleepers under static loading on ballasted railway tracks”. *The Scientific World Journal*

[2] Montalbán, L., Baydal, B., Zamorano, C. and Real Herraiz, J.I. (2014). “Experimental Modal Analysis of transverse-cracked rails. Influence of the cracks on the real track behavior”. *Structural Engineering and Mechanics*, 52(5) 1019-1032

[3] Nahvi, H.and Jabbari, M. (2005). “Crack detection in beams using experimental modal data and finite element model”. *International Journal of Mechanical Sciences*, 47 1477-1497

[4] Skrinar, M. (2009). “Elastic beam finite element with an arbitrary number of transverse cracks”. *Finite Elements in Analysis and Design*, 45(3) 181-189

[5] UIC 719: Earthworks and track bed construction for railway lines, 3rd edition, February 2008

[6] Ministerio de fomento/Secretaria de estado de Infraestructuras y transportes. (1999). “*Recomendaciones para el proyecto de plataformas ferroviarias*”. Centro de publicaciones, Spain (in Spanish)

[7] Gallego, I. and López, A., (2009). “Numerical simulation of embankment-structure transition design”. *Journal of Rail and Rapid Transit*, 223 331-343

Study of thermal flux along railway track under extreme cold conditions. Comparison of a new material with traditional configurations.

Miriam Labrado Palomo[1*], José Luis Velarte González[1], Teresa Real Herraiz[1], Julia Irene Real Herraiz[1]

[1]Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

*Corresponding author. E-mail: milabpa@upv.es. Telephone: +34 96 387 70 00

July 18, 2016

1. Introduction

In extreme cold climates, the ground is characterized by the presence of a *permafrost layer*. This, by definition, is the layer of the ground placed at a certain depth which is permanent frozen.

As may be guessed, the position of the layer will depend on climate conditions, since the surface temperature, or even the presence of rain processes, may vary the thickness of the permafrost layer.

If this happens, differential ground settlements will appear and railway track deterioration will take place. In this sense, the present investigation (involved in a R+D project called ICY-TRACK) develops a new bituminous subballast able to avoid differential track settlements due to freeze-defrost cycles in the permafrost layer. Hence, the aim of the project is to develop a new material able to protect the infrastructure of railway lines placed at extreme-cold countries. This process is supported by the development of a Finite Elements (FE) model of a railway track able to assess the variations of the thermodynamic characteristics of the ground through time.

2. Methodology

The development of the new material is based in a very important theoretical base: the minor importance of the convention processes against the great importance of conduction processes during heat transfer [1]. Thus, the equation that governs the process is (Eq. 1), where the equivalent heat capacity (C_e^*) and the thermal conductivity (λ_e^*) are considered [1, 2].

$$C_e^* \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(\lambda_e^* \frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(\lambda_e^* \frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(\lambda_e^* \frac{\partial T}{\partial z}\right) \quad (1)$$

As could be deduced from the previous expression, the heat transfer problem through the soil layers is time and space dependent. According to this, it would be possible to carry on several static analyses (for example, the month of July when the air temperature is the highest and the permafrost layer thickness is the minimum; or the month of January when the air temperature is the lowest and the thickness of permafrost layer is the highest), but this possibility would only provide the thermal state at a certain timestep. Hence, in order to provide the thermal state through time, it is necessary to develop a transient analysis.

Once the analysis type is decided, the numerical model is developed. Then, the insulating capacity of the new material and the temperature-depth distribution profile is obtained through time and some conclusions are drawn.

3. FE model development

In order to begin with the development of the model, the first step is to obtain the permafrost layer status. To do so, a 5 year transient analysis is performed in a numerical model (model 1) which contains only the natural soil layers. The idea is to obtain the permafrost layer position once the temperatures have been stabilized.

Then a full section model (model 2) which also contains the track elements (Fig. 1) is developed in order to obtain the evolution of temperature-time inside the layers in different scenarios.

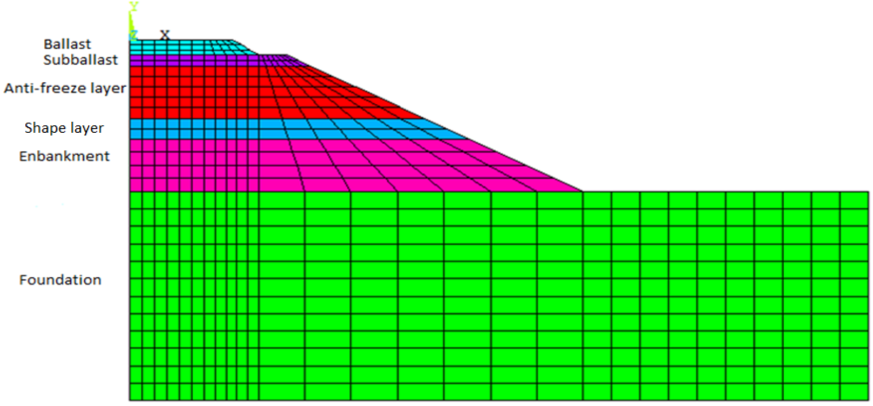


Fig. 1. Full section FE model

A total of four different scenarios have been studied. The aim of these comparisons is to determine the effectiveness of the new material developed in the context of the project against the configurations used in countries with extreme climatology:

- Scenario 0: Traditional section with 0.7m granular subbalast layer (reference section)
- Scenario 1: New section including bituminous subballast of 0,2; 0,3 and 0,7 m thickness instead of the granular subballast layer
- Scenario 2: Scenario 1 + additional insulation on embankment slope (geogrid, Fig. 2 left)
- Scenario 3: Scenario 1 + additional insulation on embankment base (EPS, Fig. 2 right)

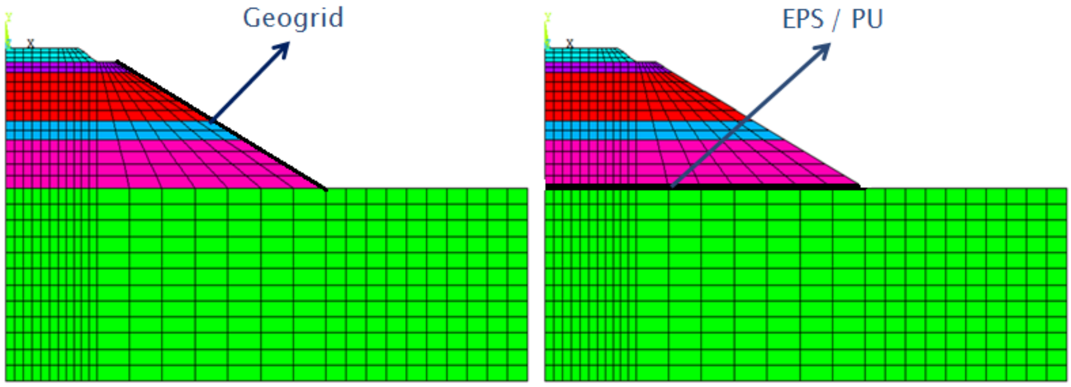


Fig. 1. Full section FE model. Scenario 3 (left) and scenario 4 (right)

All the aforementioned scenarios have been studied depending on the average air temperature (-5ºC and 0ºC) and depending on the anti-freeze layer thickness (1m and 2.4m).

For the model development, the commercial software ANSYS has been used. A typical railway section from the Spanish Railway Administrator (ADIF) has been modelled according to Spanish standards [3]. Thus, a 2-D cross-sectional model has been set since a three-dimensional model does not provide additional information but involves longer computational requirements. The section also has been simplified by symmetry.

With regard to the thermal properties, each material contains both thermal conductivity and heat capacity (C_e^*, λ_e^*) in frozen state (C_f, λ_f) and unfrozen state (C_u, λ_u), as well as the latent heat (L) according to Eqs. (2-3). These parameters are heat dependent, and should to be obtained for each temperature (T) considered.

$$C_e^* = \begin{cases} C_f & T < (T_m - \Delta T) \\ \dfrac{L}{2\Delta T} + \dfrac{C_f + C_u}{2} & (T_m - \Delta T) \leq T \leq (T_m + \Delta T) \\ C_u & T > (T_m + \Delta T) \end{cases} \tag{2}$$

$$\lambda_e^* = \begin{cases} \lambda_f & T < (T_m - \Delta T) \\ \lambda_f + \dfrac{\lambda_u + \lambda_f}{2\Delta T}[T - (T_m - \Delta T)] & (T_m - \Delta T) \leq T \leq (T_m + \Delta T) \\ \lambda_u & T > (T_m + \Delta T) \end{cases} \tag{3}$$

Boundary conditions are time-dependent and are defined for a right simulation of the phenomena. To do so, a sinusoidal function is applied over the free surface which defines the different materials of the track section. Thus, the heat of these points would change depending on the time corresponding to each step.

The expressions used in order to simulate the temperature oscillations on the model frontiers are different depending on the material considered. Such expressions are a sum of the average annual temperature of each material, the temperature variation over time and the effect of the climate change on the average increase. Anyhow, a geothermal flow of 0.06 W/m^2 is assumed, no thermal flow has been considered in the foundation side face and a gradient of 0.25 degrees has been considered as a consequence of global warming.

4. Simulations and results

Once the model is concluded, a total of 40 simulations are developed according to the aforementioned scenarios in order to determine the effectiveness of the new material developed against the configurations used in countries with extreme climatology. In each case, a time step of 24 hours is set.

Unfortunately, it would be impossible to summarize all the results obtained through the aforementioned simulations. For this reason, it is only shown the results for scenarios 0, 1, 2 and 3 in extreme cold climate (average air temperature -5^{o}C) provided with anti-freeze layer thickness of 1 meter. The other results are summarized.

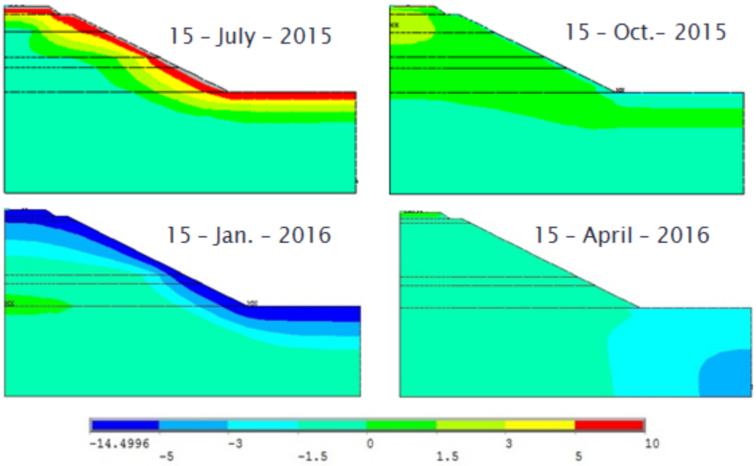


Fig. 2. Results obtained in Scenario 0 (granular subballast thickness 0.7 m) with average air temperature -5^{o}C and anti-freeze layer thickness of 1 meter

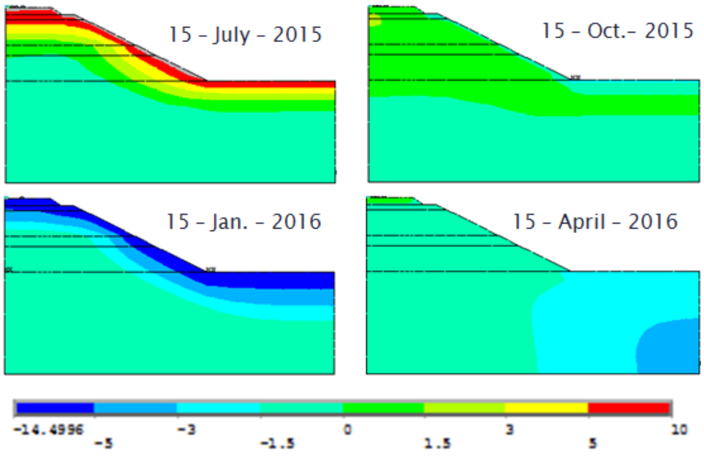


Fig. 3. Results obtained in Scenario 1 (bituminous subballast thickness 0.2 m) with average air temperature -5^{o}C and anti-freeze layer thickness of 1 meter

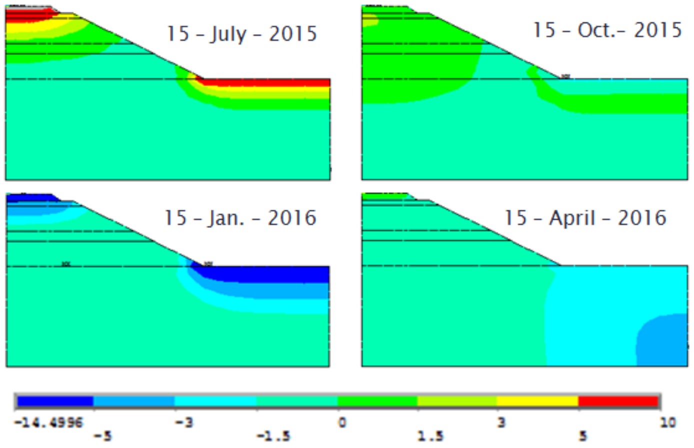


Fig. 4. Results obtained in Scenario 2 (bituminous subballast thickness 0.2 m + additional insulation on the embankment slope) with average air temperature -5^{o}C and anti-freeze layer thickness of 1 meter

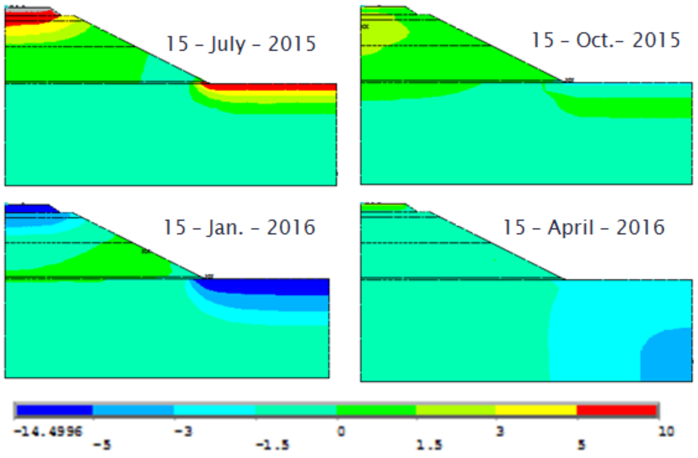


Fig. 4. Results obtained in Scenario 3 (bituminous subballast thickness 0.2 m + additional insulation on the embankment base) with average air temperature -5^{o}C and anti-freeze layer thickness of 1 meter

5. Conclusions

According to the previous results, as well as the rest of simulations, the following conclusions can be drawn:

- Similar behavior between granular subballast layer and bituminous subballast is obtained if typical properties of bituminous products are set. In both cases there is insufficient insulation to prevent successive freeze-defrost cycles in the base of the embankment.
- Bituminous subballast seems not having enough isolation capacity in extreme cold climate areas (-5ºC) to prevent freeze-defrost cycles into the embankment base if summer and winter are compared. By contrast, in typical cold climate areas (0ºC) the arrangement of layers used (with any thickness studied) avoids the appearance of freeze-defrost cycles through time.
- Using both isolation techniques together with bituminous subballast achieves that a greater part of the base become frost-free in extreme cold climate areas (-5ºC. Thus, if an additional isolation is located on the embankment base, the freeze-defrost cycles tend to disappear in almost all embankment base.

Anyhow, the development of the R+D project still continues. For this reason, it is also necessary to review in next steps of development all the models with new bituminous material properties (improved respect the current ones) in order to obtain more accurate results about isolation capacity.

References

[1] An, W. D., Wu, Z. W., & Ma, W. (1990). Interaction among temperature, moisture and stress fields in frozen soil. *Lanzhou University Press, Lanzhou*

[2] ZhiQiang, L., & Yuanming, L. Nonlinear Analysis for the Three-Dimensional Temperature Fields of the Ventilated Embankment in Qing-Tibet Railway.

[3] IGP. (2008). Instructions and recommendations for the writing of platform projects. IGP-2008. *Spanish Standard* (in Spanish).

Wheel-rail irregularities detection System using analytical and numerical methods.

Teresa Real Herraiz[1*], Silvia Marzal[1], Silvia Morales Ivorra[1], Julia Irene Real Herraiz[1]

[1]Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

*Corresponding author. E-mail: tereaher@upv.es. Telephone: +34 96 387 70 00

July 18, 2016

1. Introduction

Railway vibrations have become one of the major problems in the field of railway engineering. The great impact of this phenomenon affects to the users comfort and represents one of the most important transport externalities.

The main source of vibrations generation is the presence of irregularities in the wheel-rail contact. Thus, if both surfaces do not offer a perfect match, the contact forces growth inducing vibrations and important damages to both vehicle and track.

In order to go more deeply into this problem, this research is focused on the process to simulate defects in the wheel-rail contact and analyze the vibratory response induced. To do so, the contact forces generated on different scenarios are obtained by means of analytical methods and then incorporated to a numerical model calibrated by real data.

The work is involved in the context of a more extensive R+D Project whose aim is to automatically detect premature defects on the railway vehicles wheels by means of accelerometers and the use of Wavelet Transform technique.

2. Methodology

As previously explained, the aim of this investigation is to go more deeply into the study of vibration generation phenomenon caused by the passage of railway vehicles. This investigation will allow a further development of a new predictive maintenance system able to detect premature damages in wheels of railway vehicles.

To do so, the first step is to develop an exhaustive data gathering campaign. Thanks to this, the vibratory response of the rail caused by the passage of railway vehicles can be obtained by accelerometers. Thus, depending on where they were placed, it will be possible to recognize the location of different defects and determine the vibratory pattern of each pathology (Fig. 1).

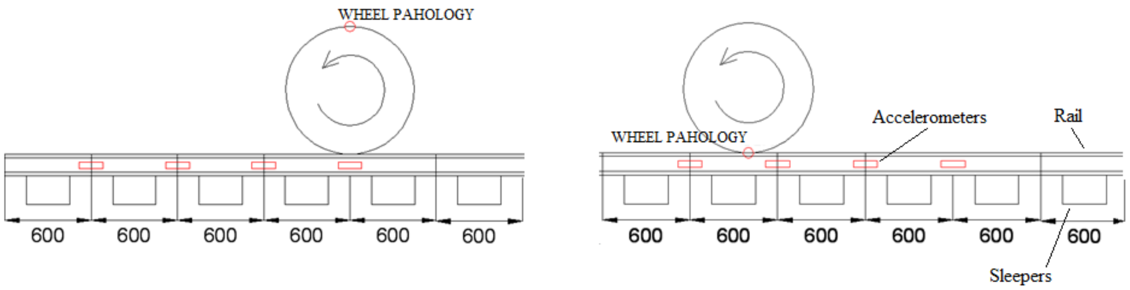


Fig. 1 Circulation of a damaged wheel over a monitored track

Since the vibratory pattern associated to each defect is different, the problem must be solved in two different stages. In first instance, an analytical model is developed in order to obtain the value of the overload produced in a certain situation. In a second stage, the overload is used as an input in a Finite Element model (FE model) which provides the vibratory response of a track in different scenarios: healthy wheels and damaged wheels.

Then, the vibratory response obtained during the data gathering campaign and the results obtained by numerical simulations are compared in order to assess the validity of the simulations. Once the FE model is calibrated and validated, it is possible to predict the existence of pathologies in wheels by comparing new vibratory responses registered in real tracks with those predicted by numerical simulations.

3. Case of study

For the data gathering campaign, eight biaxial accelerometers and four strain gauges were located at different points of the track according to Fig. (1). Thus, the accelerometers are activated once the railway vehicle circulates over the first pair of strain gauges, and are deactivated a few seconds after the passage of the vehicle over the second pair of strain gauges.

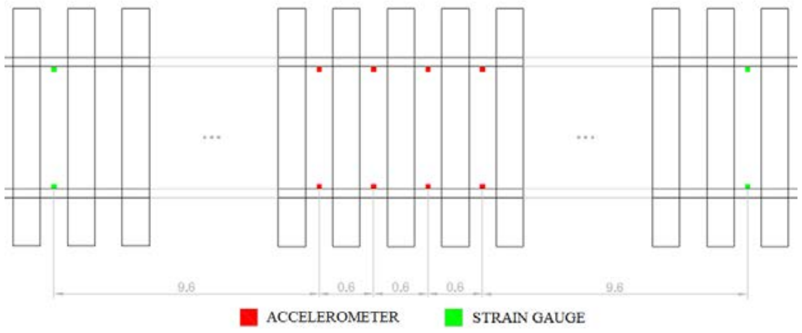


Fig. 1. Location of the measuring systems on the track.

In this case, the studied track was a slab track through which passenger trains and freight trains circulated at different speeds. Once the data gathering campaign ended, the wheel profile of each vehicle was studied detecting the presence of several wheel defects (cavities, flats, abrasions or cracks among others).

4. Simulations and results

Next, the development of both analytical and numerical models is explained.

4.1 Analytical model

Analytical models provide the time-history load for the different cases. The roughness profiles (ε) of each defect are introduced in the movement equations through different equations: wheel flats (Eq. 1); squats (Eq. 2) and rail corrugation (Eq. 3), where the equation parameters d, l, H, a and f_c are geometric data of defects.

$$\varepsilon(x) = \frac{d}{2}\left(1 - \cos\left(\frac{2\pi x}{l}\right)\right) \tag{1}$$

$$\varepsilon(x) = \frac{H}{2}\left(1 - \cos\left(\frac{2\pi x}{a}\right)\right) \tag{2}$$

$$\bar{\varepsilon}(f_c) = \begin{Bmatrix} |\bar{\varepsilon}(f_c)|\left(cos\varphi(f_c) + jsin\varphi(f_c)\right)\ if\ f_c \geq 0 \\ \varepsilon^*(f_c) \qquad if\ f_c < 0 \end{Bmatrix} \tag{3}$$

These profiles are superimposed and introduced in a three masses track model which is the responsible to obtain the overload in function of time. The equations that govern the model are shown below (Eqs. 4-6) where m_r, m_c and m_t are parameters of inertia of the wheel, the rail and the sleeper, respectively; c_r, c_c and c_t are damping parameters of the wheel, the rail and the sleeper, respectively; k_r, k_c and k_t are the parameters of stiffness of the wheel, the rail and the crossbeam, respectively; y_r, y_c y y_t are the generalized coordinates describing the vertical displacements of the wheel, the rail and the sleeper starting from its equilibrium position and K_H is the contact stiffness parameter.

$$m_r\ddot{y}_r + c_r\dot{y}_r + k_r y_r = K_H(y_c - y_r + r - \varepsilon(t))^{3/2} - \left(\frac{P}{8} + m_r g - k_r y_{r0}\right) \tag{4}$$

$$m_c\ddot{y}_c + c_c(\dot{y}_c - \dot{y}_t) + k_c(y_c - y_t) = -K_H(y_c - y_r + r - \varepsilon(t))^{3/2} \tag{5}$$

$$m_t\ddot{y}_t + c_c(\dot{y}_t - \dot{y}_c) + k_c(y_t - y_c) + c_t\dot{y}_t + k_t y_t = 0 \tag{6}$$

This equation system is solved using finite difference method.

4.2 FE model

Once the dynamic overloads are obtained by means of analytical methods, the time-history loads are introduced in a numerical model of the track (Fig. 2). This model is developed using the commercial software ANSYS. The numerical model is governed by

Eq. (7) where [K], [C] and [M] are respectively the mass, damping and stiffness matrices of the structure and $\{\ddot{\boldsymbol{u}}\}$, $\{\dot{\boldsymbol{u}}\}$ and $\{\boldsymbol{u}\}$ are the acceleration, velocity and displacement vectors:

$$[M]\{\ddot{\boldsymbol{u}}\} + [C]\{\dot{\boldsymbol{u}}\} + [K]\{\boldsymbol{u}\} = \{\boldsymbol{F}(t)\} \tag{7}$$

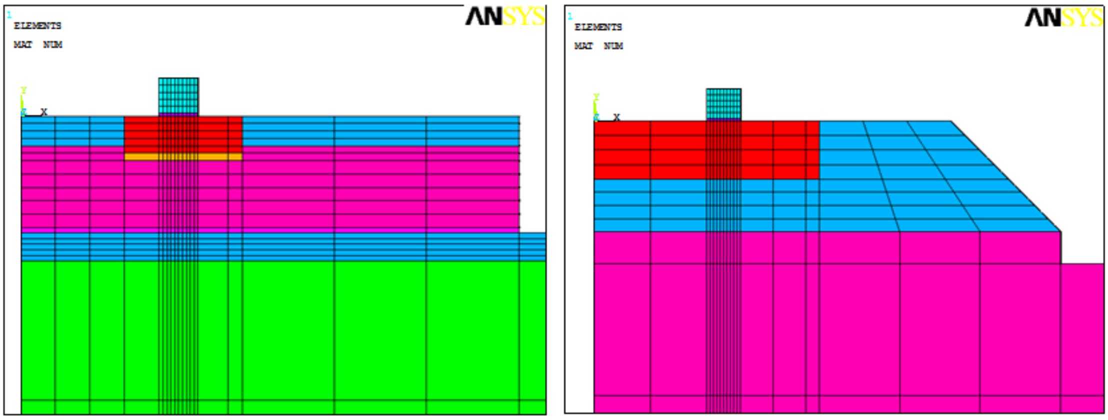


Fig. 2. Cross sections of the FE model of the track.

Once the model is calibrated and validated (Fig. 3), different defects are simulated in the model and an algorithm which allow to automatically detecting these defects is developed.

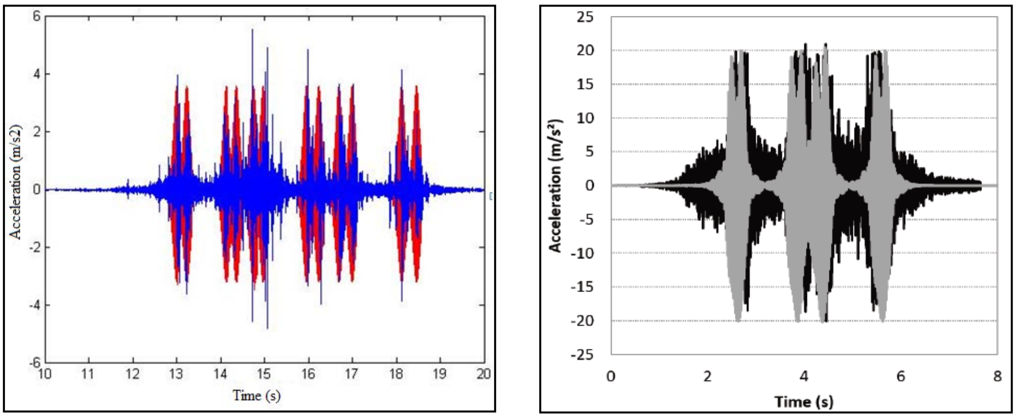


Fig. 3. Comparison between the real registers (blue and grey) and simulated registers (red and black) in a slab track (left) and ballasted track (right).

4.3 Detection algorithm

In order to develop de detection algorithm, the first step is to process the vibrations registered during the data gathering campaign (Fig. 4) by means of Wavelet Transform, since this technique allows distinguishing different levels of decomposition. In this sense, wheel flats usually present accelerations in the frequency band from 150 to 800Hz. Thus, if there is any anomaly or defect in different bandwidths, those will be detected.

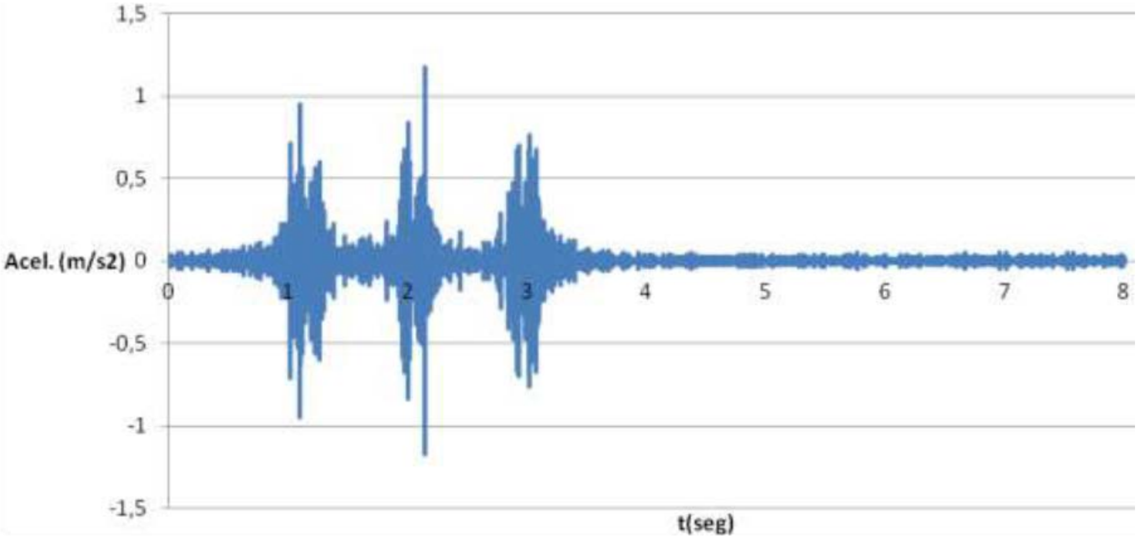


Fig. 5. One of the accelerograms registered during the data gathering campaign.

Then, the registers are processed by statistic techniques obtaining the normal distribution profile (Fig. 4 left) and those values exceeding 50% the value showing the 99% percentile shall be identified as defects. Thus, a response graph in which the value "1" indicates the presence of possible default and the value "0" no presence is obtained (Fig. 4 right).

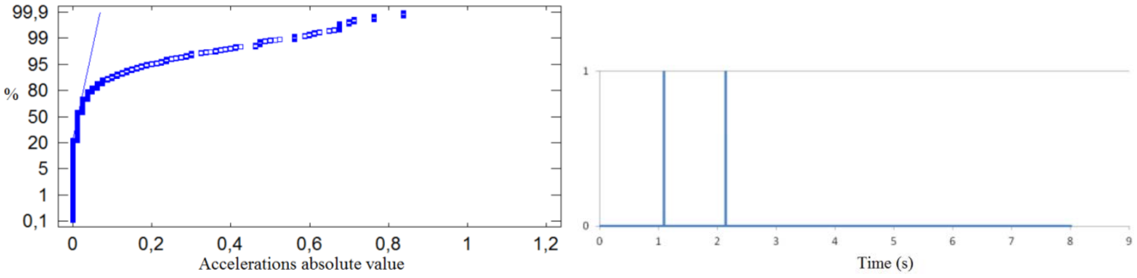


Fig. 4. Statistic analysis (left) and response graph (right).

5. Conclusions

The present investigation develops a new method able to detect damages in wheels of railway vehicles by the interaction of analytical and numerical models with real vibratory measurements. To do so, statistic techniques together with Wavelet Transform technique is applied to the vibratory signal on the time domain.

References

[1] Nieto, N., & Rojas, D. M. O. (2008). El uso de la transformada wavelet discreta en la reconstrucción de señales senosoidales. *Scientia Et Technica*, *1*(38), 381-386.

Design of a sound barrier able to reduce railway noise with olive stones.

Carlos Miñana Albanell[1], Miriam Labrado Palomo[1], Rafael Royo[2], Jorge del Pozo[3]

[1]Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

[2]Departamento de Termodinámica Aplicada, Polytechnic University of Valencia, 46022, Valencia, Spain

[3]Department of Civil Engineering, School of Architecture and Engineering, Universidad Europea de Madrid, Spain

July 18, 2016

1. Introduction

Train induced noise is considered one of the main externalities of railway transport, disturbing people and environment. In this sense, the main sources of railway noise can be divided in three groups: Aerodynamic noise (generated by the interaction between the air and the elements of the vehicle); Traction noise (generated by engines, brakes or fans among others) and Rolling noise (generated by the rail-wheel contact).

Nowadays, there are many solutions to attenuate or reflect railway traffic (as green tracks, anti-vibrations systems, improved dampers or noise barriers). Some of these solutions are expensive and/or difficult to implement. Nevertheless, the use of wave barriers is widely extended due to their relation cost/effectiveness.

In general terms, noise barriers can be divided in two groups: Reflective Noise Barriers (in which the incident energy is reflected to the emission area, modifying the trajectory of the wave) and Absorptive Noise Barriers (in which the incident energy is absorbed by the filling material, at the expense of increasing noise within track area). Thus, by the optimum design of this mitigation solution, it will be possible to act over the four phenomena related to noise propagation: reflexion, absorption, transmission and diffraction. The first three will depend on the in-filling material, while the last one will depend on the wave barrier position and dimensions.

2. Case of study

The present investigation proposes the development of a new noise barrier composed by two layers (the first one made of porous concrete with damping properties to provide absorption properties, and the second one made of structural concrete to provide acoustic isolation between the source and the receiver) able to reduce railway borne noise.

Specifically, the porous concrete material is provided with carbonized olive stones. Thanks to this additive, a very porous structure is achieved. This fact ensures a huge noise absorption capacity.

3. Methodology

In order to obtain the most appropriate concrete dosage, three types of olive stones are proposed: calcined olive stones; crushed olive stones and calcined + crushed olive stones. All of them are studied by mechanical and acoustic points of view.

With regard to the mechanical behaviour, three laboratory tests are performed: compressive strength test; density test and reaction to fire test. By contrast, the acoustic behaviour is studied by the Kundt’s tube test. All the process is supported by numerical simulations, in which both mechanical and acoustic behaviours were studied in order to obtain an optimum geometrical design.

4. Test development and results

Next, the development and main results of the laboratory tests is shown.

4.1 Compressive strength test

The aim of this test is to prove the bearing capacity to resist fresh concrete during construction process. To do so, a progressive load in a cubic specimen of 15 x 15 x15 cm until the breaking point. The results of this test shown that the calcined olive stones presented resistances the largest resistances (over 7MPa), while the crushed stones presented the lowest resistances (among 3.5MPa).

4.2 Density test

The aim of this test is to know the density (apparent and real) of the new concrete in terms of acoustic absorption. Thus, the lower the density is, the higher the absorption capacity is.

Once again, the calcined stones presented the most interesting properties, since both real and apparent densities were the lowest (963 and 771 kg/m^3, respectively). By contrast, crushed stones presented the highest densities (1323 and 1240 kg/m^3, respectively).

4.3 Reaction to fire test

The aim of this test is to assess the ignitability of specimens, transmission of the flame and loss of material. To do so, a concrete specimen has to be exposed to a flame during 15 minutes, analyzing the evolution of fire and final aspect of the specimen.

With regard to the results, all the specimens avoided the flame cross. Furthermore, no specimen was burnt. Hence, all the specimens fulfil the requirements.

4.4 Kundt's tube test

The aim of this test is to calculate the spectrum of absorption coefficients for each specimen studied, as well as their critical thickness. In this sense, the *spectrum of absorption coefficient* (α) is the parameter used to determine the absorptive capacity of a material and is defined as the ratio between the absorbed sound energy and the incident sound energy, depending on the frequency. Thus, depending on its final value, it is called *pure reflective* (if $\alpha \approx 0$) or *pure absorbent* (if $\alpha \approx 1$).

To do so, the Method of the Transference Function was used according to Spanish Standards [X]. The results shown that the calcined stones presented the highest absortion coefficient (α=0.98) followed by crushed olive stones (α=0.85) and calcined + crushed olive stones (α=0.50). The critical thickness was set around 0.09m.

5. Numerical simulations

In order to decide the optimum geometry of noise barrier, 3 solutions were studied (Fig. 2). To do so, the development of both structural and acoustical models was performed. Furthermore, the process was supported by laboratory tests following the European Standard [2].

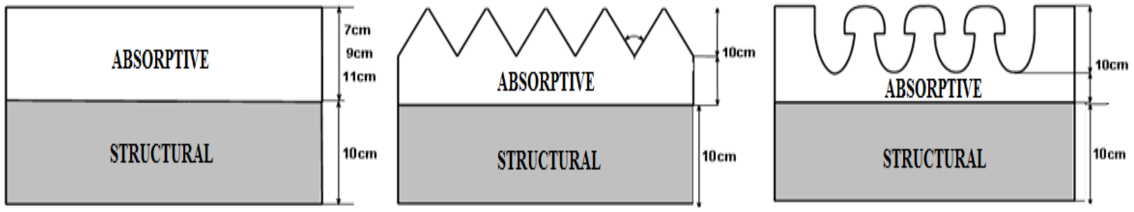


Fig. 2 Solutions proposed. Flat Module Geometry (left); Peaks Module Geometry (centre) and Fungus Module Geometry (right)

As can be noticed in Fig. 2, the absorptive layer thickness for Flat Module Geometry is studied in three different scenarios: the critical thickness obtained in Kundt's test (9cm); one over the critical thickness (11cm) and one under the critical thickness (7cm). By contrast, the Peaks Module Geometry and the Fungus Module Geometry are just studied for 10cm, since it is expected that the geometry improves the mitigation capacity.

With regard to the final dosage, all the aforementioned geometries include calcined olive stones. Nevertheless, the percentage varies. Thus, two different dosages are studied: 50 % calcined olive stones + 50 % arid 4/10 (dosage 1) and 75 % calcined olive stones + 25 % arid 4/10 (dosage 2). Hence, a total of 10 scenarios were performed.

5.1 Structural study

In order to analyze the structural behavior of the aforementioned geometrical designs, a Finite Elements (FE) model was developed in ANSYS software. To do so, static loads were applied (distinguishing between the construction phase and the in-service phase) and linear-elastic behavior of materials was adopted. The outputs of the simulations were the stresses acting on supports, the displacements of the free edge and the strains in the mid section. Three union types were studied: independent panel, smooth joint and tongue-and-groove joint (Fig. 3 left, centre and right respectively)

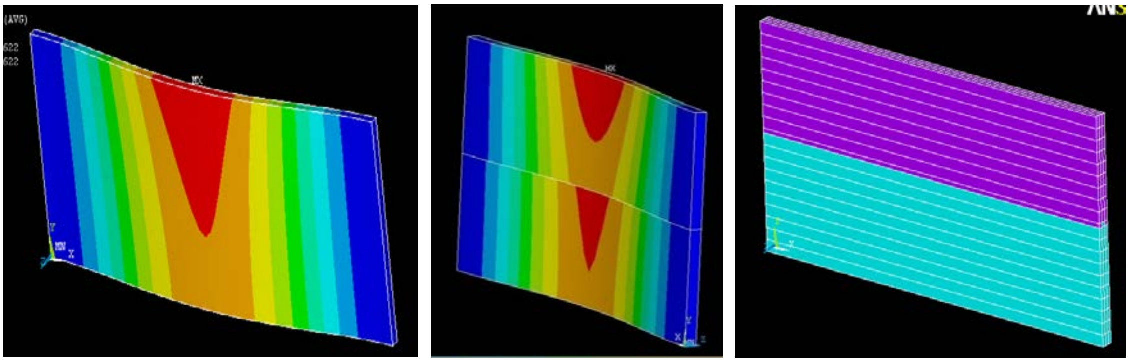


Fig. 2 Structural model simulations

According to the results, during the in-service phase all the strains and stresses were perfectly acceptable. Nevertheless, during the construction phase if a tongue-and-groove joint was selected, it should be necessary to design the panels with a minimum of 10cm thickness.

5.2 Acoustic study

The acoustic behavior of the aforementioned geometrical designs was analyzed by both analytical equations and numerical simulations (CadnaA software). To do so, the height of the elements; the relative distances to the track and 1/3 octave noise spectrums were introduced as input parameters.

With regard to the analytical model, this was developed in order to analyze the noise reduction resulting from placing a barrier between the source and receiver in different locations. Thus, the attenuation can be obtained according to the Insertion Loss Factor (Eq. 1), where *N* is the Number of Fresnel obtained according to Eq. (2), where λ is the wavelength of sound, *d* is the right distance between the source and observer and A + B is the path to save the barrier between source and observer obtained according to Fig. (3).

$$\begin{gathered} IL = 20log\frac{\sqrt{2\pi N}}{tgh\sqrt{2\pi N}} + 5 \quad if \quad N \geq -0.2 \\ IL = 0 \quad if \quad N \geq -0.2 \end{gathered} \tag{1}$$

$$N = \pm\frac{2}{\lambda}(A + B - d) \tag{2}$$

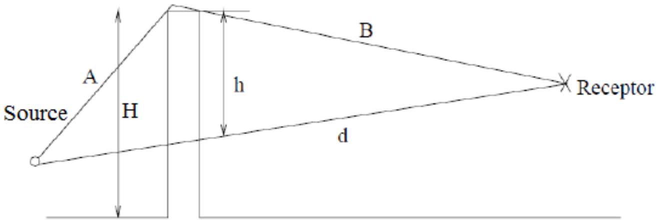


Fig. 3 Scheme representing Insertion Loss Factor parameters

According to the results, it was concluded that the most effective location would be close to the noise source, instead of close to the receptor. Nevertheless, this model is not able to study parameters as type of ground, noise absorption induced by the atmosphere or interferences with physical objects (among others). Thus, in order to solve these problems, a numerical model was developed in CadnaA commercial software. To do so, three distances between the source and the receptor were analyzed (3m, 10m and 17m), obtaining several noise maps as shown below (Figs. 4-5).

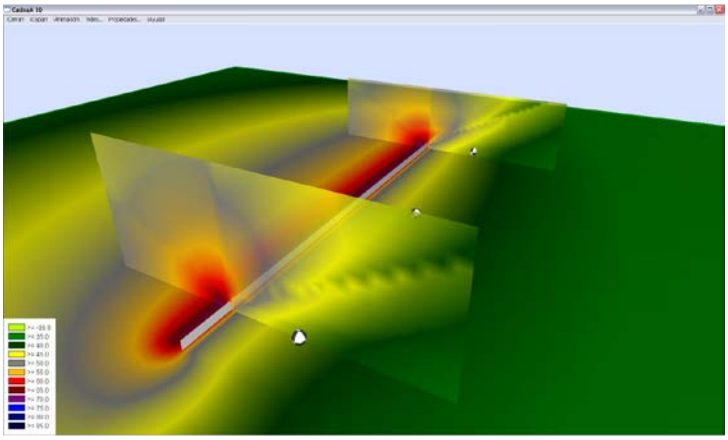


Fig. 4 Noise map obtained by CadnaA simulations

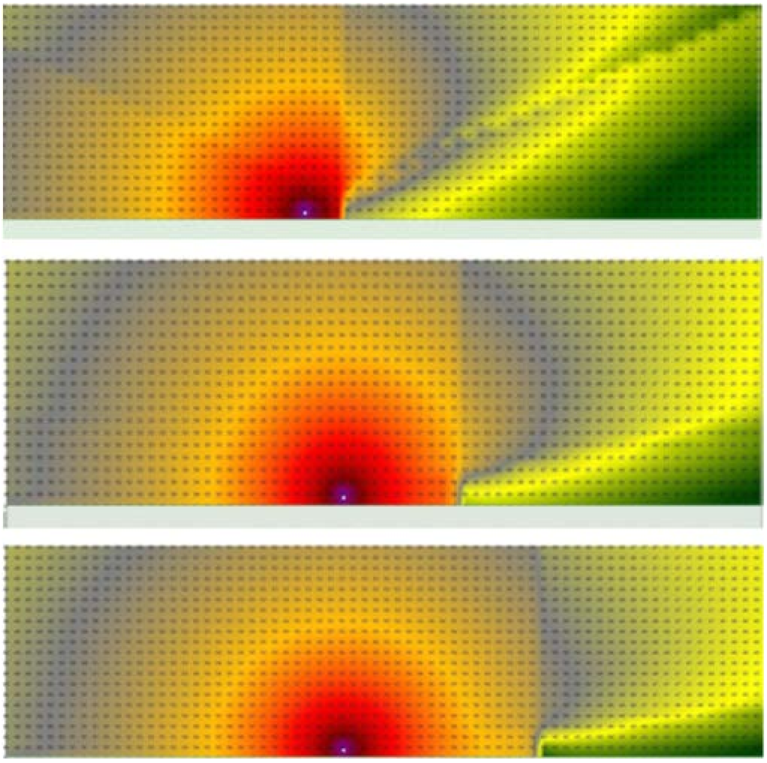


Fig. 5 Noise maps obtained by CadnaA simulations. Locations at 3m from the source (up); 10m from the source (centre) and 17m from the source (down).

According to the results, it can be stated that Noise Barriers offer greater noise attenuation when located close to the receptor, if barrier dimensions and noise spectrum are kept.

Furthermore, attending the optimum dosage studied by laboratory tests [2], it can be stated that the type A (dosage 1) has a higher absorption coefficient than the type B (dosage 2) in almost the entire spectrum, because of the greater amount of carbonized olive stones.

Finally, attending to the geometry of wave barriers, it can be stated that Flat Module Geometry of 11cm (the thickest panel) and Fungus Module Geometry (the irregular shape) are the optimum solutions for frequency bands below 630Hz.

References

[1] UNE-EN ISO 10534-2:1998. Acústica. Determinación del coeficiente de absorción acústica y de la impedancia acústica en tubos de impedancia. Parte 2: Método de la función de transferencia. Spanish Standard (in Spanish).

[2] UNE-EN ISO 354:2004. Acoustics - Measurement of sound absorption in a reverberation room. European Standard.

Influence of wheels conicity on the risk of derailment.

Silvia Morales Ivorra[1*], Fran Ribes-Llario[1], Clara Zamorano Martín[1], Julia Irene Real Herraiz[1]

[1]Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

[2]Foundation for the Research and Engineering in Railways, 160 Serrano, 28002 Madrid, Spain

*Corresponding author. E-mail: silmoiv@cam.upv.es. Telephone: +34 96 387 70 00

July 18, 2016

1. Introduction

During the last decades, the behaviour of vehicle-track system has been widely studied with the aim of optimizing the design and security of both vehicles and tracks. In this sense, vehicle-track dynamics have been considered of utmost importance (especially on curved tracks).

As can be guessed, the design of wheel and rail profile, as well as their contact, is considered a key factor in vehicle-track dynamics. For this reason, many studies have been carried out in order to optimize both rail and wheel profiles, as [1-2] among others.

A term highly related with wheel profile is the wheel conicity. This can be defined as the slope of the tread or running surface of the wheel relative to the axis of the wheelset. According to [3], high values lead to improvements in steering performance in curves, but may induce dynamic instability.

According to this, the present study analyzes the influence of wheels conicity on the loads transmitted to the track and on the risk of derailment in a curved stretch of track by means of a multi-body model.

2. Multibody model

In order to study the influence of wheel conicity on derailment risk in a curved track, a multibody model has been developed by means of the VAMPIRE Pro commercial software. The track and the vehicle have been implemented as a set of masses connected between by pairs of springs and dashpots (Table 1)

Carbody Mass (kg)	Bogie Mass (kg)	Wheelset Mass (kg)	Primary Stiffness (MN/m)	Primary Damping (MNs/m)	Secondary Stiffness (MN/m)
47600	4000	1600	1.69	5.2	3

Table 1. Main features of the vehicle

In the model, track gauge has been set to 1435 mm; the rail has been supposed to be an UIC-54 and the stiffness and damping characteristics have been detailed. Regarding the geometry of the track, a leftward curve of 2.1km radius and 100m cant has been implemented. Hence, the curve was designed for a vehicle speed of 137 km/h.

Forces in the rail-wheel contact due to train self-weight; the influence of track geometry and the conicity of wheels are obtained by means of the equation of motion (Eq. 1) where [M] is the mass matrix, [C] is the damping matrix, [K] is the stiffness matrix, *u* is the displacement vector, $\dot{u}$ is the velocity vector, $\ddot{u}$ is the acceleration vector and {F (t)} is the external forces vector

$$[M]\{\ddot{u}\} + [C]\{\dot{u}\} + [K]\{u\} = \{F(t)\} \qquad (1)$$

3. Simulations and results

The variations on vertical and lateral loads in a curved stretch are analyzed for 12 different cases of conicity: 0; 0.02; 0.025; 0.0333; 0.05; 0.1; 0.15; 0.20; 0.25; 0.3; 0.35 and 0.4 (Fig. 1) combined with 6 different vehicle speeds: 60km/h; 120km/h; 137km/h; 160km/h; 200km/h and 250km/h in four different scenarios according to Table 2. Furthermore, the influence of rail corrugation in the same stretch has been studied. In this context, the risk of vehicle derailment is calculated in terms of derailment coefficient as the quotient between lateral and vertical forces (Nadal's criterion).

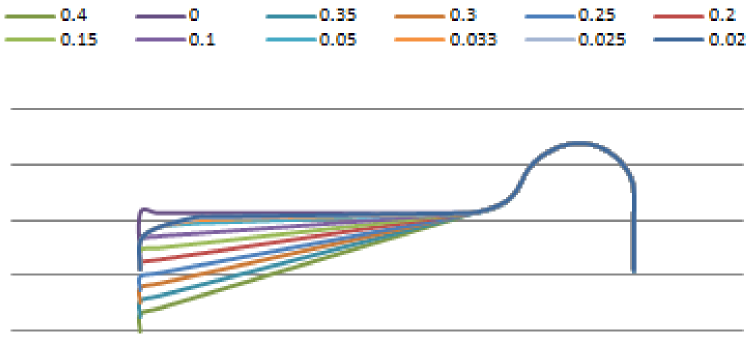


Figure 1. Wheel profile for each conicity considered

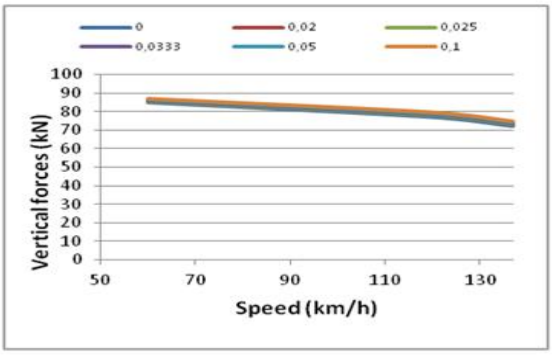
Scenario	Vehicle speed	Conicity
A	Low (60 - 137 km/h)	Low (<0.1)
B	Low (60 - 137 km/h)	High(>0.1)
C	High (137 - 250 km/h)	Low (<0.1)
D	High (137 - 250 km/h)	High (>0.1)

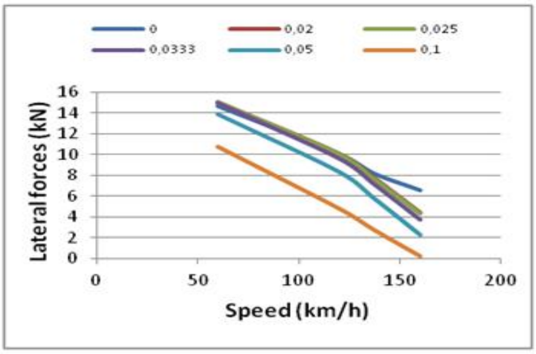
Table 2. Scenarios studied

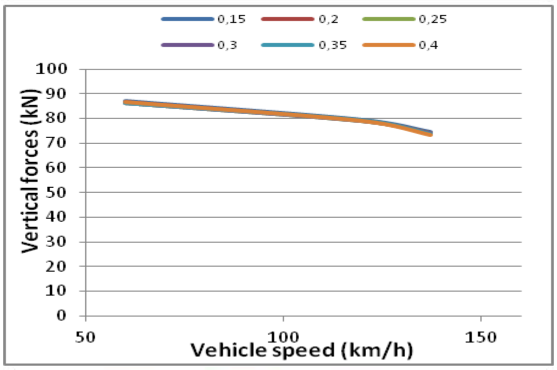
For the simmulations, the following simplifications are adopted:

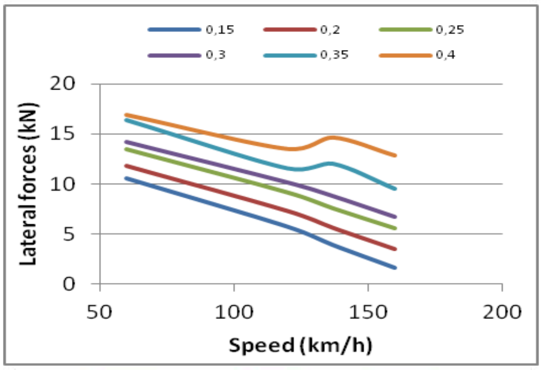
- In scenarios A and B, since vehicle speeds are lower than the critical one (this is 137km/h), the highest solicitations will appear in the low rail of the curve. Thus, results have been studied in the most unfavourable situation.
- In scenarios C and D, since vehicle speeds are higher than the critical one (this is 137km/h), the highest solicitations will appear in the high rail of the curve. Thus, results have been studied in the most unfavourable situation.

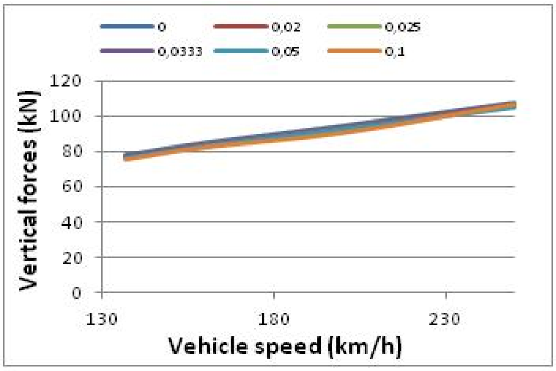
According to the aforementioned simplifications, the results are shown below

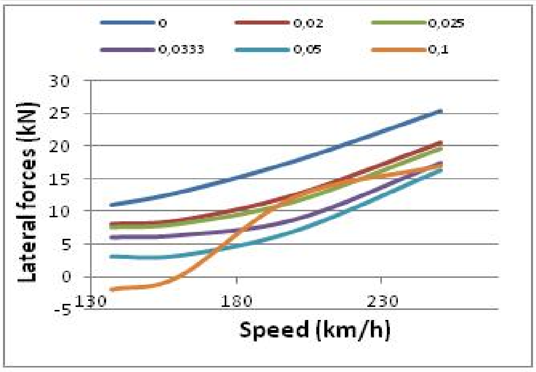


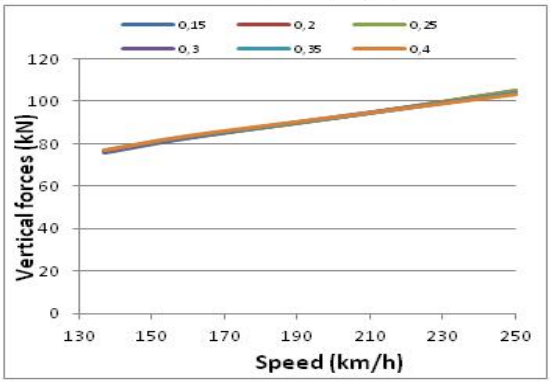












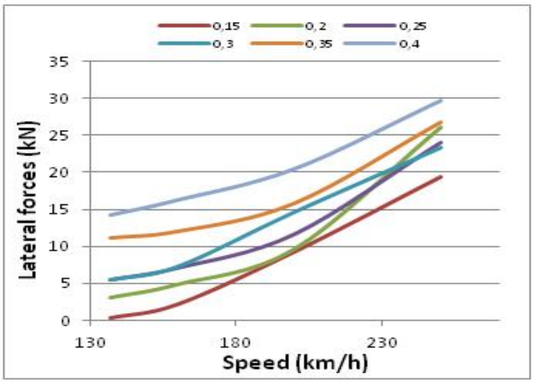


Fig. 2. Vertical (left) and lateral (right) forces in Scenarios A, B, C and D

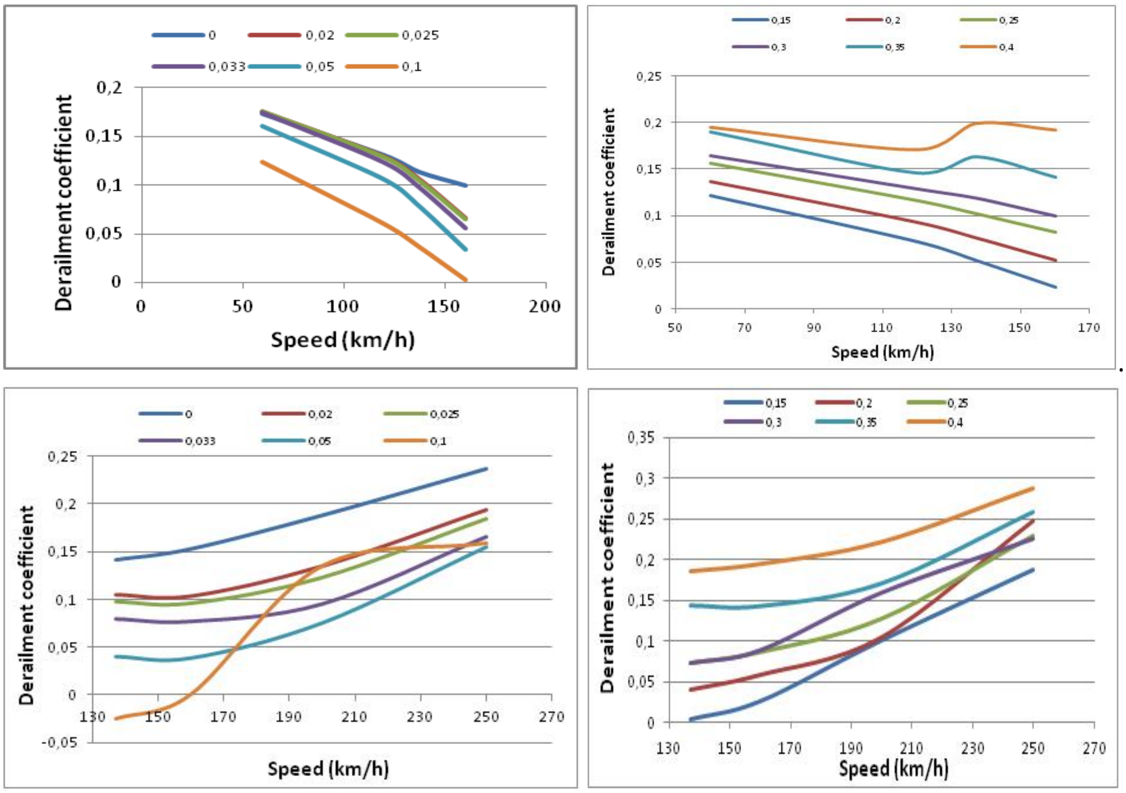


Fig. 3. Derailment coefficient in Scenarios A (up left), B (up right), C (down left) and D (down right)

3.1 Results for Scenario A

From Fig. 2, it can be noticed that vertical forces remain almost invariables both with speed and conicity variations. Regarding the lateral forces, it can be deduced that the higher the vehicle speed or, in other words, the lower the uncompensated accelerations, the lower the lateral forces in the rail-wheel contact. Furthermore, higher values of conicity compel lower lateral forces. Thus, according to Fig. 3, the lower the uncompensated accelerations and the bigger the conicity, the lower the derailment coefficient.

3.2 Results for Scenario B

From Fig. 2, it is noticeable that vertical forces are not significantly dependent on vehicle speed or conicity. Nevertheless, lateral forces present the opposite behavior: the lower the vehicle speed and the higher the conicity, the bigger the lateral forces reached in the rail-wheel contact. Regarding the conicity, it is clearly seen that wheels presenting low values of conicity induce lower lateral forces to the wheel-rail contact.

From the point of view of derailment risk (Fig. 3), derailment coefficient should follow the same trend than lateral forces, since it can be calculated as the relationship between lateral and vertical forces and vertical forces are almost invariant.

3.3 Results for Scenario C

Fig. 2, shows that vertical forces slightly increase with vehicle speed while their value is independent on conicity. Furhtermore, lateral forces significantly increase with vehicle speed and they are reduced when conicity increases. As expected, derailment coefficient follows the same trend than lateral forces (Fig. 3).

3.4 Results for Scenario D

From Fig. 2, both vertical and lateral forces rise in a significant manner, being the dependence on speed more significant in the second case. Furthermore, as presented in previous cases, conicity only is able to affect lateral forces. In this sense, the higher the conicity, the higher the lateral forces transmitted to the track.

The risk of derailment in this case grows with vehicle speed and with conicity, following the same trend as lateral forces (Fig. 3). Furthermore, it can be seen that this statement is more noticeable at lower speeds.

3.5 Results for rail corrugation

According to the four previous scenarios, it can be said that the higher the uncompensated accelerations and the further the conicity from the optimum one (0.1), the higher the risk of derailment. In this context, the amplitude of wear has been set 0.1mm and its wavelength to 0.6m, analyzing the same four scenarios:

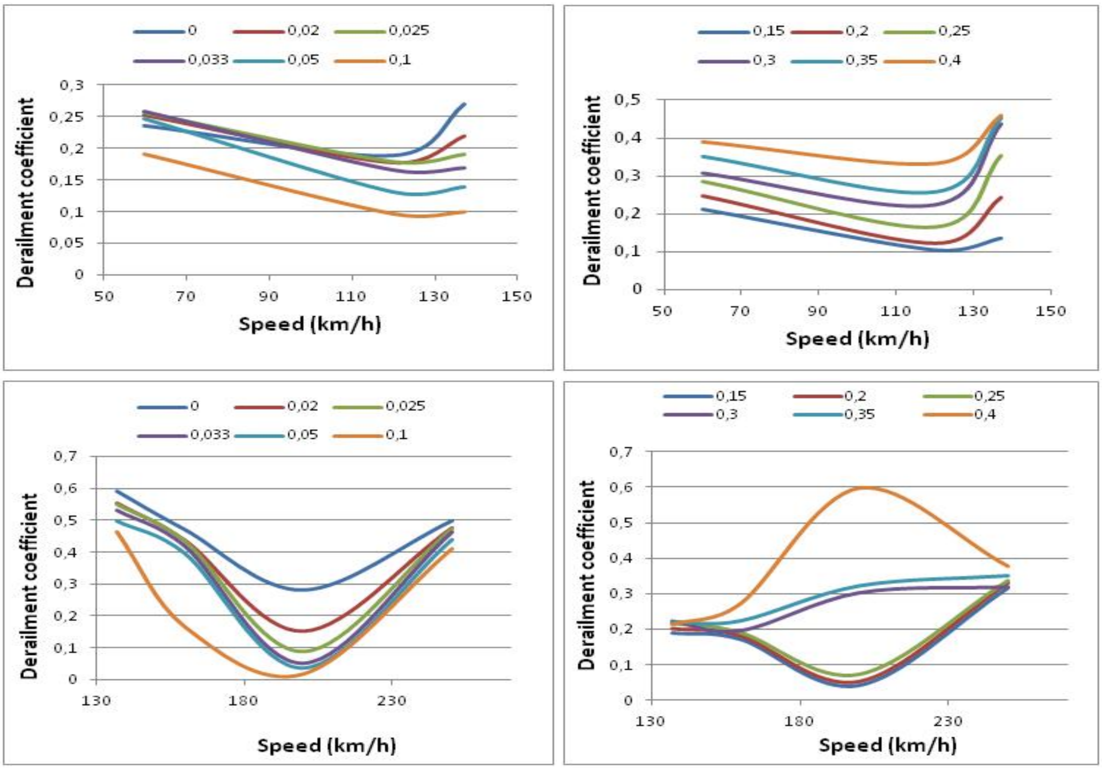


Fig. 4. Derailment coefficient in Scenarios A (up left), B (up right), C (down left) and D (down right)

As in the previous cases, derailment coefficient has been calculated in the most affected rail: low rail at low speeds and high rail at high speeds.

From the analysis performed, it can be seen that in general terms, the trend is the similar for corrugated and for healthy rails. Hence, the higher the uncompensated acceleration, the higher the derailment coefficient. Nevertheless, some irregularities in this trend may be observed for speeds close to 135 and 200km/h. These irregularities may be due to the fact that at these speeds the eigenmodes of the corrugated system are excited, leading to resonance.

Regarding to the influence of wheel conicity under these circumstances, it may be concluded that the lower the conicity for cases A and C and the higher the conicity for cases B and D, the bigger the risk of derailment.

4. Conclusions

In the present study, the influence of wheels conicity on lateral and vertical forces in the rail-wheel contact has been studied, as well as on the risk of derailment. From the analysis carried out, the following conclusions may be drawn:

- Lateral forces in the rail-wheel contact are significantly affected by wheels conicity, while changes in vertical forces are not noticeable.
- The higher the uncompensated accelerations, the higher the lateral forces in the contact as well as the risk of derailment.
- Optimum wheels conicity has been found at a value of 0.1. The further the conicity from that value, the higher the lateral forces and the risk of derailment.
- In presence of rail corrugation, the risk of derailment is increased compared to the case where rails are healthy, but the conclusions drawn for healthy rails are still valid in corrugated tracks. Thus, the higher the uncompensated acceleration and the further wheels conicity from the optimum, the higher derailment risk.

References

[1] Leary, J. F., Handal, S. N., & Rajkumar, B. (1991). Development of freight car wheel profiles—a case study. *Wear, 144*(1), 353-362.

[2] Lack, T., & Gerlici, J. Iterational method for railway wheel tread profile design.*XVIII Konferencja naukowa–Pojazdy Szynowe, Politechnika Slaska–Komitet transportu PAN. Materialy konferencyjne, 1*, 137-149.

[3] American Public Transportation Association. (2007). *Standard for definition and measurement of wheel tread taper.* APTA PRESS Task Force, 2007.

Study of attenuation capacity of new rubberized concrete in railway applications.

Antonio José Pérez[1*], Jesús Herminio Alcañiz [2], Teresa Real[3], Julia Irene Real Herraiz[3]

[1] CHM, Avd. Jean Claude Combaldieu, s/n, 03008 Alicante, Spain

[2] Laboratory of Construction, Universidad Católica de Murcia, Murcia, Spain

[3]University Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

*Corresponding author. E-mail: ajperez@chm.es. Telephone: +34 965 145 205.

November 30, 2016

1. Introduction

Along last decades, the importance of railway transport has sharply increased. Nevertheless, the transmission of vibrations induced by railways through the surrounding soil is still an issue that needs to be addressed. In this sense, mitigation measures (as wave barriers and wave impeding blocks, WIB) are ones of the most interesting solutions to this problem.

Specifically, wave impeding blocks attenuation mechanism consists on the stiffening of the natural soil layers. Thus, the propagation regime of the soft soil can be modified. This attenuation capacity could be even more increased by adding synthetic fibers to concrete. Thus, the micro-cracks appeared on the interface between fibers and the cementing matrix allows infinitesimal movements during the vibrational periods. The cycles of opening-closing of the micro cracks are transformed in a loss of energy [1].

According to this, the present investigation develops a Finite Elements (FE) model in order to simulate the pass of a train through a track, being the validity of the model assessed by real data. The FE model is then used to quantify the vibrations attenuation capacity of different types of wave barriers composed by rubberized concrete, comparing their effectiveness depending on their position within the infrastructure and their thickness. Analyses are performed both in the time and the frequency domains.

2. Case of study

In order to assess the validity of numerical simulations, a data gathering campaign was carried out in a real track placed near Barcelona (Spain). It was a ballasted track with UIC-54 rails, elastic railpads and concrete sleepers. The whole track was built over a vast layer of clays.

During the campaign, the vibrations generated by a passenger train in different points of the track were obtained by means of SEQUOIA FAST TRACER accelerometers placed in two critical points of the track: the rail web and the sleeper surface. Then, these measures were divided in two sets of data: one set used to calibrate the model and to estimate the unknown parameters of the track, and other set used to evaluate the proper performance of the model.

3. Methodology

In order to reproduce the influence of the pass-by of trains on the surrounding areas, a three-dimensional vehicle-track model is developed using commercial software ANSYS LS-DYNA V17. The model consists of two sub-models that represent the track and the vehicle.

3.1 Track sub-model

The track, the soil and the WIB are represented as a mesh of hexahedral elements whose maximum dimensions depend on the maximum wavelength transmitted [2]. In each node of the mesh, Lagrange equation is solved (Eq. 1), where [M] is the mass matrix, [C] the damping matrix, [K] the stiffness matrix, {u} the displacement vector, {u̇} the velocity vector, {ü} the acceleration vector and {F (t)} the external forces vector:

$$[M]\{\ddot{u}\} + [C]\{\dot{u}\} + [K]\{u\} = \{F(t)\} \qquad (1)$$

Since the effect of the train does not induce large strains in the soil, displacements on the system are limited to the elastic range in the stress-strain diagram. Thus, materials' behavior has been assumed linear elastic.

3.2 Multi-body vehicle model

The vehicle has been reduced to a car body, two bogies and four axles, which have been modeled as a three-dimensional multi-body system. Thus, the vehicle's equation of motion can be reduced to Eq. (2), where $\ddot{x}_i$, $\dot{x}_i$ and x_i respectively represent the accelerations, velocities and displacements of the i element denoted by the subscript c (car body), b (bogie), u (unsprung mass) or r (rail). Furthermore, g represents the gravity acceleration and F_C represents the wheel-rail contact forces.

$$\begin{bmatrix} M_c & 0 & 0 & 0 \\ 0 & M_b & 0 & 0 \\ 0 & 0 & M_u & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \ddot{x}_c \\ \ddot{x}_b \\ \ddot{x}_u \\ \ddot{u}_C \end{bmatrix} + \begin{bmatrix} c_s & -c_s & 0 & 0 \\ -c_s & c_p + c_s & -c_p & 0 \\ 0 & -c_p & c_p & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_c \\ \dot{x}_b \\ \dot{x}_u \\ \dot{u}_c \end{bmatrix} +$$

$$\begin{bmatrix} k_s & -k_s & 0 & 0 \\ -k_s & k_p + k_s & -k_p & 0 \\ 0 & -k_p & k_p + k_H & -k_H \\ 0 & 0 & -k_H & k_H \end{bmatrix} \begin{bmatrix} x_c \\ x_b \\ x_u \\ u_c \end{bmatrix} == \begin{bmatrix} M_c \\ M_b \\ M_u \\ 0 \end{bmatrix} g + \begin{bmatrix} 0 \\ 0 \\ 0 \\ F_C \end{bmatrix} \qquad (2)$$

For Finite Element representation, 8-nodes hexahedral elements were selected for both car body and bogies, while point elements were selected for wheels. Meanwhile, springs and dampers were selected to represent primary and secondary suspension. Finally, the wheel/rail contact is modeled by means of a Hertzian spring.

3.3 Vehicle-track interaction

In the multi-body system representing the vehicle, the wheel–rail interaction is modeled as a node-to-beam contact allowing for sliding and loss of contact, using the Penalty algorithm. The contact elements provide an elastic support between the rail and the wheel simulating the hertzian contact.

Vehicle submodel has also been solved by the equation of motion. To solve the non-linear equations of the problem, full Newton–Raphson method has been used, while Newmark implicit time integration method has been used to solve the transient dynamic equilibrium equations.

4. Simulations and results

In the current section, the attenuation capacity of two different configurations of WIBs is analyzed. In both cases, the impeding barrier is made by a 4.7 m wide and 0.3m thick concrete slab, placed directly under the ballast layer in Case 1 and immediately above the clays layer (at 0.7 m from the bottom of the ballast layer) in Case 2 (Fig. 1).

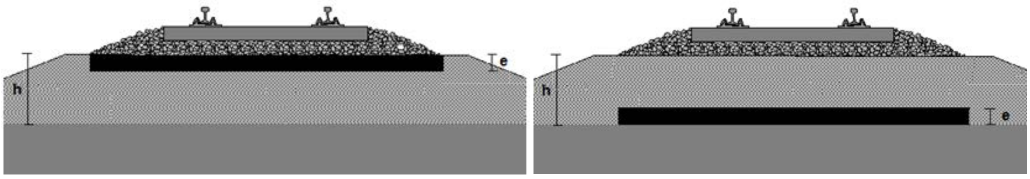


Fig. 1 Sketch of the cross section of the track. Case 1 (left) and Case 2 (right)

This analysis is carried out in terms of wave accelerations, evaluating vibrations in four critical points: At the rail web, on the surface of a sleeper, and in two points of the ground, set at a distance of the track of 2 m, and 4 m, respectively.

4.1 Case 1 (time domain)

The attenuation capacity of the proposed wave impeding block is assessed by comparing the vibrations generated with and without any mitigation measure in four critical points: at the rail web, on the surface of a sleeper in one point of the ground set at 2 m from the track and in one point of the ground set at 4 m from the track.

Fig. 2 shows how the presence of the WIB does not significantly affect vibrations on the rail, while vibrations at the sleeper are slightly reduced. Meanwhile, the isolation effect of the wave impeding block on the ground is highly remarkable, especially at short distances, where vibrations are almost a half when the WIB is placed.

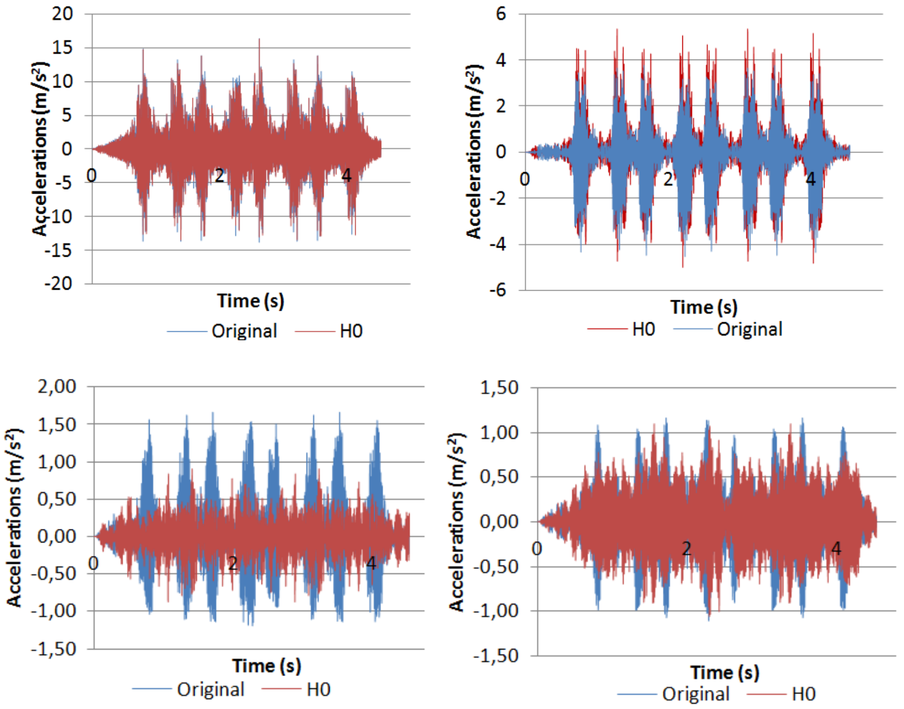


Fig. 2 Comparison between vertical accelerations at the rail web (up left), the sleeper surface (up right), the ground at 2 m (down left) and the ground at 4 m (down right) in Case 1

4.2 Case 2 (time domain)

The effect of case 2 on the track vibrations is shown in Fig. 3. Thus, the rail and the sleeper show no influence of the WIB in terms of vibrations isolation. However, it can be observed how acceleration levels on the ground are increased, in spite of being reduced, by the effect of the WIB, especially far from the track. This result, which opposes with the main purpose of the wave impeding blocks, is in accordance with the research carried out by [3], who stated that when the depth at which the WIB is embedded is larger than a threshold one, train induced vibrations may be amplified.

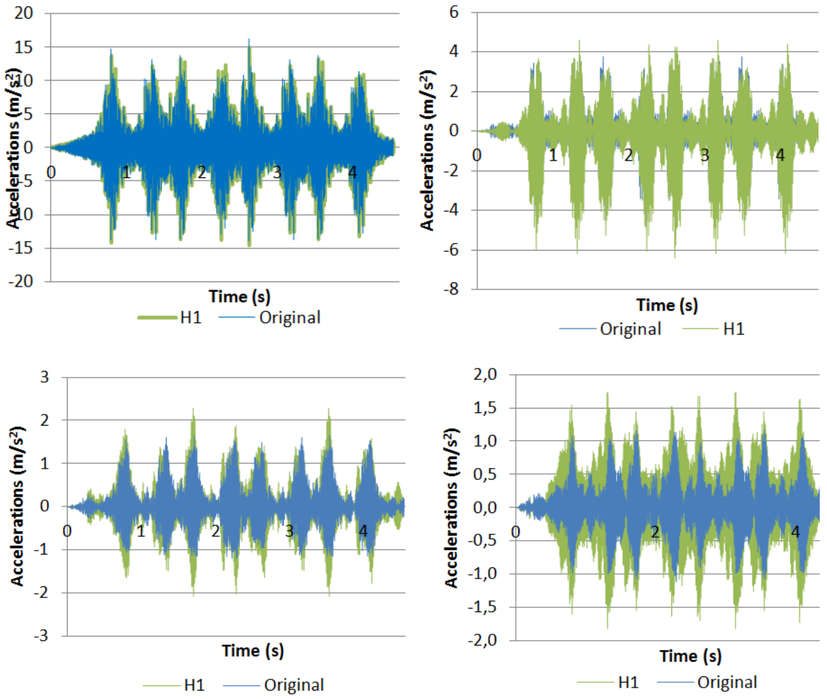


Fig. 3 Comparison between vertical accelerations at the rail web (up left), the sleeper surface (up right), the ground at 2 m (down left) and the ground at 4 m (down right) in Case 2.

4.3 Comparison between cases 1 and 2 (frequency domain)

A spectral comparison has been performed between both cases in the frequency domain, in four critical points (Fig. 4). In this picture, H0 and H1 represent the depth at which the wave impeding block is buried in Cases 1 and 2, respectively.

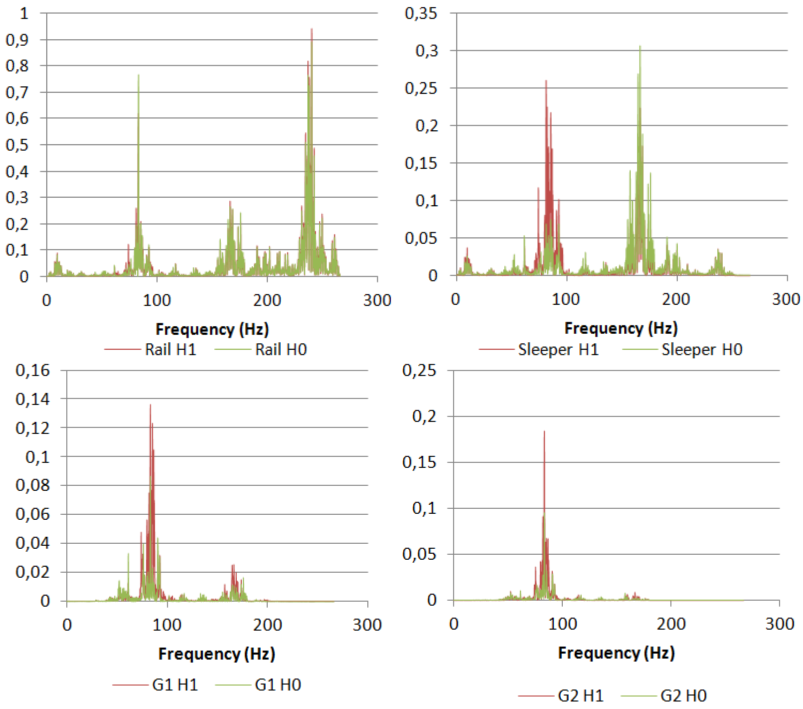


Figure 9: Comparison between Case 1 (green) and Case 2 (red) in a rail (up left), a sleeper (up right); on the ground at a distance of 2 m (down left) and 4 m (down right)

In general terms, the results shown the same trend than those obtained in the time domain. Thus, attenuation capacity of both dispositions does not present significant differences when comparing rail registers. Regarding results obtained over a sleeper, it can be affirmed that low frequencies vibrations are highly mitigated in Case 1. In contrast, results for medium frequency accelerations present the opposite behavior. Meanwhile, high-frequency vibrations are efficiently reduced in both cases.

Furthermore, if vibration levels in the ground at two different distances from the track are compared, the wave impeding block placed at H0 is clearly more efficient than that buried at a depth H1. Moreover, vibrations transmitted from the sleeper to the ground are mainly close to 100 Hz, and vibrations mitigated by the effect of the ground are predominantly of mid and high frequencies.

4.4 Optimum thickness of the WIB

Three WIB thicknesses have been simulated: 0.1, 0.3 and 0.5 meters (all of them underneath the superstructure of the track). The frequency and time domain analysis have been repeated concluding that the thicker the wave impeding block, the larger its shielding capacity is. However, there is a threshold value at which the influence of slab thickness decays, leading to small differences on the mitigation capacity of different slabs. Since the thicker the slab, the higher the construction costs, these costs can be optimized finding that threshold value which, in this case, has been set to e=0.3m.

5. Conclusions

According to the simulations performed, it has been concluded that impeding blocks placed just underneath the ballast layer are efficient tools to reduce vibrations on the surrounding ground. Nevertheless, if a WIB is buried below a threshold depth, vibration levels may be amplified rather than reduced. Anyhow, WIBs buried at a shallow depth present a better isolation capacity of low-frequency waves. Meanwhile, it has been concluded that the thicker the slab, the higher its mitigation capacity is. However, there is a threshold value at which slab thickness becomes less significant.

References

[1] Yan, L., Jenkins, C. H., & Pendleton, R. L. (2000). Polyolefin fiber-reinforced concrete composites: Part I. Damping and frequency characteristics. *Cement and concrete research*, *30*(3), 391-401.

[2] Real Herráiz, J. I., Zamorano, C., Hernandez, C., Comendador, R., & Real, T. (2014). Computational considerations of 3-D finite element method models of railway vibration prediction in ballasted tracks. In Journal of Vibroengineering (Vol. 16, No. 4, pp. 1709-1722).

[3] Gao, G., Li, N., & Gu, X. (2015). Field experiment and numerical study on active vibration isolation by horizontal. *Soil Dynamics and Earthquake Engineering*, *69*, 251-261.

Influence of the friction coefficient on the squeal noise frequency appearing using FE models.

Julia Irene Real Herraiz[1], Silvia Morales Ivorra[1], Clara Zamorano Martín[2], Teresa Real Herraiz[1]

[1]Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

[2]Foundation for the Research and Engineering in Railways, 160 Serrano, 28002 Madrid, Spain

*Corresponding author. E-mail: jureaher@tra.upv.es. Telephone: +34 96 387 70 00.

July 18, 2016

1. Introduction

In the field of transport engineering, the phenomenon of railway noise has become one of the most worrying problems in recent times.

Specifically, the *squealing noise* is one of the most annoying sources of noise generation. This takes place in the wheel-rail contact when the vehicle is travelling on a curve by the displacement produced between both surfaces, and generally grows in inverse proportion to the radius of curvature. Thus, this phenomenon is usually shown in small radii curves (<200 meters) but it is unusual on curves over 500m radius [1]. The frequency range associated to this phenomenon is set between 600 and 10000Hz [2] and generally matches with the eigenfrequencies of wheels. Thus, the vibrations generated are manifested as structural oscillations of the wheel and rail that derive in high noise intensities.

Friction coefficient is highly related with this phenomenon, since depending on its value, the friction forces generated in the wheel-rail contact will growth or drop affecting to the wheel-rail displacements. Thus, the present investigation analyzes the influence of friction coefficient on the squealing noise generation through numerical simulations in frequency domain of complex eigenvalues. This study is involved in a R+D project which purpose is to reduce the rolling and squealing noise.

2. Methodology

In order to analyze the complex eigenvalues, two possibilities may be adopted: analytical methods (for simple geometries) or finite element methods (for complex geometries).

This analysis is most convenient when solution is close to the sliding state, since it is a linear approximation. However, if the frictional forces are coupled with two degrees of freedom, the model can be unstable (even with constant friction coefficients). Hence, in order to have a better knowledge of the different behavior of the wheel-rail system and its influence on the squeal noise, a Finite Element (FE) model has been developed in ANSYS software. This tool allows obtaining the values of the frictional coefficients at which squeal noise appears.

The settlements of this problem require the approach of an asymmetric stiffness matrix (Eq. 1) in which k_c is the stiffness of the wheel-rail contact, f_i are the frictional forces, n_i are normal forces to both surfaces and m is the friction coefficient considered.

$$\begin{Bmatrix} f_1 \\ n_1 \\ f_2 \\ n_2 \end{Bmatrix} = \begin{bmatrix} 0 & mk_c & 0 & -mk_c \\ 0 & k_c & 0 & -k_c \\ 0 & -mk_c & 0 & mk_c \\ 0 & -k_c & 0 & k_c \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{Bmatrix} \tag{1}$$

If this asymmetric matrix is solved, complex eigenvalues are generated. The real part indicates the existence of stability (if negative) or instability (if positive), and the imaginary part indicates the frequency at which squeal occurs. Thus, the existence of squealing noise can be detected when the imaginary part of the eigenvalues is positive.

3. Simulations and results

For the analysis of complex eigenvalues, a modal analysis of the structure is required. Thus, the equation to solve is Eq. (2), where [M] and [K] are the mass and stiffness matrices and {ü}, {u} and {F} are the accelerations, displacements and friction forces vectors. By modal analysis principle, no damping is considered.

$$[\mathrm{M}]\{\ddot{\mathrm{u}}\} + [\mathrm{K}]\{\mathrm{u}\} = \{\mathrm{F}\} \tag{2}$$

For the simulations, the pair's nodes at the wheel-rail contact are matched for coupling the movements of the wheel and the rail, and the normal and frictional forces are inputs. In this study, the sliding forces are assumed to be saturated and its value is equal to the frictional force (Eq. 3), where F_r and F_c are the lateral forces acting over the wheel and the rail respectively. Thus, the forces are proportional to normal forces (N) according to the frictional coefficient (μ) in the wheel-rail contact.

$$F_r = -F_c = \mu \cdot N \tag{3}$$

Normal force is simulated using springs and it is expressed by Eq. (4), where k_c is the contact stiffness and u_{ry} and u_{cy} are the vertical displacements of corresponding contact nodes.

$$N = -k_c \cdot (u_{ry} - u_{cy}) \tag{4}$$

If Eqs. (3, 4) are combined, then Eq. (5) is obtained (which can be rewritten as Eq. (6) for all frictional forces $\{F_f\}$ in the contact nodes being $[K_f]$ the friction matrix).

$$\begin{Bmatrix} F_r \\ F_c \end{Bmatrix} = \mu\, k_c \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{Bmatrix} u_{ry} \\ u_{cy} \end{Bmatrix} \tag{5}$$

$$\{F_f\} = [K_f] \cdot \{u\} \tag{6}$$

Replacing the forces vector in the main equation, the problem becomes as Eq. (7). The solution of the second order differential equation is Eq. (8), where s is the eigenvalue and $\{\phi\}$ is the eigenvector of the equation.

$$[\mathrm{M}]\{\ddot{\mathrm{u}}\} + [\mathrm{K} - K_f]\{\mathrm{u}\} = 0 \tag{7}$$

$$\{u\} = \{\phi\}\, e^{st} \tag{8}$$

If Eqs. (7, 8) are combined, the problem can be solved according to Eq. (9). Hence, the solution is a complex eigenvalue problem which might be expressed as Eq. (10) where α_i is the real part and ω_i is the imaginary part of the solution.

$$\left([M]\, S^2 + [K - K_f]\right)\{\phi\} = 0 \tag{9}$$

$$S = \alpha_i \pm j\, \omega_i \tag{10}$$

The numerical method chosen to solve Eq. (9) in the ANSYS model is the *Unsymetric* method. For the model's development, MATRIX27 elements have been chosen for coupling wheel-rail movements and consider the effect of friction in the contact zone. For the wheel and rail mesh, SOLID185 elements have been set. Finally, in order to simulate the springs located under the rail, COMBBIN14 elements have been set. The final geometry and mesh is shown below (Fig. 1).

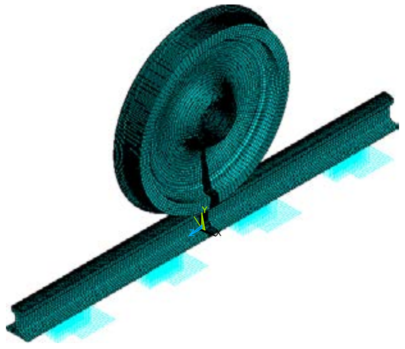


Fig. 1. General model.

Once the FE model is developed, it is calibrated using squealing noise measurements from a real data gathering campaign (in which a friction coefficient of μ=0.9 was registered). After the model calibration, this is used for studying the relation between the friction coefficient and the frequency at which squeal noise appears in three scenarios: an straight track with no cant (scenario A); a curved track with 55mm cant(scenario B) and a curved track with 110mm cant (scenario C). All the aforementioned scenarios include 1/20 tilt at rails.

Thus, as previously explained, the existence of squealing noise can be detected when the imaginary part of the eigenvalues is positive (first way). Anyhow, the existence of squealing noise can also be detected if two different vibration frequencies match (second way). Next, the results of the aforementioned simulations are shown.

3.1 Results for Scenario A

First of all, the cross section of wheel-rail contact on both rails is shown. In the case of high rail (outer wheel to the curve), the contact position may be located between the flange and the lateral edge of the rail (Fig. 1 centre), or on the railhead (Fig. 1 right). In the case of low rail (inner wheel to the curve), the contact position is set between the rail head and the central band of the wheel (Fig. 1 left).

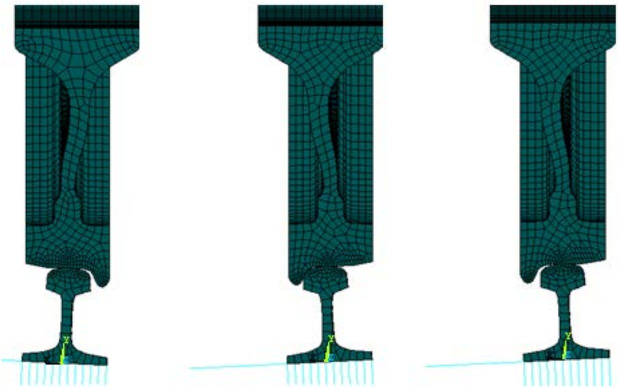


Fig. 2. Cross section on the low rail head contact (left); on the high rail flange contact (centre) and on the high rail head contact (right).

With regard to the first way to detect the presence of squealing noise, the results of simulations are shown below.

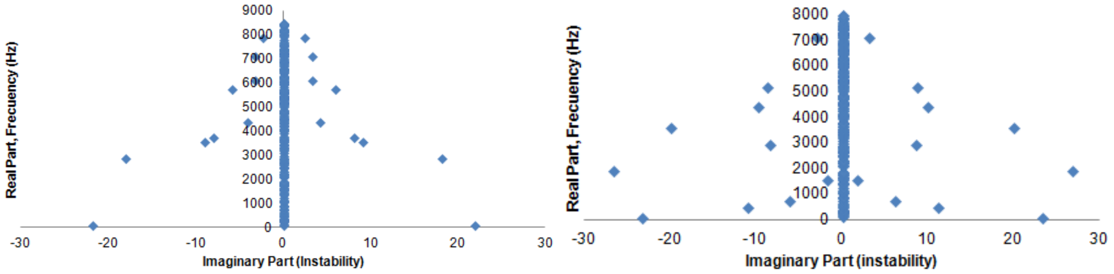


Fig. 3. Real part versus imaginary part. Results of low rail eigenmodes (left) and high rail eigenmodes (right) obtained with μ = 0.9 in scenario A.

With regard to the second way to detect the presence of squealing noise, some results of simulations are shown below.

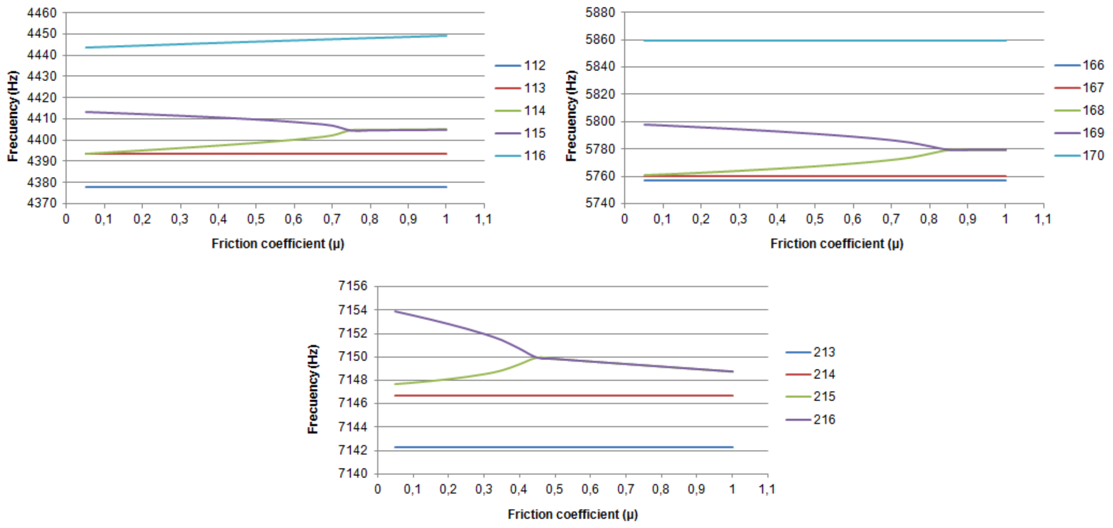


Fig. 4. Frequency versus friction coefficient. Results of low rail eigenmodes with a vibration frequency close to 4400Hz (left), 5800 Hz (right) and 7150 Hz (centre) in Scenario A.

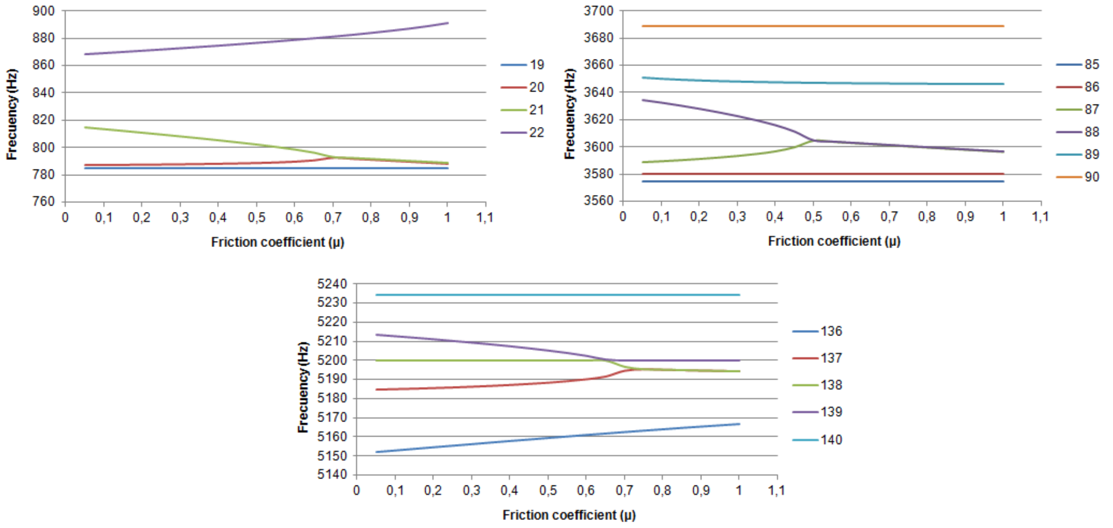


Fig. 5. Frequency versus friction coefficient. Results of high rail eigenmodes with a vibration frequency close to 800Hz (left), 3600Hz (right) and 5200Hz (centre) in Scenario A.

According to the results, Fig. 4 (low rail) shows that squealing noise phenomena take place at 4400 Hz with μ=0.75, at 5780 Hz with μ=0.85 and at 7150 Hz with μ=0.45. Meanwhile, Fig. 5 (high rail, contact zone between the flange and the lateral edge of the rail) shows that squealing noise phenomena take place at 790 Hz with μ=0.7, at 3600 Hz with μ=0.5, and at 5200 Hz with μ=0.75. The rest of results are further explained on Table 1.

3.2 Results for Scenarios B and C

Next, the complete results of simulations are summarized.

		Scenario A		Scenario B		Scenario C	
	1/3 Octave Band (Hz)	µ	Frecuency (Hz)	µ	Frecuency (Hz)	µ	Frecuency (Hz)
Low Rail. Railhead Contact	2000			0.2	1947	0.2	1948
	4000	0.75	4404	0.2	4393	0.25	4392
	6300	0.85	5779			0.5	6175
	6300			0.45	6540	0.45	6539
	8000	0.45	7149			0.55	7164
High Rail. Railhead Contact	4000			0.35	3797	0.3	3797
	4000	0.7	4406	0.85	4401		
	5000	0.45	5190				
	6300					0.35	6795
	8000			1	7155	0.8	7154
High Rail. Wheel Flange Contact	800	0.7	793	0.65	793	0.55	795
	4000	0.5	3604	0.45	3603	0.45	3602
	4000	0.6	4417	0.5	4429	0.45	4428
	5000	0.75	5195	0.75	5196	0.7	5196
	6300			0.85	6168	0.8	6168
	8000	0.9	7153	0.5	8151		

Table 1. Frequencies at which squealing noise takes place and their friction coefficient.

4. Conclusions

The present investigation develops a FE model able to reproduce the squealing noise phenomenon. This model is calibrated and validated with experimental measurements registered during a data gathering campaign of a railway track curve with friction coefficient µ=0.9. According to the results, it is demonstrated that the squealing noise generation can be controlled varying friction coefficient.

References

[1] Thompson, D. (2008). *Railway noise and vibration: mechanisms, modelling and means of control*. Elsevier.

[2] Hsu, S. S., Huang, Z., Iwnicki, S. D., Thompson, D. J., Jones, C. J., Xie, G., & Allen, P. D. (2007). Experimental and theoretical investigation of railway wheel squeal. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *221*(1), 59-73.

Probabilistic analysis of induced stresses and vibrations of a short-span bridge considering random irregularities

Fran Ribes-Llario[1], Carlos Miñana[1], Francisco José Toledo[1], Julia Real Herraiz[1]

[1]Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia,
46022, Valencia, Spain

*Corresponding author. E-mail: frarilla@cam.upv.es. Telephone: +34 96 387 70 00.

July 18, 2016

1. Introduction

Resonance is a well known phenomenon which occurs when the loading frequencies coincide with the natural frequencies of the bridges or the trains, and it is able to induce severe damages in a structure, in extreme cases, to its collapse. For this reason, many investigations have recently been focused on characterizing such an important phenomenon. The main aim of these studies is, firstly, to take measures in order to avoid the excitation of the eigenfrequencies of the system and, secondly, to minimize the effect of resonance when it occurs.

As stated by Xia et al. (2006) [1], the resonance of train-bridge systems is influenced by many factors, such as the periodically loading on the bridge induced by the moving load series formed by the wheel-axle weights of the train vehicles, the harmonic forces caused by irregularities, and the periodical actions on the moving vehicles of long bridges with identical spans and their deflections.

Within this context, train-structure interaction models have been implemented so as to deepen on vehicle and bridge responses. As an example, Wang et al. (2010) [2] simulated a two-span continuous beam and obtained the two critical velocities causing the resonance response. By means of numerical models, Lu et al. (2013)[3] studied the influence of the bridge-to-carriage length ratio and Yang & Lin (2005)[4] demonstrated that the primary frequencies in the bridge response might be caused by the driving frequencies, which are related to the time the train spent crossing the bridge, and the dominant frequencies, caused by the repeated loads.

According to Mao & Lu (2013)[5] the bridge response under a moving train is quite complicated since the excitation does not only involve characteristics from a moving load but also repeated load pulses from consecutive axles, bogies and carriages. Furthermore, authors as Yau (2001)[6] and Kwark et al. (2004)[7] studied the resonance of continuous bridges due to moving trains.

Since models based on finite elements method have been found to be a useful tool to reproduce the behavior of the track and the structure (Ju & Lin(2003)[8]), the influence of numerous parameters make it difficult to determine when the resonance occurs. In the present investigation a probabilistic methodology is proposed to obtain the induced accelerations in a short span railway bridge considering random irregularities

2. Methodology

A probabilistic methodology is proposed to obtain the induced accelerations in a short span railway bridge considering random irregularities. The purpose is to create an efficient and automatic procedure, which allows identifying the maximum accelerations induced at different points of the bridge and the maximums stress on different scenarios. The procedure was assessed through the Monte-Carlo method implemented in Ansys.

The Key functionality of Monte-Carlo simulation [9] techniques is the generation of random numbers with a uniform distribution from 0 to 1. The interpretation of the results of a Monte-Carlo simulation analysis is based on statistical methods. For the cumulative distribution function (CDF) the data is sorted in ascending order and the CDF of the ith data point, here denoted with F_i, can be derived from:

$$\sum_{k=i}^{N} \frac{N!}{(N-k)!\,k!} F_i^k (1-F_i)^{N-k} = 50\%$$

This equation is solved numerically for F_i.

3. Simulations and results

The structure-track-vehicle system has been implemented by means of a finite elements model, developed with the commercial software ANSYS LS-DYNA V17. It has been divided in two submodels: the structure-track model and the vehicle model. In the present section, both submodels will be presented, as well as the interaction between them.

First of all, the geometry of structure-track submodel has been implemented, reproducing a real stretch placed near Xativa (Spain). The track, which lies on a bridge, is provided with UIC 45 rails, wooden sleepers and a ballast layer. Further details on the real track will be presented below.

The dynamic response of the submodel is calculated by relating the internal forces of the system to the external forces. This relationship can be written by means of the equation of motion:

$$[\mathrm{M}]\{\ddot{\mathrm{u}}\} + [\mathrm{C}]\{\dot{\mathrm{u}}\} + [\mathrm{K}]\{\mathrm{u}\} = \{\mathrm{F}^{\mathrm{a}}(\mathrm{t})\}$$

Regarding the external forces, they are given by the vehicle submodel. Thus, displacements, velocities and accelerations can be calculated for each node of the model solving the equation above.

The frequency range studied with the model varies between 2 and 100 Hz. In addition, the frequency range limits determine the model dimensions and also the model elements size.

It must be pointed that material behavior is assumed to be linear elastic. This hypothesis is assumed because it has been previously checked that the dynamic wave produced by the moving train does not induce large strains in the soil in this case. Consequently, the displacements are limited to the elastic range in the stress-strain diagram.

Regarding the vehicle sub model, it has been modeled considering a three-mass system, accounting the wheel, bogie and the carbody masses. The masses are linked by springs and dampers in parallel which simulate the primary and secondary suspensions

In the three-masses system representing the vehicle, the wheel–rail interaction is modeled as a node-to-beam contact allowing for sliding and loss of contact, using the Penalty algorithm. The contact elements provide an elastic support between the rail and the wheel simulating the hertzian contact.

Vehicle submodel has also been solved by the equation of motion. To solve the non-linear equations of the problem, full Newton–Raphson method has been used, while Newmark implicit time integration method has been used to solve the transient dynamic equilibrium equations.

Track irregularities were randomly generated using power spectral density functions. In the present study the power spectral density function proposed by SNCF was selected. This function is determined by:

$$\mathrm{G}(\Omega) = \frac{10^{-6}\mathrm{A}}{\left(1 + \frac{\Omega}{\Omega_{\mathrm{r}}}\right)^{3}}$$

Where A is a parameter that depends on the track quality and varying between 160 or 550, for a good or bad quality tracks, respectively, and Ω_{r} is the reference frequency and takes the value of 0.307 m-1.

The numerical process to generate track irregularities is given by:

$$\mathrm{r}(\mathrm{x}) = \sqrt{2}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{A}_{\mathrm{i}} \cos(\Omega_{\mathrm{i}}\mathrm{x} - \theta_{\mathrm{i}})$$

Where Ω_i is the distance frequency within the wavelength range, $A_i$corresponds to the amplitude and θ_i is the independent random phase angle that is uniformly distributed in the range between 0 and 2π.

The distance frequency, , corresponds to:

$$\Omega = \frac{2\pi}{\lambda}$$

Where λ is the wavelength of the irregularity. Whereas the amplitude, Ai, can be determined through the PSD function, G (X), as follows:

$$A_i = \sqrt{\frac{1}{2\pi} G(\Omega_i)\Delta\Omega_i}$$

Where $\Delta\Omega_i$ is the frequency increment.

Usually the frequency increment is defined by establishing the upper and lower limits of the distance frequency range, $\Omega_{max}$and Ω_{min}, respectively, and selecting an adequate number of discrete frequencies:

$$\Delta\Omega = \frac{\Omega_{max} - \Omega_{min}}{N}$$

The aim of this study is to asses numerically the influence of track irregularities on the induced vibrations and the stresses on different points of the bridge.

3.1 Simulation A

The proposed methodology intends to be appealing and simple to use. It should be noted that it aims to assess the maximum accelerations of the bridge at a certain train speed, and to determine the probability of occurrence of different acceleration levels. In the next figures it is seen the relative frequency of apparition of the different acceleration values for the slab.

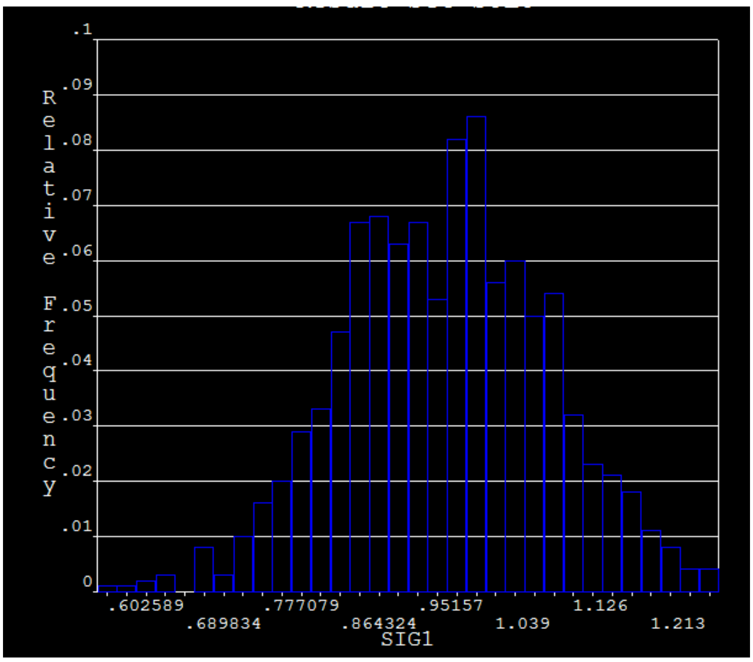


Fig. 1. Histogram of accelerations for a train speed of 90 km/h.

1.1 Simulation B

The proposed methodology was applied to the selected case study in order to assess the bridge accelerations of the trains running over the bridge at a wide range of train speeds In the next figure the maximum accelerations obtained for each simulation and for different train speeds are illustrated on a histogram, making easier to establish a relation between train speed and slab vibration.

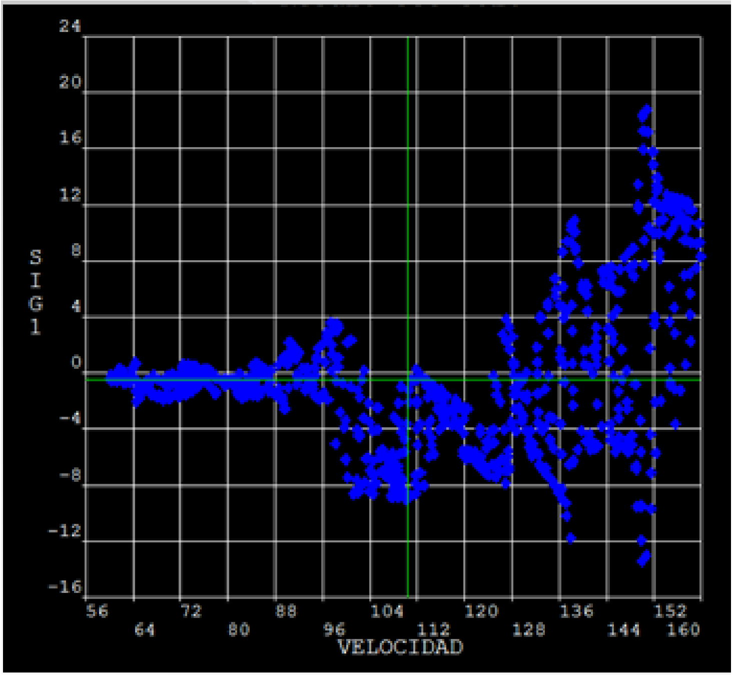


Fig. 2. Histogram of accelerations for different train speeds.

4. Conclusions

The current paper proposed a probabilistic methodology for determining the accelereations obtaind at different situations considering differente variables. This methodology combined Monte Carlo simulations with the extreme value theory for efficiency purposes. A comutter railway bridge, located at Xativa, was selected as case study for the application of the methodology. The variability of the bridge, the track and the train was accounted for, as well as the existence of track irregularities representing different track maintenance levels. The main goals of the investigation are:

- Possibility to introduce uncertainities
- Irregularities statistically introduced
- Useful for safety assessment

References

[1] Xia, H., Zhang, N., & Guo, W. W. (2006). Analysis of resonance mechanism and conditions of train-bridge system. *Journal of Sound and Vibration , 297*, 810-822.

[2] Wang, Y., Wei, Q., Shi, j., & Long, X. (2010). Resonance characteristics of two-span continuous beam under moving high speed trains. *Latin American Journal of Solids and Structures , 7* (2).

[3] Yang, Y., & Yau, J. (2005). Vehicle bridge interaction dynamics and potential applications. *Journal of sound and vibration* , 247-259.

[4] Mao, L., & Lu, Y. (2013). Critical speed and resonance criteria of railway bridge response to moving trains. *18* (2), 131-141.

[5] Yau, J. D. (2001). Resonance of continuous bridges due to high speed trains. *Jounral of marine science and technology , 9* (1), 14-20.

[6] Kwark, J. W., Choi, E. S., Kim, Y. J., Kim, B. S., & Kim, S. I. (2004). Bynamic behavior of two-spans continuous concrete bridges under moving high-speed train. *Journal of computers and structures , 82*, 463-474.

[7] Ju, S. H., & Lin, H. T. (2003). Resonance characteristics of high-speed trains passing simple supported bridges. *Journal of Sound and Vibration , 267*, 1127-1141.

[8] Reh, S., Beley, J. D., Mukherjee, S., & Khor, E. H. (2006). Probabilistic finite element analysis using ANSYS. *Structural Safety*, *28*(1), 17-43.

Multiscale CFD analysis of a partial sectorization and longitudinal ventilation safety system in railway tunnels.

Laura Sánchez Rayo[1*], Fran Ribes-Llario[1], Antonio Enrique Blanco Saura[1], Julia Irene Real Herraiz[1]

[1]University Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

*Corresponding author. E-mail: rayolau88@gmail.com. Telephone: +34 96 387 70 00.

July 18, 2016

1. Introduction

Although fires in railway tunnels are not a common situation, past events show their terrible consequences. In this sense, the development of safety ventilation systems and their optimal operation is of outmost importance.

If a tunnel catches fire, the phenomenon of *stratification* may take place. This, by definition, is the process in which the fumes are located at the upper part of the tunnel due to the difference of density between the fumes and the surrounding air. This fact has been concluded as an advantage during evacuation, since it allows having a better visibility and reduced concentration of toxic gases in the lower part.

According to this, the present investigation deals with the study of ventilation in tunnels under fire conditions. To do so, the Computational Fluid Dynamivs method (CFD method) is used. The study is involved in an R+D project of which final result is an anti-spread system composed by a partial sectorized zone and longitudinal ventilation system able to evacuate fumes safely from the sectorized zone.

2. Case of study

After the analysis of different configurations for the development of an anti-spread system, which includes sectoring system and different ventilation types, it has been concluded that the best option for evacuating the fumes consists on the installation of a longitudinal ventilation system optimized according to the sectioned zone. This allows evacuating fumes without breaks stratification.

Hence, the case of study (whose numerical simulation is further explained) consists on a set of fireproof curtains distributed along the tunnel axis each 100 m. The system also provides longitudinal ventilation system by using fans in the centre of the sectorized sections. The curtains isolate the affected area and prevent the spread of the fire. It also limits the oxygen supply, leading to the fire self-extinction.

3. Methodology

In this study, the computational fluid Dynamics (CFD) method is used to evaluate the smoke behaviour with the partial sectorization system. However, a CFD analysis of fire may require high computational resources, since the computational cost increases with the tunnel length [X]. To solve this problem, a multiscale model has been developed.

Thanks to this solution, the computational domain is divided into several computational subdomains [X], between which there is an exchange of information within interfaces. Nevertheless, domain decomposition is affected by the fact that sectors with fire sources and active fans, cannot perform two-dimensional simplification considering that the flow is highly turbulent. However, this simplification can be adopted in *far field* zones from the fire source, because the velocity components parallel to the longitudinal direction of the tunnel, is greater than the cross-directional component. The decomposition methods or computational domain discretization techniques (finite difference, finite element, finite volume) provide the theoretical basis for the development and implementation of multiscale techniques.

Hence, the full tunnel domain (Ω) is decomposed in two subdomains $\Omega_{3\mathrm{D}}$ and $\Omega_{2\mathrm{D}}$, where 3D model contains the active elements of interest (fire source, active fans and curtains) and the surrounding tunnel infrastructure, in summary, the *near field*. Meshing implemented in the subdomain $\Omega_{3\mathrm{D}}$ follows the same premise that in the case of the full model but with greater mesh resolution in the interface surface.

However, a drawback of this simplification is its inability to analyze phenomena that occur in the model cross sections. Hence, two types of simulations have been carried out from different standpoints. First, simulations have been performed using the full CFD approach. Then, simulations have been repeated by multiscale CFD approach in order to compare the results obtained by full CFD approach. The analysis is performed for two different tunnel lengths: 500 and 1100 meters. Thus, four simulations are developed: 500 meters tunnel length with full CFD (noted as 500F); 500 meters tunnel length with multiscale CFD (noted as 500M); 1100 meters tunnel length with full CFD (noted as 1100F) and 1100 meters tunnel length with multiscale CFD (noted as 1100M).

For the simulations, the fire source has been placed at the centre of the computational domain. In order to reduce the computational time, symmetry has been considered along the longitudinal axis. On both tunnel ends, two prismatic volumes simulate the outside environment conditions.

Tunnel geometry and flux characteristics in zones near the fans, fire source and a curtain (in which a high mesh resolution is necessary) requires an unstructured tetrahedral mesh. All the significant zones of the numerical model are shown in the next figure, where a, b and c shows a polyhedral mesh detail of the fireproof curtain, fan and fire source, respectively.

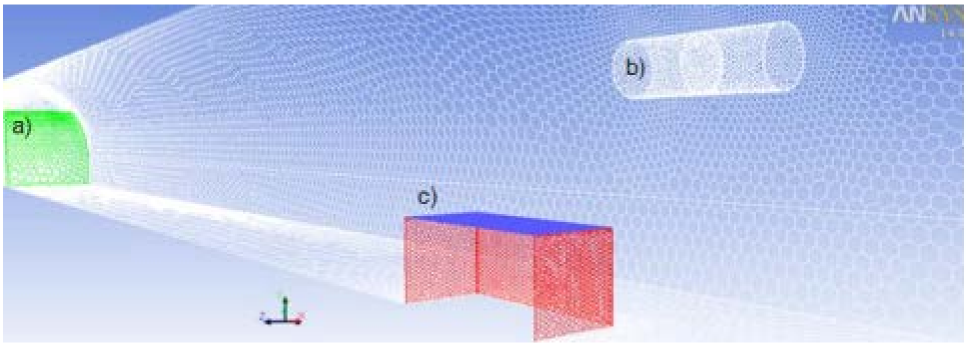


Fig. 1 Detailed polyhedral mesh of a) fireproof curtain, b) fan and c) fire source

The differential equations for mass, momentum and energy conservation are solved using the commercial software ANSYS software. To close these equations and solve the Reynolds-Averaged Navier-Stokes equations, a standard κ-ε turbulence model has been employed. This model has been validated by several researchers, including fire scenarios in tunnels [1-3].

Accordingly, the focus fire is modeled as a prismatic volume whose upper surface injects a temperature of 761.45 K (equivalent to 30 MW fire size, with ratio air mass fraction/CO_2 mass fraction 0.93/0.07) and lateral surfaces act as air sinks. To obtain the air and gases present in fire, Eqs. (1, 2) have been employed, where $\dot{m}_g$ represents the mass flow rate of the fire source, $\dot{W}$ represents the convective heat release rate, c_p is the specific heat of the mixture, T_c is the combustion temperature of gases, T_0 is the surrounding temperature, $\dot{m}_{air}$ is the air mass flow rate used in the lateral surfaces of the fire and $\dot{m}_{CO_2}$ is the mass flow rate of the carbon dioxide:

$$\dot{m}_g = \dot{W}/\left(c_p(T_c - T_0)\right) \tag{1}$$

$$\dot{m}_{air} = \dot{m}_g - 1/3\,\dot{m}_{CO_2} \tag{2}$$

The tunnel walls and the simulated fire curtains were considered completely rigid and adiabatic. The ventilators were treated as cylindrical bodies in which a difference of pressure is imposed ($\Delta P = 250\ Pa$) that resulted in a volumetric mass flow rate.

4. Simulations and results

In order to compare the results obtained by the different approaches simulated (500F, 500M, 1100F and 1100M), the contours of CO_2 concentration have been plotted of the symmetry plane.

For each approach it has been considered the same set-up of anti-spread system and fire size. The equivalent power fire has been set in 30 MW, fan pressure of 250 Pa (in all fans except the one in the sectorized zone) and the height of fireproof curtains remains constant (h_C=4.5m). The results shown have been displayed in two different time instants: 150 seconds and 300 seconds.

4.1 Results of CO_2 concentration before the anti-spread system activation

First, the evolution of fumes has been compared in the time instant before the unfolded of the curtains. During 150 seconds from the fire starts, the fire develops without the performance of the anti-spread system. Figs. 2 and 3 show the symmetry plane of tunnels, the region near to fire source since upstream fan located fire, and a part of region corresponding to the far field, located at right hand of interface.

The results show how the 30 *MW* fire causes pronounced density gradients between the fire fume and surrounding air. Hence, the smoke stratification takes place. Furthermore, in Figs. 2b-2c and Figs. 3b-3c, the subdomain Ω_{3D} provides values of velocity, temperature and mass fraction to subdomain Ω_{2D}, through the right hand side of interface Γ_a located in $x = a$. The domain descomposition method, the multiscale procedures and the multiscale iteration process does not affect to the simulations results.

Finally, according to Figs. 2c-3c, the smoke moves through the top of the tunnel with no stratification breaking.

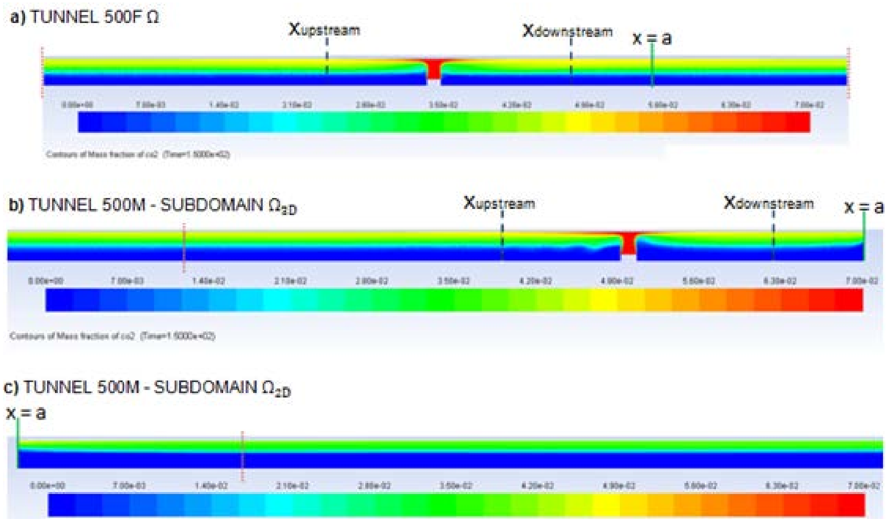


Fig 2. Mass fraction contour of CO_2 at t = 150 s, before the anti-spread system activation, in the symmetry plane for full and multiscale CFD approaches, when the tunnel length is 500 m.

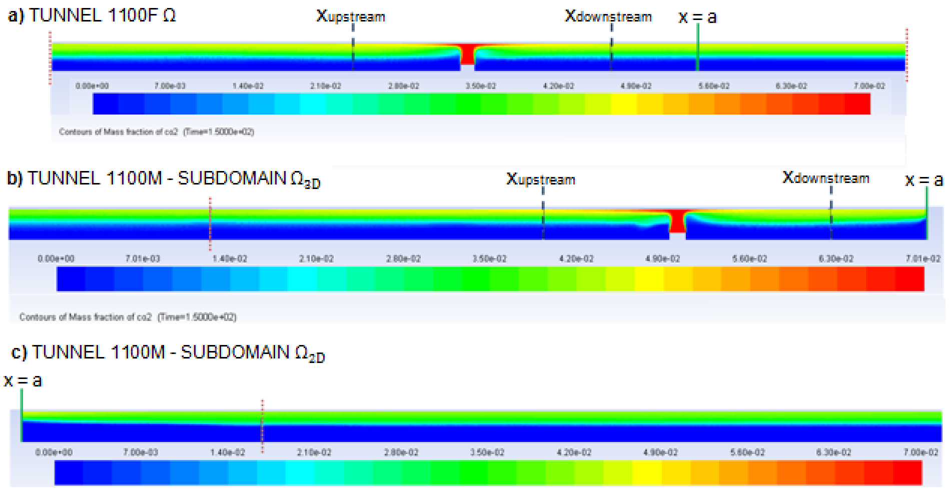


Fig 3. Mass fraction contour of CO_2 at t = 150 s, before the anti-spread system activation, in the symmetry plane for full and multiscale CFD approaches, when the tunnel length is 1100 m.

4.2 Results of CO_2 concentration after the anti-spread system activation

After 150 seconds in that the fire has developed freely, the anti-spread system comes into operation, by activating the fans upstream the fire source and unfolding the fireproof curtains of the central sector. Figs. 4-5 show how when all fans located upstream fire (except the one which remains inside the sectored zone) are activated, the high fume concentration is retained inside the sectored zone, in calculation domain of Tunnel 500F and Tunnel 1100F, respectively, at $t = 150\ s$ after anti-spread system activation. The stratification downstream of the sectored zone is maintained for both tunnel lengths in both CFD approaches (see Figs. 4c-5c), allowing a safe evacuation. As can see in Figs. 4a-4b and 5a-5b, upstream of the sectored zone, the *backlayering* phenomena occurs. This is defined as a reversal movement of smoke counter flow ventilation direction. The *backlayering* is not considered as a negative effect due to the main goal of the anti-spread system is to maintain a fume concentration as low as possible at the evacuation area.

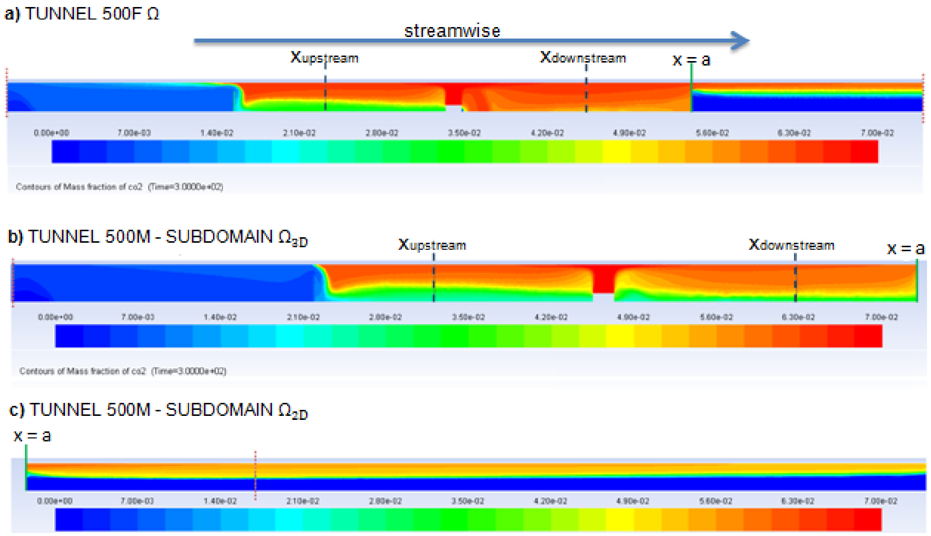


Fig 4. Mass fraction contour of CO_2 at t = 150 s, after the anti-spread system activation, in the symmetry plane for full and multiscale CFD approaches, when the tunnel length is 500 m.

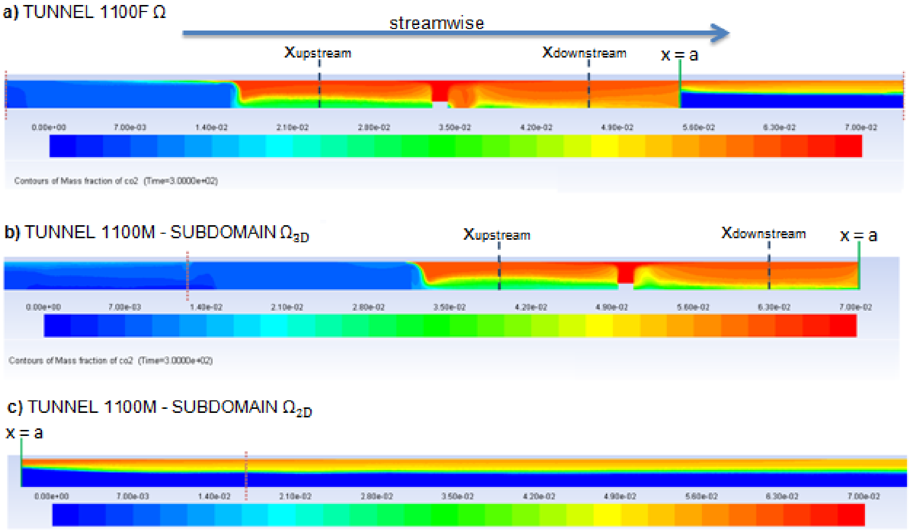


Fig 5. Mass fraction contour of CO_2 at t = 150 s, after the anti-spread system activation, in the symmetry plane for full and multiscale CFD approaches, when the tunnel length is 1100 m.

Due to the *pressure-outlet* boundary condition imposed in the cross-section area not covered at the left hand side of the interface Γ_a located in $x = a$, a less turbulent flow behaviour takes place. Thus, the sectored zone smoke stratification is kept. This boundary condition does not affect the stratification downstream of the fireproof curtain (located in $x = a$).

5. Conclusions

The simulation of multiscale approach from tunnels for both lenghts shows that the domain descomposition method, the multiscale procedures and the multiscale iteration process does not affect to these simulations results. However, after the anti-spread system is activated, the fume concentration retained inside the sectored zone varies depending on the approach used. In the case of multiscale CFD approach, due to the pressure-outlet boundary condition imposed a less turbulent flow is induced, stratification is kept. However, this boundary condition does not affect the behavior of stratification downstream fireproof curtain. The time reduction obtained using the multiscale approach is set around 40%.

References

[1] Colella F, Rein G, Verda V, Borchiellini R., 2011. Multiscale modeling of transient flows from fire and ventilation in long tunnels. Comput Fluids; 51(1):16-29

[2] Colella F, Rein G, Borchiellini R, Torero JL., 2011. A novel multiscale methodology for simulating tunnel ventilation flows during fires. Fire Technol. 47(1):221-53

[3] Vega M, Argüelles Díaz K, Fernández Oro J, B TR SM. C., 2008. numerical 3D simulation of a longitudinal ventilation system: Memorial tunnel case. Tunnel Underground Space Technol;23(5):539-51.

Evaluation of the dynamic response of a bridge under scour conditions. Development of a numerical prediction model in real bridges.

José Luis Velarte González[1*], Beatriz Baydal Giner[1], Teresa Real Herraiz[1], Julia Irene Real Herraiz[1]

[1]Institute for Multidisciplinary Mathematics, Polytechnic University of Valencia, 46022, Valencia, Spain

**Corresponding author. E-mail: jovegon@cam.upv.es. Telephone: +34 96 387 70 00

July 18, 2016

1. Introduction

The incessant economic and population growth experienced by cities around the world in recent decades has led to the need of building more and better transport infrastructures. For this reason, countries with complex orography (as Spain) have required strong economic investments to save geographical features. This has led to the construction of costly linear infrastructures (such as bridges or tunnels among others).

In this sense, the scour phenomenon at piers and buttresses is a risk factor that may cause the bridge collapse. By definition, this is the situation that occurs when the structure's foundations get exposed to the flow due to the confining materials drag. Generally, this situation takes place after turbulent processes (as floods) and compels relevant stiffness variations. Thus, the dynamic response of the pier is strongly affected by the presence of scour holes, even if filled [1].

According to this, the present investigation develops a new method based on Operational Modal Analysis techniques for scour detection. The method is based on the use of hardware components able to register the vibratory response of the structure through time and a soil-structure finite elements model. Thus, different scour levels can be studied by numerical simulations in order to predict the vibratory response of the bridge. Then, the results can be compared with the real registers in order to quantify the scour affection.

2. Case of study

The bridge of study (Fig. 1) is located in the region of Murcia (Spain). It is a road bridge supported by two sets of four piles. The foundation is composed of pile caps of 6 meters depth and the bridge deck is constructed with 5 rows of precast double T concrete beams.

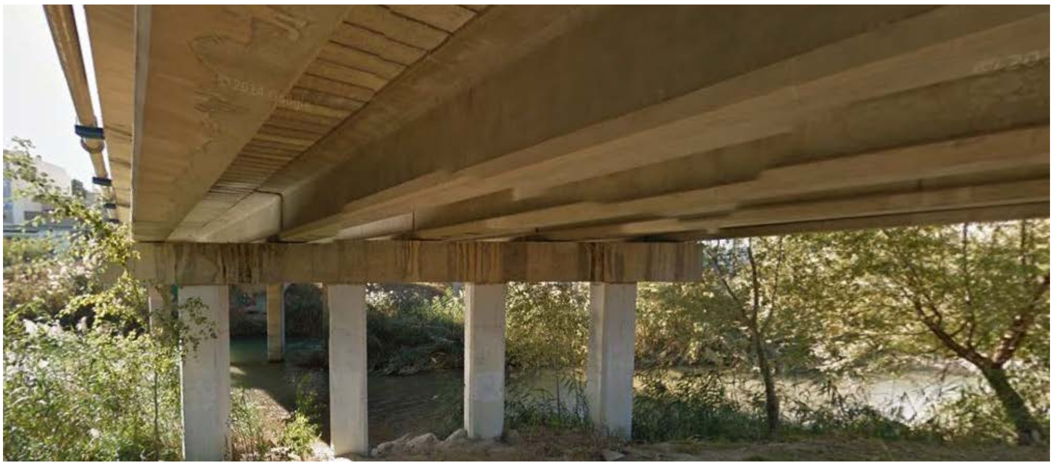


Fig. 1. Studied bridge

3. Methodology

The scour detection is performed by a continuous feedback process between a series of sensor nodes and a Finite Element Model (FEM).

3.1 Development of the sensor nodes

Each sensor (Fig. 2) is made of four components: one triaxial accelerometer (which registers the vibratory response of the structure); one biaxial inclinometer (which registers the relative spin of the structure); one Arduino Microcontroller (which operates the whole system and stores the nodes' registers) and a supply system (which provides energy).

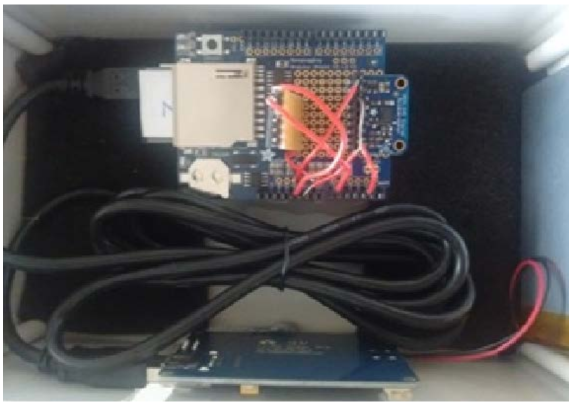


Fig. 2. Sensor node

All sensors are located at the top of buttresses and piles (out of reach of flow). For the studied bridge, 5 sensors are located: 1 sensor per buttress and 2 sensors per pile (one upstream and one downstream, Fig. 2).

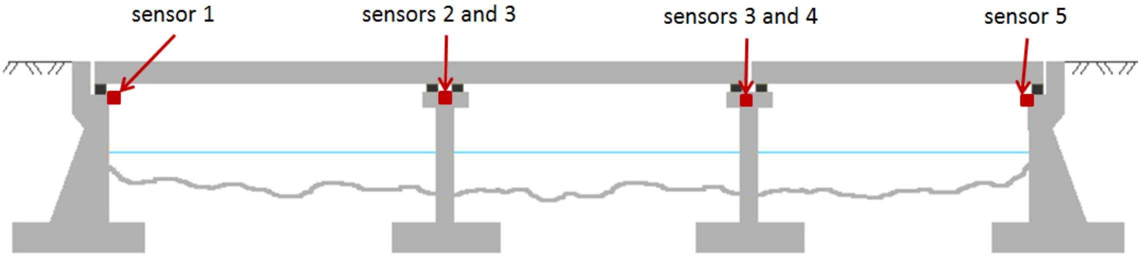


Fig. 2. Final location for the sensors at the studied bridge

3.2 Development of the Finite Element Model

The numerical model (Fig. 3) is developed in ANSYS LS-DYNA software and is used to predict the vibratory response of the bridge against different scour scenarios.

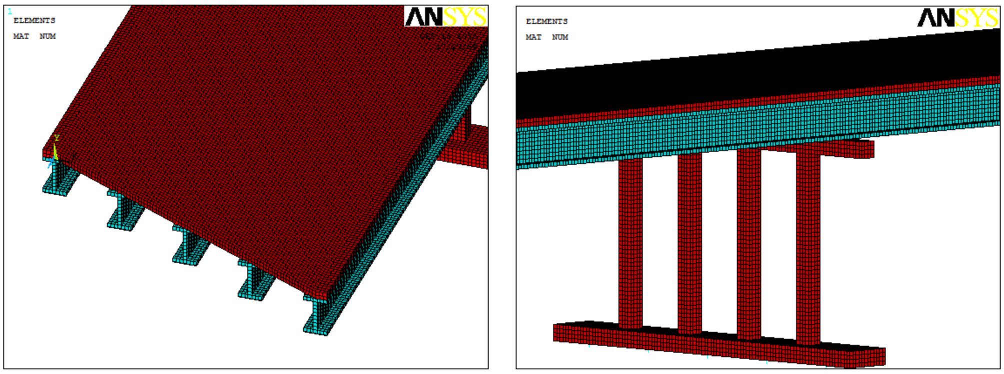


Fig. 3. Finite elements model

The elements used for the simulation of the superstructure are hexahedral 8-node type with linear interpolation (SOLID45). The foundation is simulated using BEAM3D elements. In both cases the maximum element size is 10 cm.

The pile simulation is carried out as a beam supported in an elastic medium according to the Winkler's theory [2]. Thus, the deformation of the foundation is confined to the loaded regions only. The soil-foundation interaction is simulated by means of horizontal springs which simulate the confinement of the soil. For the modal analysis, materials have been assumed to present a linear-elastic behavior (since reduced traffic loads were expected).

The analysis is based on the Modal Analysis technique, so the equation of motion (expressed in matrix form) is established according to Eq (1), where [K] and [M] are respectively the mass and stiffness matrices of the structure and $\{\ddot{\boldsymbol{u}}\}$ and $\{\boldsymbol{u}\}$ are, respectively, the acceleration and displacement vectors:

$$[M]\{\ddot{\boldsymbol{u}}\} + [K]\{\boldsymbol{u}\} = 0 \tag{1}$$

By principle of Modal Analysis, no damping is assumed. Thus, for the linear system considered, free vibration is determined by harmonic functions according to Eq. (2)

where the eigenvectors associated to the mode shape of the i-th natural frequency (Φ_i) are represented for the i-th natural frequency (ω_i) through time (t):

$$\{\boldsymbol{u}\} = \{\Phi_i\}Cos\omega_i t \tag{2}$$

Combining the previous equations, the modal analysis is performed solving Eq. (3), which can be solved for n values of the natural circular frequencies of the system (ω_i) and eigenvectors $\{\Phi_i\}$.

$$(-\omega_i^2\ [M]\ +\ [K])\,\{\Phi i\}\ =\ 0 \tag{3}$$

3.3 FEM-node feedback process

The feedback process between the sensor nodes and a Finite Element Model is developed in three steps. In the first step, the vibratory response of the structure induced by vehicle traffic is recorded. Thus, the signal can be processed by ARMAV methodology (Auto-Regressive Moving Average Vector) according to (Eq. 4), where *y[n]* is an m-dimensional matrix of temporary registers; *u[n]* corresponds to the input of the system (ie, a stationary Gaussian white noise input.) and a_k and b_k matrices are, respectively, the auto-regressive matrix (AR) and moving average matrix (MA) [3]:

$$y[n] = \sum_{k=1}^{p} a_k y[n-k] + u[n] + \sum_{k=1}^{q} b_k u[n-k] \tag{4}$$

Thus, the term AR contains all the modal information of the vibratory system and the term MA contains information concerning the excitation and external noise.

With regard to the second step of the process, once the numerical model has been fully developed, several scour scenarios are simulated. By doing this, it is possible to obtain the eigenfrequencies associated to the different modes of vibration of each scenario.

Finally, in the third step the results obtained by the ARMAV method and the results obtained by FEM simulations are compared. By doing this, it is possible to detect and quantify the severity of scour damage.

4. Simulations and results

It should be pointed out that the measurement campaign could not be fully concluded due to technical problems. For this reason, it is only shown the results obtained by the numerical model.

According to this, since the soil-foundation interaction is simulated by means of horizontal springs which simulate the confinement of the soil, the effect of scour is simulated by eliminating the springs on top of the foundation. Two scenarios are studied: Scour in one pier (Scenario 1, Table 1) and scour in both piers (Scenario 2, Table 2). Only the results of the very first vibration modes are shown below, since the higher modes are a combination of the first ones [4].

	Scour level (m)						
	0	0.25	0.5	0.75	1	2	3
Mode 1	9.30	9.25	9.16	9.10	9.02	8.78	8.61
Mode 2	12.45	12.43	12.38	12.27	12.63	12.06	12.54
Mode 3	14.99	14.32	13.42	12.95	11.80	10.00	8.93
Mode 4	15.13	15.12	15.13	15.13	15.12	15.12	15.11
Mode 5	15.74	15.64	15.04	14.54	15.75	15.61	15.93
Mode 6	16.64	16.64	16.63	16.63	16.62	16.65	16.64
Mode 7	16.85	16.84	16.82	16.81	16.77	16.90	16.88
Mode 8	17.31	17.29	17.25	17.21	17.11	17.41	17.38
Mode 9	20.18	19.91	19.67	19.59	19.51	19.40	19.37

Table 1 Frequencies (in Hz) associated to each mode of vibration in Scenario 1

	Scour level (m)						
	0	0.25	0.5	0.75	1	2	3
Mode 1	9.30	9.23	9.13	9.07	8.97	8.70	8.53
Mode 2	12.45	12.43	12.38	12.27	12.63	12.06	11.12
Mode 3	14.99	14.32	13.42	12.95	11.80	10.00	8.93
Mode 4	15.13	14.80	14.35	14.09	13.75	13.02	12.73
Mode 5	15.74	15.62	15.05	14.54	13.85	12.58	-
Mode 6	16.64	16.49	16.12	15.91	15.71	15.40	-
Mode 7	17.31	17.22	17.12	17.06	16.93	16.23	16.95
Mode 8	20.18	19.86	19.51	19.34	17.34	17.01	15.88
Mode 9	21.55	21.47	21.39	21.34	21.29	21.20	21.17

Table 2 Frequencies (in Hz) associated to each mode of vibration in Scenario 2

In view of the results, in line with studies [4, 5], as the scour level growths, some modes show a sharp decline in the natural frequencies. By contrast, others modes show a slight variation or virtually nonexistent.

The main reason lies in the classification of vibration modes. Thus, some modes can be classified as either vertical or horizontal. In the first ones, due to the limited influence of lateral stiffness of piles against vertical deformations, a slight variation of natural frequencies is shown. By contrast, those modes with predominance of horizontal component show a significant reduction in the frequency of vibration. In such case, the final magnitude will be higher or lower depending on the characteristics of each bridge.

Finally, if both tables are compared (Fig. 3), a faster drop of natural frequencies in scenario 2 can be observed. The main reason lies in the grater affection of loss of stiffness when scour appears in both piles.

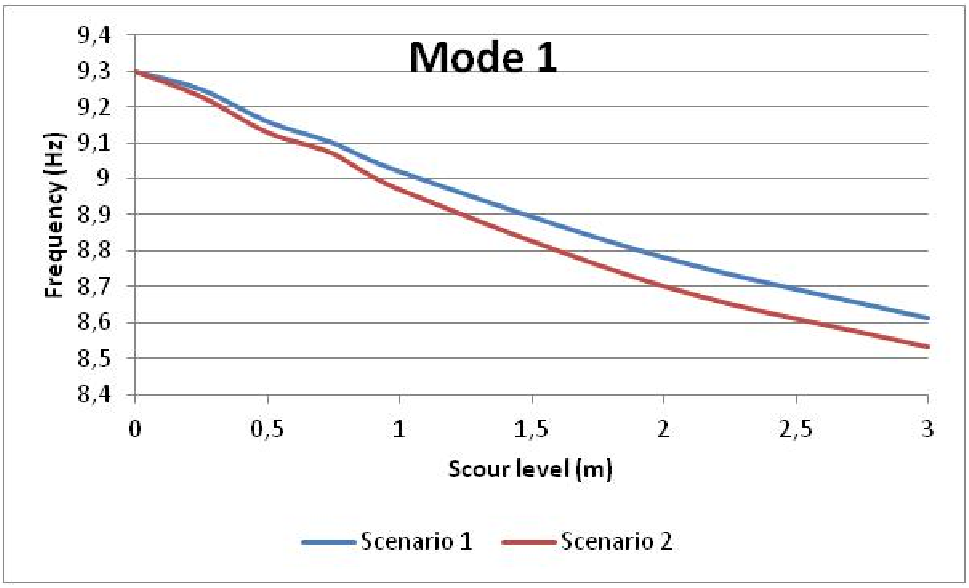


Fig. 3. Natural frequencies variations in mode 1. One pile scoured (blue line) and both piles scoured (red line).

5. Conclusions

According to the results obtained by numerical simulations, a direct relationship between the scour level and the variation of natural frequencies is observed. Thus, as the scour level growths, the natural frequencies of each natural mode of vibration tend to reduce. Furthermore, it is shown that the larger the number of affected piles is, the greater the drop is (since the overall stiffness of the system decreases faster).

References

[1] Foti, S., Sabia, (2010). D. Influence of foundation scour on the dynamic response of an existing bridge. *J Bridge Eng, ASCE 2011*, *16*(2), 295–304

[2] Winkler, E. (1867). Die Lehre von der Elasticitaet und Festigkeit: mit besonderer Rücksicht auf ihre Anwendung in der Technik für polytechnische Schulen, Bauakademien, Ingenieue, Maschinenbauer, Architecten, etc. Dominicus.

[3] M., Dabiran, N., & Taghikhany, T. (2014). STRUCTURAL DAMAGE DETECTION USING DAMAGE LOCATING VECTOR WITH WIRELESS SMART SENSORS.

[4] Prendergast, L.J., Hester, D., Gavin, K., O'Sullivan, J.J.(2013). An investigation of the changes in the natural frequency of a pile affected by scour, Journal of Sound and Vibration, 332(25) ,685-702

[5] Elsaid, A., Seracino, R.(2014). Rapid assessment of foundation scour using the dynamic features of bridge superstructure, Construction and Building Materials, 50, 42-51

Solving the random Cauchy problem for the heat model with unbounded spatial domain using a mean square finite difference scheme

J.-C. Cortés[♭], J.-V. Romero[♭], M.-D. Roselló[♭]*, and M.A. Sohaly[†]

(♭) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València,

Camino de Vera s/n, Edificio 8G, 2°, 46022 Valencia, Spain,

(†) Department of Mathematics - Faculty of Science,

Mansoura University - Egypt

November 30, 2016

1 Introduction

This paper is concerned to develop the theory of the finite difference method to the Cauchy problem for heat equation in one dimensional with unbounded spatial domain as in the form:

$$u_t = \beta u_{xx}\ ,\ t \in [0, \infty)\,; \quad -\infty < x < \infty \tag{1}$$

such that:

$$u(x, 0) = u_0(x),$$

where β is a random variable.

In this paper is proved the mean square consistency and the mean square stability for a finite difference method to solve model (1).

*e-mail: drosello@imm.upv.es

2 Preliminaries

In this section, we introduce some definitions and important results to service the technique used in the paper. A real random variable X defined on the probability space $(\Omega,\ \mathcal{F},\ P)$ and satisfying the property

$$\mathbb{E}\left[|X|^2\right] < \infty,$$

is called 2 order random variable (2 r.v.), where $\mathbb{E}\left[\cdot\right]$ denotes the expectation operator. If $X \in \mathrm{L}_2(\Omega)$, then the L_2 norm is defined as:

$$\left\|X\right\|_2 = \left[\mathbb{E}\left[|X|^2\right]\right]^{\frac{1}{2}}.$$

Definition 1 *[1] A sequence $\{X_n, n > 0\} \in \mathrm{L}_2(\Omega)$ is mean square convergent to a random variable $X \in \mathrm{L}_2(\Omega)$ if*

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^2\right] = 0.$$

Definition 2 *[2, 3] A finite difference scheme (FDS) $\mathrm{L}^n_k u^n_k = G^n_k$ ($\mathrm{L}^{\mathrm{n}}_{\mathrm{k}}$ is the discretization operator) approximating the partial differential equation (PDE) $\mathrm{L}v = G$ (L denotes the differential operator) is mean square consistent if the solution of the PDE, v, satifies:*

$$V^{n+1} = Q\ V^n + (\Delta t)G^n + (\Delta t)\tau^n,$$

and $\|\tau^n\| \to 0$ as $\Delta x, \Delta t \to 0$. V^n denotes the vector whose k-th component is $v(k\Delta x, n\Delta t)$.

Definition 3 *[2, 3] A finite difference scheme (FDS) $\mathrm{L}^n_k u^n_k = \mathrm{G}^{\mathrm{n}}_{\mathrm{k}}$ (L^n_k is the discretization operator) that approximating the partial differential equation (PDE) $\mathrm{Lv} = \mathrm{G}$ (L denotes the differential operator) is mean square stable, if there exist some positive constants ϵ, δ, non-negative constants η, ξ and u^0 is the initial data such that*

$$\left\|u^{n+1}\right\| \leq \eta \mathrm{e}^{\xi t}\left\|u^0\right\|,$$

for $t = (n+1)\Delta t$, $0 < \Delta x \leq \epsilon$, $0 < \Delta t \leq \delta$.

Definition 4 *[2] A finite difference scheme (FDS)* $\mathrm{L}_\mathrm{k}^\mathrm{n}\mathrm{u}_\mathrm{k}^\mathrm{n} = \mathrm{G}_\mathrm{k}^\mathrm{n}$ *(*$\mathrm{L}_\mathrm{k}^\mathrm{n}$ *is the discretization operator) approximating the partial differential equation (PDE)* $\mathrm{Lv} = \mathrm{G}$ *(*L *denotes the differential operator) is said to be accurate of order* (p, q) *to a given partial differential equation if:*

$$\|\tau^\mathrm{n}\| = \mathcal{O}(\Delta x^p) + \mathcal{O}(\Delta t^q),$$

where τ^n *is the truncation error.*

Remark 1 *[2] If the scheme is accurate of order* (p, q)*,* $p \geq 1, q \geq 1$ *then the scheme is consistent.*

If we consider elements in $(\mathrm{l}_2(\Omega), \|\cdot\|)$, [2],

$$\mathrm{l}_2(\Omega) = \{x = (x_{-\infty}, \ldots, x_{-1}, x_0, x_1, \ldots, x_\infty) : \ \|x\| < +\infty\},$$

the following norm will be used

$$\|x\|^2 = \mathbb{E}\left[\left(\sup_k |x_k|\right)^2\right].$$

3 Random Finite Difference Technique

The principle of finite difference technique is close to the approximation schemes used to solve deterministic (stochastic) ordinary and partial differential equations [4, 5, 6, 7]. The main idea of this technique is to replace the spatial differential in the model by a spatial difference operator at a location by using the neighbouring nodal points. Let we define the grid cells for the space to be $\Delta x = (x_k - x_{k-1})$ for $k \geq 1$ and also, define the time steps to be $\Delta t = (t_n - t_{n-1})$ for $n \geq 1$. Let us consider $u_k^n = u(k\Delta x, n\Delta t)$ the approximation of the the exact solution for problem (1), $u(x,t)$, at the point $(k\Delta x, n\Delta t)$. By replacing the first and second derivatives in (1) with the following difference formulas:

- Forward formula by two points for the time:

$$u_t(k\Delta x, n\Delta t) \approx \frac{u_k^{n+1} - u_k^n}{\Delta t}$$

- Central formula by three points for the space:

$$u_{xx}(k\Delta x, n\Delta t) \approx \frac{u_{k+1}^{n} - 2u_k^n + u_{k-1}^{n}}{(\Delta x)^2}$$

we obtain the (forward time - central space) random difference scheme

$$u_k^{n+1} = (1-2r)\, u_k^n + r u_{k+1}^n + r u_{k-1}^n, \qquad u_k^0 = u_0(k\Delta x) = u_0(x_k), \tag{2}$$

where $r = \frac{\beta \Delta t}{(\Delta x)^2}$.

Consistency, stability and convergence are important topics in deterministic and stochastic theory for many numerical methods [5]. The aim of this paper is to appropriate these topics to the scheme (2) in mean square sense as we shown in the next points.

3.1 Consistency of RFDS (2)

Definition 5 *A random finite difference scheme (RFDS)* $\mathrm{L_k^n u_k^n} = \mathrm{G_k^n}$ *that approximate the random partial differential equation (RPDE)* $\mathrm{Lv} = \mathrm{G}$ *is mean square consistent if the solution of the RPDE, v, satifies:*

$$\mathrm{V}^{n+1} = Q\mathrm{V}^n + (\Delta t)\mathrm{G}^n + (\Delta t)\tau^n \tag{3}$$

and:

$$\mathbb{E}\left[\left(\sup_k |\tau_k^n|\right)^2\right] \to 0,$$

as $\Delta x, \Delta t \to 0$. V^n *denotes the vector whose k-th component is* $v(k\Delta x, n\Delta t)$.

Theorem 1 *The RFDS* (2) *associated to problem* (1) *is mean square consistent.*

Proof Let us rewrite the RFDS (2) in the form,

$$u_k^{n+1} = u_k^n + r(u_{k+1}^n - 2u_k^n + u_{k-1}^n), \qquad r = \frac{\beta\Delta t}{(\Delta x)^2},$$

and let $u(x,t)$ be a solution to RPDE (1). Using Taylor expansions we have

$$\begin{aligned}(\Delta t)\tau_k^n &= \mathrm{u}_k^{n+1} - \left\{\mathrm{u}_k^n + r\left[\mathrm{u}_{k+1}^n - 2\mathrm{u}_k^n + \mathrm{u}_{k-1}^n\right]\right\} \\ &= \mathrm{u}_k^n + (\mathrm{u_t})_k^n(\Delta t) + \mathcal{O}\left((\Delta t)^2\right) - \mathrm{u}_k^n - r\mathrm{u}_k^n - r(\mathrm{u_x})_k^n(\Delta x) \\ &\quad - r(\mathrm{u_{xx}})_k^n\frac{(\Delta x)^2}{2} - r(\mathrm{u_{xxx}})_k^n\frac{(\Delta x)^3}{6} + \mathcal{O}\left((\Delta x)^4\right) + 2r\mathrm{u}_k^n - r\mathrm{u}_k^n \\ &\quad + r(\mathrm{u_x})_k^n(\Delta x) - r(\mathrm{u_{xx}})_k^n\frac{(\Delta x)^2}{2} + r(\mathrm{u_{xxx}})_k^n\frac{(\Delta x)^3}{6} + \mathcal{O}\left((\Delta x)^4\right)\end{aligned}$$

Hence we have:

$$\Delta t\,\tau_k^n = (u_t - \beta u_{xx})_k^n \Delta t + \mathcal{O}\left((\Delta t)^2\right) + \mathcal{O}\left(\Delta t\,(\Delta x)^2\right).$$

since, $u_t - \beta u_{xx} = 0$, then we have:

$$\tau_k^n = \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x)^2.$$

Finally, taking the supremum and the expectation, one gets,

$$\|\tau_k\|^2 = \mathbb{E}\left[\left(\sup_k |\tau_k^n|\right)^2\right] \to 0, \qquad \Delta x, \Delta t \to 0.$$

Hence, the RFDS (2) is mean square consistent.

3.2 Stability of RFDS (2)

Definition 6 *A random difference scheme* $\mathrm{L_k^n u_k^n} = \mathrm{G_k^n}$ *(*$\mathrm{L_k^n}$ *is the discretization operator) that approximating RPDE* $\mathrm{Lv} = \mathrm{G}$ *(*L *denotes the differential operator) is mean square stable, if there exist some positive constants* ϵ, δ, *non-negative constants* η, ξ *and* u^0 *is initial data such that:*

$$\mathbb{E}\left[\sup_k \left|u_k^{n+1}\right|^2\right] \le \eta e^{\xi t}\mathbb{E}\left[\sup_k \left|u_k^0\right|^2\right] \tag{4}$$

for $t = (n+1)\Delta t$, $0 < \Delta x \le \epsilon$, $0 < \Delta t \le \delta$.

Theorem 2 *The RFDS (2) that according to the problem (1) is mean square stable under the condition:* $\Delta t \le \frac{(\Delta x)^2}{2\beta_1}$ *where* $0 < \beta(\omega) \le \beta_1$

Proof From the RFDS (2) we have:

$$\mathbb{E}\left[\sup_k \left|U_k^{n+1}\right|^2\right] = \mathbb{E}\left[\sup_k \left|(1-2r)\mathrm{U_k^n} + r\mathrm{U_{k+1}^n} + r\mathrm{U_{k-1}^n}\right|^2\right]$$
$$= \mathbb{E}\left[\left[(1-2r)^2 + r^2 + r^2 + 2r\,|1-2r| + 2r\,|1-2r| + 2r^2\right]\sup_k |\mathrm{U_k^n}|^2\right].$$

If $0 < r \le \frac{1}{2}$ then we have $|1-2r| = 1-2r$. Hence,

$$\mathbb{E}\left[\sup_k \left|U_k^{n+1}\right|^2\right] \le \mathbb{E}\left[\sup_k \left|U_k^{n}\right|^2\right] \le \cdots \le \mathbb{E}\left[\sup_k \left|U_k^{0}\right|^2\right].$$

Hence, the RFDS (2) is mean square stable with $\eta = 1$, $\xi = 0$ and under the condition,

$$\Delta t \le \tfrac{(\Delta x)^2}{2\beta_1}, \text{ where } 0 < \beta(\omega) \le \beta_1,\ \forall \omega \in \Omega.$$

Acknowledgments

This work has been partially supported by the Ministerio de Economía y Competitividad grant MTM2013-41765-P. M.A. Sohaly is indebted to the Egypt Ministry of High Education Cultural Affairs for its financial support [mohe-casem(2016)].

References

[1] T. T. Soong. *Random Differential Equations in Science and Engineering.* Academic Press, New York, 1973.

[2] J. W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*, volume 22. Springer Science & Business Media, 2013.

[3] J. C. Cortés, L. Jódar, L. Villafuerte, and R. J. Villanueva. Computing mean square approximations of random diffusion models with source term. *Mathematics and Computers in Simulation*, 76(1-3):44–48, 2007.

[4] M. A. Sohaly. Mean square convergent three and five points finite difference scheme for stochastic parabolic partial differential equations. *Electronic Journal of Mathematical Analysis and Applications*, 2(1):164–171, 2014.

[5] J. W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*, volume 22. Springer Science & Business Media, New York, 1998.

[6] G. W. Recktenwald. Finite-difference approximations to the heat equation. *Class Notes*, 2004.

[7] A. R. Mitchell and D. F. Griffiths. *The Finite Difference Method in Partial Differential Equations.* John Wiley, 1980.

Using a Digital Filter-Based Synthetic Turbulence for DNS Simulation OF Sprays Atomization

F.J. Salvador[♭] [*] Marco Crialesi-Esposito[♭],
J.-V. Romero[†], and M.-D. Roselló[†],

(♭) CMT-Motores Térmicos, Universitat Politècnica de València
Camino de Vera, s/n, Edificio 6D 46022 Valencia, España

(†) Instituto de Matemática Multidisciplinar, Universitat Politècnica de València
Camino de Vera, s/n, Edificio 8G, 2o 46022 Valencia, España

November 30, 2016

1 Introduction

In the last couple of decades, sprays have been a central point of investigation for Internal Combustion Engines. In fact, spray atomization is of fundamental relevance in combustion process and pollutant formation. As regulation on pollutant emission and energy efficiency are becoming more and more restrictive, the scientific community has invested considerable time and resources addressing the combustion process from both a theoretical and a practical standpoint, both with numerical and experimental techniques. In this context, it is nowadays evident that the actual knowledge on sprays, primary and secondary atomization as well as coalescence in the injection process is far from been complete as it becomes more and more relevant for applied research, especially in a complex frame as the ICE application, where the combination of injection velocities, pressures and characteristic length is

[*] fsalvado@mot.upv.es

quite unique and difficult to replicate and study with experimental technique on the appropriate length scale.

In this work, Direct Numerical Simulation (DNS) is used to provide a detailed description of the very first millimetres downstream the nozzle: this area is of fundamental importance in the formation of the spray, as it presents the regions in which the atomization begins ([7]) due to the combination of aerodynamic drag forces and air/liquid turbulence interaction. As a simulation environment, the code Paris-Simulator, developed in [1], is been chosen.

The results provided up to now with DNS for the Near-Field region have reportedly simulate low injection velocity, therefore pressure conditions unrealistic for Diesel ICE and rare for Gasoline Direct Injection ICE. Currently, only Lebas et al. [3] have simulated turbulence at the outlet of the nozzle, accounting for the turbulence generated by the fluid inside the nozzle duct. Many studies have related cavitating [5] and non-cavitating conditions [6] inside the nozzle with non negligible effects on the turbulence distribution at the nozzle outlet. Furthermore, it has been proved in previous works [2] that the higher the turbulence at the injector outlet is, the more the atomization affects the spray shape, as the intact core length reduces significantly and the atomization process starts earlier.

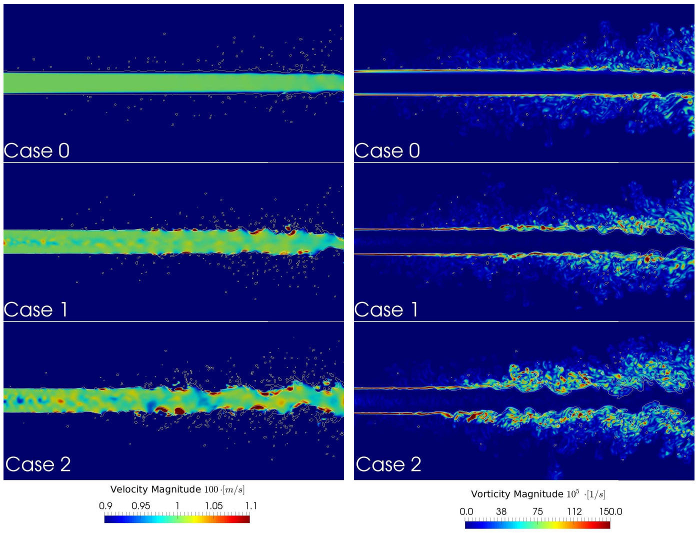
As appears evidently, still large improvement in the understanding of turbulent atomization can be achieved. This work investigate the effects of turbulence on the spray's shape and formation, while simulating the inlet turbulence with a methodology derived in [2] and applied to circular jet. In order to do so, 3 cases have been simulated with increasing inlet turbulent kinetic energy, in order to study high atomization in turbulent regimes typical of high speed jets, while minimizing the simulation domain.

2 Main results

The turbulence intensity u' and the turbulence length scale L used in the present work are listed in table 1.

	L	u'
case 0	0	0
case 1	$0.1D$	5%
case 2	$0.17D$	5%

Table 1: Inlet turbulence model parameters for all the cases simulated



(a) Velocity of the spray core (b) Vorticity effects induced by the inlet turbulence

Figure 1: Section of the early spray core at $t = 20\mu s$

Figure 1 highlights the changes generated by synthetic turbulence boundary condition. Figure 1(a) shows the velocity in the spray core and Figure 1(b) shows the vorticity effects generated by the inlet turbulence.

The synthetic turbulence in case 0 to 2 increase its area of influence as expected; in particular is clear how cases 1 and 2 presents a strong turbulent field. This variation of the velocity field in respect to case 0 generates many significant effects in the atomization processes. First, the external shape of the spray (represented in Figure 1 by the white line) gets heavily affected by the internal radial component of the velocity that force the formation of rims. The presence of liquid rims increase the effects of aerodynamic drag forces, exposing a wider region of liquid to shear stresses, therefore increasing the local vorticity as shown in Figure 1(b).

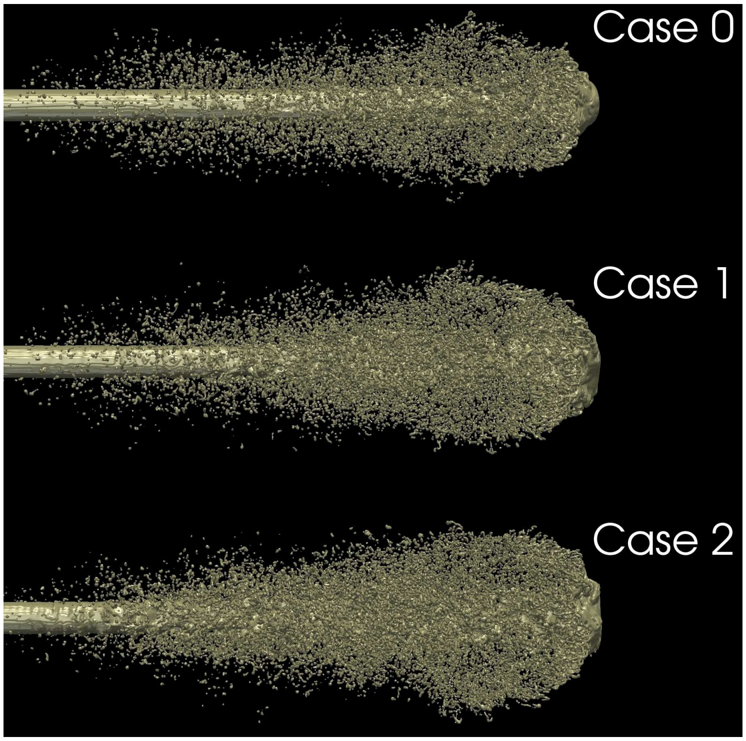


Figure 2: External aspect of the injected spray at $t = 20\mu s$

Figure 2 shows the external aspect of the spray at $t = 20\mu s$. As it can be clearly noted, the higher the turbulence induced, the sooner the atomization proces starts, shortening the *external non-perturbed length*. While comparing the 3 cases, case 2 display the formation of a earlier *atomization region*, due to the rims created in the *external non-perturbed length* that allows the creation of a dense cloud of droplets in the near-nozzle field. On the other hand it is evident that in case 0 the *intact core length* maintain an almost exact cylindrical shape up to the spray tip, due to the low nitrogen density and the low injection velocity. As a confirmation of the synthetic turbulence influence on the atomization process, in case 1 the droplet cloud (that will eventually define the spray angle) starts in an axial position between case 0 and case 2. Similar results have been obtained in [4] for a nozzle of similar size in different injection condition.

The different behaviour among the three cases can be quantified by means of the mass concentration (m_c), calculated as :

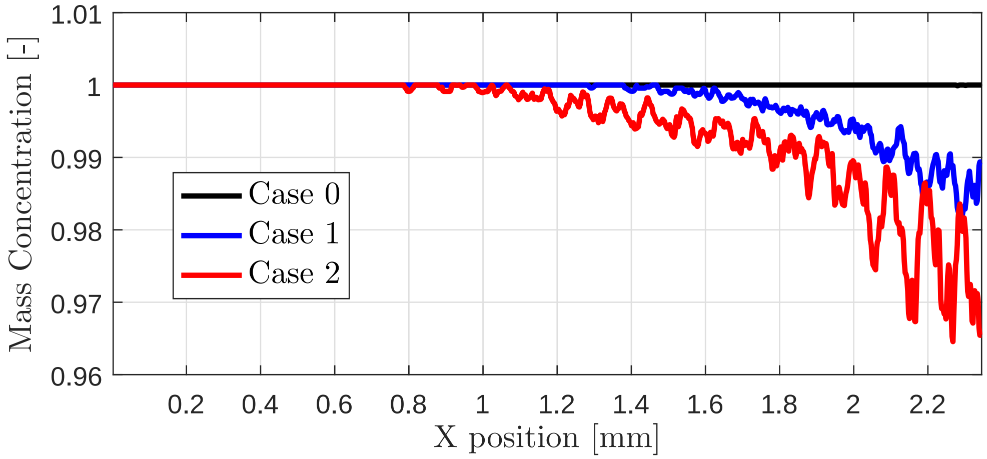


Figure 3: Axial average mass concentration

$$m_c = \frac{\rho_l \cdot C}{\rho_l \cdot C + \rho_g \cdot (1 - C)} \tag{1}$$

where ρ_l is the liquid density, ρ_g is the gas density and C is the color function defined in [1]. Figure 3 shows the time-averaged mass concentration in the spray axis. Once the spray is stabilized for the three cases, m_c is used to characterized the *intact core length*, which is directly related to the atomization intensity. As it can be seen in Figure 3 in the case 0, due to the poor atomization, the mass concentration in the axis is not perturbed, showing a value of 1 (pure liquid) in the spatial window analysed (up to 2.34 mm). However, in case 1 and especially in case 2, the intact core length drastically decreases as a result of the higher turbulence induced in the nozzle exit. This behaviour quantifies the earlier qualitative explanation of Figure 2, where the *external non-perturbed length* increase with the inlet turbulence lengthscale.

It is interesting to notice that the case with the highest turbulence level (namely case 2) in Figure 1(b) experience a core deformation that creates rims since the nozzle outlet. This generates an increase in the vorticity and in the local velocity field, finally increasing the atomization, as showed in 2. The rims are almost non existent in case 0 leading to a low vorticity field and, consequently to poor atomization, mainly focused around the spray tip where droplets are separating from the ligaments.

References

[1] Gilou Agbaglah, Sébastien Delaux, Daniel Fuster, Jérôme Hoepffner, Christophe Josserand, Stéphane Popinet, Pascal Ray, Ruben Scardovelli, and Stéphane Zaleski. Parallel simulation of multiphase flows using octree adaptivity and the volume-of-fluid method. *Comptes Rendus Mécanique*, 339(2-3):194–207, 2011.

[2] M. Klein, A. Sadiki, and J. Janicka. A digital filter based generation of inflow data for spatially developing direct numerical or large eddy simulations. *Journal of Computational Physics*, 186(2):652–665, 2003.

[3] R. Lebas, T. Menard, P.A. Beau, A. Berlemont, and F.X. Demoulin. Numerical simulation of primary break-up and atomization: DNS and modelling study. 35(3):247–260, mar.

[4] T Ménard, S Tanguy, and A Berlemont. Coupling level set/VOF/ghost fluid methods: Validation and application to 3D simulation of the primary break-up of a liquid jet. *International Journal of Multiphase Flow*, 33(5):510–524, 2007.

[5] Francisco Javier Salvador, Jorge Martínez-López, J. V. Romero, and M. D. Roselló. Computational study of the cavitation phenomenon and its interaction with the turbulence developed in diesel injector nozzles by Large Eddy Simulation (LES). *Mathematical and Computer Modelling*, 57(7-8):1656–1662, 2013.

[6] Francisco Javier Salvador, Santiago Ruiz, Jaime Gimeno, and Joaquin De la Morena. Estimation of a suitable Schmidt number range in diesel sprays at high injection pressure. *International Journal of Thermal Sciences*, 50(9):1790–1798, 2011.

[7] J. Shinjo and A. Umemura. Simulation of liquid jet primary breakup: Dynamics of ligament and droplet formation. *International Journal of Multiphase Flow*, 36(7):513–532, jul 2010.

Classification of Epidemiological Data By using ROC Curves

L. Acedo[1], R.-J. Villanueva[1], R.M. Shoucri[2]

(1) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de Valéncia, Valencia, Spain.
(2) Department of Mathematics & Computer Science,
Royal Military College of Canada, Kingston, Ontario, Canada

INTRODUCTION

Receiver Operating Characteristic (ROC) was originally introduced during World War II in radar applications for detection of signals in noise (Egan 1975). Several studies on the properties of ROC curves and the use of the area under the curve (AUC) as a comparative measure in decision-making and in medical applications are (see for instance Hanley & McNeil 1982, Bradley 1997, Fawcett 2006). There are also several interesting books on the topic (see for instance, Pepe 2003, Krzanowski & Hand 2009, Zou et al 2012).

This study presents an application of ROC curves for classification of observations of meningococcal disease between three epidemiological groups: (disease-free, endemic, endemic with oscillations). Data from three networks have been obtained from 45 simulations classified into three groups that give a relation between epidemic growth and time: Group 1 (disease-free) consists of 8 simulation curves, group 2 (endemic) consists of 16 curves, and group 3 (endemic with simulations) consists of 21 curves. Given an observation of disease variation over a period one or two months, how to classify it in one of the three groups indicated. In this study we use two methods to calculate the true positive percentage (tp, correct hit (also called true positive rate tpr)) and the false positive percentage (fp, false alarm (also called false positive rate fpr)) that correspond respectively to the y-axis and x-axis of the ROC curve, namely a Gaussian model (Marzban 2004, Brown 2006), and a Bootstrap method with cross-validation based on the likelihood ratio (Martinez & Martinez 2008).

MATHEMATICAL BACKGROUND AND RESULTS

We have 45 curves divided into three groups corresponding to three patterns of epidemiology:

Group 1 : curves 1, 15, 27, 28, 29, 38, 39, 42.
Group 2 : curves 2, 5, 11, 21, 23, 26, 31, 32, 34, 35, 40, 41, 43, 44, 45, 46.
Group 3 : curves 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 16, 17, 18, 19, 20, 22, 24, 25, 30, 33, 36.

For each curve in the three groups, we can calculate x_{k1}, x_{k2}, x_{k3}, x_{k4}, x_{k5},…….x_{kt} that correspond to the number of contaminations along curve k at increasing time t. The curve number k corresponds to one of the curves shown above, and the time t is time expressed in months. Figure 1 shows respectively the average values of the curves for each of the three groups drawn between t = 12 months and t = 164 months, the different trend of the three groups is evident.

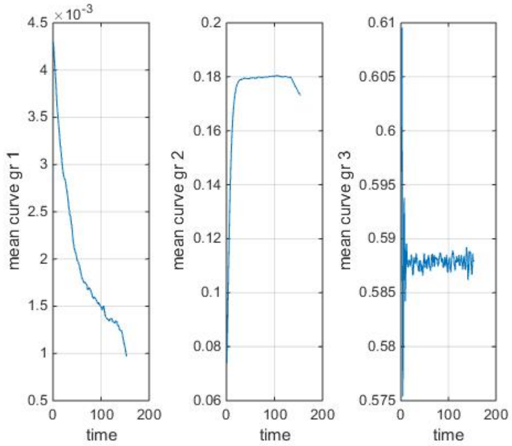


Figure 1: average curve for each of the three groups of data for t = 12 months to t = 164 months; group 1 = disease-free behaviour, group 2 = endemic behaviour, group 3 = cyclic behavior

A) Gaussian Model.

To understand how the algorithm works, choose for example a time tc = 100 months on curve number 27 (group 1), and we calculate xc = $x_{27,100} - x_{27,99}$ (or xc = $(x_{27,101} - x_{27,99})/2$) (or the absolute values). We want now to see if we can develop an algorithm that would allow us to classify xc correctly in group 1. It turns out the algorithm, described below, gives consistent correct results for all the trials that have been carried out.

We calculate the difference $Xd_k = x_{k,100} - x_{k,99}$ (or the absolute values) for all the curves in each of the three groups. We then calculate the differences $Xd = Xd_k - Xd_n$ for all $n \neq k$ at tc =100, for each of the three groups, as well as $dx = Xd_n - xc$ for the n curves of each group. The values of each row of the matrix Xd and the vector dx form approximately a binormal distribution as shown in Figs 2 to 4 (left).

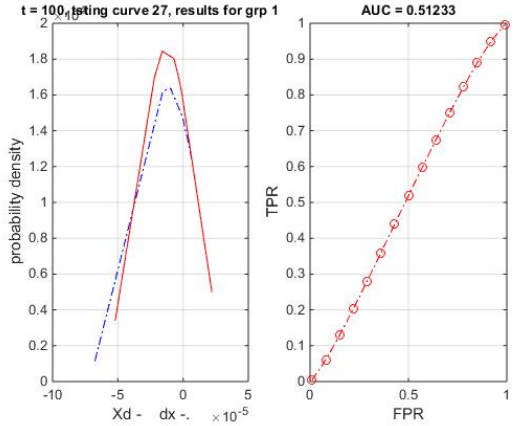


Figure 2: (left) Curves for Xd (-) and dx (-.) (approximate Gaussian distribution) corresponding to group 1; (right) ROC curve corresponding to classification of dx. Notice that the area AUC = 0.51233 under the ROC curve is minimum (correct classification) compared to AUC areas given in Figs. (3) and (4).

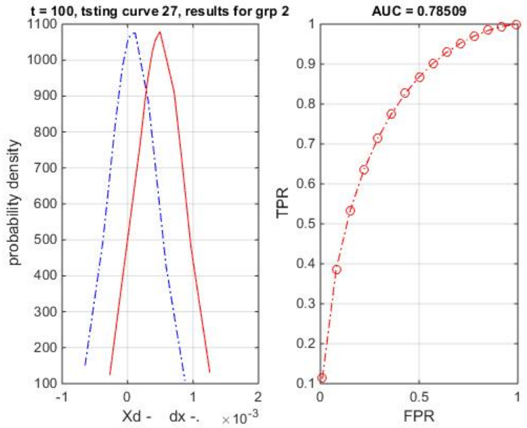


Figure 3: (left) Curves for Xd (-) and dx (-.) (approximate Gaussian distribution) corresponding to group 2; (right) ROC curve corresponding to classification of dx. Notice that the area AUC = 0.78509 under the ROC curve is greater than in Fig. 2.

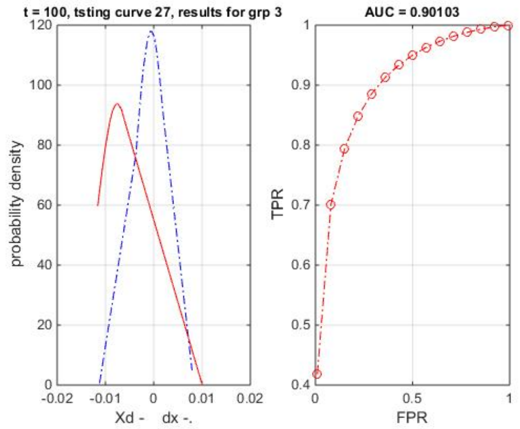


Figure 4: (left) Curves for Xd (-) and dx (-.) (approximate Gaussian distribution) corresponding to group 3; (right) ROC curve corresponding to classification of dx. Notice that the area AUC = 0.90103 under the ROC curve is greater than in Fig. 2.

The ROC curve is obtained by plotting tp (vertical axis) against fp (horizontal axis) as explained in the appendix. The algorithm works correctly if it indicates that dx belongs to group 1 as in our example, which corresponds to the minimum area AUC under the ROC (see Figs 2 to 4, right). Maximum area corresponds to maximum difference, minimum area corresponds to correct classification. The calculation based on the assumption of Gaussian distribution of data is outlined in the appendix and gives consistent results. It has been observed that keeping xd constant at tc = 100, but changing the value of t at which $Xd_k = x_{k,t} – x_{k,t-1}$ is calculated around tc still results in correct classification. This result indicates the possibility to classify dx in the correct group even if Xd is calculated at a value of tc that is not precisely known.

B) Bootstrap model

An approach based on maximum-likelihood calculation gives also consistent results, it is based on Bayes decision rule. In a two-class decision, Bayes rule can be written in the form

$$P(x|\omega_1)\ P(\omega_1) > P(x|\omega_2)\ P(\omega_2)\ \rightarrow\ x \text{ in } \omega_1$$

Rearranging terms, we get an expression of the likelihood ratio

$$L(x) = P(x|\omega_1) / P(x|\omega_2) > P(\omega_2) / P(\omega_1) = t_c$$

$L(x) > t_c \rightarrow$ x is in ω_1 $L(x) < t_c \rightarrow$ x is in ω_2

$P(\omega)$ = prior probability t_c = threshold

$P(x|\omega)$ = class-conditional probability (assume a normal distribution)

Calculation of dx and Xd as like in the previous section, the MATLAB code of the Bootstrap algorithm based on Bayes decision rule is given in Martinez & Martinez (2008). Results are shown in Figs 5 and 6.

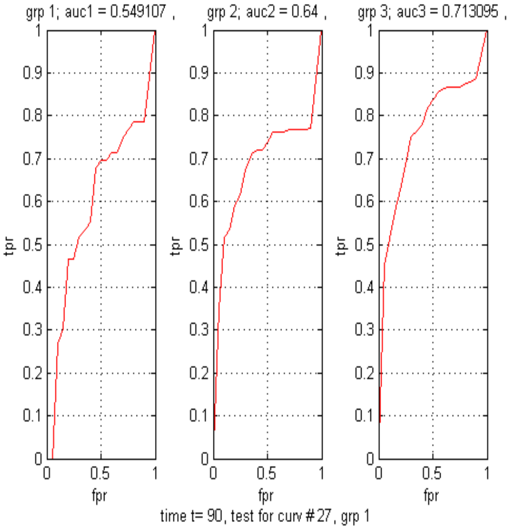


Figure 5: ROC curve based on the Bootstrap algorithm, calculated at tc = 90 months. Group 1 for curve 27 is the correct classification (minimum AUC).

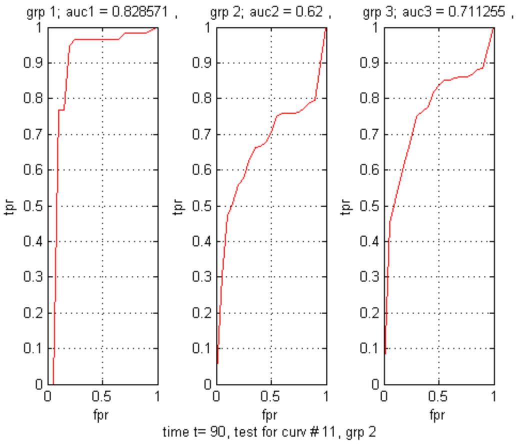


Figure 6: ROC curve based on the Bootstrap algorithm, calculated at tc = 90 months. Group 2 for curve 11 is the correct classification (minimum AUC).

APPENDIX

Calculation of the ROC curve

In order to draw the ROC curve and to calculate the area AUC under the curve, we need to calculate the true positive percentage tp = H and the false positive percentage fp = F) as follows (Marzban 2004)

$$H = \int_{t}^{\infty} L_1(x)dx \qquad F = \int_{t}^{\infty} L_0(x)dx \tag{1}$$

where the parameter *t* represents the decision threshold. We use a Gaussian distribution that is widely used in classification applications; in this case the likelihood of the classification in the *i*-th is given by

$$L_i(x) = \frac{1}{\sigma_i\sqrt{2\pi}}\exp(-\frac{1}{2}[\frac{x-\mu_i}{\sigma_i}]^2) \tag{2}$$

where μ_i is the mean and σ_i the standard deviation of the normal distribution. By substituting (2) in (1) we get

$$H = \Phi(\frac{\mu_1 - t}{\sigma_1}) \qquad F = \Phi(\frac{\mu_0 - t}{\sigma_0}) \tag{3}$$

where $\Phi(x)$ is the standard cumulative normal distribution that can be expressed as follows:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{\infty}^{x}\exp(-\frac{1}{2}z^2)dz \tag{4}$$

with

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2) \tag{5}$$

The ROC curve is thus the curve given by the relation

$$[F(t), H(t)] = \left[\Phi(\frac{\mu_0 - t}{\sigma_0}), \Phi(\frac{\mu_1 - t}{\sigma_1})\right] \tag{6}$$

By eliminating the parameter *t* from Eq. (3) we get the following expression for the ROC curve

$$H = \Phi(a + b\Phi^{-1}(F)) \tag{7}$$

where

$$a = \frac{|\mu_1 - \mu_0|}{\sigma_1} \text{ and } b = \frac{\sigma_0}{\sigma_1} \tag{8}$$

$F \in [0,1]$ is drawn along the horizontal axis and $H \in [0,1]$ is drawn along the vertical axis. The area AUC under the ROC curve can be expressed as follows

$$
\begin{aligned}
AUC &= \int_{-\infty}^{\infty} H(t)dF(t) \\
&= \int_{-\infty}^{\infty} \Phi(\frac{\mu_1 - t}{\sigma_1})\varphi(\frac{\mu_0 - t}{\sigma_0})(-\frac{1}{\sigma_0})dt \\
&= \Phi(\frac{a}{\sqrt{1+b^2}}) \qquad (9)
\end{aligned}
$$

The MATLAB code for the calculation can be summarized as follows (Brown 2006):

```
muo = mean(yo);                         % calculate mean and standard deviation
so = std(yo);

mu1 = mean(y1);                         % calculate mean and standard deviation
s1 = std(y1);

a = abs(mu1-muo)/s1;                    % parameters of the binormal curve
b = so/s1;

fpr = linspace(0.01:0.02:0.09);         % abscissae axis of the ROC curve
PHIim1 = norminv(fpr,0,1);              % calculation of $\Phi^{-1}$(fpr)
tpr = normcdf(a+b*PHIm1);               % calculation of tpr

auc = normcdf(a//sqrt(1+b*b));          % calculation of area under ROC
```

REFERENCES

1) Acedo L, Shoucri RM and Villanueva RJ (2015), A proposal to classify epidemiological behavior of a network model of meningococcal C using ROC curve, Modelling for Engineering &Human Behaviour, Sept 9-11, p. 276-285; Instituto Universitario de Matematica Multidisciplinar , Univ. Politècnica de Valéncia. http://jornadas.imm.upv.es/Modelling2015

2) Bradley AP (1997), The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognition,* 30, p. 1145-1159.

3) Brown CD and Davis HT (2006), Receiver operating characteristics curves and related decision measures: a tutorial, *Chemometrics and Intelligent Laboratory Systems*, 80, p. 24-38.

4) Egan JP (1975), *Signal detection theory and ROC analysis*. Academic Press.

5) Fawcett T (2006), An introduction to ROC analysis, *Pattern Recognition Letters,* vol. 27, p. 861-874.

6) Hanley JA & McNeil BJ (1982), The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143, p. 29-36.

7) Krzanowski WJ & Hand DJ (2009), *ROC Curves for Continuous Data*, CRC Press.

8) Martinez WL and Martinez AR (2008): *Computational Statistics Handbook with Matlab*, Chapman & Hall/CRC, chap. 10.

9) Marzban C (2004): The ROC Curve and the Area under it as Performance Measures, *Weather and Forecasting*, 19, p. 1106-1114.

10) Pepe MS (2003): *The Statistical evaluation of medical Tests for Classification and prediction*, Oxford Univ. Press.

11) Zou KH, Liu A, Bandos AI, Ohno-Machado L & Rockette HE (2012), *Statistical Evaluation of diagnostic performance Topics in ROC Analysis,* CRC Press.

Ball convergence and the dynamics for a novel iterative method free from the second derivative for solving equations in Banach spaces

I. K. Argyros[♭], Í. Sarría Martínez de Mendivil[†]
and J. A. Sicilia[†*]

(♭) Cameron University, Department of Mathematics Sciences
Lawton, OK 73505, USA.

(†) Universidad Internacional de La Rioja, Escuela de Ingeniería
Avenida Gran Vía Rey Juan Carlos I, 41, 26002 Logroño, Spain.

November 30, 2016

1 Introduction

Many problems in different disciplines can be brought in a form like

$$F(x) = 0,$$

using mathematical modelling [1, 2, 7]. The solutions of these equations can rarely be found in closed form. That is why most solution methods for these equations are iterative. The practice of Numerical Functional Analysis for finding such solutions is essentially connected to Newton-like methods [1, 2, 3]. Newton's method converges quadratically to x^* if the initial guess is close enough to the solution. Iterative methods of convergence order higher than two such as Chebyshev-Halley-type methods, require the evaluation of

[*] e-mail:juanantonio.sicilia@unir.net

the second Fréchet-derivative, which is very expensive in general. However, there are integral equations where the second Fréchet-derivative is diagonal by blocks and inespensive or for quadratic equations, the second Fréchet-derivative is constant. Moreover, in some applications involving stiff systems, high order methods are usefull. That is why it is important to study the convergence of high-order methods. We study the local convergence of the method defined for each $n = 0, 1, 2, \dots$ by

$$\begin{aligned} y_n &= x_n - \alpha F'(x_n)^{-1}F'(x_n) \\ x_{n+1} &= x_n - 2[I - \frac{1}{4}(F'(x_n)^{-1}F'(y_n) - I) \\ &+ \frac{3}{4}(F'(x_n)^{-1}F'(y_n) - I)^2](F'(x_n) + F'(y_n))^{-1}F(x_n), \end{aligned} \tag{1}$$

where x_0 is an initial point, α is a parameter and F' denotes the Fréchet-derivative of operator F. Method 1 was studied by Babajee, Cordero,Soleymani and Torregrosa in [4] in the special case when $X = Y = R^m$ and $\alpha = \frac{2}{3}$.

2 Local convergence analysis

We present the local convergence of method 1 in this Section. Let $L_0 > 0, L > 0, M > 0$ and $\alpha \in (-\infty, +\infty)$ be given parameters. It is convenient for the local convergence analysis of method 1 that follows to define some scalar functions and parameters. Define functions g_1, ρ, h_ρ on the interval $[0, \frac{1}{L_0})$ by

$$g_1(t) = \frac{1}{2(1 - L_0 t)}(Lt + |1 - \alpha|M),$$

$$\rho(t) = \frac{L_0}{2}(1 + g_1(t))t$$

$$h_\rho(t) = \rho(t) - 1$$

and parameters r_1, r_A by

$$r_1 = \frac{2(1 - M|1 - \alpha|)}{2L_0 + L}$$

and

$$r_A = \frac{2}{2L_0 + L}$$

Suppose that $M|1-\alpha| < 1$.

Then, we have that $g_1(r_1) = 1, 0 \leq g_1(t) < 1$ for each $t \in [0, r_1)$ and $0 < r_1 < r_A$. We also have that $h_\rho(0) = -1 < 0$ and $h_\rho(t) \to +\infty$ as $t \to \frac{1}{L_0}^-$. Then, it follows from the intermediate value theorem that function h_ρ has zeros in the interval $(0, \frac{1}{L_0})$. Denote by r_ρ the smallest such zero. Moreover, define functions g_2, h_2 on the interval $[0, r_\rho)$ by

$$g_2(t) = \frac{1}{2(1-L_0t)}[L + \frac{3L_0M^2(1+g_1(t))}{(1-L_0t)(1-\rho(t))}]t$$

and

$$h_2(t) = g_2(t) - 1.$$

We have that $h_2(0) = -1 < 0$ and $h_2(t) \to +\infty$ as $t \to r_p^-$. Denote by r_2 the smallest zero of function h_2 in the interval $(0, r_\rho)$. Set

$$r = min\{r_1, r_2\}. \tag{2}$$

Theorem 1 *Let $F : D \subset X \to Y$ be a Fréchet-differentiable operator. Suppose that there exist $x^* \in D, L_0 > 0, L > 0, M > 0$ and $\alpha \in (-\infty, +\infty)$ such that for each $x, y \in D$*

$$F(x^*) = 0, F'(x^*)^{-1} \in L(X, Y), \tag{3}$$

$$||F'(x^*)^{-1}(F'(x) - F'(x^*))|| \leq L_0||x - x^*||, \tag{4}$$

$$||F'(x^*)^{-1}(F'(x) - F'(y))|| \leq L||x - y||, \tag{5}$$

$$||F'(x^*)^{-1}F'(x)|| \leq M, \tag{6}$$

$$M|1-\alpha| < 1 \tag{7}$$

and

$$\bar{U}(x^*, r) \subseteq D, \tag{8}$$

where the radius r is defined by (2). Then, the sequence $\{x_n\}$ generated for $x_0 \in U(x^*, r) - \{x^*\}$ by method 1 is well defined, remains in $U(x^*, r)$ for each $n = 0, 1, 2, \ldots$ and converges to x^*. Moreover, the following estimates hold

$$||y_n - x^*|| \leq g_1(||x_n - x^*||)||x_n - x^*|| < ||x_n - x^*|| < r \tag{9}$$

and

$$||x_{n+1} - x^*|| \leq g_2(||x_n - x^*||)||x_n - x^*|| < ||x_n - x^*||. \tag{10}$$

Furthermore, for $T \in [r, \frac{2}{L_0})$ the limit point x^* is the only solution of equation $F(x) = 0$ in $\bar{U}(x^*, T) \cap D$.

3 Dynamical study

It is a well-known fact [4, 6, 7] that the study of the orbits of the critical points gives rise to the dynamical behavior of an iterative method. In particular, to determine if there exists any attracting strange fixed point or periodic orbit, the following question must be answered: For which values of the parameter, the orbits of the free critical points are attracting periodic orbits? In order to answer this question we draw the parameter space and we study it. Some of the results given by this study are that there exist several behaviors such us:

- Convergence to different cycles.
- Divergence to infinity.
- Convergence to strange fixed points.
- Or even chaotical behavior.

References

[1] Amat, S., Busquier, S., Gutiérrez, J.M., Geometric constructions of iterative functions to solve nonlinear equations, J. Comput. Appl. Math., 157 (2003), 197–205.

[2] I.K. Argyros, Computational Theory of Iterative methods, Series, Studies in Computational Mathematics, 15, Editors: C.K. Chui and L. Wuytack, Elsevier Publ. Co., New York, 2007.

[3] I.K. Argyros, D. Chen, Results on the Chebyshev method in Banach spaces, Proyecciones 12(2)(1993), 119-128.

[4] I.K. Argyros, Á. A. Magreñán, On the convergence of an optimal fourth-order family of methods and its dynamics, Appl. Math. Comput. 252, 336–346 (2015).

[5] D.K.R. Babajee, A. Cordero, F. Soleymani, J.R. Torregrosa, On a novel fourth-order algorithm for solving systems of nonlinear equations, Journal of Applied Mathematics, 2012, (2012), doi:10.1155/2012/165452.

[6] Magreñán, Á. A., Argyros, I.K., On the local convergence and the dynamics of Chebyshev?Halley methods with six and eight order of convergence, Journal of Computational and Applied Mathematics 298 (2016), 236–251.

[7] Magreñán, Á. A., Different anomalies in a Jarratt family of iterative root-finding methods, Applied Mathematics and Computation 233 (2014), 29–38.

Computing prices for the American option problems with multi assets

R. Company, V. Egorova, L. Jódar, F. Soleymani[*]

Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain

November 30, 2016

1 Introduction

Options with multi assets are based upon more than one underlying asset, unlike the well-known standard vanilla options. In this situation, due to the curse of dimensionality which is of exponential growth, the complexity of the problem grows when the dimensionality increases. That is to say, the number of unknowns for solving the corresponding PDE simultaneously grows exponentially [1].

In this extended abstract, American basket option pricing problem is considered for the general case of N underlying assets S_i in the following form [1]

$$\frac{\partial P}{\partial \tau} = \frac{1}{2}\sum_{i=1,j=1}^{M} \rho_{i,j}\sigma_i\sigma_j S_i S_j \frac{\partial^2 P}{\partial S_i \partial S_j} + \sum_{i=1}^{M}(r-d_i)S_i\frac{\partial P}{\partial S_i} - rP + F(P),$$

where $P(S_1, .., S_N, \tau)$ is the option price, $\tau = T - t$ is the time to maturity of the option, σ_i is the volatility of S_i, $\rho_{i,j}$ is the correlation between S_i and S_j, r is the risk free rate, d_i is the constant dividend yield of i-th asset and $F(P)$ is the rationality term [2]. The solution of the problem with rationality

[*] Speaker. E-mail address: fazso@upvnet.upv.es

parameter converges to the solution of the classical American option price when the free rationality parameter tends to infinity. Accordingly, in this work, we consider the rationality term as follows [3]:

$$F(P) = \lambda \left(P(\mathbf{S}, 0) - P(\mathbf{S}, \tau)\right)^{+}, \tag{1}$$

The initial condition for this class of basket options (for put) can be described by a general equation for the contract function [4] as follows:

$$P(\mathbf{S}, 0) = \left(E - \sum_{i=1}^{M} \alpha_i S_i \right)^{+}, \tag{2}$$

where E is the exercise price of the complete basket and α_i are the percentages in the set of assets.

Here our aim is to apply a numerically stable method of the cross derivative term removing and construct a stable numerical solution. To this end, for the correlation matrix $R = \{\rho_{i,j}\}$ which is positive semi-definite, LDL^T decomposition is constructed, where L is a lower triangular matrix and D is a diagonal matrix with positive diagonal elements.

Several conditions on the stability and positivity of the numerical solution are given. An example including three underlying assets are provided to show the convergence behavior of the proposed semi-discretization technique.

2 Removing the mixed derivatives

Using the dimensionless logarithmic substitution

$$x_i = \frac{1}{\sigma_i} \ln \frac{S_i}{E},\ i = 1, \ldots, M, \quad V(\mathbf{x}, \tau) = \frac{P(\mathbf{S}, \tau)}{E}, \tag{3}$$

where $\mathbf{x} = [x_1, \ldots, x_M]^T$, we attain

$$\begin{gathered} \frac{\partial V}{\partial \tau} = \frac{1}{2} \sum_{i=1,j=1}^{M} \rho_{ij} \frac{\partial^2 V}{\partial x_i \partial x_j} + \sum_{i=1}^{M} \delta_i \frac{\partial V}{\partial x_i} - rV + \frac{1}{E} F(EV), \\ x_i \in \mathbb{R}, \quad i = 1, ..., M, \quad 0 < \tau \leq T, \end{gathered} \tag{4}$$

where $\delta_i = \frac{r - q_i - \frac{\sigma_i^2}{2}}{\sigma_i}$.

Now by applying the linear transformation discussed before based on the LDL^T factorization of the correlation matrix [5]

$$\mathbf{y} = [y_1, \ldots, y_M]^T = C\mathbf{x}, \quad U(\mathbf{y}, \tau) = V(\mathbf{x}, \tau), \tag{5}$$

where $C = (c_{ij})_{1 \leq i,j \leq M} = L^{-1}$, we can obtain the following simplified transformed nonlinear PDE for multi-asset option pricing problem

$$\frac{\partial U}{\partial \tau} = \frac{1}{2}\sum_{i=1}^{M} D_{ii}\frac{\partial^2 U}{\partial y_i^2} + \sum_{i=1}^{M}\left(\sum_{j=1}^{M}\delta_j c_{ij}\right)\frac{\partial U}{\partial y_i} - rU + \frac{1}{E}F(EU), \tag{6}$$

where the cross derivative terms have been removed. Under transformations (3) and (5) the initial condition (2) takes the form

$$U(\mathbf{y}, 0) = \left(1 - \sum_{i=1}^{M}\alpha_i e^{\sigma_i x_i}\right)^+, \tag{7}$$

where

$$\mathbf{x} = [x_1, ..., x_M]^T = C^{-1}\mathbf{y}. \tag{8}$$

For dealing with the above time-dependent PDEs, one way is the method of lines based on the semi-discretization with respect to spatial variables which results in a system of (linear or nonlinear) ordinary differential equations (ODEs) in time with the corresponding matrix of coefficients A.

3 The semi-discretized system

The semi-discretization of the equation (6) can now be constructed using the central difference approximation for the spatial derivatives, resulting in the system of (nonlinear) ODEs of the form

$$\begin{aligned}\frac{du_{j_1,\ldots,j_M}}{d\tau} &= \frac{1}{2}\sum_{i=1}^{M} D_{ii}\frac{u_{j_1,\ldots,j_i-1,\ldots,j_M} - 2u_{j_1,\ldots,j_i,\ldots,j_M} + u_{j_1,\ldots,j_i+1,\ldots,j_M}}{h_i^2} \\ &+ \sum_{i=1}^{M}\left(\sum_{j=1}^{M}\delta_i c_{ij}\right)\frac{u_{j_1,\ldots,j_i+1,\ldots,j_M} - u_{j_1,\ldots,j_i-1,\ldots,j_M}}{2h_i} \\ &- ru_{j_1,\ldots,j_M} + \frac{1}{E}F(Eu_{j_1,\ldots,j_M}),\end{aligned} \tag{9}$$

which its stencil has only $2M + 1$ mesh points in contrast to $M^2 + M + 1$ mesh points based on the recent finite difference method given in [6].

4 Full discretization

Here we can introduce the following notations $i = 1, \dots, M$:

$$h_i = \beta_i h, \qquad c_i = \sum_{j=1}^{M} \delta_j c_{ij}, \tag{10}$$

$$d_i = \frac{D_{ii}}{\beta_i^2}, \quad d = \sum_{i=1}^{M} d_i, \qquad a_0 = -\frac{1}{h^2}\left(d + rh^2\right), \tag{11}$$

$$a_{-i} = \frac{1}{2h^2}\left(d_i - \frac{h}{\beta_i} c_i\right), \qquad a_{+i} = \frac{1}{2h^2}\left(d_i + \frac{h}{\beta_i} c_i\right), \tag{12}$$

Now the system (9) with the boundary and initial conditions can be presented in the following vector form

$$\begin{cases} \frac{d\mathbf{u}}{d\tau}(\tau) = A\mathbf{u}(\tau) + \lambda\left(\mathbf{u}(0) - \mathbf{u}(\tau)\right)^+, \\ \mathbf{u}(0) = [u_0(0), \dots, u_N(0)]^T, \end{cases} \tag{13}$$

where $u_j(0) = U(\xi_j, 0) = \left(1 - \sum_{i=1}^{M} \alpha_i e^{\sigma_i x_i(\xi_j)}\right)^+$, wherein $x_i(\xi_j) = \left(C^{-1}\xi_j\right)_i$ is the i-th entry of $C^{-1}\xi_j$.

If $k = \frac{T}{N_\tau}$, so $\tau^n = nk$, $n = 0, \dots, N_\tau$. Thus for full discretization we have [7]:

$$\mathbf{u}(\tau^{n+1}) = e^{Ak}\mathbf{u}(\tau^n) + \lambda \int_0^k e^{As}\left(\mathbf{u}(0) - \mathbf{u}(\tau^{n+1} - s)\right)^+ ds. \tag{14}$$

Now, by replacing $\mathbf{u}(\tau^{n+1} - s)$ by the known value $\mathbf{u}(\tau^n)$ corresponding to $s = k$, we attain

$$\int_0^k e^{As}\left(\mathbf{u}(0) - \mathbf{u}(\tau^{n+1} - s)\right)^+ ds \approx \left(\int_0^k e^{As} ds\right)\left(\mathbf{u}(0) - \mathbf{u}(\tau^n)\right)^+. \tag{15}$$

We use the accurate Simpson's rule $\int_0^k e^{As} ds \approx k\varphi(A, k)$, where $\varphi(A, k) = \frac{1}{6}\left(I + 4e^{A\frac{k}{2}} + e^{Ak}\right)$.

Denoting $\mathbf{u}^n = \mathbf{u}(\tau^n)$, we get the final explicit scheme

$$\mathbf{u}^{n+1} = e^{Ak}\mathbf{u}^n + k\lambda\varphi(A, k)\left(\mathbf{u}^0 - \mathbf{u}^n\right)^+, \quad \tau^n = nk, \ n = 0, \dots, N_\tau - 1. \tag{16}$$

Coefficients a_{-i} and a_{+i}, $i = 1, \ldots, M$, depend on d_i and c_i, see e.g. (10). If step size h is chosen as

$$h \leq \min_{1 \leq i \leq M} \frac{d_i}{|c_i|}, \tag{17}$$

then a_{-i} and a_{+i} are non-negative.

We remark that $u_i^0 \leq 1$, and from (16) u_i^{n+1} is a function g_i on the arguments $u_0^n, \ldots, u_N^n$, given by

$$u_i^{n+1} = g_i(u_0^n, \ldots, u_N^n) = \left(e^{Ak}\right)_i \mathbf{u}^n + k\lambda \left(\varphi(A,k)\right)_i \left(\mathbf{u}^0 - \mathbf{u}^n\right)^+ . \tag{18}$$

And furthermore by the non-negativity of e^{Ak} and $\varphi(A,k)$ one gets

$$\frac{\partial g_i}{\partial u_j^n} \geq \left(e^{Ak}\right)_{ij} - k\lambda \left(\varphi(A,k)\right)_{ij}, \quad 0 \leq i, j \leq N. \tag{19}$$

Now, we attain the following bound for the temporal step size

$$k < \frac{h^2}{d + (r+\lambda)h^2}. \tag{20}$$

Theorem 1. *With previous notation under conditions (17) and (20) the numerical solution $\mathbf{u}^n$ of the scheme (16) is non-negative and $\|\cdot\|_\infty$-stable, with $\|\mathbf{u}^n\|_\infty \leq 1$ for all values of $\lambda \geq 0$ and any time level $0 \leq n \leq N_\tau$.*

5 An experiment

Example 1. *As a numerical example we consider the European basket call option with no dividends and the following parameters (see [8], p. 76)*

$$\sigma_1 = 0.3,\ \sigma_2 = 0.35,\ \sigma_4 = 0.4,\ r = 0.04,\ \rho_{ij} = 0.5, \alpha_i = \frac{1}{3},\ T = 1,\ E = 100. \tag{21}$$

The spot price is chosen to be $S_1 = S_2 = S_3 = E$. The reference value $P_{ref} = 13.245$ is computed by using an accurate Fast Fourier Transform technique (see [8], Chapter 4). Since the considered option is of European style, penalty term is not necessary and λ is chosen to be zero.

The numerical results of the proposed method P_h are presented in the following table and compared with the sparse grid solution technique P_l on an equidistant grid of [8] and the method of [6] denoted by KM with rationality approach [3]. The numerical comparisons show the efficiency of the proposed scheme for multi-asset option pricing problems.

n	P_h	P_l	KM (with rationality)
8	11.4957	12.862	12.394
16	13.3457	13.150	13.055
32	13.3272	13.221	13.235
64	13.2470	13.239	13.241
Reference value (P)	13.245		

Table 1: Option price for a 3D case in Example 1.

References

[1] D. Tavella, C. Randall, Pricing Financial Instruments: The Finite Difference Method. New York: John Wiley and Sons, 2007.

[2] R. Company, V. Egorova, L. Jódar, and C. Vazquez, Finite difference methods for pricing american put option with rationality parameter: Numerical analysis and computing, J. Comput. Appl. Math., 304 (2016), 1-17.

[3] K.S.T. Gad, J.L. Pedersen, Rationality parameter for exercising American put, Risks, 3 (2015), 103-111.

[4] B. F. Nielsen, O. Skavhaug, A. Tveito, Penalty methods for the numerical solution of American multi-asset option problems," J. Comput. Appl. Math., 222 (2008), 3-16.

[5] R. Company, V. Egorova, L. Jódar, F. Soleymani, A mixed derivative terms removing method in multi-asset option pricing problems, Appl. Math. Lett., 60 (2016), 108-114.

[6] M. Yousuf, A.Q.M. Khaliq, R. Liu, Pricing american options under multi-state regime switching with an efficient L-stable method, Int. J. Compute. Math., 92 (2015) 2530-2550.

[7] S. Cox, P. Matthews, Exponential time differencing for stiff systems, J. Comput. Phys. 176 (2002) 430-455.

[8] C.C.W. Leentvaar, Pricing multi-asset options with sparse grids, PhD thesis, TU Delft, 2008.

Generalized centro-invertible matrices

L. Lebtahi[♭] [*], O. Romero[†], and N. Thome[‡]

(♭) Universitat de València, Facultat de Matemàtiques
C/ Dr. Moliner, 50, 46100 Burjassot, Valencia.

(†) Universitat Politécnica de València, Departamento de Comunicaciones
Camino de Vera s/n, 46022, Valencia,

(‡) Universitat Politécnica de València, Instituto Universitario de Matemática Multidisciplinar
Camino de Vera s/n, 46022, Valencia.

November 30, 2016

1 Introduction

Centrosymmetric matrices are those complex matrices $A \in \mathbb{C}^{n\times n}$ such that $AJ_n = J_nA$ where $J_n \in \mathbb{C}^{n\times n}$ is the matrix with 1's on the secondary diagonal and 0's otherwise. An important property of centrosymmetric matrices says that if A is a centrosymmetric matrix with γ linearly independent eigenvectors, then γ linearly independent eigenvectors of A can be chosen to be symmetric or skew-centrosymmetric. The particular case of matrices that commute with a permutation matrix was studied in [11] by Stuart and Weaver. This class of matrices has been widely studied considering their applications [3, 12, 13]. In [1], Abu-Jeib has used them to analyze some spectral properties of regular magic squares. After applying the Sinc collocation method to Sturm-Liouville Problems, the resulting matrices are centrosymmetric. Eigenspectrum properties of symmetric centrosymmetric matrices presented in [5] are applied for solving a generalized eigensystem of smaller dimensions than the original ones. Moreover, algorithms for solving

[*] e-mail: leila.lebtahi@uv.es, oromero@dcom.upv.es, njthome@mat.upv.es

centrosymmetric linear systems of equations are presented in [4]. Not only direct problems have been solved by using centrosymmetric matrices, but also the inverse eigenproblem and its approximation have been considered by Bai in [2]. By generalizing centrosymmetric matrices, centrohermitian matrices are defined as those matrices $A \in \mathbb{C}^{m\times n}$ satisfying $J_m A J_n = \overline{A}$, where $\overline{A}$ means the conjugate of the corresponding entries of A. Each square centrohermitian matrix is similar to a matrix with real entries and full information about the spectral properties of square centrohermitian matrices is given for instance in [10].

On the other hand, a more general situation where the equation $KA = A^{s+1}K$ is studied for involutory matrices K has been analyzed in [7, 8] for nonnegative integer values of s. They are called $\{K, s+1\}$-potent matrices and when $s = 0$ it is called a $\{K\}$-centrosymmetric matrix. Complementing this study, the case $s = -2$, that corresponds to generalized centro-invertible matrices, has been considered in [9]. In addition, some applications for image blurring/deblurring were developed. Clearly, $\{K, s+1\}$-potent matrices generalize centrosymmetric ones to a wider class of matrices.

2 Direct and inverse problems for generalized centro-invertible matrices

This article deals with generalized centro-invertible matrices. Specifically, we treat both problems: one of them is to generate generalized centro-invertible matrices and the other one solves the inverse problem.

The first algorithm solves the direct problem from a numerical point of view. In other words, this is a procedure to obtain examples of generalized centro-invertible matrices in an easy way.

Algorithm 1

Input: An involutory matrix $K \in \mathbb{C}^{n\times n}$.

Output: A matrix A belonging to $\mathbf{GCI}_K$.

Step 1 Generate an arbitrary nonsingular matrix $P \in \mathbb{C}^{n\times n}$.

Step 2 Fix the number of 1's corresponding to the parameter r; the number of -1's is $n - r$.

Step 3 Compute the matrix $A = KP \begin{bmatrix} I_r & O \\ O & -I_{n-r} \end{bmatrix} P^{-1}$.

End.

Algorithm 1 solves the problem of finding matrices A such that $KAK = A^{-1}$ for a given involutory matrix K. The next aim is the study of the inverse problem, that is, to find matrices K satisfying the previous matrix relation for a given matrix A.

ALGORITHM 2

Input: A matrix $A \in \mathbb{C}^{n\times n}$ for some integer $n \geq 2$.

Outputs: All the involutory matrices $K \in \mathbb{C}^{n\times n}$ such that $A \in \mathbf{GCI}_K$, if any such K exist.

Step 1 Compute A^{-1} by solving linear systems $Ax = e_i$, where e_i are the canonical vectors of $\mathbb{C}^n$.

Step 2 Construct the matrix $M := A^T \otimes I_n - I_n \otimes A^{-1}$.

Step 3 Find the general solution v to $Mv = \mathbf{0}$. The $n^2 \times 1$ vector v will depend on $d = \dim(\ker(M))$ arbitrary parameters.

Step 4 If $v = \mathbf{0}$, or equivalently, if $d = 0$, then go to Step 8.

Step 5 Treating v as $v = v(K)$ for an $n \times n$ complex matrix K containing d parameters, recover K from v.

Step 6 Determine the allowed values for the d arbitrary parameters so that $K^2 = I_n$. If there are no allowed parameter values, then go to Step 8.

Step 7 The output is the set of all of the matrices K whose parameter values are allowed. Go to End.

Step 8 "There is no involutory matrix $K \in \mathbb{C}^{n\times n}$ such that $A \in \mathbf{GCI}_K$."

End

Both algorithms can easily be implemented on a computer. For that, we have used the MATLAB package R2016a version 9.0.

References

[1] Abu-Jeib I. Centrosymmetric matrices: Properties and an alternative approach *Canadian Applied Mathematics Quarterly*, Volume(10), 4, 429–445, 2002.

[2] Bai Z. The inverse eigenproblem of centrosymmetric matrices with a submatrix constraint and its approximation *SIAM Journal on Matrix Analysis and Applications*, Volume(26), 4, 1100–1114, 2005.

[3] Cantoni A., and Butler P. Eigenvalues and Eigenvectors of Symmetric Centrosymmetrlc Matrices it Linear Algebra and its Applications, Volume(13), 275–288, 1976.

[4] El-Mikkawy M., and Atlan F., On Solving Centrosymmetric Linear Systems *Applied Mathematics*, Volume(4), 21–32, 2013.

[5] Gaudreau P., and Safouhi H. Centrosymmetric Matrices in the Sinc Collocation Method for Sturm-Liouville Problems arXiv:1507.06709v1 [math.NA].

[6] N. Higham, Function of matrices: Theory and Computation, Philadelphia, SIAM, 2008.

[7] Lebtahi L., Romero O., and Thome N. Characterizations of $\{K, s+1\}$-Potent Matrices and Applications *Linear Algebra and its Applications*, Volume(436), 293-306, 2012.

[8] Lebtahi L., Romero O., and Thome N. Algorithms for $\{K, s+1\}$-potent matrix constructions, *Journal of Computational and Applied Mathematics*, Volume(249), 157–162, 2013.

[9] Lebtahi L., Romero O., and Thome N. Generalized centro-invertible matrices with applications *Applied Mathematics Letters*, Volume(38), 106–109, 2014.

[10] Lee A. Centrohermitian and Skew-Centrohermitian Matrices *Linear Algebra and its Applications*, Volume(29), 205–210, 1980.

[11] Stuart J., and Weaver J. Matrices that commute with a permutation it Linear Algebra and its Applications, Volume(150), 255–265, 1991.

[12] Weaver J. Centrosymmetric (cross-symmetric) matrices, their basic properties, eigenvalues, and eigenvectors *American Mathematical Monthly*, Volume(92), 711–717, 1985.

[13] Zhongyun L. Some properties of centrosymmetric matrices and its applications *Numerical Mathematics*, Volume(14), 2, 136–148, 2005.

Multidimensional high-order methods for solving nonlinear models: applications to heat conduction equation [*]

A. Cordero[♭] [†], Esther Gómez[†], and Juan R. Torregrosa[♭]

(♭) Instituto de Matemáticas Multidisciplinar, Universitat Politècnica de València,,

Cno. de Vera s/n 46022 Valencia Spain,

(†) Facultat de Ciències Matemàtiques, Universitat de València,

November 30, 2016

1 Introduction

The construction of iterative methods for approximating the solution of systems of nonlinear equations, $F(x) = 0$, where $F : D \subset \mathbb{R}^n \to \mathbb{R}^n$ is a multidimensional function defined in an open convex D, is an important and interesting task in numerical analysis and applied scientific branches. With the improvement of computers, the problem of solving nonlinear systems by numerical methods has gained importance.

In 1990, Moré proposed a collection of nonlinear problems and most of them are phrased in terms of $F(x) = 0$. On the other hand, Grosan and Abraham also discussed the applicability of the systems of nonlinear equations in problems of neurophysiology, kinematics syntheses, chemical equilibrium, combustion and economics modeling. In addition, the reactor and steering problems are solved by phrasing them in the form of $F(x) = 0$. Moreover,

[*]This research was partially supported by Ministerio de Economía y Competitividad MTM2014-52016-C2-2-P and Generalitat Valenciana PROMETEO/2016/089.

[†]e-mail: acordero@mat.upv.es, ester.daih@gmail.com, jrtorre@mat.upv.es

Lin et al. also discussed the applicability of the systems of nonlinear equations in transport theory. These and other more examples allow us to affirm that finding the solution ξ of a nonlinear system $F(x) = 0$ is a classical and difficult problem with many applications in sciences and engineering. The best known method for finding a solution $\xi \in D$ is Newton's scheme,

$$x^{(k+1)} = x^{(k)} - [F'(x^{(k)})]^{-1}F(x^{(k)}),\ k = 0, 1, 2, \ldots,$$

where $F'(x^{(k)})$ is the Jacobian matrix of function F evaluated in the kth iteration.

Based on Newton's or Newton-like iterations, some higher order methods for computing a solution of nonlinear system $F(x) = 0$ have been proposed in the literature. The aim of these new schemes is to accelerate the convergence or to improve the computational efficiency. Sometimes researchers approximated the solution of a system of nonlinear equations with the help of schemes designed for nonlinear equations. Although it is a simple way to develop new schemes for systems of nonlinear equations, it is not always possible. So, different authors also tried some other approaches and procedures to develop new and higher-order methods. Recently, Sharma et al. [6] proposed fourth and six-order iterative methods based on weighted-Newton iteration, Artidiello et al. [1] proposed fourth-order methods based on the weight function approach, Wang et al. in [7] proposed seventh-order derivative-free iterative schemes based on the first order divided difference operator $[x, y; F]$. Other researchers have used quadrature formulae, Adomian polynomial, divided difference approach, ... for constructing iterative schemes to solve nonlinear systems.

In order to compare the different methods under the point of view of the computational cost, we recall the computational efficiency index, CI, introduced by the authors in [3], which combine the efficiency index defined by Ostrowski [5] and the number of products-quotients required per iteration. We define this index as $CI = p^{1/(d+op)}$, where p is the order of convergence, d is the number of functional evaluations and op is the number of products-quotients per iteration. Let us remark that for evaluating function F we need n scalar functional evaluations (the coordinate functions of F), whilst for evaluating Jacobian F' it is necessary to evaluate n^2 functions (all the entries of matrix F'). On the other hand, all the iterative methods for solving nonlinear systems require one or more matrix inversion, that is, one or more linear systems must be solved. So, the number of operations needed for solving a linear system plays in this context an important role.

We recall that the number of products and quotients required for solving a linear system by Gaussian elimination is $\frac{1}{3}n^3 + n^2 - \frac{1}{3}n$, where n is the size of the system. In addition, for solving q linear systems, with the same matrix of coefficients, by using LU decomposition we need $\frac{1}{3}n^3 + qn^2 - \frac{1}{3}n$ products-quotients. By using this information, in this paper we compare the computational efficiency index of our method and others known ones.

The main objective of this paper is to develop high-order iterative methods in such a way that they utilize as lower number of functional evaluations as possible as well as they have good stability properties. Specifically, we construct the following three-step iterative method

$$\begin{array}{rcl} y^{(k)} & = & x^{(k)} - [F'(x^{(k)})]^{-1}F(x^{(k)}), \\ z^{(k)} & = & y^{(k)} - [\alpha_1 I + \alpha_2 M + \alpha_3 M^2]\,[F'(y^{(k)})]^{-1}F(y^{(k)}), \\ x^{(k+1)} & = & z^{(k)} - [\beta_1 I + \beta_2 M + \beta_3 M^2]\,[F'(y^{(k)})]^{-1}F(z^{(k)}), \end{array} \quad (1)$$

where $M = [F'(y^{(k)})]^{-1}F'(x^{(k)})$, I denotes the identity matrix of size $n \times n$, and α_i, β_i, $i = 1, 2, 3$, are real parameters.

Under certain conditions of function F, we find values of the parameters for which an iterative method of order of convergence eight is constructed. Let us observe that, in each iteration, we need to evaluate function F three times and twice the Jacobian matrix F'. On the other hand, all the linear systems that we need to solve in each iteration (except Newton's step) have the same matrix of coefficients $F'(y^{(k)})$. These facts allow us to affirm that method (1) has a competitive computational efficiency index.

We check the numerical behavior of our method on a nonlinear one-dimensional heat conduction equation. A heat transfer problem is said to be one-dimensional if the temperature in the medium varies in one direction only and thus heat is transferred in one direction, and the variation of temperature and thus heat transfer in other directions are negligible or zero. For example, heat transfer through the glass of a window can be considered to be one-dimensional since heat transfer through the glass occurs predominantly in one direction (the direction normal to the surface of the glass) and heat transfer in other directions (from one side edge to the other and from the top edge to the bottom) is negligible.

To describe a heat transfer problem completely, an initial condition ($t = 0$) and two boundary conditions must be given for each direction of the coordinate system along which heat transfer is significant. Therefore, we need to specify two boundary conditions for one-dimensional problems, four boundary conditions for two-dimensional problems, and six boundary conditions for three-dimensional problems. Different authors have approximated the solution of these problems by means of numerical techniques, see for example [2, 4] and the references therein.

In our study a particular case is used, corresponding to the following heat conduction equation

$$u_t = \alpha(x)u_{xx} + \alpha'(x)u_x + u^2 + h(x,t), \quad 0 \leq x \leq 1, \quad t \geq 0, \tag{2}$$

where $\alpha(x) = (x-3)^2$ and $h(x,t) = -7(x-3)^2e^{-t} - (x-3)^4e^{-2t}$. The initial condition is $u(x,0) = (x-3)^2$ and the boundary conditions are

$$u(0,t) = 9e^{-t}, \quad u(1,t) = 4e^{-t}.$$

By applying an implicit method of finite differences we can transform problem (2) in a family of nonlinear systems, which provide the approximated solution in a time t_k from the approximated solution in t_{k-1}. We compare the numerical results given by our method and the corresponding obtained by others known ones. We also compare them with the exact solution $u(x,t) = (x-2)^2e^{-t}$ in order to analyze the stability and consistence of the different used methods.

References

[1] S. Artidiello, A. Cordero, J.R.Torregrosa, M.P. Vassileva, Multidimensional generalization of iterative methods for solving nonlinear problems by means of weight-function procedure, Appl. Math. Comput. 268 (2015) 1064–1071.

[2] M. Christou, C. Sophocleous, C. I. Christov, Numerical investigation of the nonlinear heat diffusion equation with high nonlinearity on the boundary, Appl. Math. Comput. 201(1-2) (2008) 729–738.

[3] A. Cordero, J.L. Hueso, E. Martínez, J.R. Torregrosa, A modified Newton-Jarratt's composition, Numer. Algor. 55 (2010) 87–99.

[4] R. Kouhia, On the solution of non-linear diffusion equation, Rakenteiden Mekaniikka (Journal of Structural Mechanics), 46 (4) (2013) 116–130.

[5] A.M. Ostrowski, Solution of equations and systems of equations, Prentice-Hall, Englewood Cliffs, New York, 1964.

[6] J.R. Sharma, R.K. Guna, R. Sharma, An efficient fourth order weighted-Newton method for systems of nonlinear equations, Numer. Algor. 2 (2013) 307–323.

[7] X. Wang, T. Zhang, W. Qian, M. Teng, Seventh-order derivative-free iterative method for solving nonlinear systems, Numer. Algor. 70 (2015) 545–558.

Modelling Acoustics on the Poincaré Half-Plane

Michael M. Tung$^{\flat *}$

($\flat$) Instituto de Matemática Multidisciplinar,
Universitat Politècnica de València,
Camino de Vera, s/n, 46022 Valencia, Spain.

November 30, 2016

1 Introduction

In recent years metamaterials have provided researchers and engineers with unprecedented tools for the design and construction of artificial devices with properties exceeding the possibilities found in nature. While optical metamaterials have been the focus of continued interest for the last decade, acoustic metamaterials have only recently drawn the attention of researchers [1]. The simulation of optical and acoustic phenomena with curved background spacetimes not only poses challenges in engineering, but also raises fundamental questions beyond their possible experimental verification, see *e.g.* Ref. [2].

In this work, we demonstrate the use of a novel technique based on a fundamental variational principle in combination with powerful differential-geometric methods to model acoustic wave propagation on a curved spacetime [3, 4, 5]. In particular we show how to implement acoustic wave propagation on the Poincaré half-plane model, $\mathbb{H}^2_+ = \{(x,y) \in \mathbb{R}^2 : y > 0\}$ endowed with the Poincaré metric [6]. It is the simplest and one of the most thoroughly investigated non-Euclidean models of two-dimensional hyperbolic geometry (see *e.g.* [7]), which makes it a suitable spacetime candidate for the implementation and study of an acoustic metamaterial.

[*] e-mail: mtung@imm.upv.es

We will comment on the design and implementation of such a spacetime with acoustic metadevices via the corresponding constitutive equations which relate the physical acoustic parameters to the underlying curved spacetime.

Finally, we outline how to derive within this framework the partial differential equation for the acoustic potential which describes wave propagation on the Poincaré half-plane. Apart from the harmonic time and x-dependence, it is possible to solve analytically the emerging Sturm-Liouville problem for the y-dependence and formally describe the solutions for the acoustic potential in terms of a modified Neumann series.

2 Field formulation of acoustics and variational principle

Variational principles are powerful methods in classical and field mechanics—this includes optics as an electromagnetic field theory—to define in a very concise manner the laws which govern their physical playground. The corresponding equation of motions are extremal solutions of the postulated action integrals and completely determine the physical behaviour of the system. Much of the mathematical charm and sophistication of variational principles lies in its coordinate-frame independent formulation. Moreover, Noether's theorem allows with almost no effort to shed light on the underlying symmetries of the theoretical model. In this formalism, physical laws have their equivalent in equations of motion with self-adjoint differential operators acting on the related field variables [8]. This gives rise to separable partial differential equations that frequently comprise Sturm-Liouville problems for one of the variables, so that its solutions may be obtained in an analytical or at least semi-analytical way.

For acoustics within a smooth spacetime M (endowed with Lorentzian metric $\mathbf{g}$ and negative signature such that $g = \det \mathbf{g} < 0$), we postulate [3] that the action integral $\mathscr{A}$ is stationary with respect to variations of the acoustic potential $\phi : M \to \mathbb{R}$:

$$\frac{\delta}{\delta\phi}\mathscr{A}[\phi] = \frac{\delta}{\delta\phi}\int\limits_{\Omega} d\mathrm{vol}_g \mathscr{L}(\phi_{,\mu}) = 0, \tag{1}$$

where the integration domain Ω is a bounded, closed set of spacetime and the invariant volume element is denoted by $d\mathrm{vol}_g = \sqrt{-g}\, dx^0 \wedge \ldots \wedge dx^3$.

The explicit form of the Lagrangian density function $\mathscr{L} : M \times TP \to \mathbb{R}$, P denoting the ambient space of the acoustic potential, is constrained by several symmetry requirements (energy-momentum conservation, locality and free-wave propagation), so that the simplest possible choice is [3]

$$\mathscr{L}(\phi_{,\mu}) = \tfrac{1}{2}\sqrt{-g}\, g^{\mu\nu}\phi_{,\mu}\phi_{,\nu}. \tag{2}$$

Note that if $\mathbf{v}$ denotes the local fluid velocity, p the acoustic pressure, ϱ_0 the density, and $c > 0$ the time-independent wave speed of the acoustic metamaterial, the gradient[1] appearing in Eq. (2) is

$$\phi_{,\mu} = \begin{pmatrix} p/c\varrho_0 \\ -\mathbf{v} \end{pmatrix}. \tag{3}$$

This expression encapsulates elementary relations of acoustics [9] and holds within a fixed laboratory frame.

Finally, substituting Eq. (2) into Eq. (1) yields the Euler-Lagrange equation for the acoustic potential. It is the wave equation which fully determines the dynamics of the acoustic system with underlying spacetime $(M, \mathbf{g})$.

For the actual implementation of such spacetime $(M, \mathbf{g})$ the acoustic engineer requires to fine-tune the mass-density tensor ϱ and bulk modulus κ in the laboratory—also called *physical space*—and relate them to their magnitude in the corresponding space with known acoustic wave propagation—called *virtual space*. Both spaces are connected by the *constitutive relations* [3]:

$$\kappa = \frac{\sqrt{-g}}{\sqrt{-\bar{g}}}\,\bar{\kappa}, \qquad \rho_0\rho^{ij} = \frac{\sqrt{-\bar{g}}}{\sqrt{-g}}\,\bar{g}^{ij}, \tag{4}$$

where without loss of generality $\bar{\rho}/\rho_0 \equiv 1$. For most cases the quantities in virtual space may be conveniently chosen $\bar{\kappa} = 1$ and $\bar{g}^{ij} = \delta^{ij}$.

3 Acoustic wave propagation on the Poincaré half-plane

Poincaré's half-plane $\mathbb{H}^2_+ = \{(x, y) \in \mathbb{R}^2 : y > 0\}$ is the upper 2D half-plane endowed with the Poincaré metric, the simplest case of two-dimensional

[1] Greek tensor indices indicate the full range of spacetime values, whereas Latin will only refer to the spatial values. Comma and semicolon are standard notation for partial and covariant derivatives, respectively.

hyperbolic geometry or, alternatively, a surface with a constant negative Gaussian curvature [6]. The line element of $\mathbb{R} \times \mathbb{H}^2_+$ spacetime in terms of the nonholonomic basis 1-forms θ^μ is given by

$$ds^2 = -\underbrace{(\,c\,dt\,)}_{\theta^0} \otimes (c\,dt) + \underbrace{\frac{dx}{y}}_{\theta^1} \otimes \frac{dx}{y} + \underbrace{\frac{dy}{y}}_{\theta^2} \otimes \frac{dy}{y}. \tag{5}$$

Cartan's structure equations allow to efficiently compute the curvature 2-form in the nonholonomic frame. From $\Omega^1{}_2 = \hat{R}^1{}_{212}\,\theta^1 \wedge \theta^2 = -d\theta^1 = (-1)\theta^1 \wedge \theta^2$, it then follows that the only independent component of the Riemann tensor is $\hat{R}^1{}_{212} = -1$. All other components vanish. As usual, the Ricci tensor and scalar are obtained by contraction. The components of the Ricci tensor in the nonholonomic frame are $\hat{R}_{00} = 0$ and $\hat{R}_{11} = \hat{R}_{22} = -1$. Thus, the Ricci scalar in both frames, nonholonomic and coordinate frame, is $R = \hat{R} = -2$.

Putting this information together, the associated Einstein tensor is also frame-independent and has only the following non-zero component

$$G_{00} = R_{00} + \eta_{00} = -1. \tag{6}$$

Matching Eq. (6) with the stress-energy tensor of a perfect fluid implies that for an observer falling along a geodesic the spacetime $\mathbb{R} \times \mathbb{H}^2_+$ is pressure-free and only consists of *exotic matter*.[2] Precisely this mass-energy distribution would generate $\mathbb{R} \times \mathbb{H}^2_+$ in physical spacetime.

The analogous acoustic space is implemented by a suitable choice of the physical parameters ϱ and κ. For this we only require the components of metric $\mathbf{g}$, immediately read off from Eq. (5), and the constitutive equations, Eqs. (4). Thus, we obtain the following simple prescription for the acoustic analogue of $\mathbb{R} \times \mathbb{H}^2_+$:

$$\kappa = \left(\frac{y}{y_0}\right)^2 \bar{\kappa}, \qquad \rho_0 \rho^{ij} = \left(\frac{y_0}{y}\right)^2 \delta^{ij}, \tag{7}$$

where $y_0 > 0$ is just a constant to fix the dimension.

[2] More concretely, it has constant negative mass-energy density $\rho_0 = -c^2/8\pi G$, where c is the speed of light and G is the gravitational constant.

The acoustic wave equation on the Poincaré half-plane agrees with the geodesics for field ϕ on a curved spacetime with the underlying metric provided by Eq. (5). Applying the variational principle, Eq. (1), for this spacetime gives the associated Euler-Lagrange equation

$$\Delta_{\mathbb{R}\times\mathbb{H}^2_+}\phi = -\frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} + y^2\,\Delta_{\mathbb{R}^2}\phi = 0, \qquad (8)$$

where Δ_M denotes the Laplace-Beltrami operator on manifold M.

The free-wave solution $\phi(t, x, y)$ for Eq. (8) displays a harmonic dependence in the time variable t and for the propagation along the x-axis. All of the non-trivial behaviour is contained in the propagation along the y-axis, as expected. Standard techniques [10] yield as a solution for the nontrivial y-dependence of Eq. (8) the modified Bessel functions $I_\alpha(y)$ and $K_\alpha(y)$. The numerical evaluation has been worked out by a finite-element analysis and directly by the expansion in terms of a modified Neumann series [11].

4 Conclusions

The equivalent of the Poincaré half-plane in transformation acoustics is a particularly fascinating approach, since the original model was among the first with hyperbolic geometry to be throughly studied in history. In oder to create such a spacetime, in Einstein's theory of gravity space would have to be filled with exotic matter, which will perhaps never be attainable. However, we have shown that the implementation within an acoustic metamaterial requires a bulk modulus and isotropic mass-energy density which display a specific dependence on height, *viz.* Eq. (7). This might be a workable alternative in the near future.

To derive the corresponding wave equation for the acoustic potential on the Poincaré half-plane, we have introduced a covariant variational principle and arrived at a complete description of acoustic free-wave phenomena with the underlying spacetime $\mathbb{R} \times \mathbb{H}^2_+$.

Acknowledgements

M. M. T. wishes to thank the Spanish *Ministerio de Economía y Competitividad* and the European Regional Development Fund (ERDF) for financial support under grant TIN2014-59294-P.

References

[1] S. A. Cummer, Transformation Acoustics. *Acoustic Metamaterials*, R. V. Craster, S. Guenneau (eds.), Springer Series in Materials Science **166**, pp. 197–218. Berlin, Springer Verlag, 2013.

[2] V. M. Redkov and E. M. Ovsiyuk, Quantum Mechanics in Spaces of Constant Curvature. New York, Nova Science Publishers Inc., 2012.

[3] M. M. Tung. A fundamental Lagrangian approach to transformation acoustics and spherical spacetime cloaking, *Europhys. Lett.*, **98**:434002–34006, 2012.

[4] M. M. Tung, E. B. Weinmüller. Gravitational Frequency Shifts in Transformation Acoustics, *Europhys. Lett.*, **101**:54006–54011, 2013.

[5] M. M. Tung, J. Peinado. A covariant spacetime approach to transformation acoustics. *Progress in Industrial Mathematics at ECMI 2012*, M. Fontes, M. Günther, N. Marheineke, (eds.), Mathematics in Industry **19**, pp. 335–340. Berlin, Springer Verlag, 2014.

[6] H. Poincaré. Théorie des Groupes Fuchsiens, *Acta Mathematica* **1**:1–62, 1882.

[7] J. Abardia Bochaca, *Tools to work with the Half-Plane Model*, Departament de Matemàtiques, Universitat Autònoma de Barcelona, `http://mat.uab.es/~juditab/enllac3.htm`, June 2015.

[8] C. Lanczos, The Variational Principles of Mechanics. New York, Dover Publications, 1970.

[9] F. P. Mechel, Formulas of Acoustics. Berlin, Springer Verlag, 2002.

[10] E. G. Kalnins, Separation of Variables for Riemannian Spaces of Constant Curvature. New York, Longman Scientific & Technical, 1986.

[11] C. G. Neumann, Theorie der Besselschen Funktionen, Leipzig, Teubner, 1867.

Homogenization techniques for the neutron transport equation using the finite element method

A. Vidal-Ferràndiz$^\flat$, S. González-Pintor†,
D. Ginestar$^{\ddagger *}$, G. Verdú$^\flat$, C. Demazière$^\natural$

($\flat$) Instituto de Seguridad Industrial: Radiofísica y Medioambiental,
Universitat Politècnica de València, Camino de Vera, s/n, 46022, Valencia,
($\dagger$) Department of Mathematical Sciences, Chalmers University of Technology
and the University of Gothenburg, SE-412 96 Gothenburg, Sweden.
($\ddagger$) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Camino de Vera, s/n, 46022, Valencia.
($\natural$) Department of Physics, Chalmers University of Technology,
SE-412 96 Gothenburg, Sweden.

1 Introduction

The neutron transport equation describes the distribution of neutrons inside a nuclear reactor. Solving this equation in a reactor core remains computationally challenging because of the complexity of the spatial domain, the energy dependence of the cross sections, and the fine angular discretization required. In order to achieve the necessary accuracy for the solution at a reduced computational cost, this problem is solved with a scheme consisting in a two-stage calculation through an homogenization process. First, the transport equation is solved accurately in isolated domains with proper assumptions regarding the boundary conditions. Then, these solutions are connected through a full domain calculation with a homogenized equation [1].

[*] e-mail:dginesta@mat.upv.es

This approach is justified because of two main reasons. Firstly, the solution domain shows a pattern with similar substructures repeated periodically, for which a subdomain representing one of these substructures with appropriate reflective boundary conditions is used to capture the fine scale behavior of the solution. Secondly, homogenizing these substructures reduces the angular dependence of the solution at the coarser scale, where the problem now is well represented by a lower order operator [2].

A first step in a homogenization methodology is to choose reactor properties that should be reproduced when the homogenized problem is solved. Usually, these quantities are the node averaged reaction rates, the surface-averaged net currents and the multiplicative constant of the reactor, which is implicitly conserved if the two aforementioned quantities are preserved. In the generalized equivalence theory [1], flux Discontinuity Factors (DFs) are introduced to preserve these quantities by imposing suitable discontinuities to the neutronic flux in the interfaces of the homogenized regions.

Typically, the preferred lower order operator used for the whole core calculations has been the neutron diffusion equation, while the homogenized regions are axially discretized assemblies of the core. As the computational resources available have increased, the homogenized regions have become smaller, down to pin size, while the low order operator has increased its order to take into account some transport effects that occur at these smaller scales, such as the P_N formulation. In this work, we focus on the implementation of the DFs for the correction of the homogenization error within a finite element method (FEM), when using P_N as a low order operator for pin-by-pin homogenization.

2 The one-dimensional P_N equations

We consider the eigenvalue problem associated with the multi-group, steady-state, neutron transport equation in slab geometry,

$$\begin{aligned}\left(\mu\frac{\mathrm{d}}{\mathrm{d}x}+\Sigma_t^g(x)\right)\psi^g(x,\mu) &= \sum_{g'=1}^{G}\int_{-1}^{1}\Sigma_s^{gg'}(x,\mu_0)\psi^{g'}(x,\mu')\mathrm{d}\mu' \\ &\quad +\frac{1}{\lambda}\sum_{g'=1}^{G}\frac{\chi^g(x)}{2}\nu\Sigma_f^{g'}(x)\int_{-1}^{1}\psi^{g'}(x,\mu')\mathrm{d}\mu', \qquad (1)\end{aligned}$$

where θ is the angle between the neutron and the x axis, $\mu = cos(\theta)$, θ_0 is the angle between neutrons and the scattered ones, $\mu_0 = cos(\theta_0)$.

A spherical harmonics approximation to the neutron transport equation in slab geometry, assumes that the angular dependence of the neutron flux distribution can be expanded in terms of $N+1$ Legendre polynomials,

$$\psi^g(x,\mu) = \sum_{n=0}^{N} \frac{2n+1}{2} \phi_n^g\left(x\right) P_n\left(\mu\right) . \tag{2}$$

The P_N equations can be expressed in matrix notations as [4],

$$\begin{gathered} \frac{\mathrm{d}\,\Phi_1}{\mathrm{d}x} + \mathbf{\Sigma_0}\Phi_0 = \frac{1}{\lambda}\mathbf{F}\Phi_0 \,, \qquad (3) \\ \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{n}{2n+1}\Phi_{n-1} + \frac{n+1}{2n+1}\Phi_{n+1}\right) + \mathbf{\Sigma_n}\Phi_n = 0 \,, \qquad \text{for } n = 1,\ ....,\ N. \end{gathered}$$

where,

$$\mathbf{\Sigma_n} = \begin{bmatrix} \Sigma_t^1 - \Sigma_{sn}^{11} & \dots & -\Sigma_{sn}^{1G} \\ \vdots & \ddots & \vdots \\ -\Sigma_{sn}^{G1} & \dots & \Sigma_t^G - \Sigma_{sn}^{GG} \end{bmatrix}, \mathbf{F} = \begin{bmatrix} \chi^1\nu\Sigma_f^1 & \dots & \chi^1\nu\Sigma_f^G \\ \vdots & \ddots & \vdots \\ \chi^G\nu\Sigma_f^1 & \dots & \chi^G\nu\Sigma_f^G \end{bmatrix}, \Phi_n = \begin{bmatrix} \phi_n^1 \\ \vdots \\ \phi_n^G \end{bmatrix}.$$

Using a linear change of variables the system of equation (3) can be expressed in a system of second order elliptic diffusive-like equation for the even moments. For example, the set of P$_5$ equations is

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathbf{D}\frac{\mathrm{d}}{\mathrm{d}x}U\right) + \mathbf{A}U = \frac{1}{\lambda}\mathbf{M}U \,, \tag{4}$$

where the effective diffusion matrix, $\mathbf{D}$, the absorption matrix, $\mathbf{A}$, and the fission matrix, $\mathbf{M}$, are defined as,

$$\mathbf{D} = \begin{bmatrix} \frac{1}{3}\mathbf{\Sigma_1}^{-1} & 0 & 0 \\ 0 & \frac{1}{5}\mathbf{\Sigma_3}^{-1} & 0 \\ 0 & 0 & \frac{1}{7}\mathbf{\Sigma_5}^{-1} \end{bmatrix}, \quad A_{ij} = \sum_{m=1}^{3} \mathbf{c_{ij}^{(m)}}\mathbf{\Sigma_m}, \quad M_{ij} = \mathbf{c_{ij}^{(1)}}\mathbf{F} \,, \tag{5}$$

and the coefficients matrix, $\mathbf{c^{(m)}}$,

$$\mathbf{c^{(1)}} = \begin{bmatrix} 1 & -\frac{2}{3} & \frac{8}{15} \\ -\frac{2}{3} & \frac{4}{9} & -\frac{16}{45} \\ -\frac{8}{15} & -\frac{16}{45} & \frac{64}{225} \end{bmatrix}, \mathbf{c^{(2)}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{5}{9} & -\frac{4}{9} \\ 0 & -\frac{4}{9} & \frac{16}{45} \end{bmatrix}, \mathbf{c^{(3)}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{9}{25} \end{bmatrix}. \tag{6}$$

3 Homogenization Method

In the generalized equivalence theory [1], flux discontinuity factors for diffusion are already introduced. For a given interface limiting two adjacent homogenized regions, the DFs are defined as interface constants f_j^-, f_j^+, such that

$$f_j^- u^{h-} = f_j^+ u^{h+} \ , \tag{7}$$

where u^{h-} and u^{h+} are the lateral (directional) limits of the homogenized flux in the interface. Thus, a definition of the discontinuity factors is

$$f_j^- = \frac{u^-}{u^{h-}} \, , \qquad f_j^+ = \frac{u^+}{u^{h+}} \, , \tag{8}$$

so continuity for the heterogeneous reconstructed flux is enforced [1].

The angular flux in one-dimensional geometries, $\psi(x,\mu)$, can be reconstructed with the different angular moments, ϕ_n, obtained from the P_N equations. Then, an homogeneous problem must be solved in the homogenized subdomain using odd reference flux moments as boundary conditions to calculate the homogeneous even flux moments. To calculate the discontinuity factors for the P_N equations, equation (8) can be extended to

$$f_n^+ = \frac{u_n^+}{u_n^{h+}}, \qquad f_n^- = \frac{u_n^-}{u_n^{h-}}, \qquad \text{for } n = 0, 2, \ldots, N \, , \tag{9}$$

where u_n^- and u_n^+ are the values at left and right extremes of the heterogeneous flux moments extracted from the transport solution and u_n^{h-} and u_n^{h+} are the left and right extremes of the homogeneous flux moments calculated with the P_N approximation in the homogenized region.

Using a triangulation $\mathcal{T}_h$ to split the original domain Ω into subdomains, and naming the set of all the edges and the set of interior edges of this triangulation as $\mathcal{E}_h$ and $\mathcal{E}_h^0$, respectively, problem (4), together with the continuity and discontinuity for the moments and its derivatives can be rewritten as

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\mathbf{D}\frac{\mathrm{d}}{\mathrm{d}x}U\right) + \mathbf{A}U = \frac{1}{\lambda}\mathbf{M}U \quad \text{in each } T \in \mathcal{T}_h, \tag{10}$$

$$[\![U]\!]_{f|_e} = 0 \quad \text{on each } e \in \mathcal{E}_h, \tag{11}$$

$$\left[\!\!\left[D\frac{\mathrm{d}}{\mathrm{d}x}U\right]\!\!\right]_{|e} = 0 \quad \text{on each } e \in \mathcal{E}_h^0. \tag{12}$$

where the jumps $[\![\cdot]\!]_f$ are defined as follows

$$[\![u]\!]_f = (f_n^+ u_n^+ - f_n^- u_n^-) \quad \text{on } e \in \mathcal{E}_h, \tag{13}$$

f defining the jumps imposed to the solution, u, over each particular edge, e, for a coarse mesh triangulation $\mathcal{T}_{h_0}$. A scheme to approximate the problem defined by equations (10), (11) and (12) has been implemented using an Interior-Penalty Finite Element Method (IP-FEM) as similar to the one developed for the diffusion theory presented in [3].

4 Numerical Results

To study the performance of the homogenization method exposed for the P_N equations, a one-dimensional reactor configuration based in the C5G7 benchmark is defined. The reactor is solved in a heterogeneous way using different spherical harmonics approximations. The pin power averages have an error less than 2.5% while the eigenvalue error is above 650 pcm. From these results, it can be seen that low order spherical harmonics approximations cannot reproduce accurately sub-pin heterogeneities and homogenization methods are necessary to compute accurately this problem.

Homogenized results at pin level are presented in Table 1, both without using Discontinuity Factors (No DFs), and using Pin-level Discontinuity Factors (PDFs). They provide accurate results for pin power averages, specially if the proposed pin discontinuity factors are used. Increasing the order of the spherical harmonics approximation, N, in the homogenized problem the eigenvalue error and pin averaged errors are reduced. In this way, the eigenvalue error can be reduced to 38 pcm and the pin average error for the neutronic power to less than 1%.

Acknowledgements

The work has been partially supported by the Ministerio de Economía y Competitividad under projects ENE 2014-59442-P and MTM2014-58159-P, the Generalitat Valenciana under PROMETEO II/2014/008 and the Universitat Politècnica de València under FPI-2013. The work has also been supported partially by the Swedish Research Council (VR-Vetenskapsrådet) within a framework grant called DREAM4SAFER, research contract C0467701.

Table 1: Homogenized results at pin level.

Transport	**Homogenization**	**Eigenvalue**		**Pin RMS (%)**
Approx.	**Method**	k_{eff}	Δk_{eff} (pcm)	Neutronic Power
P_1	No DFs	1.12458	854	1.66
	PDFs	1.13331	19	1.43
P_3	No DFs	1.12795	517	1.20
	PDFs	1.13350	55	0.92
P_5	No DFs	1.13053	259	0.94
	PDFs	1.13350	38	0.78
Transport Reference		1.13312		

References

[1] Smith K. S. Assembly homogenization techniques for light water reactor analysis. *Progress in Nuclear Energy*, 17.3:303-335, 1986.

[2] Sanchez, R. Assembly homogenization techniques for core calculations. *Progress in Nuclear Energy*, 51.1:14-31, 2009.

[3] Vidal-Ferràndiz A., González-Pintor S., Ginestar D., Verdú G., Asadzadeh M., Demazière C., Use of discontinuity factors in high-order finite element methods, *Annals of Nuclear Energy*, 87:728-738, 2016.

[4] Hamilton, Steven P., and Thomas M. Evans, Efficient solution of the simplified PN equations, *Journal of Computational Physics*, 284:155-170, 2015.

Eigenvalue problems associated with the neutron diffusion equation

A. Carreño[♭], A. Vidal-Ferrándiz[♭], D. Ginestar[†][*], G. Verdú[♭]

(♭) Instituto de Seguridad Industrial: Radiofísica y Medioambiental,

Universitat Politècnica de València, Camino de Vera, s/n, 46022, Valencia,

(†) Instituto Universitario de Matemática Multidisciplinar,

Universitat Politècnica de València, Camino de Vera, s/n, 46022, Valencia

November 30, 2016

1 Introduction

Modal analysis has been efficiently used to study different problems in reactor physics. In this sense, several eigenvalue problems can be defined for neutron transport equation, each one with a particular application [1]. However, due to the complexity of transport equation, the multigroup neutron diffusion equation is widely used as an approximation to this equation. For two energy groups and assuming that the delayed neutron precursors are in a steady state, the equations for λ, γ and α eigenvalue problems are

$$\begin{pmatrix} -\vec{\nabla}(D_1\vec{\nabla}) + \Sigma_{a_1} + \Sigma_{12} & 0 \\ -\Sigma_{12} & -\vec{\nabla}(D_2\vec{\nabla}) + \Sigma_{a_2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \frac{1}{\lambda} \begin{pmatrix} \nu\Sigma_{f1} & \nu\Sigma_{f2} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \tag{1}$$

$$\begin{pmatrix} -\vec{\nabla}(D_1\vec{\nabla}) + \Sigma_{a_1} + \Sigma_{12} & 0 \\ 0 & -\vec{\nabla}(D_2\vec{\nabla}) + \Sigma_{a_2} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \frac{1}{\gamma} \begin{pmatrix} \nu\Sigma_{f1} & \nu\Sigma_{f2} \\ \Sigma_{12} & 0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \tag{2}$$

[*] e-mail:dginesta@mat.upv.es

$$\left(- \begin{pmatrix} -v_1\vec{\nabla}(D_1\vec{\nabla}) + v_1(\Sigma_{a_1} + \Sigma_{12}) & 0 \\ -v_2\Sigma_{12} & -v_2\vec{\nabla}(D_2\vec{\nabla}) + v_2\Sigma_{a_2} \end{pmatrix} \right. \\ \left. +(1-\beta)\begin{pmatrix} v_1\nu\Sigma_{f1} & v_1\nu\Sigma_{f2} \\ 0 & 0 \end{pmatrix} \right) \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \alpha \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \tag{3}$$

respectively. In the case of $\beta = 0$ in (3) we obtaining the intermediate alpha modes, otherwise the prompt alpha modes are obtained. A high order finite elements method is used to discretize equations (1), (2) and (3). The finite element method used in this work has been implemented using the open source finite elements library Deal.II [2]. After the discretization all the modes problems can be reduced to an algebraic eigenvalue problem of the form

$$A\bar{\psi} = \delta\bar{\psi}, \text{ with } \delta = \lambda, \alpha, \gamma, \tag{4}$$

which has to be solved to find a set of dominant eigenvalues and their corresponding eigenfunctions.

2 Solution of eigenvalues problems

A first approach to solve the partial eiganvalue problem (4) is to use the Krylov-Schur method implemented in the library SLEPc [3]. This method uses the Arnoldi process to construct a Krylov subspace and performs a Krylov-Schur decomposition. Also an alternative method to solve the partial eigenvalue problem (4) is proposed. Thus, given a partial eigenvalue problem of the form

$$AV = V\Lambda, \tag{5}$$

where Λ is a diagonal matrix with the desired eigenvalues. It is assumed that the eigenvectors can be factorized as

$$V = ZS \ , \tag{6}$$

where $Z^{\mathrm{T}}Z = I$. Then the Newton method is used to solve the problem (7)

$$F(Z, \Lambda) := \begin{pmatrix} AZ - Z\Lambda \\ W^{\mathrm{T}}Z - I_q \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \ . \tag{7}$$

where W is a fixed matrix. The resulting system of equations is uncoupled using a Rayleigh–Ritz procedure (see [5] for more details).

3 Numerical results

To study the performance of methods to determine the λ-modes, α-modes and γ-modes, two benchmarks reactors have been considered.

3.1 Homogeneous reactor

The simplest theoretical reactor is one consisting of a 3D prismatic reactor with homogeneous material since it can be solved analytically for all its eigenvalues and compared with finite element method results. Considering the error

$$\varepsilon_{eig}(pcm) = 10^5 \left(\frac{|\delta_i - \delta_i^*|}{|\delta_i|} \right), \quad \delta = \lambda, \alpha, \gamma,$$

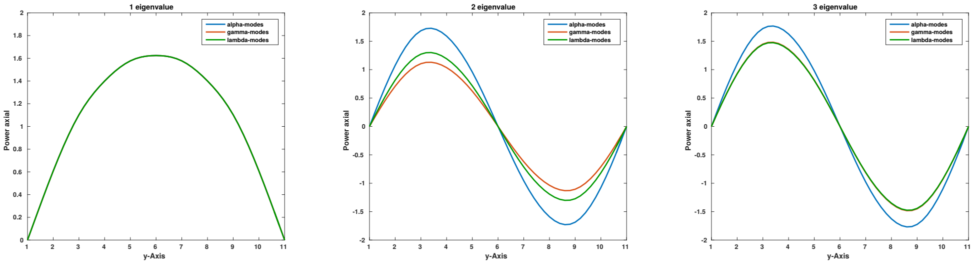
the errors between the analytical and numerical solution are displayed in Table 1. This Table shows that a good approximation is obtained using polynomials of degree two in the finite element method and one refinement of the mesh.

Table 1: Eigenvalues and errors for the homogeneous reactor.

FED	Refi.	λ_1	ε_{eig}	α_1	ε_{eig}	γ_1	ε_{eig}
2	0	0.0656174	7.09	-36550.56	40.5485	0.22761	38.1135
2	1	0.0656609	$3.7\,10^{-2}$	-36539.18	9.4009	0.22769	0.16780
3	0	0.0656608	1.2	-36539.18	9.4009	0.22769	0.1680
3	1	0.0656609	$2.3\,10^{-3}$	-36539.13	9.264	0.22769	0.0077
Anal. solut.:		0.0656609		-36535.74		0.22769	

3.2 Langenbuch 3D reactor

The Langenbuch 3D benchmark [6] is chosen to compare the different modes and eigenvalue solvers in a more realistic case. The different values of λ, γ and α with finite element degree equal to 2 and 1 refinement of the mesh are displayed in Table 2. It is observed that this reactor is quasi-critical since the dominant λ and γ are near 1 and α is near to 0. The radial profiles for different modes are represented in 1. For the first eigenvalue, the functions are equal, and for the second one and third one, they are very similar. So, near criticality the eigenfunctions for all modes are similar.



(a) First eigenvalue (b) Second eigenvalue (c) Third eigenvalue

Figure 1: Radial profiles

Table 2: Eigenvalues in Langenbuch reactor

Modes	1st Mode	2nd Mode	3rd Mode	4th Mode
λ-modes	0.999695	0.967682	0.967682	0.95165
γ-modes	0.999828	0.981648	0.981648	0.972488
α-modes	-8.1666	-876.144	-876.144	-1285.58

The γ eigenvalues have the advantage that they are not limited to systems with fissions like the λ eigenvalues, however it has the limitation that if $|\gamma - 1|$ is lower than $|\lambda - 1|$ for the same configuration, this implies that the eigenvalue is less sensitive, and its numerical calculation requires more computational time to reach the convergence than the λ-modes calculations.

Regarding α-eigenvalue, it seems that it is the best one for transient analysis using modal decomposition methods, due to the nature of this kind of modes. Nevertheless, as the γ-modes problem, the computational time for α-modes is larger than for the computation of the λ-modes.

So, it is proposed computing α-modes and γ-modes from λ-modes using the modified Newton method (MBNM). First, it is supposed that the eigenvectors of α and γ-modes problem can be written as the lineal combination of eigenvectors of the λ modes and a Rayleigh-Ritz process is applied to obtain a initial approximation for each kind of modes, which is used to initiate the modified block Newton method. The computational time with both methods are in Table 3. It is observed that the MBNM is competitive for computing α and γ modes if the λ modes have been previously computed.

Table 3: Computational time (s) using Krylov-Schur and MBNM methods.

Method	Krylov-Schur	MBNM
α-modes	367.7	179.56
γ-modes	249.3	170.2

Acknowledgments

This work has been partially supported by Spanish Ministerio de Economía y Competitividad under projects ENE2014-59442-P, MTM2014-58159-P and BES-2015-072901.

References

[1] G. Velarde, C. Ahnert, J.M. Aragones. Analysis of the eigenvalue equations in k, lambda, gamma, and alpha, applied to some fast and thermal-neutron systems Nuclear Science and Engineering, 66(3), 284–294, 1978.

[2] W. Bangerth and R. Hartmann and G. Kanschat. deal.II – a General Purpose Object Oriented Finite Element Library. ACM Trans. Math. Softw., 33(4), 24/1–24/27,2007.

[3] Vicente Hernandez and Jose E. Roman and Vicente Vidal. SLEPc: A Scalable and Flexible Toolkit for the Solution of Eigenvalue Problems. ACM Trans. Math. Software, 31(3),351–362,2005.

[4] R. Lösche, H. Schwetlick and G. Timmermann. A modified block Newton iteration for approximating an invariant subspace of a symmetric matrix. Linear Algebra and its Applications, 275, 381–400, 1998.

[5] S. González-Pintor, D. Ginestar, G. Verdú. Modified Block Newton method for the lambda modes problem. Nuclear Engineering and Design, 259, 230–239, 2013.

[6] S. Langenbuch, W. Maurer, W. Werner. Coarse-Mesh Flux-Expansion Method for the Analysis of Space-Time Effects in Large Light Water Reactor Cores. Nuclear Science and Engineering, 63(4), 437–456, 1977.

A STOCHASTIC DYNAMIC MODEL TO EVALUATE THE INFLUENCE OF ECONOMY AND WELL-BEING ON UNEMPLOYMENT CONTROL

Joan C. Micó[1], María T. Sanz[2], Antonio Caselles[3] and David Soler[4]

Abstract

This paper presents a stochastic dynamic mathematical model to study the evolution of the unemployment rate and other relevant related variables in a country. This model is composed by three basic interrelated subsystems: demographic, economic and well-being ones. A key aspect of this model is that it considers three UN well-being variables simultaneously: Human Development Index, Gender Empowerment Index and Gender Differentiation Index. These variables involve key concepts for human development, as Health, Education, Economy and Female Labor. The model has been fitted for the case of Spain in the 2002-2014 period. Finally, several tentative scenarios and strategies have been tested to reduce the unemployment rate in Spain in the horizon of year 2025.

Keywords: Unemployment rate; United Nations well-being variables; sex/age-structured population dynamics; stochastic model; forecasting.

1. Introduction

Unemployment is one of the most important problems today in the world, and particularly in many southern European countries. For instance, Greece had a 24.14% unemployment rate in March 2016 and Spain had a 21% unemployment rate also in March 2016. But there are countries with more than 40% of jobless rate, like Bosnia and Herzegovina, Congo or Haiti [1].

This is not only an economic problem, but also a social and a physical and mental health problem. In addition, politicians and professionals do not succeed in reducing the current unemployment rate and in the way to create jobs without reducing the well-being of a country. In order to progress in the way to provide a tool to tackle this problem, a mathematical model that relates the unemployment rate with demographic, well-being and economic variables is presented in this paper. Some of the input variables can be controllable and others do not. With this information, governments can design feasible strategies to be simulated under different scenarios. The results of the simulations lead to assign suitable values to the controlled input variables in order to decrease the unemployment rate respecting well-being.

The reviewed literature about the problem here treated reveals several papers that present some kind of relation between demographic or well-being variables and the unemployment rate ([2, 3, 4]). Also, recent studies that build models trying to connect unemployment with other socioeconomic variables consider a low number of variables but do not simulate and test different strategies to control unemployment ([5, 6])

[1] Departament de Matemàtica Aplicada, Universitat Politècnica deValència.
Camino de Vera, s/n, 46022 Valencia, Spain. e-mail: jmico@mat.upv.es

[2] Departament de Didàctica de les Matemàtiques, Universitat de València
Avda. Tarongers, 4. 46022 València, Spain. e-mail: m.teresa.sanz@uv.es

[3] Departament de Matemàtica Aplicada, Universitat de València.,
C/ Doctor Moliner, 50. 46100 Burjassot-València, Spain. e-mail: antonio.caselles@uv.es

[4] Institut Universitari de Matemàtica Pura i Aplicada, Universitat Politècnica de València.
Camino de Vera, s/n, 46022 Valencia, Spain. e-mail: dsoler@mat.upv.es

On the other hand, the use of the United Nations (*UN*) well-being variables seems a universal way to measure the evolution of well-being in the context of a mathematical model. These variables allow studying well-being by countries with general models whose equations depend only on the particular historical data ([7], [8], [9])

In the present paper the model proposed by Caselles et al. [9] is generalized (widen) to study unemployment control in a country through the demographic, economic and well-being variables. Thus, the model becomes a richer tool than the classical approaches that use only economic variables to face the reduction of the unemployment rate. This new stochastic model has been fitted for the case of Spain in the 2002-2014 period, and several tentative scenarios and strategies have been tested in order to reduce this rate in the horizon of year 2025.

The methodology used to construct the model is the General Modeling Methodology (*GMM*), proposed by Caselles [10, 11], which is implemented by using *SIGEM*: the Complex-Systems Based Programmes' Generator. The *GMM* is a hypothetic-deductive methodology to construct complex models which contains Data Mining Methods in some of its phases.

2. The mathematical model

The mathematical model is divided into three subsystems: demographic, well-being and economic. The links between these three subsystem are given in Fig. 1.

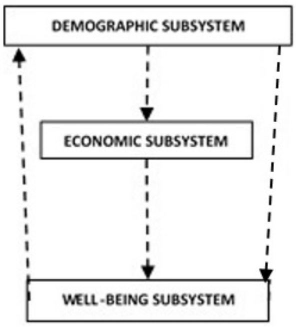


Fig. 1: Relationships between well-being, demographic and economic subsystems.

Note that the disaggregation of the well-being subsystem is already known [7]. Note also that due to the limitation to the length of this work, we are unable to show the considerable number of equations that relate the different variables in the other two subsystems. That is why we will only show here the relationships between them through Forrester diagrams [12].

2.1. The demographic subsystem

The basis of our demographic subsystem is the one provided by Caselles et al. [9], where a sex and age-structured von Foerster-McKendrick model is presented.

Fig. 2 shows a part of the Forrester Diagram of the model, which provides the demographic subsystem disaggregated and linked with the well-being variables, and Table 1 shows the demographic variables used in this diagram. Note that the sub-indices *i=1* and *i=2* represents the variables related with females and males, respectively.

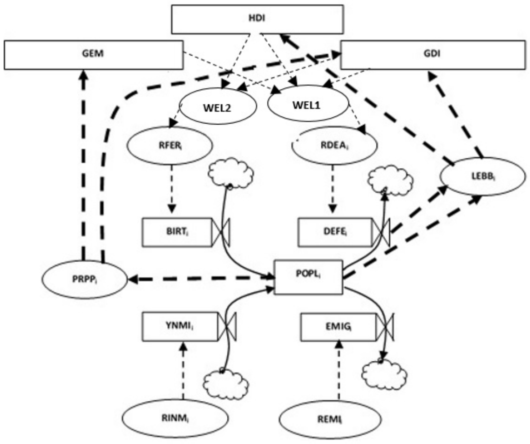


Fig. 2: Forrester Diagram of the demographic subsystem and the link with the well-being variables.

Table 1
Variables used in the demographic subsystem.

Variable	Definition
BIRT	Birth by year
DEFE	Death by year
EMIG	Emigrates by year
LEBB	Life Expectancy at Birth
POPL	Total Population
PRPP	Population proportion (Population by sex/Total Population)
RDEA	Death Rate
REMI	Emigration Rate
RFER	Fertility Rate
RINM	Immigration Rate
YNMI	Immigrates by year
WEL1	Well-being variables with *HDI*, *GDI* and *GEM*
WEL2	Well-being variables with *HDI* and *GDI*

2.2. The economic subsystem

Fig. 3 shows the corresponding Forrester Diagram to this subsystem, where its variables are emphasized in relation to the other subsystems. Note that the unemployment rate is calculated through Demographic subsystem variables and, the demographic variables are calculated with well-being variables. Table 2 defines the variables shown in Fig. 3.

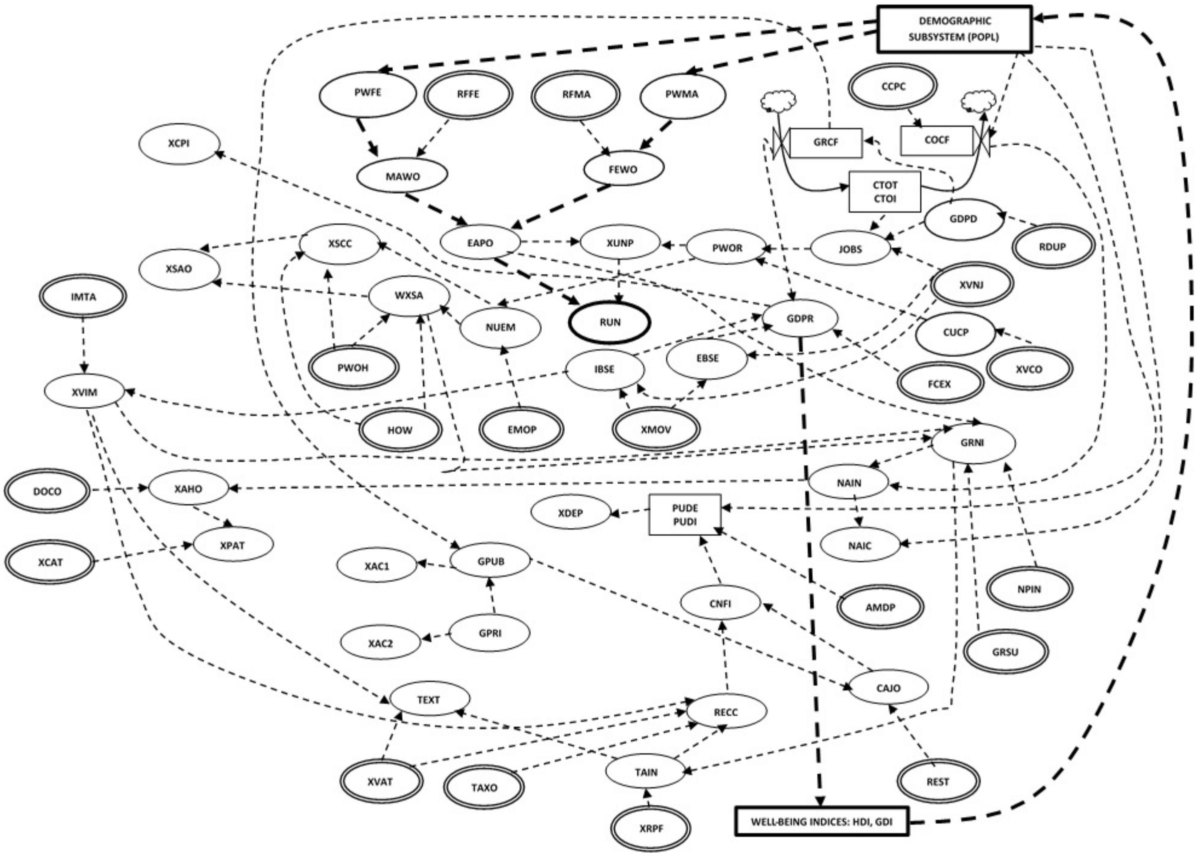


Fig. 3: Forrester Diagram of the economic subsystem.

Table 2
Variables used in the economic subsystem.

Variable	Definition	Variable	Definition
AMDP	Amortization	*NAIN*	National income
CAJO	Capital jobs	*PWFE*	Female population at working age
CCPC	Consumption of fixed capital per capita	*PWMA*	Male population at working age
CNFI	Capacity or need of financing	*PWOH*	Average price per worked hour
COCF	Consumption of fixed capital	*PWOR*	Number of people working
CTOI	Initial consumption of fixed capital	*RECC*	Capital resources
CTOT	Total fixed capital at the end of the current year	*REST*	Current transfers and acquisition of land and assets
CUCP	Coefficient of utilization of productive capacity	*RFFE*	Female labor force rate
DOCO	Spending on domestic consumption	*RFMA*	Male labor force rate
EAPO	Economically active population	*RNDP*	Net property incomes
EBSE	Goods and services exports	*RPUD*	Interest rate on public debt
FCEX	Final consumption expenditure	*RUNM*	Unemployment rate
FEWO	Number of female workers	*TAAS*	Percentage of employees on the occupied population
GDPD	Gross domestic product deflactor	*TAIN*	Taxes related to income
GDPR	Gross domestic product	*TAXO*	Other taxes
GPRI	Gross formation of private fixed capital	*TAXT*	Total taxes
GPUB	Gross formation of public fixed capital	*XAC1*	PUIN/XOIN
GRCF	Gross capital formation	*XAC2*	PVIN/XOIN
GRNI	Gross national income	*XAHO*	National savings
GRSU	Gross surplus	XCAT	Capital transfer
HOWO	Number of hours worked per year and per person	*XMOV*	Monetary value ($ for 1€)
IBSE	Goods and services imports	*XPAT*	National patrimony
IMTA	Import tax rate	*XRPF*	Income tax rate
JOBS	Number of jobs	*XUCO*	Unemployment compensation
MAWO	Number of male workers	*XUNP*	Number of the unemployed people
NAIC	National income per capita	*XVAT*	Taxes related to VAT
NUEM	Number of employees	*XVIM*	Taxes related to imports
PUDE	Public debt	*XVNJ*	Average value of a new job
PUDI	Initial Public debt	*XWSA*	Gross wages and salaries

3. The model validation

The historical data used in this article to fit the model have been obtained from the Spanish National Statistics Institute database [13] in the 2002-2014 period. The software tool used for the model verification is *SIGEM*. On the one hand, the deterministic validation is considered successful for three reasons: the visual evaluation of the graphic overlapping of the historical data and the calculated data is satisfactory, the determination coefficients, R^2, are very high and the randomness of the residuals is verified by the maximum relative error, which do not exceed the 5% (Fig. 4). On the other, the stochastic formulation is also successful for two reasons: all results have a normal distribution and a 99% confidence intervals is creating and is checking that 11 out of 12 points are within the interval, and that the only point that is outside the interval, is very close to it (Fig. 5).

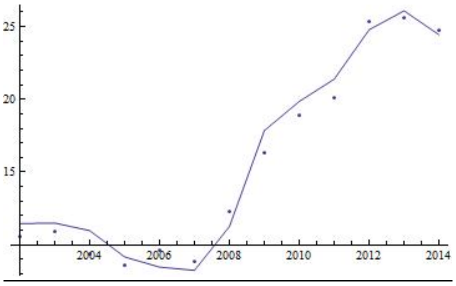


Fig. 4: Deterministic validation. Variable *RUNM (%)*, real data (dots), tendency values (line), for 2002-2014 period in Spain. R^2=0.963313.

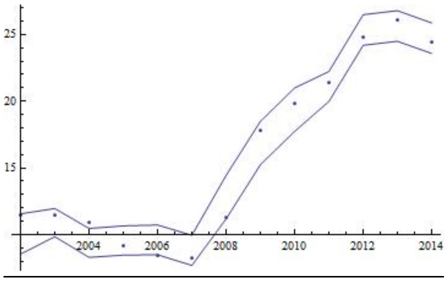


Fig. 5: Stochastic validation. Variable *RUNM (%)*, real data (dots), maximum and minimum values 99% confidence (lines), for 2002-2014 period in Spain.

4. Forecasts

As a tentative exercise to try to reduce the unemployment rate trend we classify the input variables into control variables and scenario variables (non controllable ones); we design four strategies and three scenarios, (Tables 7 and 8).

Table 7:
Strategies. ↑ increase, ↓ decrease and ≈ keep the tendency

Variable	1	2	3	4
XVNJ	↑	↓	↑	≈
XIDE	↑	↓	↓	≈
XMOV	↓	↓	↑	≈
PWOH	↑	↓	↓	≈
GPRI	↑	↓	↓	≈
AMDP	↑	↓	↓	≈
HOWO	↓	↑	↑	≈
GPUB	↓	↑	↑	≈
XVAT	↓	↑	↑	≈
XRPF	↓	↑	↑	≈
DOCO	≈	≈	≈	≈
XCAT	≈	≈	≈	≈
REST	≈	≈	≈	≈

Table 8
Scenarios. ↑ increase, ↓ decrease and ≈ keep the tendency.

Variable	1	2	3
GRFE	↑	≈	≈
RLIF	↑	≈	≈
GRMA	↑	≈	≈
RLIM	↑	≈	≈
RPUD	↓	↑	≈
EPIF	↑	≈	≈
PAEF	↑	≈	≈
PPPF	↑	≈	≈
FCEX	↓	↑	≈
RFFE	↑	≈	≈
RFMA	↑	≈	≈
EMOP	↑	≈	≈
CUCP	↑	≈	≈
NPIN	↑	↓	≈
GRSU	↑	↓	≈
CCPC	↑	↓	≈

At a starting point, the concrete objective has been to decrease the unemployment rate in the horizon of the 2025 year. The corresponding calculations are performed with the simulator generated by *SIGEM* and with *MATHEMATICA* 10.2 [14]. The optimal strategy to reach our goal is chosen by observing the evolution of the unemployment rate. The conclusion is that Strategy 2 makes minimum the unemployment rate in the horizon of the 2025 year, and the greatest decrease in unemployment rate corresponds to Scenario 1.

5. Conclusion

An abstract complex demographic-economic model has been presented and used to attempt to control the unemployment rate evolution in a country as well as its main related variables. The model includes three well-being variables defined by the United Nations (*HDI*, *GDI* and *GEM*) that consider key concepts for human development, as Health, Education, Economy and Female Labor, as well as detailed demographic and economic sub-models. Some tentative strategies and scenarios have been designed to attempt to find the best feasible strategy (within the control variables considered in the model) to reduce unemployment in a country and to determine the collateral consequences (over the other variables considered in the model: economical, demographical and wellbeing ones) of these strategy.

The study has been performed with two model formulations: deterministic (for simplicity) and stochastic (to obtain confidence intervals for forecasts). Both formulations of the model have been fitted and verified with the corresponding criteria and real data from Spain in the 2002-2014 period. Finally, three scenarios and four strategies have been considered to reduce the unemployment rate in Spain in 2025 year.

References

[1] http://www.tradingeconomics.com (accessed 22.08.2016)

[2] M. Pompili, M. Innamorati, C. Di Vittorio (2014). Unemployment as a risk factor for completed suicide: a psychological autopsy study. Archives of Suicide Res., 18, 2. 181-192.

[3] A. Milner, A. Page, A. D. LaMontagne, (2014). Causes and effect studies on unemployment, mental health and suicide: a meta-analytic and conceptual review. Psychological Med., 44, 5. 909-917.

[4] H. Worach-Kardas, S. Kostrzewski (2014). Quality of Life and Health State of Long-Term unemployed in older production age. Appl. Res. in Qual. of Life, 9, 2. 335-353.

[5] J.P. De Nicco (2015). Employment-At-Will Exceptions and jobless recovery. J. of Macroecon. 45, 245–257.

[6] B. Volna (2015). Existence of chaos in the plane R2 and its application in macroeconomics. Appl. Math. and Comput. 258, 237–266.

[7] UNDP (1990-2010). Human Development Report. New York: Oxford University Press. http://hdr.undp.org/en/ (accessed 22.08.2016)

[8] M.T. Sanz, J.C. Micó, A. Caselles, D. Soler (2014). A stochastic model for population and well-being dynamics. J. of Math. Sociol. 38-2, 75-94.

[9] A. Caselles, D. Soler, M.T. Sanz, J.C. Micó (2014). Simulating Demography and Human Development Dynamics. Cybern. & Syst.: An Int J. 45:6, 465-485.

[10] A. Caselles (1994). Improvements in the Systems Based Program Generator SIGEM. Cybern. and Syst.: An Intern. J. 25, 81-103.

[11] A. Caselles (2008). Modelización y simulación de sistemas complejos (Modeling and simulation of complex systems). Valencia (Spain). Ed. Universitat de València. (Available in http://www.uv.es/caselles (accessed 22.08.2016) as well as SIGEM).

[12] J. W. Forrester (1961). Industrial dynamics. Cambridge: MIT Press.

[13] http://www.ine.es (accessed 22.08.2016)

[14] Wolfram Research, Inc., Mathematica, Versión 10.4, Champaign, IL (2016) http://www.wolfram.com/mathematica/ (accessed 22.08.2016)

An algorithm to obtain directional communities in a directed graph.

A. Hervás$^\flat$ [*], P.P. Soriano-Jiménez[†], R. Capilla[‡] and J. Peinado[††]

($\flat$) Instituto de Matemática Multidisciplinar ,

Universitat Politecnica de Valencia, 46022 Valencia, Spain.

(†) Universitat Politecnica de Valencia, 46022 Valencia, Spain.

, (‡) Universitat Politecnica de Valencia, 46022 Valencia, Spain.

(††) Instituto de Instrumentacion para la Imagen Molecular,

Universitat Politecnica de Valencia.46022 Valencia, Spain.

November 30, 2016

1 Introduction

Graph theory has been used to establish and solve many problems, [1]. In recent years, the importance of problems related to growth of social networks, medical network model and the growing interest in big data problems, the study of complex networks has become an object of interest of work for scientists. It allowed applications in many fields, from the area of biomedicine to the social sciences: genomics, study of epidemics, coauthoring publications, social relations, etc.

The study of the structure of these networks, characterized by the analysis of the degree of the vertices and the existence of paths or cycles between pairs of vertices, helps us to understand how they work and allows us to create new models, opening new expectations to the the classical theory of graphs. A complete review about the structure and dynamics of complex networks can be found in [2].

[*] e-mail:ahervas@imm.upv.es

Special importance acquires the study of those elements that are closely related. It is created a cluster of elements in the structure that are highly connected with each other and few connected with the rest. We will call these groups communities. Thus, the study of communities in a graph becomes an object of our interest. Remember that his concept is related to the high density of connections in the graph, [5], [6], [7].

The finding of communities on a network brings us closer to the knowledge of its structure and its properties. Hence the importance of the design of algorithms that allow us to obtain the communities in a network. In this regard, we must highlight the work of [5], an excellent and complete review of the state of the art in modularity, clustering and its applications, and some classical references as [6], [7], and [8].

2 Communities

In a graph, a community composes a set of vertices that are highly interrelated, meaning that there are many edges between them. In contrast, there are few edges that connect the community vertices with the rest of the graph.[7], [4].

In other words, there is a high density of connections within each community and a low density of connections between communities. The reason for using this technique is given by the fact that: *"Community structure methods normally assume that the network of interest divides naturally into subgroups and the experimenter's job is to find those groups. The number and size of the groups are thus determined by the network itself and not by the experimenter."*, [7].

Given a network, a good division in communities is the one that gets a large number of edges within the communities, against a small number of inter-community edges (the incidents vertices of these edges belong to different communities).

Modularity is a property that indicates how good this division is. It is a function that evaluates the goodness of partitions of a graph into clusters.[5], [8]. Takes value between 0 and 1, and according to [8], in practice, networks that have a strong community structure have a modularity between $0,3$ and $0,7$. Higher values are strange.

There are several algorithms that allow us to get the communities on a graph, and show us modularity. Nonetheless, they offer different results depending on the criteria for group vertices used. Most of them only apply to undirected graphs.[5], [6], [3], [13] and [9].

Walktrap, proposed by [11],works by joining communities through random walks.

In label propagation, proposed by [13], the algorithm assigns labels to the vertices that are updated at each iteration. Although it provides good computational results, it does not offer a unique solution.

The edge betweenees, proposed by [8], begins with only one community, and divides it, until obtaining n communities.

The fast greedy, by [9], improves the computational results of edge betweenees. Assumes that each vertex is a community, grouping them at each, ending with n communities

We apply all these algorithms to our graph. Despite the community obtained looking similar, the real results are very different.

Most of these algorithms are designed for undirected graphs, although they have been successfully used in some cases for directed graphs. Unfortunately, applied to directed graphs, give us results that do not fit properly to the original problems. This is the reason why we discard the use of these algorithms, since our goal is to establish a method to obtain the communities so that we can detect those vertices on which to act on them alter the system in a directed graph.

3 Proposed Algorithm.

So, we propose an algorithm that finds communities in a directed graph, in which all vertices can reach the same vertex, or might be reached. Considering the directionality of the graph, it necessarily leads us to study the two possibilities: community vertices that can be reached from a given one, and communities of vertices from which it has reached the vertex considered.

ALGORITHM 3.1 (Construction of directional communities) With this algorithm we will obtain the community sets of vertices that allow us to generate subgraphs making up the communities.

1. *From a given graph:*

2. *Applying a search algorithm we obtain the matrices of accessibility, R, and access, Q, of the graph.*

3. *We obtain the input and output degree of each vertex.*

4. *We order the vertices from highest to lowest output degree, when two or more with the same output degree, sort them by the input degree, from low to high*

5. *We take v_1, and create the first set of community vertices $C_1 = C_1 \cup \{v_1\}$*

   (a) *For every vertex v_k,*

   (b) *If $v_k \in C_1 \vee C_2 \vee .... \vee C_{k-1}$ take the next vertex.*

   (c) *Otherwise we define a new set of Community vertices $C_k = C_k \bigcup \{v_j; v_j \in R(v_k), v_j no \in C_1 \vee C_2 \vee .... \vee C_{k-1}\}\dagger$*

6. *We order the vertices from highest to lowest input degree, if there are two or more vertices with the same input degree, sort them by the output degree, from lowest to highest.*

7. *We take v_1, and created the first set of community vertices $C'_1 = C'_1 \cup \{v_1\}$*

   (a) *For every vertex v_k,*

   (b) *If $v_k \in C'_1 \vee C'_2 \vee .... \vee C'k - 1$ take the next vertex.*

   (c) *Otherwise we define a new set of Community vertices $C'_k = C'_k \bigcup \{v_j; v_j \in R(v_k), v_j no \in C'_1 \vee C'2 \vee .... \vee C'_{k-1}\}\ddagger$*

8. *We built $G_{OUT} = \bigcup_{i \in I} G_{OUT(C_i)}$ as the graph generated by each community sets, C_i obttained in $\dagger$.*

9. *We built $G_{IN} = \bigcup_{j \in J} G_{IN(C_j)}$ as the graph generated by each community sets, C_j obtained in $\ddagger$.*

With Graph OUT we obtain the communities formed by highly emitters vertices and its receiving environment. An alteration in the demand for these highly emitter vertices, affects all vertices of the community.

With Graph IN we identify vertices receptors and the community that provide traffic vertices to that vertex. These vertices receivers will be sensitive to any variation in their contributors.

The Graph IN algorithm provides us less information, but complements the one given by Graph OUT.

Using the Graph IN algorithm, we can detect vertices or groups of vertices that only have incoming traffic. The supply modification in these vertices would allow us to control some flows.

The edges that connect communities obtained with Graph IN, indicate that traffic is sent from a community to another whose offer is lower. In this way, we can detect secondary degrees related to the demand for large communities.

4 Conclusions.

We have proposed an algorithm that allows us to group the vertices of the graph in directed communities, with high modularity, that guarantees a good level of relationship between the vertices.

The algorithms are programmed using R Project, citeR, and can be found in the directions:
"http://www.upv.es/orgpeg/pub/CONSTRUCTION-OF-DIRECTIONAL-COMMUNITIES.r"
for the Algorithm 2
"http://www.upv.es/orgpeg/pub/The-graph-for-a-system-250-deg-x-25000-stu.pdf"
"http://www.upv.es/orgpeg/pub/Metodos-comparados-1011.pdf"
"http://www.upv.es/orgpeg/pub/Metodos-comparados-0102.pdf"
"http://www.upv.es/orgpeg/pub/Metodos-comparados-03050608.pdf"
"http://www.upv.es/orgpeg/pub/Graph-In-and-Graph-Out-for-250-deg-x-25000-stu.pdf"

The algorithm has been applied to several regions and the results are similar and consistent with the problem being analyzed.

References

[1] Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. Network flows. Upper Saddle River, Prentice Hall. 1993.

[2] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U.. Complex networks: Structure and dynamics. *Physics reports*, 424(4), 175-308.2006

[3] Clauset,A. Newman, M. E. and Moore, C. Finding community structure in very large networks *Physical review E*, vol. 70, no. 6, p. 066111, 2004.

[4] Csermely, P. (2008). Creative elements: network-based predictions of active centres in proteins and cellular and social networks. *Trends in biochemical sciences*, 33(12), 569-576.

[5] Fortunato,S. Community detection in graphs, *Physics Reports*, vol. 486, no. 3,pp. 75-174, 2010.

[6] M. Girvan and M. E. Newman Community structure in social and biological networks *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821-7826, 2002.

[7] Newman, M.E. Modularity and community structure in networks. *Proceeding of theNational Academy of Sciences*. 103 (23), 8577-8582. 2006

[8] Newman M. E. and Girvan M. Finding and evaluating community structure in networks *Physical review E*, vol. 69, no. 2, p. 026-113, 2004.

[9] Newman M. E. Finding community structure in networks using the eigenvectors of matrices *Physical review E*, vol. 74, no. 3, p. 036-104, 2006.

[10] Newman M. E. Fast algorithm for detecting community structure in networks,*Physical review E*, vol. 69, no. 6, p. 066133, 2004.

[11] Pons, P. and Latapy,M. Computing communities in large networks using random walks.*J. Graph Algorithms Appl.*, vol. 10, no. 2, pp. 191-218, 2006.

[12] The R Project for Statistical Computing. www.r-project.org. (5-2016)

[13] Raghavan,U. N. Albert,R. and Kumara. S. Near linear time algorithm to detect community structures in large-scale networks.*Physical Review E*, vol. 76, no. 3,p. 036-106, 2007.

Managing Dependence in Reliability Models by Markovian Arrival Processes

C. Santamaría[†] [*], B. García-Mora[†], and G. Rubio[†]

(†) Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València

November 30, 2016

1 Introduction

The most common assumptions in reliability studies is that *failures* occur independently and with the same distribution. However these two assumptions are unrealistic in practice since *inter-failures* times are usually correlated and not identically distributed. There is a constant need to get arrival models with these two properties. The *Markov Arrival Process (MAP)* is an active research field for dealing with not identically distributed and correlated inter-failures times ([1], [2], [3]). We study two versions of the *MAP* approach with simulated data of devices undergoing three failures each one (Figure 1). The first one is our model, that we began to develop in [2] and [4]. The second one is a recent more standard approach that we consider in order to validate.

2 The *MAP* model

It is simulated a sample of operational random times from the 2*–state non-stationary MAP* for 100 devices with three failures for each one [5]. They

[*] e-mail: crisanna@imm.upv.es

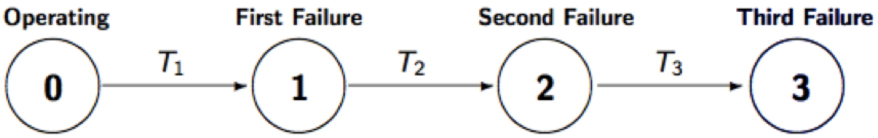


Figure 1: Device with three failures

are independent. The sample is

$$t^{(1)} = (t_1^{(1)}, t_2^{(1)}, t_3^{(1)}),\, t^{(2)} = (t_1^{(2)}, t_2^{(2)}, t_3^{(2)}), \dots, t^{(100)} = (t_1^{(100)}, t_2^{(100)}, t_3^{(100)}) \tag{1}$$

where in each device we have *the times of the three failures.* Now let T_k be the random variable representing the operational time between the (k–1)–th failure and the k–th failure. The times in columns represent the three variables T_1, T_2, T_3 of the *inter–failure times*, correlated and not identically distributed (Figure 1).

The idea is to make up a *MAP* model with representation (π, D_0, D_1) for modeling the process $0 \to 1 \to 2 \to 3$ of three failures for the 100 independent devices. A Markovian Arrival Process (MAP) [6] is an irreducible Markov chain with a finite state space S, initial vector π and a generator matrix Q which can be represented as $Q = D_0 + D_1$ where

- $D_1 \geq 0$, $D_1 \neq 0$.
- $D_0(i,j) \geq 0$ for $i \neq j$.
- (π, D_0) is a phase-type distribution.

D_0 is obtained in a first step from the convolution of two functions and in a second step the obtained result is convoluted with another function [6]. These functions are mixtures of three *Erlangs* distributions in each transition [7].

D_1 is fitted by maximizing the *likelihood function* for the 3 interarrival operational times T_1, T_2 *and* T_3 of the 100 devices with the following expression

$$f(t^{(1)}, t^{(2)}, \dots t^{(100)}) = \pi e^{D_0 t_1^{(i)}} D_1 e^{D_0 t_2^{(i)}} D_1 e^{D_0 t_3^{(i)}} D_1 \mathbf{e}$$

The representation of the Cumulative Distribution Functions for the interarrival times T_1, T_2, T_3 estimated by the *MAP* model with the empirical distribution is shown in the Figure 2.

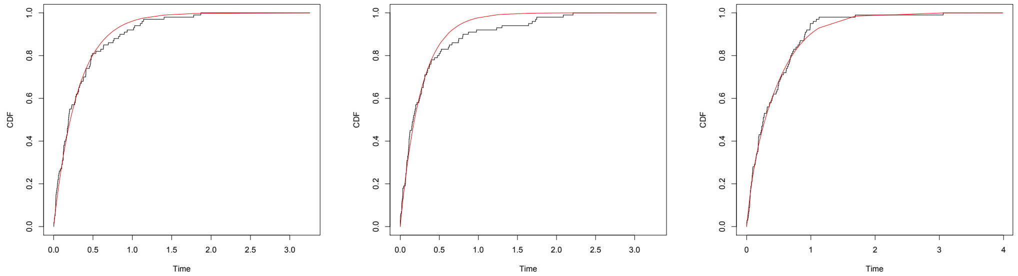


Figure 2: CDF (*smooth line*) and empirical distribution (*step function*) for the *inter–failure* times T_1, T_2, T_3 in the MAP model.

3 The 2–state non–stationary Markovian Arrival Process (MAP_2)

In order to compare our *MAP model* with the state of the art we choose the most recently model using the MAP methodology [1]. We apply a 2–state non–stationary Markovian Arrival Process to our data, denoted by MAP_2. It is a doubly stochastic process $\{J(t),\, N(t)\}$ where

- $J(t)$ represents an irreducible, continuous, Markov process with state space $S = \{1,\, 2\}$.
- The counting process $N(t)$ represents the number of failures in the interval $(0,\, t]$.
- The initial state $i_0 \in S$ is generated according to an initial probability $\alpha = (\alpha,\, 1 - \alpha)$.

The MAP_2 can be characterized by M $= \{\alpha,\, D_0,\, D_1\}$ where D_0 and D_1 are rate matrices. $D = D_0 + D_1$ is the generator of $J(t)$, with stationary vector ϕ, calculated as $\phi P^* = \phi$. P^* is the transition probability matrix, given by $P^* = (-D_0)^{-1} D_1$. The cumulative density function (CDF) and the moments of the *variables* T_1, T_2 *and* T_3 are defined by the expressions

$$F_{T_k}(t) = 1 - \alpha_k e^{D_0 t}\mathbf{e}$$

where $\alpha_{\mathbf{k}} = \alpha (P^*)^{k-1}$ and $T_k \sim PH\{\alpha_{\mathbf{k}},\, D_0\}$ represent different phase–type distributions for the correlated variables T_1, T_2 *and* T_3.

The goal is to estimate the model parameters in $\{\tilde{\alpha},\, \tilde{D_0},\, \tilde{D_1}\}$ in the MAP_2. For it we use an optimization problem (P) [8], that is solved using the local search *MATLAB's routine* **fmincon** (Optimization toolbox). A *multistart approach* (100 different starting points randomly selected of the simulated data) is performed and we keep the solution with the minimum objective function $\gamma_\tau(\tilde{\alpha},\, \tilde{D_0},\, \tilde{D_1})$ in the optimization problem (P).

We resolve the problem (P) for two canonical representations of the MAP_2

- The expression of the first canonical representation is

$$\tilde{\alpha} = (\tilde{\alpha},\, 1-\tilde{\alpha}),\;\; \tilde{D_0} = \begin{pmatrix} \tilde{x} & \tilde{y} \\ 0 & \tilde{u} \end{pmatrix},\;\; \tilde{D_1} = \begin{pmatrix} -\tilde{x}-\tilde{y} & 0 \\ \tilde{v} & -\tilde{u}-\tilde{v} \end{pmatrix} \tag{2}$$

- The second canonical representation

$$\tilde{\alpha} = (\tilde{\alpha},\, 1-\tilde{\alpha}),\;\; \tilde{D_0} = \begin{pmatrix} \tilde{x} & \tilde{y} \\ 0 & \tilde{u} \end{pmatrix},\;\; \tilde{D_1} = \begin{pmatrix} 0 & -\tilde{x}-\tilde{y} \\ -\tilde{u}-\tilde{v} & -\tilde{v} \end{pmatrix} \tag{3}$$

and we select the estimated *parameters* $\{\tilde{\alpha},\, \tilde{x},\, \tilde{y},\, \tilde{u},\, \tilde{v}\}$ under the *canonical representation* with the *highest log–likelihood* given in

$$\log f(\mathbf{t}^{(1)},\, \mathbf{t}^{(2)}, \ldots \mathbf{t}^{(N)} | D_0,\, D_1) = \sum_{i=1}^{N} \log f(\mathbf{t}^{(i)} | D_0,\, D_1) \tag{4}$$

which provides evidence in favor of the estimation given by the second canonical estimated representation by its highest log–likelihood value (-11.3338 vs -19.2334). The representation of the Cumulative Distribution Functions for the interarrival times T_1, T_2, T_3 estimated by the MAP_2 is shown in the Figure 3.

4 Concluding remarks

Both models *(MAP and MAP_2)* fit well to our simulated data and both models introduce the *dependence between interarrival times.* In [3] we showed that the *MAP* model can deal with *covariates* and *censored data.* However the dimension of the problem increases [2] with the number of events. On the other hand MAP_2 works perfectly with *a lot of events* but to our knowledge there is any work with *covariates* or *censored data.*

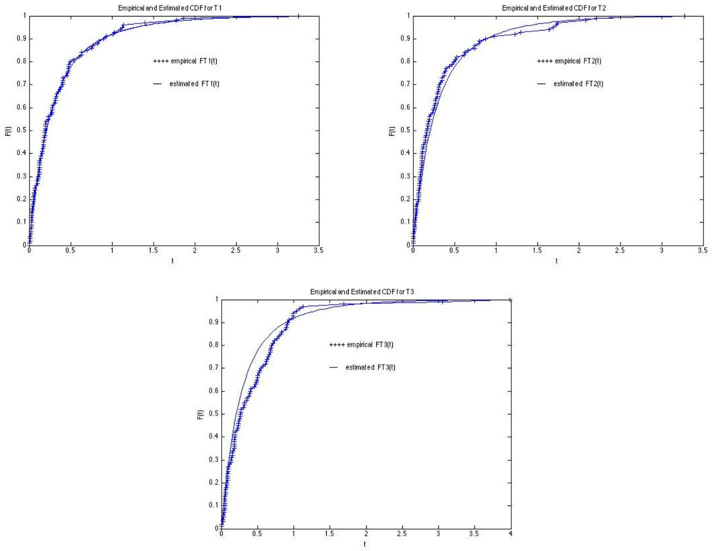


Figure 3: CDF (*smooth line*$-$) and empirical distribution ($++++$) for the inter–failure times T_1, T_2, T_3 in the MAP_2 model.

References

[1] Rodríguez J., Lillo R.E., Ramírez–Cobo P. Failure modeling of an electrical N–component framework by the non–stationary Markovian arrival process. *Reliability Engineering and System Safety*, (134): 126–133, 2015.

[2] Santamaría C., García–Mora B., Rubio G and Pérez–Ocón R. Managing dependence in Flowgraphs models. An application to Reliability Engineering. Modelling for Engineering & Human Behaviour 2015, Valencia, Instituto Universitario de Matemática Multidisciplinar, 2015.

[3] Rubio G, García–Mora B., Santamaría C., Pontones J.L. Modeling dependence in multistate processes. International Work–Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2016. 20–22 April, 2016. Granada (Spain).

[4] Rubio G, García–Mora B., Santamaría C., Pontones J.L. Incorporating multiple recurrences in a Flowgraph model for bladder carcinoma.

International Work–Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2015. 15–17 April, 2015. Granada (Spain).

[5] Buchholz P., Kriege J. and Felko I. Input Modeling with Phase–Type Distributions and Markov Models. Theory and Applications, Springer Cham Heidelberg New York Dordrecht London, 2014.

[6] Neuts M. F. Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach., John Hopkins University Press, 1981.

[7] Pérez-Ocón, R., Segovia. M.C. Modeling lifetimes using phase-type distributions. In: Aven, T., Vinnem, J.E. (eds.) Risk, Reliability and Societal Safety, Proceedings of the European Safety and Reliability Conference 2007 (ESREL), Taylor & Francis, 3rd edition (2007).

[8] Carrizosa E., Ramírez–Cobo P. Maximum likelihood estimation for the two–state Markovian arrival. arXiv: 1401.3105v1.

Studying the evolution over the next few years of the meningococcal genogroups in Spain using a competition Lotka-Volterra model

R. Abad[†], L. Acedo[♭], J. Díez-Domingo[‡],
J. A. Moraño[♭] [*], J. Vázquez[†], and R. J. Villanueva[♭]

(†) Institute of Health Carlos III, Madrid, Spain,

(♭) Instituto de Matemática Multidisciplinar
Universitat Politècnica de València, Valencia, Spain,

(‡)Centro Superior de Investigación en Salud Pública (CSISP), Valencia, Spain.

1 Introduction

Neisseria meningitidis (NM), or meningococcus, is a bacterium that can cause several forms of meningococcal disease. Symptoms are headache, stiff neck, fever, nausea, vomiting, etc., which can lead to coma [1].

The bacteria are spread by exchanging respiratory and throat secretions (saliva or spit) during close contacts (coughing or kissing) [1].

There are 12 types or genogroups based on the capsular polysaccharides: A, B, C, H, I, K, L, X, Y, Z, E and W but 90% of all infections are caused by types A, B, C, Y and W, being A, B and C the most common (W-cases are only the 4% in US). On the other hand, the type A has been the most prevalent in Africa and Asia, but is rare in North America and Europe [2, 3].

Meningococcus is a part of the common flora in the nasopharynx of up to $5 - 15\%$ of adults and the genogroups are in competition in this ecosystem with humans. So, changes in health habits or prophylactic measures may change the distribution of the genogroups in this ecosystem.

[*] e-mail: jomofer@imm.upv.es

Some years ago, in Chile, there was an outbreak of type W with a strong impact in the public health due to its lethality, so the Public Health System is aware about possible outbreaks of new genogroups, because the population is not protected against them and, as in Chile, the consequences may be disastrous.

In the western countries, most of the cases are produced by groups B and C. Serogroups Y and W are less frequent although there are differences in their incidence in some countries. The serogroup W135 is associated to cases and outbreaks after traveling to Mecca [4].

This is a first approach of predicting the evolution dynamics of meningococcus in order to find out if it is possible to detect meningococcus outbreaks in advance.

2 Data

We have data of 8 genogroups of meningococcus in Spanish population, but only from 2 time instants (December 2011 and December 2012) that we can see in the Table 1.

x^i	B	C	Y	W	X	Z	E	ND*
Dec-2011 ($x^i_{t=0}$)	44.92	2.14	3.74	4.28	1.60	1.07	6.42	35.83
Dec-2011 ($x^i_{t=12}$)	30.74	1.10	5.49	5.49	0.11	4.39	8.78	43.90

Table 1: Percentage of Spanish people where each one of the types of meningococcus is predominant in December of 2011 and December of 2012. *ND means unclassified meningococcus.

This data have been provided by the Reference Laboratory for Meningococci belonging to the National Microbiology Centre of the Institute of Health Carlos III.

3 Modeling

As we consider the 8 meningococcus genogroups are in competition for human ecosystem resources we propose a Lotka-Volterra model that is widely used in ecology.

The formula for the discrete model of this type is:

$$X_{t+1}^i = X_t^i + r_i X_t^i (K_i - X_t^i) - r_i X_t^i \sum_{j=1}^{8} \alpha_{ij} X_t^j, \qquad i = 1,,8$$

where X_t^i is the number of bacteria of genogroup i at instant t, r_i is the growth rate for serogroup i and α_{ij} is the effect of the number of bacteria of the group j has on the number of bacteria of group i, being $\alpha_{ii} = 0$.

Besides, as the data are in unit fractions we need to scale the model, so we do the change

$$x_t^i = \frac{X_t^i}{K_i}, \qquad i = 1,,8$$

where K_i is the carrying capacity for the group i, x_t^i is the fraction of ecosystem occupied by bacteria of genogroup i at time t, $0 \leq x_t^i \leq 1$ and $\sum x_t^i = 1$. In this way we obtain

$$x_{t+1}^i = x_t^i + r_i x_t^i K_i (1 - x_t^i) - r_i x_t^i \sum_{j=1}^{8} \alpha_{ij} x_t^j K_j, \qquad i = 1,,8, \text{ and}$$

$$x_{t+1}^i = x_t^i + H_i x_t^i (1 - x_t^i) - x_t^i \sum_{j=1}^{8} M_{ij} x_t^j, \qquad i = 1,,8$$

where $H_i = r_i K_i$ and $M_{ij} = r_i \alpha_{ij} K_j$ are 64 unknown parameters ($M_{ii} = 0$).

4 Method

The method used for fitting the model with the data has been the following:

- We generate 64 random parameters (H_i and M_{ij}), allowing small oscillations in their components.
- Based on 2011-data, x_0^i, we simulated with our parameters until 2012, $\tilde{x}_{12}^i$, in order to compare to reference data x_{12}^i.
- We search the 64 parameters that applied to the model produce an output as close as possible to the data, so we applied Nelder-Mead method to find the best 64 parameters that better fit to the last data.

5 Results and graphs

We have performed several fittings of the model and we take the model parameters such that the model output minus the data is less in the mean square sense.

Once the best model parameters have been calculated, we predict until 120 months (2021) the evolution over the time of the different genogroups of meningococcus. The result can be seen in the Figure 1.

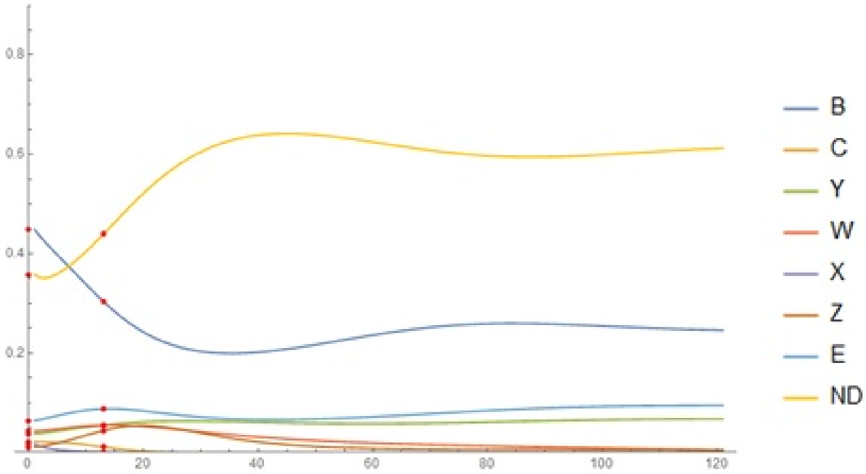


Figure 1: Model fitting and prediction until 2012. We can see that the different genogroups are in a certain stable scenario.

A zoom of the low part of the Figure 1 can be seen in Figure 2.

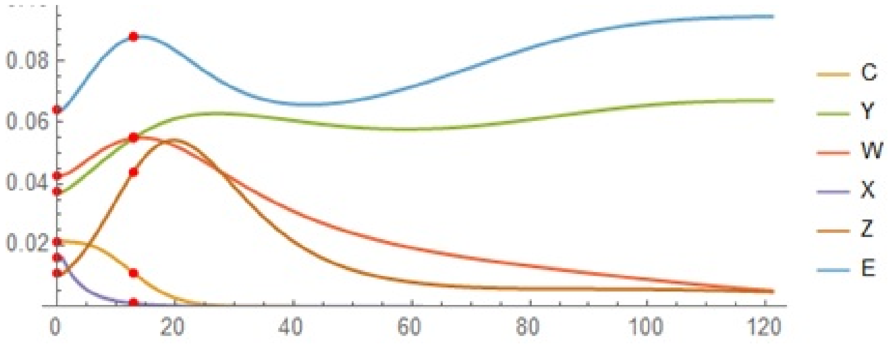


Figure 2: Zoom of the lower part of Figure 1.

Also in Figure 3 we can see the individual fitting and evolution of each one of the genogroups of meningococcus over the time.

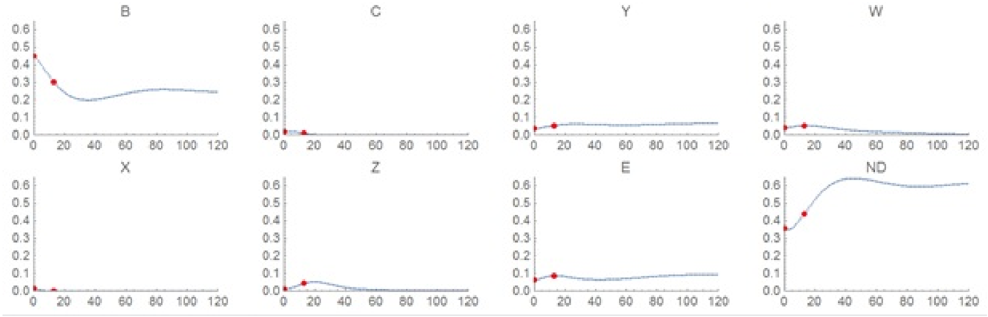


Figure 3: Model fitting and prediction until 2012 of the genogroups individually.

6 Conclusions

In this paper, we present a Lotka-Volterra model to describe the competition dynamics of 8 genogroups of meningococcus. Using data provided by the Reference Laboratory for Meningococci, we fit the model according to the data. Once the model has been fitted, we used the model parameters to predict in short and medium term the evolution dynamics of each meningococcus in the human ecosystem.

This is a first approach of predicting the evolution dynamics of meningococcus in order to find out if it is possible to detect meningococcus outbreaks in advance. Our prediction says that is not likely to have an outbreak in the next years (until 2021) because there are not high variations in all meningococcus since Dec-2011. In fact, until now, since Dec-2011, there is not any outbreak of any type of meningococcus.

7 Acknowledgements

This work has been partially supported by the FIS grant PI13/01459.

References

[1] http://www.cdc.gov/meningococcal/index.html

[2] Cartwright K.,Meningococcal carriage and disease. In: Cartwright K., editor. Meningococcal disease. Chichester, UK: John Wiley & Sons; p.71- 114, 1995.

[3] De Walls P. Immunization strategies for the control of serogroup C meningococcal disease in developed countries. Expert Rev Vaccines 5: 269-75, 2006.

[4] Protocolos de la Red Nacional de Vigilancia Epidemiológica, http://www.isciii.es/ISCIII/es/contenidos/fd-servicios-cientifico-tecnicos/fd-vigilancias-alertas/fd-procedimientos/PROTOCOLOS_RENAVE-ciber.pdf