ICAMΣ'21

International Conference on Applied Mathematics in Engineering

http://icame.balikesir.edu.tr

PROCEEDINGS

ICAMΣ'21

September 1-3, 2021 – Balikesir, TURKEY

2nd International Conference on Applied Mathematics in Engineering

Proceedings Book

Edited by

Assoc. Prof. Dr. Firat EVIRGEN

Assoc. Prof. Dr. Ibrahim KUCUKKOC

December 2021, Balıkesir

Cite articles as follows:

Author, A., Author, B. (2021). Paper Title. In F. Evirgen, I. Kucukkoc (Eds.), Proceedings of the Second International Conference on Applied Mathematics in Engineering (ICAME'21), Balikesir, Turkey, September 1-3, Page number.

ISBN: 978-605-70379-8-5

Preface

We have the pleasure of successfully organising "*The 2nd International Conference on Applied Mathematics in Engineering (ICAME'21)*", between 1-3 September 2021 in Burhaniye/Balikesir, Turkey. Although it has been planned to be held on site in 2020 (two years after ICAME'18), the conference had to be postponed to 2021 and held online due to severe Covid-19 pandemic affected our planet.

The conference provided an ideal academic platform for researchers to present the latest research and evolving findings of applied mathematics on engineering, physics, chemistry, biology, and statistics. The conference also offered the opportunity of discussing advances in the field of applied mathematics, its effect on engineering and real-life problems. Especially, the most current applied mathematical problems (i.e. fractional calculus and its real-life applications, operational research, mathematical modeling in health science and engineering, optimization and control in engineering, non-linear dynamical systems and chaos, optimization and control problems) have been discussed via 143 oral presentations given to an audience with over 120 participants from 27 countries.

Authors of the high-quality presentations have been invited to present the greatly extended versions of their papers to the special issues organized in prestigious journals indexed in SCIE/ESCI and/or Scopus, such as *Journal of Computational and Applied Mathematics (Elsevier)*, *Numerical Algebra, Control and Optimization (AIMS)*, and *An International Journal of Optimization and Control: Theories & Applications (IJOCTA).*

This book contains the full-text articles of some of the papers (not submitted to journal special issues) presented at ICAME'21. All submissions to this book have been peer-reviewed via a single-blind reviewing process and the final decision whether to publish the papers has been given by the book editors.

With this opportunity, we would like to thank again to our plenary speakers *Albert C. J. Luo (Southern Illinois University Edwardsville, USA)*, *Sverre Holm (University of Oslo, Norway)*, *Gerhard-Wilhelm Weber (Poznan University of Technology, Poland)* and *Praveen Agarwal (Anand International College of Engineering, Jaipur, India)*; invited speakers *Carla Pinto (School of Engineering, Polytechnic of Porto, Portugal), Huseyin Merdan (TOBB University of Economy and Technology, Turkey)* and *Amin Jajarmi (Department of Electrical Engineering, University of Bojnord, Iran)*; organizers of the special sessions, authors, as well as the members of the international scientific committee and organizing committee for their exceptional cooperation.

Hope to see you all in the next event of the ICAME conference series in 2023.

December 2021

Book Editors

Assoc. Prof. Dr. Firat Evirgen
Department of Mathematics
Balikesir University, Turkey

Assoc. Prof. Dr. Ibrahim Kucukkoc
Department of Industrial Engineering
Balikesir University, Turkey

Committee

Honorary Chair

Ilter Kus, Balikesir University (Rector), Turkey

Chair

Ramazan Yaman, Istanbul Atlas University, Turkey

Co-chairs

J. A. Tenreiro Machado, Polytechnic Institute of Porto, Portugal
Necati Ozdemir, Balikesir University, Turkey
Dumitru Baleanu, Cankaya University, Turkey
Jordan Hristov, Univ. of Chemical Tech. and Metallurgy, Bulgaria

International Scientific Committee

Abdeljawad, T., Saudi Arabia
Agarwal, R.P., USA
Agarwal, P., India
Ahmad, I., Saudi Arabia
Akin, O., Turkey
Andrade, J.M., UK
Area, I., Spain
Atanackovic, T., Serbia
Atangana, A., South Africa
Ayaz, F., Turkey
Bagirov, A., Australia
Baskonus, H.M., Turkey
Baleanu, D., Turkey
Bohner, M., USA
Bulut H., Turkey
Caputo, M., Italy
Cattani, C., Italy
Celik, E., Turkey
Chen, W., China
Dassios, I., Ireland
Debbouche, A., Algeria
El-Khazali, R., United Arab Emirates
El-Sayed, A., Egypt
Erdogan, F., Turkey
Erturk, V.S., Turkey
Fernandez, A., Turkey
Golmankhaneh, A.K., Iran
Gulsu, M., Turkey
Gurbuz, B., Turkey
Hammouch, Z., Morocco
Holm, S., Norway
Hristov, J., Bulgaria
Inc, M., Turkey
Jafari, H., South Africa
Jarad, F., Turkey
Karapinar, E., Turkey
Kaya, D., Turkey
Kazmi, I., Pakistan
Keedwell, E.C., UK
Keskin, G., Turkey
Klimek, M. Poland
Konuralp, A., Turkey
Kumar, D., India
Kumar, S., India
Leblebicioglu, K., Turkey
Li, C.P., China
Liu, F., Australia
Luo, A.C.J, USA
Machado, J.A.T., Portugal
Magin, R. L., USA
Marin, M., Romania
Mandzuka, S., Crotia
Markowski, K.A, Poland
Merdan, H., Turkey
Momani, S., Jordan
Nigmatullin, R. R., Russia
Pinto, C.M., Portugal
Podlubny, I., Slovakia
Povstenko, Y., Poland
Prodanov, D., Belgium
Sari, M., Turkey
Shahzad, M., Pakistan
Singh, J., India
Shimizu, N., Japan
Sproessig, W., Germany
Sulaiman, S., Malaysia
Tabucanon, M.T., Thailand
Tank, F., Turkey
Tarasov, V.E., Russia
Thuto, M.V., Botswana
Torres, D.F.M., Portugal
Townley, S, UK
Trujillo, J.J., Spain
Weber, G.W., Poland
Wu, G.C., China
Yang, X.J., China
Zhang, D.Z., UK

Organizing Committee

Yaman, R., Turkey (Chair)
Machado, J.A.T., Portugal (Co-chair)
Baleanu, D., Turkey (Co-chair)
Ozdemir, N., Turkey (Co-chair)
Jordan Hristov, Bulgaria (Co-chair)

Local Organizing Committee

Akdemir, M., Turkey
Avci, D., Turkey
Demirtas, M., Turkey
Ergun, K., Turkey
Evirgen, F., Turkey
Inan, D., Turkey
Iskender, B.B., Turkey
Karaoglan, A.D., Turkey
Kasemi, E.C., Kosova
Kaymak, O.O., Turkey
Kucukkoc, I., Turkey
Ucar, E., Turkey
Ucar, S., Turkey
Yaman, G., Turkey
Yavuz, M., Turkey
Yildirim, H.H., Turkey

CONFERENCE PROGRAM OUTLINE

Day 1: Wednesday1st September 2021		Day 2: Thursday, 2nd September 2021		Day 3: Friday, 3rd September 2021	
09:30-10:00	**Opening Ceremony** (Main Hall) *Prof. Dr. Ramazan Yaman* *Prof. Dr. Dumitru Baleanu* *Prof. Dr. Necati Özdemir*	09:30-10:15	**Plenary Lecture-5** (Main Hall) Prof. Dr. Dumitru Baleanu *Chair: Prof. Dr. Jordan Hristov*	09:30-10:15	**Plenary Lecture-6** (Main Hall) Prof. Dr. Praveen Agarwal *Chair: Prof. Dr. Necati Özdemir*
10:00-10:45	**Plenary Lecture-1** (Main Hall) Prof. Dr. J. A. Tenreiro Machado *Chair: Prof. Dr. Dumitru Baleanu*	10:20-10:50	**Invited Speaker-1 (Main Hall)** Prof. Dr. Jordan Hristov *Chair: Prof. Dr. Dumitru Baleanu*	10:20-10:50	**Invited Speaker-4 (Main Hall)** Prof. Dr. Amin Jajarmi *Chair: Prof. Dr. Praveen Agarwal*
10:45	*Coffee Break*	*10:50*	*Coffee Break*	*10:50*	*Coffee Break*
11:00-11:45	**Plenary Lecture-2** (Main Hall) Prof. Dr. Sverre Holm *Chair: Prof. Dr. Dumitru Baleanu*	11:00-12:40	**Parallel Sessions - 4** 11:00-11:20 11:20-11:40 11:40-12:00 12:00-12:20 12:20-12:40	11:00-12:40	**Parallel Sessions - 7** 11:00-11:20 11:20-11:40 11:40-12:00 12:00-12:20 12:20-12:40
11:50-13:30	**Parallel Sessions - 1** 11:50-12:10 12:10-12:30 12:30-12:50 12:50-13:10 13:10-13:30	12:45-13:15	**Invited Speaker-2 (Main Hall)** Prof. Dr. Hüseyin Merdan *Chair: Prof. Dr. Gerhard W. Weber*	13:00	**Closing Ceremony** (Main Hall)
13:30	*LUNCH*	*13:15*	*LUNCH*		
14:00-14:45	**Plenary Lecture-3 (Main Hall)** Prof. Dr. Gerhard W. Weber *Chair: Prof. Dr. Necati Özdemir*	13:45-14:15	**Invited Speaker-3 (Main Hall)** Prof. Dr. Carla Pinto *Chair: Prof. Dr. Zakia Hammouch*		
14:45	*Coffee Break*	*14:15*	*Coffee Break*		
14:55-16:35	**Parallel Sessions - 2** 14:55-15:15 15:15-15:35 15:35-15:55 15:55-16:15 16:15-16:35	14:25-16:05	**Parallel Sessions - 5** 14:25-14:45 14:45-15:05 15:05-15:25 15:25-15:45 15:45-16:05		
16:40-17:25	**Plenary Lecture-4 (Main Hall)** Prof. Dr. Albert C. J. Luo *Chair: Prof. Dr. J. A. Tenreiro Machado*	*16:05*	*Coffee Break*		
		16:15-17:55	**Parallel Sessions - 6** 16:15-16:35 16:35-16:55 16:55-17:15 17:15-17:35 17:35-17:55		
17:25	*Coffee Break*	*17:55*	*Coffee Break*		
17:35-18:55	**Parallel Sessions - 3** 17:35-17:55 17:55-18:15 18:15-18:35 18:35-18:55	18:05-18:35	**Round-table Discussion**		

Contents

Bistatic Detection Performance: Random vs Coordinated Search

Mumtaz Karatas*, Levent Eriskin

Department of Industrial Engineering, National Defence University, Turkish Naval Academy
34640, Istanbul, Turkey
mkaratas@dho.edu.tr, leriskin@dho.edu.tr

Abstract

The problem of assessing the coverage quality of underwater sensor networks for search, detection and surveillance is an important research topic both for practitioners and researchers. Among different underwater sensor types, multistatic and bistatic sonars are effectively used by navies worldwide for protecting maritime zones and friendly units against hostile submerged targets. Such systems consist of sources and receivers which need not be collocated. If a multistatic underwater surveillance system is composed of single independent source and receiver at different locations, these systems are called bistatic sensors. In this study, we consider comparing the performance of random and coordinated deployment strategies of bistatic sonobuoys against stationary and mobile targets. To achieve this, using underwater acoustic theory, we first model the coverage zone of a bistatic sensor couple as a group of Cassini ovals. Next, using simple analytical geometry, we map the problem to a two-dimensional geometric problem. Finally, we approximate the expected coverage area of bistatic sensors by using equations derived from this mapping and analyse the relationship between the dimensions of the search field and coverage. We also use results from previous work obtained for the coordinated deployment of bistatic sensors against mobile targets and compare the results. The results obtained from this work can be utilized by decision-makers and practitioners for back-of-the-envelope analysis to estimate the search performance of bistatic sensors.

Keywords: Area coverage, bistatic sonar, search theory

1. Introduction

A monostatic sonar consists of a co-located source and receiver. A bistatic sonar, on the other hand, is a variant of the traditional monostatic sonar where the source and receiver are not co-located. A multistatic system consists of multiple sources and receivers distributed over the search area. Each source-receiver couple is named as a bistatic sonar system. In other words, a bistatic system is a special case of a multistatic system which consists of a single source and receiver. The basic operating concept of a multistatic sonar sensor network is to emit sound energy from a source into the water and listen for the reflected echoes returning across the receivers to detect the targets of interest.

Multistatic systems have several advantages over monostatic systems. One advantage is the covertness of the receive platforms which will make taking countermeasures difficult for the target (Craparo and Karatas, 2018; Craparo and Karatas, 2020). In addition, multistatic systems enable multi-angle observations and therefore improve tracking accuracy. It also allows multi-platform operations such as an airplane deployed receivers with a surface ship or a dipping helicopter source. The main disadvantage that multistatic systems possess is the increased system complexity and unusual coverage patterns determined by the transmission losses (Craparo et al., 2019; Craparo et al., 2017; Fügenschuh et al., 2020).

* *Corresponding Author*

Target detection (or area coverage) performance of a monostatic sonar mainly depends on the sensing range of the sensor and the distance between the sensor and target. For a bistatic sonar, on the other hand, this probability is based on the distance between the source and target as well as the distance between the target and receiver. In particular, for a given environmental condition and target type, the sensing zone of a bistatic sonar is characterized by a set of distinct ovals named as Cassini ovals (Wang et al., 2008). Hence, the problem of developing optimal (or effective) sensor location layouts for bistatic sonars is considerably harder than it is for the monostatic sonars.

In this study our main ambition is to compare the performance of random and coordinated bistatic sonobuoy deployment strategies in terms of area coverage (or stationary target detection) and mobile target detection. To achieve this, we use simple analytical geometry and acoustic theory to map the problem of determining expected area coverage of random deployed sensors to a two-dimensional geometric problem. While doing this we use special features of the Cassini ovals that are used to model the sensing zone of a bistatic couple. Next, we adopt the results in (Karatas et al., 2018) obtained for coordinated bistatic deployment scenarios both for area coverage and mobile target detection. In Section 2, we provide preliminaries on the concept of bistatic theory. In Section 3 we define our problems and derivations for both deployment strategies. Finally, we conclude with a few remarks in Section 4.

2. Preliminaries

In bistatic sonar systems, detection occurs if the emitted acoustic energy from the source is reflected off by the target and this energy generates an echo at the receiver such that it is greater than the receiver's detection threshold. Hence, since the energy travels from the source and target, and from the target to the receiver, the detection probability depends on both distances as well as other factors such as environmental conditions, receiver sensitivity, false alarm setting. Based on acoustic theory and results in (Urick 1983; Washburn and Karatas, 2015) for a bistatic system detection happens if:

$$R_1 R_2 \leq b^2 \tag{1}$$

where R_1 and R_2 represent source-target and target-receiver distances (as shown in Figure 1(left)) and b is the equivalent monostatic detection range. In other words, the parameter b represents the detection range of the bistatic sensor couple when the source and receiver are co-located.

The inequality (1) defines the interior of the well-known ovals named as "Cassini ovals" which is also used for representing the sensing zone of a bistatic couple (Cox, 1989). These distinct ovals are used in a variety of applications such as nuclear physics, military, acoustics, and bio-sciences (Karatas, 2013). Let d denote the separation distance and a the semi-distance between the source and receiver, i.e. $d=2a$. Then given that the sensors are located at $(\pm a,0)$ the Cartesian equation of a Cassini oval is $\left[(x-a)^2+y^2\right]\left[(x+a)^2+y^2\right]=b^4, (a,b)\in\mathbb{R}$. Then the shape of the Cassini oval categorized by the ratio a/b is depicted in Figure 1(right). Note that in the figure we fix $b=1$ for simplicity. The figure shows that when the sensors are co-located, the bistatic sensing zone reduces to a monostatic sensor sensing zone and is modelled as a disk of radius b. For $a/b=1$ the oval becomes a lemniscate of Bernoulli and when $a/b>1$ it comes two disjoint ovals (Willis, 2008).

3. Random and Coordinated Use of Bistatic Sonobuoys

We now analyse and compare the outcomes of random and coordinated deployment of bistatic sonobuoys against stationary and mobile targets. In our analysis we adopt a deterministic sensing model such that a target is detected only if it lies within the sensing zone (denoted by C) of the sensors and not detected if it is outside of C (Karatas, 2017; Karatas, 2018; Karatas and Eriskin, 2021). We denote the area of C as A_C.

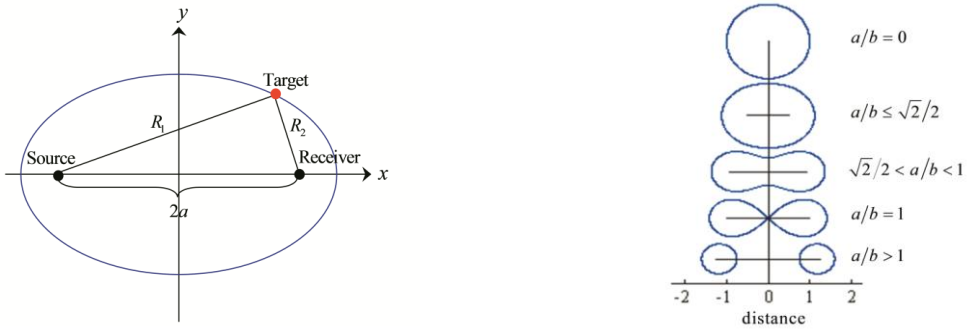


Figure 1. (left) Bistatic triangle, (right) A family of Cassini ovals for b=1.

3.1. Stationary Target Detection

Suppose that a source and receiver is deployed in a rectangular region F of area A_F with sizes $m \times n$ in which a stationary target is distributed uniformly. This assumption means that there is no prior information on the location of the target in the region. We can simply express the detection probability as the ratio area covered by the sensor pair to the area of the region, i.e. A_C/A_F. Adopting results derived in (Karataş and Akman, 2015) we can express the area A_C as follows:

$$A_C = \begin{cases} 4\int_0^{\sqrt{a^2+b^2}} f(x)dx, & a/b \le 1 \\ 4\int_{\sqrt{a^2-b^2}}^{\sqrt{a^2+b^2}} f(x)dx, & a/b > 1 \end{cases} \tag{2}$$

where $f(x) = \pm\sqrt{-a^2 - x^2 \pm \sqrt{4x^2a^2 + b^4}}$. The results in (Karatas et al., 2018) reveal that for a given area A and range b, the detection probability is maximized for the area A_C obtained for a/b=0. That is, the area of the Cassini oval is maximum when the sensors are co-located, i.e., the area is $A_C=\pi b^2$. Therefore, for coordinated search the optimal strategy is to co-locate the sensors to achieve maximum area coverage and target detection probability.

$$g(s) = \begin{cases} -2\dfrac{\sqrt{s}}{m^2 n} - 2\dfrac{\sqrt{s}}{mn^2} + \dfrac{\pi}{mn} + \dfrac{s}{m^2 n^2}, & 0 < s \le m^2 \\ -2\dfrac{\sqrt{s}}{m^2 n} - \dfrac{1}{n^2} + \dfrac{2}{mn}\arcsin\left(\dfrac{m}{\sqrt{s}}\right) + \dfrac{2}{m^2 n}\sqrt{s - m^2}, & m^2 < s \le n^2 \\ \dfrac{1}{n^2} + \dfrac{2}{mn}\arcsin\left(\dfrac{m}{\sqrt{s}}\right) + \dfrac{2}{m^2 n}\sqrt{s - m^2} & \\ -\dfrac{1}{m^2} + \dfrac{2}{mn}\arcsin\left(\dfrac{n}{\sqrt{s}}\right) + \dfrac{2}{mn^2}\sqrt{s - n^2} - \dfrac{\pi}{mn} - \dfrac{s}{m^2 n^2}, & n^2 < s \le m^2 + n^2 \end{cases} \tag{3}$$

Now suppose that bistatic sensors are randomly deployed in A. Since the area A_C is a function of the separation distance d, we first need to derive the expected distance between the sonobuoys. To achieve this, we first map the bistatic deployment problem two a geometric problem such that each sensor denotes a point in a two-dimensional rectangle of sizes $m \times n$ and by using results the given in (Philip, 2007) we can express the probability distribution of the distance between these two random points as in equation (3). In the equation $s=d^2$ and $g(d) = g(d^2)\partial s/\partial d = 2d.g(s)$. Figure 2 depicts the probability density function of distance between a source and receiver couple that are randomly deployed in a square field of size 4×4, 5×5, and 6×6. Using this result, we can express the expected separation distance as:

$$\mathrm{E}[d] = \int_0^{m^2+n^2} \sqrt{s}.g(s).\partial s = \int_0^{\sqrt{m^2+n^2}} d.g(d).\partial d \tag{4}$$

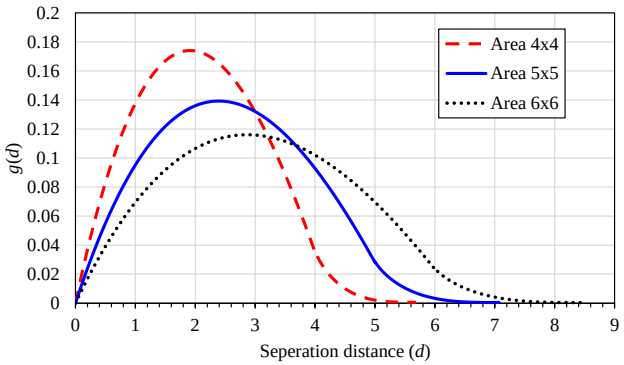


Figure 2. Pdf for the distance for sensors randomly deployed in a field of sizes 4×4, 5×5, and 6×6.

Upon computing E[d], the expected area coverage E[A_C] can be computed as:

$$E[A_C] = \int_{-\infty}^{\infty} A_C.g(A_C).\partial A_C = 4\int_0^{\sqrt{m^2+n^2}} \int_0^{\sqrt{s/4+b^2}} \sqrt{-s-x^2+\sqrt{4x^2s+b^4}}.g(s).\partial x\partial s \tag{5}$$

3.2. Mobile Target Detection

We now assume that a mobile target moving on a straight line of infinite length transits the area A such its trajectory crosses A equiprobably. Using the results in (Karatas et al., 2018) which map the mobile target detection with bistatic sensors to a line-set intersection problem, we can define the detection probability of a target as $Pd = P_C^{eff}/P_F$ where P_C^{eff} and P_F represent the effective perimeter of the Cassini oval and perimeter of the search area. Below we list the formulas used to calculate the effective perimeter of the Cassini oval for a given separation distance and b.

$$P_C^{eff} = \begin{cases} P_C, & a/b \le \sqrt{2}/2 \\ P_C^e, & \sqrt{2}/2 < a/b \le 1 \\ P_C^e - P_C^i + P_C, & a/b > 1 \end{cases} \tag{6}$$

$$P_C = \begin{cases} 4\int_0^{\sqrt{a^2+b^2}} g_C(x)dx, & a/b \le 1 \\ 4\int_{\sqrt{a^2-b^2}}^{\sqrt{a^2+b^2}} g_C(x)dx, & a/b > 1 \end{cases}, \quad P_C^e = 4\left(\frac{\sqrt{-b^4+4a^4}}{2a} + \int_{\frac{1}{2}\frac{\sqrt{-b^4+4a^4}}{a}}^{\sqrt{a^2+b^2}} g_C(x)dx \right) \tag{7}$$

$$P_C^i = 4\left(\sqrt{-a^2+\sqrt{2a^4-b^4+2\sqrt{a^4(a^4-b^4)}}} + \int_{\frac{1}{2}\frac{\sqrt{2a^4-2b^4+2\sqrt{a^8-b^4a^4}}}{a}}^{\sqrt{a^2+b^2}} g_C(x)dx \right) \tag{8}$$

where $g_C(x) = \sqrt{\dfrac{b^4(a^2-\sqrt{4x^2a^2+b^4})}{(4x^2a^2+b^4)(a^2+x^2-\sqrt{4x^2a^2+b^4})}}$.

The results for show that the effective perimeter takes its maximum value when a/b=1, i.e. when the oval is a lemniscate. Hence for coordinated search the optimal policy for detecting a transiting target is setting the semi-distance a value equal to b.

3.3. Comparison of Random and Coordinated Deployments

Using the analytic results obtained in previous subsection we now display a comparison of both deployment approaches for static and mobile target detection. Figure 3(left) shows a comparison of stationary target detection probabilities of the two strategies with respect to different search areas A_F ranging from 1 unit^2 to 50 unit^2. Similarly, Figure 3(right) shows the performance of both strategies for mobile targets.

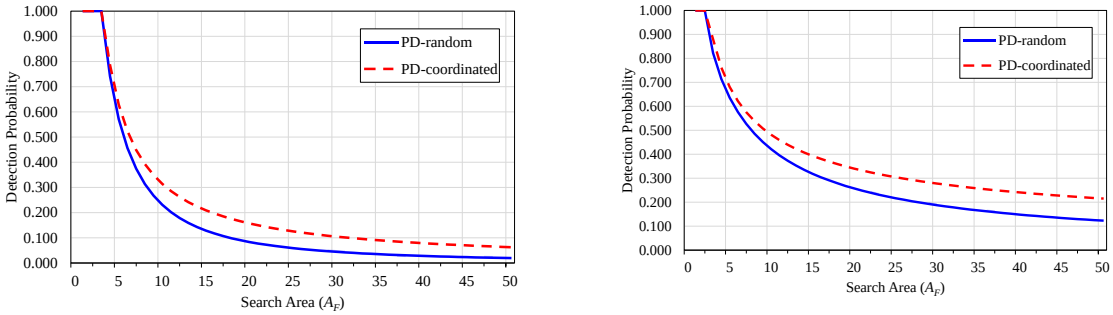


Figure 3. (left) Detection probability of stationary targets by random and coordinated deployments, (right) Detection probability of mobile targets by random and coordinated deployments.

Figures reveal that for both cases the coordinated placement of sensors outperforms the random approach in terms of detection probability performance. One important observation is that, for both target models the difference between the performance of the two strategies is negligible for small areas, e.g. for A_F<10 unit^2. This is an expected result considering that small areas yield shorter distances between randomly deployed sensors, and hence relatively large coverages. As the area increases, coordinated deployment performs almost twice better than the random strategy.

4. Conclusion

We have considered two deployment strategies for bistatic sensors: random and coordinated. For both cases we studied stationary and mobile target models and using analytic geometry, acoustic theory and results obtained from previous work, we compared the performance of the two strategies for both target models. For the coordinated search, a natural choice for stationary targets is to co-locate the source and receiver whereas for transiting targets the optimal choice is to set a/b=1. For the random case we have derived analytic results to estimate the coverage of the sensors. In its current form our analysis provides a "back of the envelope" type estimate for both cases as a function of the ratio a/b. As a future work, the comparison can be carried out for more realistic cases which incorporates the direct blast zone effect (Craparo and Karatas, 2015), noise and interference, more realistic gradual cover sensing models.

References

Cox, H. (1989). Fundamentals of bistatic active sonar. In Underwater acoustic data processing (pp. 3-24). Springer, Dordrecht.

Craparo, E. M., Fügenschuh, A., Hof, C., & Karatas, M. (2019). Optimizing source and receiver placement in multistatic sonar networks to monitor fixed targets. European Journal of Operational Research, 272(3), 816-831.

Craparo, E. M., Karatas, M., & Kuhn, T. U. (2017). Sensor placement in active multistatic sonar networks. Naval Research Logistics (NRL), 64(4), 287-304.

Craparo, E. M., & Karatas, M. (2018). A method for placing sources in multistatic sonar networks. Technical Report, Naval Postgraduate School, Monterey, United States.

Craparo, E., & Karatas, M. (2020). Optimal source placement for point coverage in active multistatic sonar networks. Naval Research Logistics (NRL), 67(1), 63-74.

Fügenschuh, A. R., Craparo, E. M., Karatas, M., & Buttrey, S. E. (2020). Solving multistatic sonar location problems with mixed-integer programming. Optimization and Engineering, 21(1), 273-303.

Karatas, M., & Craparo, E. (2015). Evaluating the direct blast effect in multistatic sonar networks using Monte Carlo simulation. In 2015 IEEE Winter Simulation Conference (WSC) (pp. 1184-1194).

Karatas, M., Craparo, E., & Akman, G. (2018). Bistatic sonobuoy deployment strategies for detecting stationary and mobile underwater targets. Naval Research Logistics (NRL), 65(4), 331-346.

Karataş, M. (2013). A multi foci closed curve: Cassini oval, its properties and applications. Doğuş Üniversitesi Dergisi, 14(2), 231-248.

Karatas, M. (2018). Optimal deployment of heterogeneous sensor networks for a hybrid point and barrier coverage application. Computer Networks, 132, 129-144.

Karataş, M., & Akman, G. (2015). Bistatic sonobuoy deployment configuration for stationary targets. Journal of Naval Sciences and Engineering, 11(2), 1-10.

Karatas, M. (2017). A multi-objective facility location problem in the presence of variable gradual coverage performance and cooperative cover. European Journal of Operational Research, 262(3), 1040-1051.

Karatas, M., & Eriskin, L. (2021). The minimal covering location and sizing problem in the presence of gradual cooperative coverage. European Journal of Operational Research.

Philip, J. (2007). The probability distribution of the distance between two random points in a box (pp. 1-13). KTH mathematics, Royal Institute of Technology.

Urick, R. J. (1983) Principles of Underwater Sound, 3rd ed. McGraw-Hill, New York, NY.

Wang, I. J., Lim, J. H., & Terzis, A. (2008, October). Energy-efficient sensor management in multi-static active sonar networks. In 2008 42nd Asilomar Conference on Signals, Systems and Computers (pp. 1611-1616). IEEE.

Washburn, A., & Karataş, M. (2015). Multistatic search theory. Military Operations Research, 20(1), 21-38.

Willis, N. J. (2008). Bistatic radar. Radar Handbook. M. I. Skolnik (Editor in Chief), McGraw-Hill Professional pp 23.4

Analysis of Tsunami and Tsunami-Structure Interaction Parameters by Compressive Sensing

Cihan Bayındır[1,2*]

[1]*İstanbul Technical University, Civil Engineering Department, İstanbul, Turkey*
[2]*Boğaziçi University, Civil Engineering Department, İstanbul, Turkey*
cbayindir@itu.edu.tr

Abstract

Tsunamis are among the most devastating hazards that can be observed in nature. Observation, sensing, recording, and analysis of the tsunami and tsunami-structure interaction parameters are of crucial importance for the safety of the coastal zone and communities. These parameters include but are not limited to tsunami water surface fluctuations, particle velocities, inundation, runup, sediment deposit, their dynamics pressures on structures. Efficient sensing, data recording, and analysis of these parameters is critically important for the reconnaissance, assessment, early warning, and avoidance of catastrophic consequences of tsunamis. One of the most successful sensing algorithms of the big data era is the compressive sensing technique (CS), which can outperform classical sampling methodologies by using far fewer samples while achieving exact recovery (Candes et al., 2006a, 2006b). In this paper, we investigate the possible usage of the CS for the effective measurement and reconstruction of the tsunami parameters of water surface fluctuation, particle velocities, and tsunami-induced wave pressures. Using the data sets of the Japanese Tohoku Tsunami occurred in 2011 after a major earthquake of M_w 9.0 (Koshimura et al., 2015), provided by the USA's National Oceanic and Atmospheric Administration (NOAA)'s Deep-Ocean Assessment and Reporting of Tsunamis (DART) portal (NOAA, 2011), we show that CS can be used as an effective tool for the measurement, analysis, and reconstruction of the tsunami and tsunami-structure interaction parameters. Although we limit ourselves with the reconstruction of water surface fluctuations and tsunami-induced dynamic pressures (Goda, 2010), the CS can be applied for monitoring of the tsunami parameters in more general settings including the effects of vortices and shorter waves (Bayındır, 2019, Bayındır et al., 2021). We discuss our findings and comment on their possible applicability and usage.

Keywords: Tsunamis, tsunami-structure interaction, compressive sensing

1. Introduction

On the 11th of March, 2011 a major earthquake with M_w 9.0 hit the Honshu region of the Japan trench. After the earthquake, a tsunami wave reached the shore in approximately 30 minutes. This 2011 Tohoku earthquake and tsunami event, which was the largest earthquake recorded in Japan, caused the killing of more than 18000 people and great volumes of overtopping from seawalls, which caused significant damage to the infrastructures including the damage and shut-downs of three nuclear reactors (Koshimura et al., 2015, NOAA, 2011). The epicentre and tsunami travel time map of the Tohoku earthquake is taken from the NOAA's DART portal (https://www.ngdc.noaa.gov/hazard/dart/2011honshu_dart.html) and depicted in Fig. 1. As this figure confirms, there were some DART stations located in the Pacific for the recording of the wave field parameters such as the water surface fluctuations (NOAA, 2011). Along the Pacific, significant changes in the wavefield and water surface fluctuations are observed even

* *Corresponding Author*

after 20 hours. Such observation times are also confirmed by various software available in the literature. FUNWAVE model is a Boussinesq equation solver and the simulation of the Tohoku 2011 tsunami event depicted in Fig. 2 confirms this finding. It is possible to download the data for different stations using the DART portal. In this study, we use the data given for station 21418, which is one of the closest stations to the epicenter of the Tohoku 2011 earthquake.

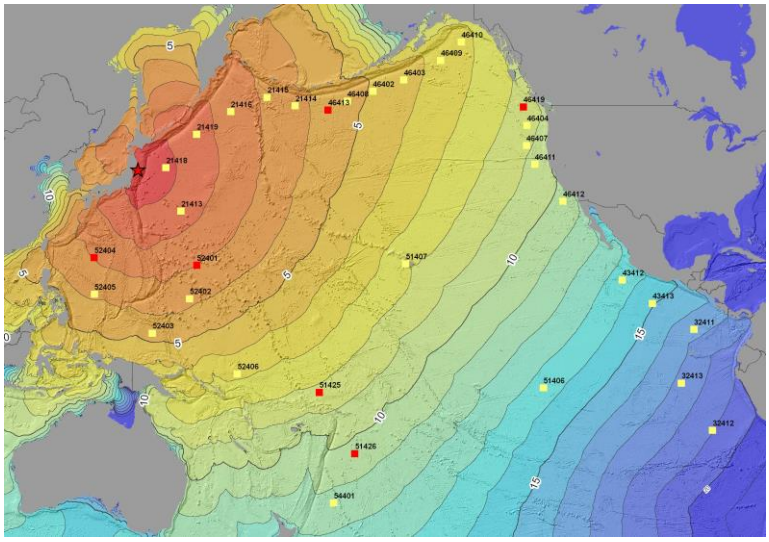


Figure 1. Map for the Tohoku earthquake travel times generated by TTT software developed by Paul Wessel, figure courtesy of NOAA (2011).

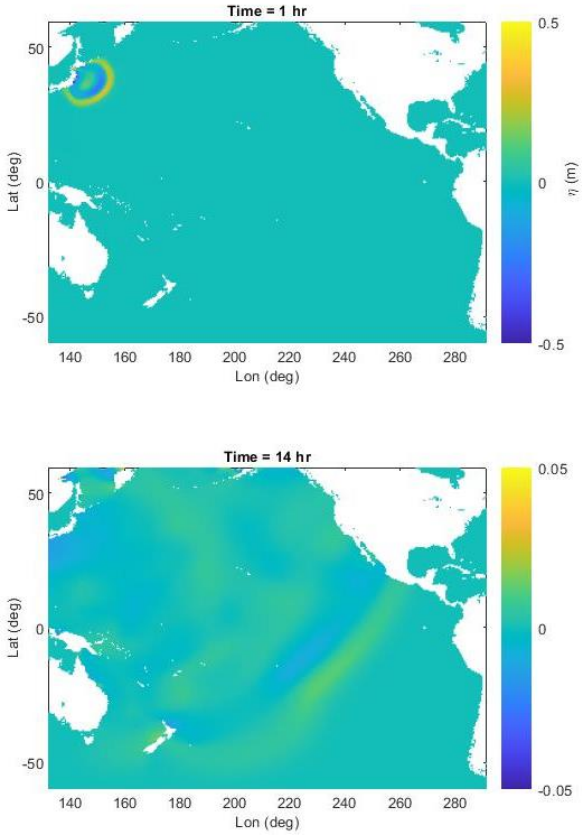


Figure 2. Water surface levels after 14hrs obtained by Boussinesq equation solver FUNWAVE model

2. Observation of Tsunami Parameters by Compressive Sensing

2.1. Review of the Compressive Sensing

Compressive sampling (CS) is one of the most revolutionary algorithms of the era (Candes et al., 2006a, 2006b). A brief summary of the CS is tried to be given in this section. Consider the signal η which is *K*-sparse. That is only *K* of the *N* entries are nonzero. It is possible to transform η into an orthogonal domain, such as the spectral Fourier domain by means of a transformation matrix Ψ. Thus the representation of the signal becomes $\eta = \psi\hat{\eta}$, where $\hat{\eta}$ denotes the coefficient vector. One can get $\eta_s = \psi\hat{\eta}_s$, where η_s is the signal with non-zero entries, after eliminating the zero entries. CS algorithm states that a *K*-sparse signal η having *N* entries can exactly be reconstructed by using $M \geq C\mu^2(\phi,\psi)K\log(N)$ samples. In this formulation, ϕ denotes the sensing and ψ denotes the transformation basis (Candes et al., 2006a, 2006b) and $\mu^2(\phi,\psi)$ is their mutual coherence. *C* shows a positive constant. Taking *M* random projections and by using the sensing matrix, one can get $g = \phi\eta$. Thus, the CS problem can be formulated as

$$\min \|\hat{\eta}\|_{l_1} \text{ subjected to } g = \phi\psi\hat{\eta} \quad (1)$$

where $\|\hat{\eta}\|_{l_1} = \sum_i |\hat{\eta}_i|$. The l_1 solution of the CS problem can be calculated as $\eta_{CS} = \psi\hat{\eta}$ (Candes et al., 2006a, 2006b). The reader is referred to Candes et al. (2006a, 2006b) for a comprehensive discussion of the CS.

3. Results and Discussion

In order the discuss and prove the applicability of the CS for successful measurements and reconstruction of the tsunami parameters, we first use the Tohoku 2011 tsunami water surface fluctuation time series recorded at station 21418 and depicted in Fig. 3. On March 11th, 2021 at around 05.48 am the initial peak and around 06.13 am the main peak appears in the recordings of station 21418. The peak wave height recorded at this station is observed to be around *H=2.15m*.

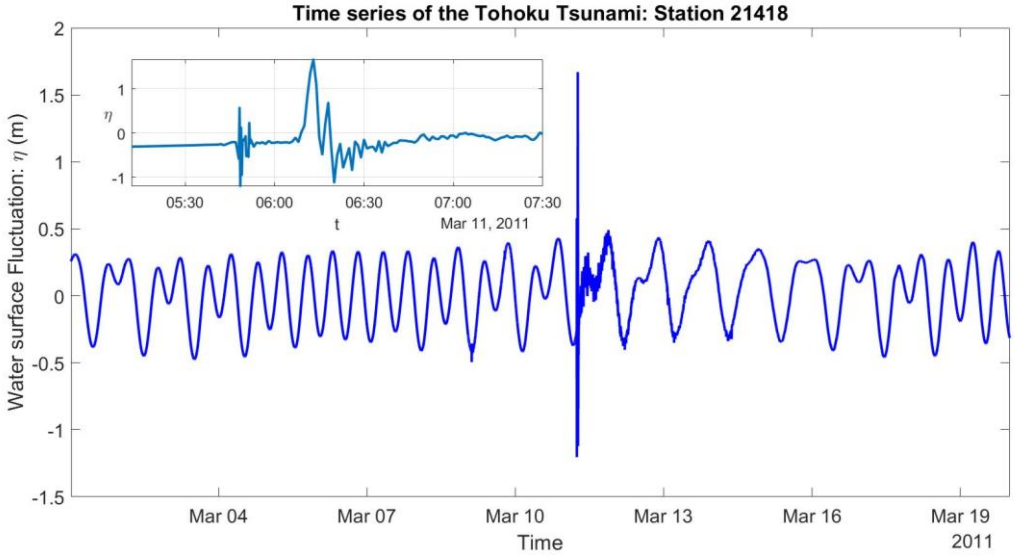


Figure 3. Time series of the Tohoku 2011 tsunami event recorded at station 21418.

3.1. Results for the Tsunami Water Surface Fluctuations

In this section, we discuss the possible usage of the CS for the recovery of the tsunami water surface fluctuation data measured at the station and depicted in Fig. 3. The Tohoku 2011 tsunami time series

data depicted in Fig. 3 exhibits a fluctuating behavior before and after the tsunami event. Thus, the spectra of this time series have a sparse representation in the Fourier domain, as illustrated in Fig. 4.

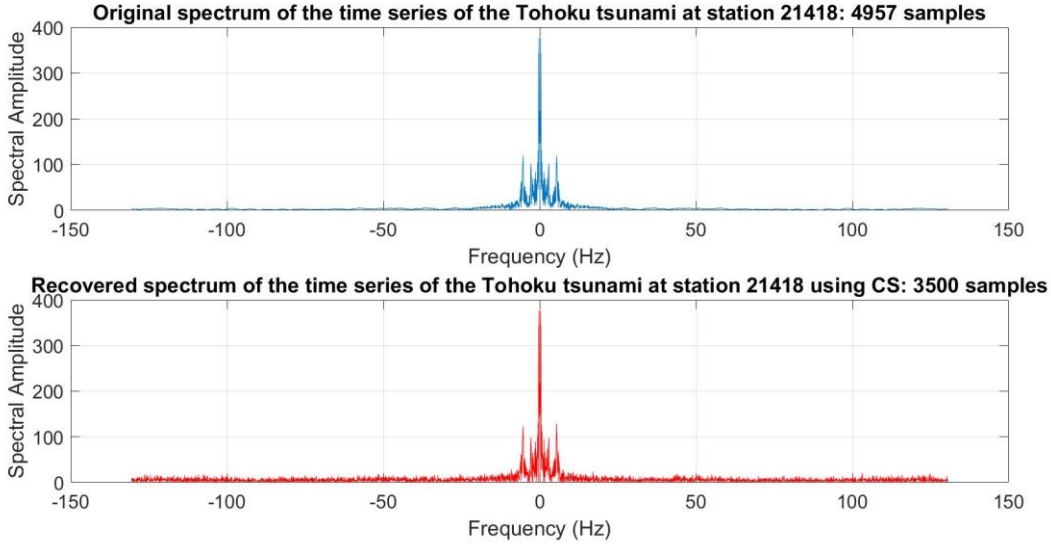


Figure 4. Spectra of the Tohoku 2011 tsunami time series measured at station 21418: a) classical sampling with 4957 samples b) CS reconstruction with 3500 samples.

As illustrated in Fig. 4, the time series of the Tohoku 2100 tsunami has a sparse spectrum. The classical spectrum has 4957 samples and by taking 3500 random measurements and solving the l_1 problem of the CS, we construct the spectra depicted in the second subfigure of Fig. 4. As one can observe from the figure, the spectral information around the central wavenumbers is well-preserved during the CS process. However, at high absolute wavenumbers, some mismatch is present in the form of white noise due to the random sampling process. The time series corresponding to the spectra depicted in Fig. 4 are constructed by an inverse Fourier transform operation and displayed in Fig.5.

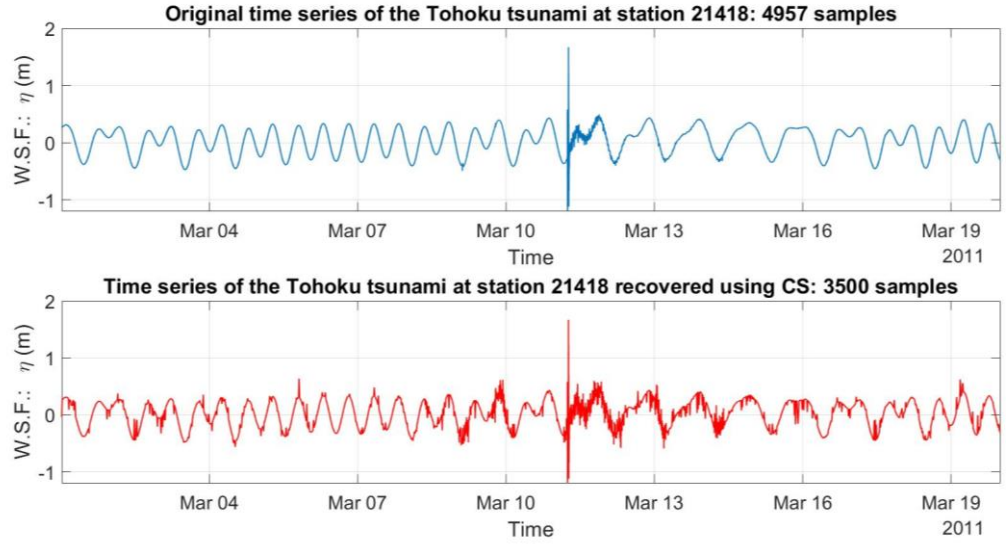


Figure 5. Time series of the Tohoku 2011 tsunami event recorded at station 21418: a) classical measurements b) compressive sampling reconstruction.

In Fig.5, it is clear that the effect of white noise is present. This causes some distortion of the time series, but it can be easily filtered out using a low-pass filter. The CS algorithm can successfully recover the initial peak and main peak of the water surface fluctuation of the Tohoku 2011 tsunami. The significant

advantage of the CS-based method is the number of samples and the undersampling ratio. The number of compressive samples is 3500 for this case, however, depending on the behavior of the time series, it is possible to achieve an exact recovery using an even fewer number of samples. The undersampling ratio is very beneficial for many measurements, analysis, interpolation/extrapolation, and structural health monitoring applications. The number of sensors as well as the temporal resolution of the measurements can be reduced to minimize measurement and data storage requirements. Although the findings presented above are given for the water surface fluctuations, they can be easily extended for the reconstruction of other parameters such as nonlinear tsunami and rogue wave interactions and tsunami-induced vortices (Bayındır, 2019, Bayındır et al., 2021).

3.2. Results for the Tsunami Induced Long-Wave Pressures

In order to design coastal protection structures, the loads induced by long waves need to be determined. There are various formulas proposed in the literature for this purpose (Goda, 2010). One of the most commonly used formulas for the estimation of the long-wave induced pressures is Hiroi's formula given as

$$P = 1.5\rho g H \tag{2}$$

Here P is the long-wave induced pressure, ρ is the seawater density, g is the gravitational acceleration and H is the long-wave height. This formula provides a uniform distribution along with depth in accordance with the depth-averaged long-wave celerity. Coastal and other sea defence structures are equipped with measurement tools can measure the wave-induced pressures in a real structural health monitoring setting or in the laboratory environment. In order to discuss the typical form of the loading we use the parameters as ρ=1025 kg/m^3, g=9.81m/s^2, *H=2.15 m* as the representative values of the Tohoku 2011 tsunami event at station 21418, and plot the corresponding depth variations in Fig. 6.

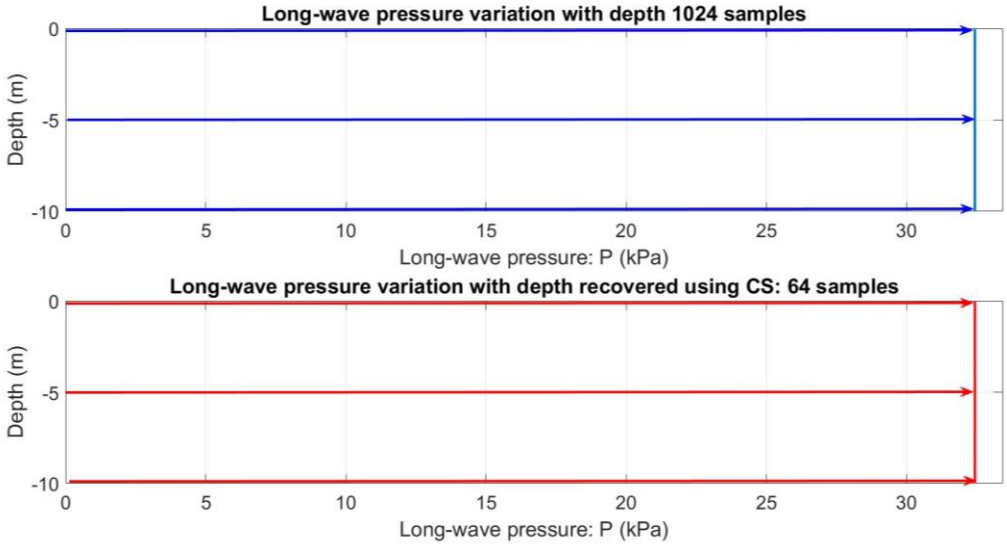


Figure 6. Compressive sampling reconstruction of the time series of the wave moment exerted by a monochromatic wave on a pile.

As Fig. 6 confirms, the depth variation can be successfully reconstructed using 64 compressive samples, where the classical measurements require 1024 samples. This strategy also works if the variation of the parameter in consideration is small along with the depth. More importantly, it is possible to keep the temporal variation of the wave height, *H(t)*, and use this time series in Hiroi's formula to obtain the time series of the long wave-induced force. Thus CS can be effectively used to construct the time series and

depth variation of the wave-induced parameters effectively. One can also extend our results to construct the 2D wavefield properties and use CS for efficient monitoring and assessment of the health of coastal structures.

4. Conclusion

In this study, we studied the problem of efficient sensing of the tsunami and tsunami-structure interaction parameters by compressive sensing. Using a data set recorded during the Tohoku 2011 tsunami event, we showed that water surface fluctuations of long waves can be efficiently measured and recorded using the compressive sensing technique. We also showed that long-wave induced pressure profiles can be efficiently recorded using a far fewer number of sensors than the classical sampling theory dictates. Our finding can also be extended for the enhancement of other tsunami-related technologies, such as the tsunami early warning systems. It is also possible to measure, analyze and interpolate the aforementioned and other types of tsunami data including horizontal excursion velocities, diffraction, and vortex fields. If a tsunami event occurs in a calm environment with no wave present, such a signal may also be treated as a sparse signal in the time/spatial domain, thus random compressive sampling can be performed in the Fourier domain.

Acknowledgements

The author gratefully acknowledges the support of the İstanbul Technical University. This work was supported by the Research Fund of the İstanbul Technical University. Project Code: MGA-2020-42544. Project Number: 42544.

References

Bayındır, C. (2019). Early detection of rogue waves using compressive sensing, TWMS Journal of Applied and Engineering Mathematics, 9 (2), 198-205. http://jaem.isikun.edu.tr/web/images/articles/vol.9.no.2/03.pdf

Bayındır, C., & Namlı, B. (2021). Efficient sensing of the von Karman vortices using compressive sensing, Computers & Fluids, 104975. https://doi.org/10.1016/j.compfluid.2021.104975

Candes, E. J., Romberg, J., & Tao, T. (2006a). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. on Inf. Theory, 52 (2), 489-509. https://doi.org/10.1109/TIT.2005.862083

Candes, E. J. (2006b). Compressive sampling, Proc. Int. Congr. Math., 3, 1433-1452. https://doi.org/10.4171/022-3/69

Goda, Y. (2010). Random Seas and Design of Maritime Structures, Advanced Series on Ocean Engineering-Volume 33, World Scientific, New Jersey. https://doi.org/10.1142/7425

Koshimura, S., & Shuto, N. (2015). Response to the 2011 great east Japan earthquake and tsunami disaster, Phil. Trans. R. Soc. A., 373, 20140373. https://doi.org/10.1098/rsta.2014.0373

NOAA Dart Portal, https://www.ngdc.noaa.gov/hazard/dart/2011honshu_dart.html Accession Date: 05 August 2011.

A Goal Programming Approach for Resource Dependent Assembly Line Balancing Problem

Yakup Atasagun*, Alper Döyen

Konya Technical University, Industrial Engineering. Department, Konya, Turkey
yatasagun@ktun.edu.tr, alperdoyen@ktun.edu.tr

Abstract

In this study, a pre-emptive goal programming model is proposed for resource dependent assembly line balancing problem in order to provide flexibility for decision makers based on their decision environment and preferred priorities. Three conflicting goals namely total cost of workstation utilization (total number of utilized workstations), cycle time and total cost of additional resources such as equipment and assistant workers are considered. The proposed model is validated on an illustrative example and a scenario analysis is performed with different priority levels of the goals. The results show that the proposed goal programming formulation is valid and useful for balancing resource dependent assembly lines.

Keywords: Goal programming, line balancing, resource dependent assembly lines

1. Introduction

Assembly line balancing (ALB) is determining which task will be assigned to which workstation in such a way that precedence relations among tasks are not violated, sum of processing times of tasks in each workstation does not exceed cycle time and a performance measure is optimized. This performance measure is usually minimization of the number of workstations utilized over the line.

ALB problems was first studied by Salveson (1955) and they are interested by many researchers to date. The literature review studies of Baybars (1986), Ghosh and Gagnon (1989), Erel and Sarin (1998), Becker and Scholl (2006), Scholl and Becker (2006), Battaïa and Dolgui (2013) and Sivasankaran and Shahabudeen (2014) are useful for the interested readers.

In most of the ALB studies it is assumed that there is only one and fixed processing time for each task. However, this situation could not always be practical especially when different resource alternatives such as particular equipment or an assistant worker are available to process a task with different durations. Some tasks in practice cannot be processed by only one worker and may necessarily need to have additional assistant worker or particular equipment. Alternatively, assistance of another worker or use of an equipment type can reduce the processing time of a task even though these additional resources are not necessary. In this case, different processing alternatives (resource combinations) have to be used in the line (Kara et al., 2011).

Faaland et al. (1992) have defined this problem as resource dependent assembly line balancing problem (RDALB). Kara et al. (2011) have addressed the problem from a wide point of view and have adapted the RDALB approach to U-shaped assembly lines (RDULB) with some new practice-oriented assumptions.

In practice, assembly line managers may prefer to obtain compromise solutions among several conflicting objectives rather than optimising a single objective. The objectives and the priority levels of

* *Corresponding Author*

these objectives may be different with regard to the decision maker and decision making environment (Kara et al., 2010).

In this study a goal programming (GP) approach is proposed for RDALB in order to provide flexibility for decision makers based on their preferred priorities.

2. A GP Model for RDALB

The GP concept was introduced by Charnes and Cooper (1957) and has been widely used as an important modelling technique for multi-criteria decision making problems (Kara et al., 2010).

In this section, a pre-emptive GP model for RDALB is proposed by adhering to the assumptions of Kara et al. (2011) for RDALB. The proposed model is structured on Kara et al. (2011)'s mathematical formulation. It is also assumed that the decision-maker is able to determine goal values precisely and following three conflicting goals are included to the model: Goal 1 (G1): total cost of utilized workstations (worker + fixed cost), Goal 2 (G2): cycle time and Goal 3 (G3): total cost of additional resources (operating cost of equipment and employment cost of assistant workers).

G1 can also be thought as the total number of utilized workstations depending on the assumption that utilization costs of all workstations are equal to each other.

2.1. Notation

Indices, Parameters and Sets

i, r, s	: task
j	: workstation
e	: equipment
t^0_{ie}	: completion time of task i with equipment e without assistant
t^1_{ie}	: completion time of task i with equipment e with assistant
N	: set of tasks
E	: set of equipment
E_i	: set of equipment which can be used to process task i
NE_e	: available number of equipment e
NA	: available number of assistants
W	: set of workstations
PR	: set of precedence relations
$(r,s) \in PR$	: a precedence relation; task r is an immediate predecessor of task s
M	: a big number
CW	: utilization cost of a workstation (worker + fixed costs)
CA	: employment cost of an assistant
c_e	: operating cost of equipment e
$CostW$	: goal value for the total cost of workstation utilization (goal 1)
CT	: goal value for the cycle time (goal 2)
$CostR$	: goal value for the total cost of additional resources (goal 3)

Variables

x_{ij}	: 1, if task i is assigned to workstation j; 0, otherwise
p_{ije}	: 1, if task i is assigned to workstation j with equipment e without assistant; 0, otherwise
q_{ije}	: 1, if task i is assigned to workstation j with equipment e with assistant; 0, otherwise
z_{je}	: 1, if equipment e is assigned to workstation j; 0, otherwise
u_j	: 1, workstation j is utilized; 0, otherwise
k_j	: 1, if an assistant is assigned to workstation j; 0, otherwise
d^-	: under achievement of the goal 1
d^+	: over achievement of the goal 1
f_j^-	: under achievement of the goal 2
f^+	: over achievement of the goal 2

g^- : under achievement of the goal 3
g^+ : over achievement of the goal 3

2.2. Mathematical Formulation

$$Min \quad d^+ \tag{1}$$

$$Min \quad f^+ \tag{2}$$

$$Min \quad g^+ \tag{3}$$

$$\sum_{j\in W} x_{ij} = 1 \qquad \forall i \in N \tag{4}$$

$$\sum_{e\in E_i} (p_{ije} + q_{ije}) = x_{ij} \qquad \forall i \in N; \forall j \in W \tag{5}$$

$$\sum_{j\in W} (\|W\| - j + 1)(x_{rj} - x_{sj}) \geq 0 \qquad \forall (r,s) \in PR \tag{6}$$

$$\sum_{i\in N} x_{ij} \leq \|N\| u_j \qquad \forall j \in W \tag{7}$$

$$\sum_{i\in N} (p_{ije} + q_{ije}) \leq M z_{je} \qquad \forall e \in E_i; \forall j \in W \tag{8}$$

$$\sum_{j\in W} z_{je} \leq NE_e \qquad \forall e \in E \tag{9}$$

$$\sum_{i\in N} \sum_{e\in E_i} q_{ije} \leq M k_j \qquad \forall j \in W \tag{10}$$

$$\sum_{j\in W} k_j \leq NA \tag{11}$$

$$\sum_{j\in W} CWu_j + d^- - d^+ = CostW \tag{12}$$

$$\sum_{i\in N} \sum_{e\in E_i} (t^0_{ie} p_{ije} + t^1_{ie} q_{ije}) + f_j^- - f^+ = CT \qquad \forall j \in W \tag{13}$$

$$\sum_{j\in W} CAk_j + \sum_{j\in W} \sum_{e\in E} c_e z_{je} + g^- - g^+ = CostR \tag{14}$$

$$x_{ij}, p_{ije}, q_{ije}, u_j, k_j, z_{je} \in \{0,1\} \qquad \forall i \in N; \forall j \in W; \forall e \in E \tag{15}$$

$$f_j^- \geq 0 \qquad \forall j \in W \tag{16}$$

$$d^-, d^+, f^+, g^-, g^+ \geq 0 \tag{17}$$

The objective functions defined in (1), (2) and (3) minimizes the over achievements of the goals 1, 2 and 3 respectively. As an example, the goal of total workstation utilization cost (G1) is achieved when d^+ is found to be zero in the solution of the model. If d^+ is greater than zero, it means that the G1 is not achieved. Equation (4) ensures that each task is assigned to at least and at most one workstation. Equation (5) determines the resources (equipment type and assistant) allocated to a workstation. Precedence relationships among tasks are satisfied by the set of constraints given in equation (6). Equation (7) determines whether workstation *j* is utilized or not. Equation (8) determines whether equipment *e* is allocated to workstation *j* or not. Equation (9) restricts the allocated number of equipment type *e* by the available number of this equipment type. Equation (10) determines whether an assistant is assigned to workstation *j* or not. Equation (11) ensures that the number of assistants assigned to workstations does not exceed the available number of assistants. Equations (12), (13) and (14) are the goal constraints for the goals 1, 2 and 3 respectively. Finally, equations (15) to (17) are sign constraints.

3. Illustrative Example and Scenario Analysis

In this section, the proposed GP formulation is validated on an illustrative problem. Firstly, required data of an RDALB problem with 10 tasks is given. Then, the problem is solved using the proposed pre-emptive GP model and results are presented. Finally, a scenario analysis is performed by solving the mentioned illustrative problem with six different goal priority orders.

3.1. Problem Data

Table 1 presents the precedence relations and the tasks processing times of the illustrative problem with resource alternatives of the tasks. In the Table 1, i and IP_i columns denote the task number and the immediate predecessors of task i respectively. The other cells of the table are about processing alternatives of the tasks. For example, task #5 has only one processing alternative without any assistance and any equipment. The processing time of the task #5 is 6 minutes. On the other hand, task #4 has two processing alternatives. This task can be completed manually (without any equipment) by one worker in 5 minutes. But the processing time of the task #4 can be reduced to 3 minutes by assistance of an assistant worker. Another example is task #3. As is seen in the Table 1, task #3 has only one processing alternative with assistant and can be completed in 13 minutes. This means that, task #3 cannot be completed without assistance. In addition, Table 1 indicates that, task #2 has four different resource alternatives. Task #2 can be completed by one worker in 10 minutes manually, in 7 minutes using equipment #1 and in 8 minutes using equipment #2. Alternatively, this task can be completed in 7 minutes with assistance of an assistant worker without using any of the equipment.

Table 1. Illustrative Problem Data

i	IP_i	Task Completion Times					i	IP_i	Task Completion Times				
		Assistant	Equipment						Assistant	Equipment			
			No	1	2	3				No	1	2	3
1	-	Yes					6	2,5	Yes				
		No	5			4			No	8	6		6
2	-	Yes	7				7	6	Yes				
		No	10	7	8				No	7			
3	-	Yes	13				8	7	Yes				
		No							No	4			3
4	1	Yes	3				9	3,7	Yes	5			
		No	5						No				
5	4	Yes					10	9	Yes				
		No	6						No	13			

The other parameters defined in the model are taken as CW=10, CA=6, c_1=3.3, c_2=1.7 and c_3=1.5 money units. Available number of assistants, equipment #1, equipment #2 and equipment #3 are 2, 1, 1 and 2, respectively. 'No equipment' case is also defined by labelling the equipment type 0. All of the processing alternatives, task times, costs and available number of the resources are generated randomly.

It is also assumed that the assembly line manager desires to achieve the following precise goals with their priorities:

- Goal with priority level 1: total cost of utilized workstations should not exceed 40 money units ($CostW$ = 40).
- Goal with priority level 2: cycle time should not exceed 15 minutes (CT = 15).
- Goal with priority level 3: total cost of additional resources should not exceed 16 money units ($CostR$ = 16).

3.2. Solution of the Illustrative Problem

Based on the priority levels of the goals, the problem is solved in seconds using CPLEX 12.5 on a workstation with an Intel Xeon E5-1650 (6 Core) 3.20 GHz processor with 16 GB RAM.

After the solution of the model with the objective of minimizing the over achievement of G1, the deviational variable d^+ was found to be 0. This means that G1 is achieved. The total cost of utilized workstations will be 40 money units, in other words four workstations will be utilized.

Then the yielded value of d^+ was fixed by adding a new constraint to the model such as $d^+=0$ and the model was re-solved with the objective of minimizing the over achievement of G2. In this case, the deviational variable f^+ was found to be 3. This means that G2 is not achieved, and the assembly line will be operated at 15+3=18 minutes of cycle time.

Finally, the yielded value of f^+ was fixed by adding a new constraint to the model such as $f^+=3$ and the model was re-solved with the objective of minimizing the over achievement of G3. In the solution, the deviational variable g^+ was found to be 0.8. This means that G3 is not achieved, and the total cost of additional resources will be 16+0.8=16.8 money units. The final solution of the model is given in Table 2.

Table 2. Final Results of the Illustrative Problem

Workstation	Assigned Tasks	Workload	Assigned equipment	Assistant worker	Cost of workstation utilization	Cost of additional resources
1	1,3	18	-	Yes	10	6
2	2,4,5	18	#1	No	10	3.3
3	6,7,8	16	#3	No	10	1.5
4	9,10	18	-	Yes	10	6
Total cost of workstation utilization					**40**	
				Total cost of additional resources		**16.8**

3.3. Scenario Analysis

Above mentioned three precise goals of the proposed GP model can be ordered in 3!=6 different ways depending on their priority levels. The illustrative problem is solved for these six different scenarios and results are presented in Table 3.

Table 3. Scenario Analysis by Changing the Priority Levels of the Goals

Scenario	Priority Order	d^+	f^+	g^+	Unsatisfied Goals	Cost of workstation utilization	Cycle Time	Cost of additional resources
1	G1-G2-G3	0	3	0.8	G2, G3	40	18	16.8
2	G1-G3-G2	0	4	0	G2	40	19	15
3	G2-G1-G3	10	0	2.3	G1, G3	50	15	18.3
4	G2-G3-G1	20	0	0	G1	60	15	13.5
5	G3-G1-G2	0	4	0	G2	40	19	15
6	G3-G2-G1	20	0	0	G1	60	15	13.5

Table 3 shows that either G1 or G2 is not satisfied in all of the scenarios. G2 is not satisfied in the case that G1 has a higher priority level compared to G2 (Scenarios #1, #2 and #5). Similarly, G1 is not satisfied in the case that G2 has a higher priority level compared to G1 (Scenarios #3, #4 and #6). This means that if the total cost of utilized workstations namely the number of workstations is limited by a smaller value, the assembly line will be operated with a longer cycle time. On the other hand, if the cycle time is limited by a shorter value, a greater number of workstations should be utilized. In addition, it is seen in Table 3 that, G3 is not satisfied in two of the scenarios (#1 and #3) only if it has the lowest priority level. This means that if the number of workstations and cycle time are both limited, the allocated budget for the additional resources should be increased.

4. Conclusion

In this study, a pre-emptive goal programming model for the RDALB problem is proposed. Three conflicting goals such as total cost of utilized workstations (total number of utilized workstations), cycle time and total cost of additional resources are considered. The proposed model is validated on an illustrative example and a scenario analysis performed with different priority orders of the goals. With the consideration of multiple objectives and priority selections, the proposed model provides flexibility for decision makers.

Development of efficient heuristic methods for larger sized problems can be considered as future research due to the NP-hard nature of the RDALB problems. Fuzzy GP approaches can also be adapted to the problem for providing more flexibility to decision makers based on their decision environments.

References

Battaïa, O., & Dolgui, A. (2013). A taxonomy of line balancing problems and their solutionapproaches [Conference Paper]. International Journal of Production Economics, 142(2), 259-277. https://doi.org/10.1016/j.ijpe.2012.10.020

Baybars, I. (1986). Survey Of Exact Algorithms For The Simple Assembly Line Balancing Problem. Management Science, 32(8), 909-932.

Becker, C., & Scholl, A. (2006). A survey on problems and methods in generalized assembly line balancing. European Journal of Operational Research, 168(3), 694-715. https://doi.org/http://dx.doi.org/10.1016/j.ejor.2004.07.023

Charnes, A., & Cooper, W. W. (1957). Management models and industrial applications of linear programming. Management Science, 4(1), 38-91.

Erel, E., & Sarin, S. C. (1998). A survey of the assembly line balancing procedures [Review]. Production Planning and Control, 9(5), 414-434. http://www.scopus.com/inward/record.url?eid=2-s2.0-0032117097&partnerID=40&md5=4959e3a6e57ae6ec1142eca9f2226634

Faaland, B. H., Klastorin, T. D., Schmitt, T. G., & Shtub, A. (1992). Assembly Line Balancing with Resource Dependent Task Times*. Decision Sciences, 23(2), 343-364. https://doi.org/10.1111/j.1540-5915.1992.tb00393.x

Ghosh, S., & Gagnon, R. J. (1989). A comprehensive literature review and analysis of the design, balancing and scheduling of assembly systems [Article]. International Journal of Production Research, 27(4), 637-670. https://doi.org/10.1080/00207548908942574

Kara, Y., Gökçen, H., & Atasagun, Y. (2010). Balancing parallel assembly lines with precise and fuzzy goals [Article]. International Journal of Production Research, 48(6), 1685-1703. https://doi.org/10.1080/00207540802534715

Kara, Y., Özgüven, C., Yalçin, N., & Atasagun, Y. (2011). Balancing straight and U-shaped assembly lines with resource dependent task times [Article]. International Journal of Production Research, 49(21), 6387-6405. https://doi.org/10.1080/00207543.2010.535039

Salveson, M. E. (1955). The assembly line balancing problem. Journal of Industrial Engineering, 6(3), 18-25.

Scholl, A., & Becker, C. (2006). State-of-the-art exact and heuristic solution procedures for simple assembly line balancing [Conference Paper]. European Journal of Operational Research, 168(3), 666-693. https://doi.org/10.1016/j.ejor.2004.07.022

Sivasankaran, P., & Shahabudeen, P. (2014). Literature review of assembly line balancing problems. The International Journal of Advanced Manufacturing Technology, 73(9), 1665-1694. https://doi.org/10.1007/s00170-014-5944-y

Effects of scanning strategies on thermal behaviour and stress fields during selective laser melting of 316L stainless steel

Zezhou Kuai [1], Zhonghua Li [2*], Bin Liu [1*], Yang Shuai [1]

[1] *North University of China, School of Materials Science and Engineering, Taiyuan, China*
[2] *North University of China, School of Mechanical Engineering, Taiyuan, China*
lizhonghua6868@163.com, linbin3y@nuc.edu.cn

Abstract

Simulations of temperature fields and stress fields during selective laser melting (SLM) of 316L stainless steel powder were performed using the finite element method. The effects of the scanning strategies on the SLM thermal behaviour, stress evolution, and residual stress were investigated. The commercial finite element analysis software ANSYS (APDL) was used to establish a single-layer multi-track three-dimensional transient numerical model of SLMed 316L stainless steel. The model considers the temperature-dependent material properties, including thermal conductivity, density, enthalpy, yield stress, thermal expansion coefficient, and Young's modulus. Three partition scanning strategies were designed. In addition to the different sizes of the divided areas, the three scanning strategies had the same total scanning area and other settings.

Keywords: Selective laser melting, Numerical simulation, Thermal behaviour, 316L stainless steel

1. Introduction

Because of its excellent corrosion resistance, high-temperature creep performance, and work hardening performance, 316L stainless steel is widely used in various fields, such as automobile manufacturing and aerospace. (Alvi, Saeidi, & Akhtar, 2020). However, with the increasingly complex geometry of 316L parts, traditional manufacturing processes have been unable to meet the production requirements. Therefore, a new processing method is needed to meet the needs of parts with complex structures.

The selective laser melting (SLM) technology is an advanced metal additive manufacturing technology based on the principle of dispersion and accumulation. The SLM technology combines computer-aided design (CAD) three-dimensional (3D) modelling, computer module management, laser, and other related technologies. A solid model is established by the CAD software, and the model slices are processed in layers. The two-dimensional (2D) contour information of the layered slices is imported into the computer. Under the control of the computer, the powder in the selected area is melted and solidified layer by layer by using laser, and the three-dimensional solid is formed layer by layer (Wang, Lei, Zhu, Chen, & Fang, 2019; Zhang, Zhu, Qi, Hu, & Zeng, 2016). The SLM technology uses the metal powder to directly obtain solid parts with arbitrary complex shapes without any fixture or mould, and thus is particularly suitable for manufacturing titanium alloy superalloy parts with a complex inner cavity structure, which has a wide range of applications in aerospace, automobile, medical moulds, and other fields (Kimura et al., 2019). The transient thermal behaviour in the SLM process is largely affected by the manufacturing process parameters. In order to obtain ideal SLM manufacturing parts, a considerable amount of experimental research is often required. Therefore, the numerical simulation method is selected as a typical method to solve these problems.

[*] *Corresponding Author*

Furumoto et al.(Furumoto et al., 2017) established a 2D finite element model to simulate and study the thermal deformation behaviour of a chromium-molybdenum-based metal powder and an iron-nickel-based metal powder during the SLM process, and carried out the corresponding experimental verification. Ilin et al.(Ilin et al., 2014) used MSC.Marc to establish a 2D model of SLM manufacturing 316L, which simulated the effects of laser power, scanning speed, and tilt angle on the width and depth of the molten pool, which were consistent with the experimental data. Chen et al.(Chen et al., 2019) established a multi-material finite element model of a Ti6Al4V layer and a TiB2 layer, and studied the influence of the process parameters on the maximum temperature of different material interfaces, the temperature gradient, and the existence time of the molten pool during the SLM process. Li et al.(Li & Gu, 2014) used the ANSYS finite element software to establish a 3D finite element model of the relationship between the AlSi10Mg powder's SLM process parameters and thermal behaviour. The effects of laser power and laser scanning speed on the thermal behaviour and structure of the molten pool were analysed. Although some numerical simulations have been conducted to study the thermal behaviour of metal powder during laser processing, the research on the thermal behaviour of 316L stainless steel powder during laser processing is still limited. There are few studies on the influence of different scanning strategies on the temperature and stress fields of 316L stainless steel. In the metal 3D printing process, the selection of the scanning strategy affects not only the surface quality and dimensional accuracy of the parts but also the manufacturing time of the parts. Therefore, the finite element analysis software ANSYS was used to conduct a numerical simulation of SLM manufacturing 316L stainless steel, and the influence of the scanning strategy on the thermal behaviour stress evolution and residual stress distribution of 316L stainless steel during SLM manufacturing was analysed.

2. SLM Theoretical Model

2.1. Temperature Control Equation

The SLM process is a process of rapid metal melting and solidification. With the movement of the laser beam acting on the powder layer, the temperature of the entire powder bed changes with a change in the laser action time and position. SLM processing can be regarded as a closed adiabatic system, so the temperature field variation can be expressed by the 3D transient heat conduction formula:

$$pc\frac{\partial T}{\partial t}=\frac{\partial}{\partial x}\left(k_x\frac{\partial T}{\partial x}\right)\pi r^2+\frac{\partial}{\partial y}\left(k_y\frac{\partial T}{\partial y}\right)\pi r^2+\frac{\partial}{\partial z}\left(k_z\frac{\partial T}{\partial z}\right)\pi r^2+Q \quad (1)$$

where ρ, c, T, k, and Q are the density, specific heat capacity, temperature, thermal conductivity, and heat generation of the internal heat source per unit time per unit volume, respectively; t is the time, and c and k vary with the temperature.

The initial temperature of the powder bed at t = 0 can be defined by the following formula:

$$\mathrm{T}(x,y,z,t)|_{\,\mathrm{t}=0}=T_0\,,\quad (x,y,z)\in \mathrm{D} \quad (2)$$

Where T_0 is 25℃.

Before laser loading, the initial temperature of the powder and substrate was the same as the temperature of the ambient gas; thus, T_0 = 298 K. The powder layer and the outer surface of the substrate were in contact with the surrounding free-flowing gas and meet the following radiation and convection heat dissipation boundary conditions:

$$k\frac{\partial T}{\partial n}+h(T-T_0)+\sigma\varepsilon\left(T^4-{T_0}^4\right)=0 \quad (3)$$

where n is the vector along the normal direction of the powder bed surface; T is the surface temperature of the powder bed or molten pool; h is the surface convection coefficient; ε is the thermal radiation coefficient; σ is the Stefan–Boltzmann constant, equal to $5.67\times10^{-8}\mathrm{W\cdot m^{-2}\cdot K^{-4}}$. In the calculation, the convection heat transfer coefficient and the radiation coefficient were fitted to the total heat transfer coefficient to reduce the degree of non-linearity, and its magnitude was $80\mathrm{W\cdot m^{-2}\cdot K^{-1}}$.

2.2. FE Model Set-up

The ANSYS finite element software was used to establish the model and calculate the temperature and stress fields. The finite element model of SLM processing is shown in Figure 1(a). The model was composed of a powder bed and a base plate. The size of the 316L powder bed was 3.96 mm × 3.48 mm × 0.04 mm, and the substrate material was forged 316L stainless steel with a size of 3.96 mm × 3.48 mm × 2 mm. Considering the accuracy and efficiency of the simulation, the model adopted the mapping meshing method, and the solid 70 element was used for meshing. The monitoring point node 1 in Figure 1(b) was located on the interface between the powder bed and the substrate at the end of the first pass. The partition scanning strategy setting is shown in Figure 2. In this figure, the whole scanning area is divided into several small regions. The number indicates the scanning sequence of the regions, and the yellow arrow represents the first track scanned in each region. Strategy 1, Strategy 2, and Strategy 3 are different in terms of the sizes of the regions, but the total area scanned (1.96 mm × 1.48 mm) and the other settings were the same.

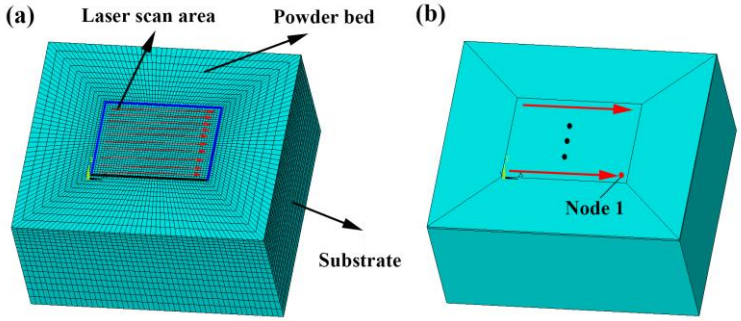


Figure 1. (a)Finite element model and scanning strategy, (b) location of monitoring point node 1

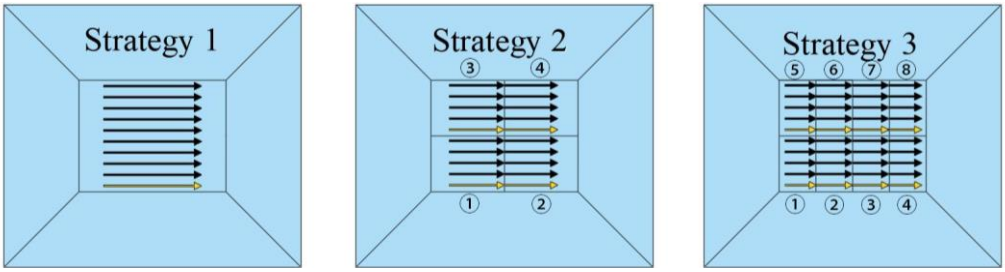


Figure 2. Partition scanning strategy

2.3. Heat Source Model

The heat source is related to the power of the laser, the diameter of the spot, and the size of the distribution, which determines the difference in the temperature distribution. The laser beam distribution modes mainly include Gaussian laser distribution and uniform distribution. The laser divergence angle of the Gaussian energy distribution is small, and it can focus on a smaller laser focus compared with a uniformly distributed laser beam. In the SLM process, the laser beam irradiated the metal powder in the form of a spot, with a small spot diameter, high laser energy density, and strong penetration. The heat source model of the Gaussian laser distribution is shown in Figure 3, which satisfies the following Gaussian distribution formula:

$$\mathrm{e} = \frac{2AP}{\pi r^2} \exp\left(-\frac{2r^2}{R^2}\right) \tag{4}$$

where A is the absorption rate of the laser beam on the surface of the material, P is the laser power, r is the distance from a point in the spot to the spot centre, and R is the spot radius. When the energy distribution mode of the laser beam is uniform, its heat flux density can be expressed as follows:

$$q = \frac{AP}{\pi {r_1}^2} \tag{5}$$

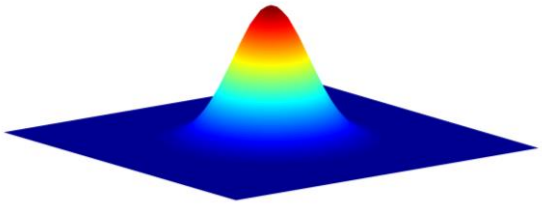


Figure 3. Gaussian heat source model

3. Results and Discussion

Figure 4 shows the temperature change diagram and the temperature change diagram of Node 1 with time for different scanning strategies. It can be seen from the figure that the maximum temperature for the different scanning strategies was basically the same. As the scanning proceeded, the final temperature tended to be around 250°C. However, the heat accumulation effect under the different scanning strategies was obviously different. From Strategy 1 to Strategy 3, the divided areas increased, and the length of the scan track in each area decreased, which caused the heat accumulation effect in each area to be even greater. As the scan length decreased, the second peak-to-peak value of both Strategy 2 and Strategy 3 increased significantly. In particular, for strategy 3, the scanning length was sufficiently short for the second peak to be close to the highest peak, which effectively reduced the cooling rate and the temperature gradient of the node. In addition, because of the shortening of the scanning length, the preheating effect was more obvious. The preheating temperature of Strategy 3 was significantly higher than that of Strategy 1 and Strategy 2, which reduced the temperature difference at this point and thus reduced the temperature gradient. As shown in Figure 4(b), there was little difference between the maximum cooling rates of Strategy 1 and Strategy 2 (approximately 8.5106°C /s), while both the heating rate and the cooling rate of Strategy 3 decreased significantly, among which the maximum cooling rate was 5.7106°C /s, showing a decrease of 33%. This was because the scanning length was very short. While the cooling of the previous one was not completed, the heating of the next one began, which had a considerable impact on the grain growth. For partition scanning, the size of partition scanning largely determined the heat accumulation in the SLM processing process, thus affecting the cooling speed and the temperature gradient.

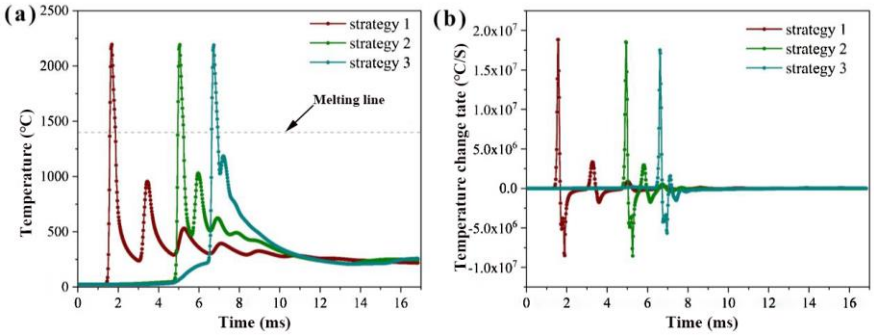


Figure 4. (a) Temperature history graph and (b) temperature change rate history graph of node 1 for different scanning strategies

Figure 5 shows the equivalent stress (von Mises stress) history of node 1 for different scanning strategies. The scanning strategy played an important role in the evolution of stress, and the stress evolution under different scanning strategies was considerably different. It can be seen from the figure that the stress change curve and the temperature change curve presented an obvious corresponding relationship. Although node 1 had not been heated, the equivalent stress appeared because the selected

node 1 was located on the interface between the powder bed and the substrate. As the heat source moved, the stress increased continuously until the heat source reached node 1; then, the equivalent stress dropped to 0. Thereafter, the equivalent stress fluctuated with the constant fluctuation of the temperature. Finally, after 100 s of cooling, the equivalent stress tended to be stable. The residual stresses of Strategy 1, Strategy 2, and Strategy 3 were 305 MPa, 296 MPa, and 286 MPa, respectively. It can be seen that the partition scanning reduced the residual stress of node 1, and as the partition increased, the residual stress decreased more significantly.

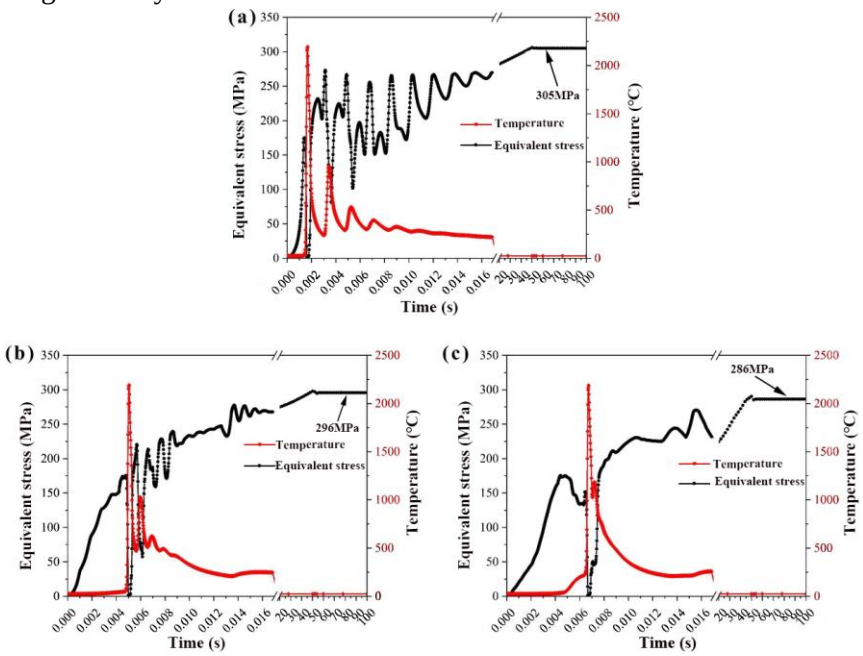


Figure 5. Von Mises stress history diagram of node 1 for different scanning strategies: (a) Strategy 1, (b) Strategy 2, and (c) Strategy 3

Figure 6 shows the residual stress distribution of the finite element model after cooling to room temperature. It can be clearly seen from the figure that the scanning strategy had a greater impact on the residual stress distribution of the model, and the residual stress distribution under different scanning strategies was considerably different. Partition scanning reduced the residual stress of the model to a certain extent. Strategy 2 reduced the maximum residual stress of Strategy 1 by 2%, and Strategy 3 reduced the maximum residual stress of Strategy 1 by 6%.

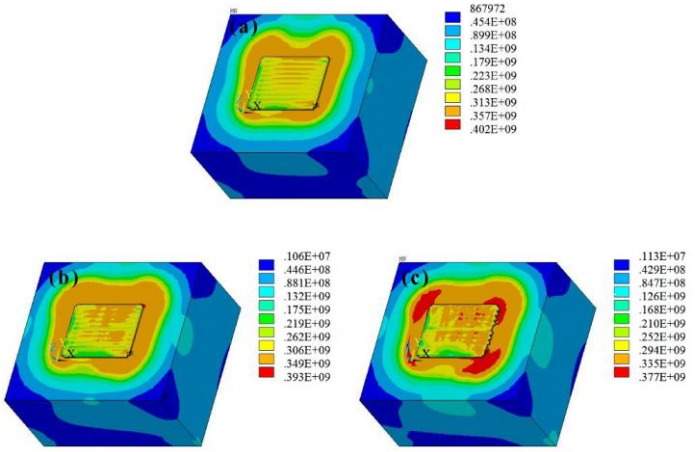


Figure 6. Residual stress distribution of the finite element model: (a) Strategy 1, (b) Strategy 2, (c) Strategy 3

Note that there was a problem of overlap between each small area in the partition scanning, and the scanning sequence of each small area affected the temperature field, which in turn affected the final residual stress distribution of the part. Therefore, the partition scanning strategy still needs to be studied in depth.

4. Conclusion

In this study, the effects of scanning strategy on the thermal behaviour, stress evolution, and residual stress distribution of the 316L stainless steel SLM manufacturing process were analysed using simulations. The main conclusions were as follows:

The results showed that for partition scanning, the partition size largely determined the heat accumulation during SLM processing, which in turn affected the cooling rate and the temperature gradient. Partition scanning reduced the residual stress of the part to a certain extent, and as the partition increased, the scanning length decreased and the residual stress decreased more significantly. The effect of heat accumulation under different scanning strategies was obviously different. As more regions were divided, the length of the scan track in each region decreased, which resulted in more significant heat accumulation effects in each region. The reduction of the scan length could effectively reduce the cooling rate and the temperature gradient of the node. Strategy 2 reduced the maximum residual stress of Strategy 1 by 2%, and Strategy 3 reduced the maximum residual stress of Strategy 1 by 6%.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (grant no. 51905497), the Major Science and Technology Projects of Shanxi Province, China (no. 20181102012), the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (no. 2019L0559), the Opening Foundation of Shanxi Key Laboratory of Advanced Manufacturing Technology (no XJZZ201805), and the Support Program for Young Academic Leaders of North University of China (no. QX201902).

References

Alvi, S., Saeidi, K., & Akhtar, F. (2020). High temperature tribology and wear of selective laser melted (SLM) 316L stainless steel. *Wear, 448-449*, 203228. doi:10.1016/j.wear.2020.203228

Chen, C., Gu, D., Dai, D., Du, L., Wang, R., Ma, C., & Xia, M. (2019). Laser additive manufacturing of layered TiB2/Ti6Al4V multi-material parts: Understanding thermal behavior evolution. *Optics & Laser Technology, 119*, 105666. doi:10.1016/j.optlastec.2019.105666

Furumoto, T., Ogura, R., Hishida, K., Hosokawa, A., Koyano, T., Abe, S., & Ueda, T. (2017). Study on deformation restraining of metal structure fabricated by selective laser melting. *Journal of Materials Processing Technology, 245*, 207-214. doi:10.1016/j.jmatprotec.2017.02.017

Ilin, A., Logvinov, R., Kulikov, A., Prihodovsky, A., Xu, H., Ploshikhin, V., . . . Bechmann, F. (2014). Computer Aided Optimisation of the Thermal Management During Laser Beam Melting Process. *Physics Procedia, 56*, 390-399. doi:10.1016/j.phpro.2014.08.142

Kimura, T., Nakamoto, T., Ozaki, T., Sugita, K., Mizuno, M., & Araki, H. (2019). Microstructural formation and characterization mechanisms of selective laser melted Al–Si–Mg alloys with increasing magnesium content. *Materials Science & Engineering A, 754*, 786-798. doi:10.1016/j.msea.2019.02.015

Li, Y., & Gu, D. (2014). Parametric analysis of thermal behavior during selective laser melting additive manufacturing of aluminum alloy powder. *Materials & Design, 63*, 856-867. doi:10.1016/j.matdes.2014.07.006

Wang, P., Lei, H., Zhu, X., Chen, H., & Fang, D. (2019). Influence of manufacturing geometric defects on the mechanical properties of AlSi10Mg alloy fabricated by selective laser melting. *Journal of Alloys and Compounds, 789*, 852-859. doi:10.1016/j.jallcom.2019.03.135

Zhang, H., Zhu, H., Qi, T., Hu, Z., & Zeng, X. (2016). Selective laser melting of high strength Al–Cu–Mg alloys: Processing, microstructure and mechanical properties. *Materials Science and Engineering: A, 656,* 47-54. doi:10.1016/j.msea.2015.12.101

Vibration Monitoring of Coastal and Ocean Structures with Pile Foundations Using Compressive Sensing

Cihan Bayındır[1,2*]

[1]*İstanbul Technical University, Civil Engineering Department, İstanbul, Turkey*
[2]*Boğaziçi University, Civil Engineering Department, İstanbul, Turkey*
cbayindir@itu.edu.tr

Abstract

Structural health monitoring of the maritime structures is a rapidly developing research area. One of the research directions followed within this context is the development of efficient mathematical data and signal processing techniques and their possible usage methods for the health monitoring of coastal and ocean structures. In this paper, we investigate the possible usage of one of such methods, namely the compressive sensing technique (CS), for the measurement and reconstruction of the vibration data of the coastal and ocean structures. CS algorithm outperforms the classical sampling theory by using far fewer measurements for the reconstruction of signals having sparse representation in different orthogonal domains. The aforementioned maritime structures are continuously subjected to harmonic loads in the marine environment, as well as impact loads such as shiploads, earthquakes. Thus, CS algorithm can be used for the reconstruction of vibration velocities, acceleration, and similar parameters under such loadings, which have sparse representations in Fourier or temporal/spatial domains. Implementing a circular cylinder and hollow elastic circular cylinder model for the modeling of the pile foundations, we show that CS can be effectively used for the monitoring and reconstruction of such vibration parameters under cyclic harmonic loads and impact loads including shiploads and earthquakes. We discuss our findings and their possible applicability and usage.

Keywords: Coastal and ocean structures, structural vibration monitoring, compressive sensing

1. Introduction

Structural health monitoring is an active area of research, however, its application in the marine environment is still very limited. In its applications, better sensors and sampling strategies are always desired. With this motivation, in this paper, we investigate the possible usage of the CS for the efficient sensing of structural health monitoring parameters in maritime engineering. More specifically, we consider the efficient measurement of the wave-induced force and moments, as well as displacements induced by point harmonic loads. We construct the time series of the wave-induced force and moments using the Morison equation. We construct the time series of the displacements induced by point harmonic loads using an elastic hollow cylinder model described by the Flügge equations of elasticity. We show that all of these parameters can be constructed using far fewer samples than the classical Shannon's theorem states, thus efficient measurement and analysis of such parameters with significant undersampling ratios are possible.

2. Mathematical Formulation

2.1. Review of Forces and Moments Exerted on a Pile by Monochromatic Waves

* *Corresponding Author*

In this section, we summarize the Morison formula used for the calculation of forces and moments exerted by flows on objects with different cross-sections. We consider a circular cylinder, which can be a model for a pile foundation. For a pile of length *dz*, the Morison formula gives

$$dF = dF_D + dF_I = \frac{1}{2} C_D \rho A |u| u + C_M \rho V \frac{Du}{Dt} \tag{1}$$

where the subscripts *D* and *I* refer to the drag and inertia forces, respectively, C_D, is the drag coefficient, C_M is the inertia coefficient, ρ is the fluid density, *A* is the area perpendicular to flow, *V* is the volume of the element and *u* is the horizontal flow excursion velocity (Dean et al., 2000). After an integration over the depth, the total force can be computed by

$$F = \int_{-h}^{\eta} dF \tag{2}$$

where *h* is the depth and η is the water surface fluctuations. Linearization of this equation and its calculation for a pile with a diameter D subjected to horizontal excursion velocity of a monochromatic wave yields

$$F = C_D DnE \cos(kx_1 - \omega t)\left|\cos(kx_1 - \omega t)\right| + C_M \pi DE \frac{D}{H} \tanh(kh) \sin(kx_1 - \omega t) \tag{3}$$

Here *n* is the ratio of group velocity to wave celerity which is taken as 1 considering shallow water conditions, $E=1/8\rho g H^2$ is the wave energy, *H* is the wave height, *k* is the wavenumber, x_1 is the location of the pile and $\omega=2\pi/T$ is the angular frequency and *T* is the period of the wave (Dean et al., 2000). Similarly, the total moment exerted by a monochromatic can be computed using

$$M = \int_{-h}^{\eta} dM = \int_{-h}^{\eta} (h+z) dF \tag{4}$$

which yields

$$\begin{aligned} M = {} & C_D DnE \cos(kx_1 - \omega t)\left|\cos(kx_1 - \omega t)\right| \left[h\left\{1 - \frac{1}{2n}\left(\frac{\cosh 2kh - 1 + 2(kh)^2}{2kh \sinh 2kh} \right)\right\}\right] + \ldots \\ & \ldots + C_M \pi DE \frac{D}{H} \tanh(kh) \sin(kx_1 - \omega t) \left[h\left\{1 - \frac{\cosh kh - 1}{kh \sinh kh}\right\}\right] \end{aligned} \tag{5}$$

The parameters for computational purposes are selected as typical parameters in a real setting, namely $C_D=1$, $C_M=2$, $D=1m$, $H=1m$, $\rho=1025\ kg/m^3$, $g=9.81m/s^2$, $h=10m$, $T=15s$, $x_1=0$. Using shallow-water limit of the dispersion equation, the wavenumber is calculated using $k = \omega / \sqrt{gh}$.

2.2. Review of an Elastic Hollow Cylinder Model Subjected to Point Harmonic Load

In order to simulate more a realistic scenario, we model pile foundation as a long hollow cylinder as illustrated in Fig.1 and we solve the Flügge equations of theory of elasticity. The pile is subjected to a horizontal harmonic force of P which can be a representation of a dynamic ship or anchoring load on a bollard or a similar structural element. For the sake of brevity, we do not give the full governing equations but the Flügge equations of elastic theory and its solutions in terms of Bessel and modified Bessel function can be seen in Forrest et al. (2006a, 2006b). We follow the approach described in Forrest et al. (2006a, 2006b), and obtain the horizontal displacement time series of an elastic long pile foundation under the effect of a harmonic load with many spectral components. The computational parameters in Forrest et al. (2006a, 2006b), are used to represent a large pile and the horizontal magnitude of the horizontal harmonic point load is taken as $P_{mag}=1\ kN$.

2.3. Review of the Compressive Sensing

Compressive sampling (CS) completely revolutionized the field of signal processing. Here, we try to give a brief sketch of the CS. Consider a *K*-sparse signal *r*, that is only *K* out of its *N* elements are nonzero. Using an orthogonal transformation matrix Ψ, *r* can be transformed into an orthogonal domain, such as Fourier or wavelets domains. In such domains, the signal can be represented by $r = \psi\hat{r}$, where $\hat{r}$ denotes the coefficient vector. Eliminating the zero entries, it is possible to get, $r_s = \psi\hat{r}_s$. Here, r_s denotes the signal with non-zero components.

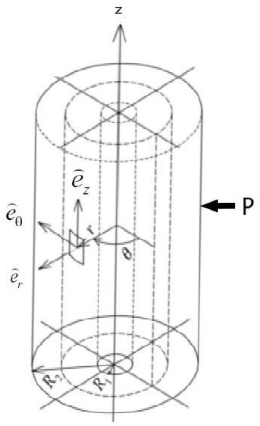


Figure 1. Model of a hollow pile

Using CS algorithm a *K*-sparse signal *r* with *N* entries can exactly be reconstructed using $M \geq C\mu^2(\phi,\psi)K\log(N)$ measurements. Here, *C* is a positive constant, ϕ is the sensing basis and ψ is the transformation basis (Candes et al., 2006a, 2006b, 2006c). $\mu^2(\phi,\psi)$ is the mutual coherence between these two. *M* random projections and the sensing matrix ϕ can be used to obtain $g = \phi r$. Thus, one can state the problem as

$$\min \|\hat{r}\|_{l_1} \text{ subjected to } g = \phi\psi\hat{r} \quad (6)$$

with $\|\hat{r}\|_{l_1} = \sum_i |\hat{r}_i|$. Among all possible solutions satisfying the given constraints, the l_1 solution of the problem can be calculated as $r_{CS} = \psi\hat{r}$. Other optimization techniques such as the re-weighted l_1 minimization or greedy pursuit algorithms can also be used for this purpose (Candes et al., 2006a, 2006b, 2006c). We refer the reader to Candes et al. (2006a, 2006b, 2006c) for a comprehensive discussion of the CS. In hydrodynamics, there are some recent applications of the CS (Bayındır, 2016, 2019, Bayındır et al., 2021).

3. Results and Discussion

3.1. Results for Forces and Moments Exerted on a Pile by Monochromatic Waves

In Fig. 2, we depict the time history of the wave force exerted by a monochromatic wave on a pile with the parameters given above. The original time series consists of 2041.

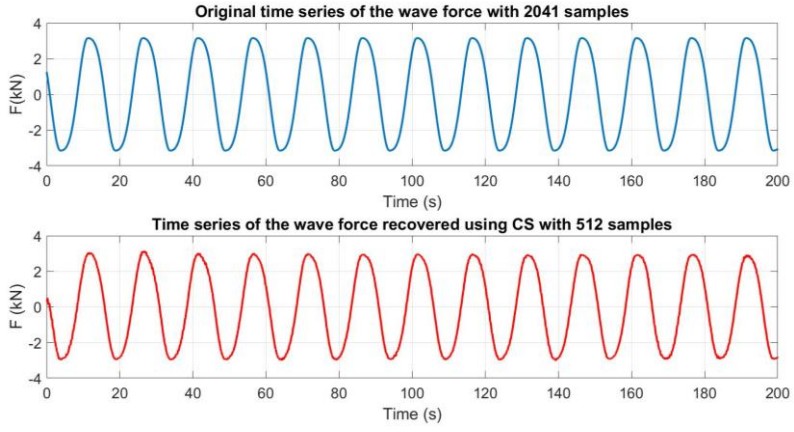


Figure 2. Compressive sampling reconstruction of the time series of the wave force exerted by a monochromatic wave on a pile.

As indicated in Fig. 3, such a time history has a sparse representation in the frequency domain. Thus, by taking random projections in the time domain and solving the l_1 problem of the CS, we construct the original time series and spectrum by using only 512 samples. Those CS reconstructions are depicted in Fig.2 and Fig.3.

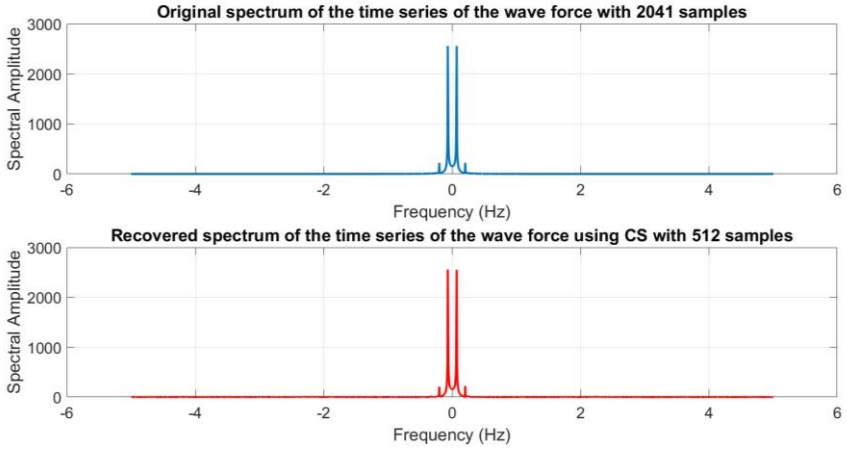


Figure 3. Compressive sampling reconstruction of the spectrum of the wave force exerted by a monochromatic wave on a pile.

In Fig.4, we show that the time series of the wave-induced moment on a pile with parameters given above can also be reconstructed using the same approach. The spectra of the time history of the moment are similar to the spectra of the wave force depicted in Fig. 3. Again, reconstruction is achieved using 512 samples which brings a significant undersampling ratio which is very beneficial for many structural health monitoring purposes.

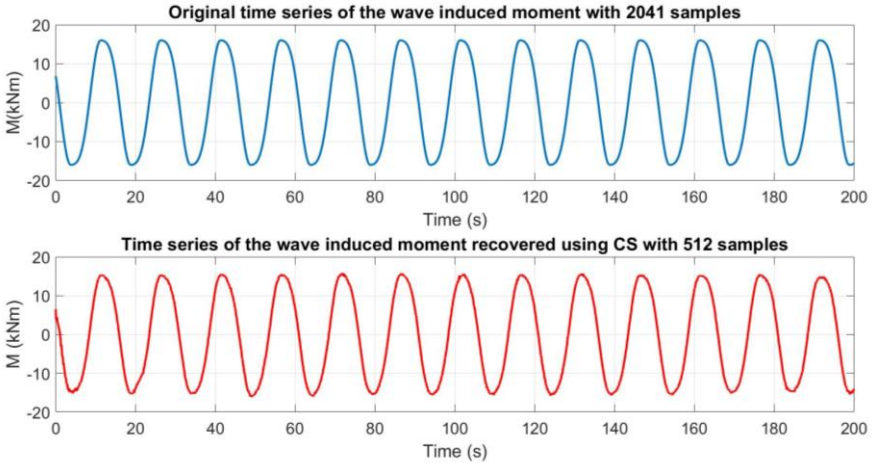


Figure 4. Compressive sampling reconstruction of the time series of the wave moment exerted by a monochromatic wave on a pile.

Although the findings presented above are given for monochromatic sea waves, our findings can easily be generalized to a more general and realistic setting. The wave excursion velocity *u* can be computed for the superposition of sea states which has a spectrum in JONSWAP or PM form. Then, using the Morison equation, the time history of the wave force and wave moment can be calculated. It is known that CS is successful for the reconstruction of the wavefields that have JONSWAP spectral representations (Bayındır, 2016).

3.2. Results for Displacements Induced by a Point Harmonic Load on Elastic Hollow Cylindrical Pile

Solving the Flügge equation of elasticity and following the analysis described in Forrest et al. (2006a, 2006b), we obtain the horizontal displacement spectra of long hollow elastic cylinder subjected to a horizontal time-harmonic point load and depict our findings in Fig. 5. The *n* numbers on this figure refers to various different resonant modes of the cylinder cross-sectional areas (Forrest et al., 2006a, 2006b).

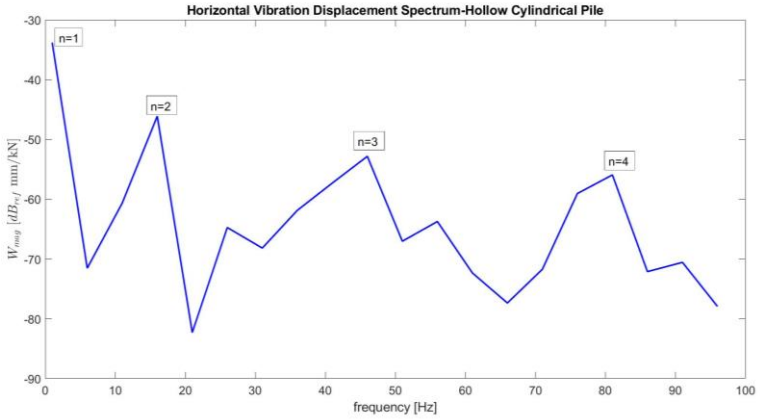


Figure 5. Horizontal displacement spectrum of an elastic hollow cylinder subjected to point harmonic load.

In order to construct the time-series of the horizontal displacement due to a time-harmonic point load, we use the spectrum depicted in Fig.5 and we inject uniformly distributed random phases and construct the time series of the displacement via an IFFT routine. The time series with 2001 samples constructed by such an approach is depicted in Fig. 6.

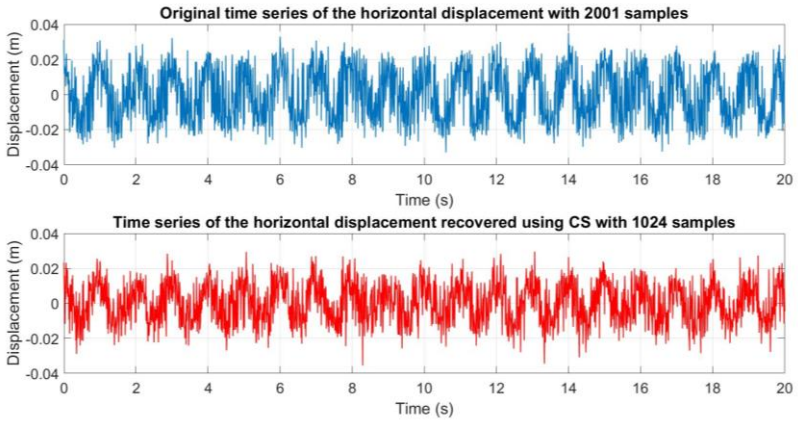


Figure 6. Compressive sampling reconstruction of the time series of the horizontal pile displacement exerted by a horizontal harmonic point load.

After the implementation of the CS algorithm with 1024 samples the time series is reconstructed and depicted in Fig. 6. As one can realize from the figure, the reconstruction is not exact, however, the range of amplitudes and wavelength are in very good agreement. Thus, the representative statistical parameters of the displacements such as rms or significant displacements can be successfully reconstructed by CS. Depending on the sparsity of the signal, the exact recovery can also be possible with an even smaller number of spectral components. This is also true for the time series of the force and moments discussed in the preceding sections. Thus, CS can be used as a very efficient algorithm for the measurement,

analysis, and interpolation/extrapolation of the time series acquired using structural health monitoring applications.

4. Conclusion

In this study, we investigated the possible usage of compressive sensing for the efficient sensing of the force, moment, and displacement induced by monochromatic sea waves and point loads. For this purpose, we have used the Morison equation and an elastic hollow cylinder model described by Flügge equations of the elastic theory. We showed the compressive sensing can be significantly advantageous for the measurement and analysis of such parameters. Although the main advantage is the significant downsampling ratio of the data, interpolation/extrapolation of the missing data is also possible by compressive sensing. Our approach would also be useful to the analysis of earthquake or ship impact effects, which have sparse representations in the time domain thus their random sampling should be performed in the spectral domain.

Acknowledgements

The author gratefully acknowledges the support of the İstanbul Technical University. This work was supported by the Research Fund of the İstanbul Technical University. Project Code: MGA-2020-42544. Project Number: 42544.

References

Bayındır, C. (2016). Compressive spectral method for the simulation of nonlinear gravity waves, Scientific Reports, 22100. https://doi.org/10.1038/srep22100

Bayındır, C. (2019). Early detection of rogue waves using compressive sensing, TWMS Journal of Applied and Engineering Mathematics, 9 (2), 198-205.
http://jaem.isikun.edu.tr/web/images/articles/vol.9.no.2/03.pdf

Bayındır, C., & Namlı, B. (2021). Efficient sensing of the von Karman vortices using compressive sensing, Computers & Fluids, 104975. https://doi.org/10.1016/j.compfluid.2021.104975

Candes, E. J., Romberg, J., & Tao, T. (2006a). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. on Inf. Theory, 52 (2), 489-509. https://doi.org/10.1109/TIT.2005.862083

Candes, E. J. (2006b). Compressive sampling, Proc. Int. Congr. Math., 3, 1433-1452. https://doi.org/10.4171/022-3/69

Candes, E. J., Romberg, J., & Tao, T. (2006c). Stable signal recovery from incomplete and inaccurate measurements, Comm. Pur. Appl. Math., 59 (8), 1207-1223. https://doi.org/10.1002/cpa.20124

Dean, R. G., & Dalrymple, R. A., (2000). Water wave mechanics for engineers and scientists. Advanced Series on Ocean Engineering, Volume 2, World Scientific. https://doi.org/10.1142/1232

Forrest, J. A., & Hunt, H. E. M. (2006a). A three-dimensional tunnel model for calculation of train-induced ground vibration, Journal of Sound and Vibration, 294 (4-5), 678-705. https://doi.org/10.1016/j.jsv.2005.12.032

Forrest, J. A., & Hunt, H. E. M. (2006b). Ground vibration generated by trains in underground tunnels, 294 (4-5), 706-736. https://doi.org/10.1016/j.jsv.2005.12.031

Optimization of Extrusion Process by using Response Surface Methodology

Aslan Deniz Karaoglan*, Gulistan Bicen, Rabia Durak

Balikesir University, Industrial Engineering Department, Balikesir, Turkey
deniz@balikesir.edu.tr, gulistanbicen14@gmail.com, rabiaduurak@gmail.com

Abstract

This study was carried out in a manufacturer that meets the plastic cups requirement of the food sector. The aim of the study is to find the mathematical relationship between the parameters affecting the thickness of the sheet in the extrusion process by the regression equation and to find out the optimum factor levels in order to obtain the target sheet thickness. To perform the optimization, response surface methodology (RSM) is used. RSM is one of the widely used modeling and optimization method. For this purpose, Minitab statistical analysis program was used. In order to determine whether the number of factors constituting the regression equation is sufficient or not, R^2 determination coefficient is calculated and it is seen that it is quite close to 1. Then, analysis of variance (ANOVA) results were examined and it was concluded that the regression equation was significant at 95% confidence level (which means α=5%=0.05). According to Minitab ANOVA results; P-value is calculated as 0.047 which is lower than α=0.05. This means the regression equation is significant. After finding an available regression equation, the final step was optimization with the help of "Minitab Response Optimizer" module. Verification of the optimum result was performed by field tests and it was found that there was no significant difference between the expected output value and the observed and the results were quite successful.

Keywords: Response surface methodology, process optimization, plastic extrusion method

1. Introduction

Plastic extrusion method is a manufacturing method used especially in the production of plastic materials such as pipes, hoses, cables, profiles. The molten plastic is shaped and cooled along the mold and the production takes place. In the extrusion process different factors affect the product quality. Optimization of these parameters aims to provide the desired quality of the final product (Karaoglan, 2021).

Extrusion process optimization was investigated by many researchers. Artificial neural network (ANN), genetic algorithm (GA), response surface methodology (RSM), Taguchi, grey wolf optimizer (GWO), simulation-based optimization, goal programing, and etc. are used for optimizing different type of extrusion processes. The summarized literature review is presented in Table 1. In this study extrusion speed, roller rotation speed, pump speed, pressure, withdrawal speed of the molded product are selected as the factors and RSM is used for the optimization. This factor combination is used previously for optimizing the plastic extrusion process by using GWO (Karaoglan, 2021). However RSM did not used previously for this selected factor combination for optimizing plastic extrusion process and this is the novelty aspect of this research. In this study, the same process optimization problem with new raw material content is solved by RSM instead of GWO. The details and the differences of the raw material data is not given because of the commercial confidentiality.

* *Corresponding Author*

Table 1. Literature review

Author (s)	Subject	Method
Zhou & Paik (2004)	Optimizing food extrusion process parameters	ANN, GA
Fowler et al. (2010)	Optimizing the polymer extrusion filter layering configurations	Simulation-based optimization
Chen et al. (2013)	Optimizing Ti-6Al-4V titanium alloy equal-channel angular extrusion (ECA)	Taguchi
Iqbal et al. (2016)	Optimizing the forming parameters of the twist extrusion process	RSM
Al-Refaie & Musallam (2019)	Optimizing the polyethylene extrusion process parameters	Goal programing
Karaoglan (2021)	Optimizing plastic extrusion process parameters	GWO

In the following section the materials and methods are presented. Experimental results and the related discussions are discussed in Section 3. Conclusions are given in Section 4.

2. Materials and Methods

In this study experimental runs are performed at plastic sheet production line (machine brand: Suzhou 2014 with 400 kg/h capacity). Then regression analysis is used for mathematical modeling. The R^2 and ANOVA are also calculated. Minitab statistical package is used for this purpose. The extrusion line is same as the line used in the study presented by Karaoglan (2021), However the raw material content and environmental conditions are different. So the observed response values and the mathematical model is differs. "Minitab Response Optimizer" module is used for the optimization. After the optimization process, confirmations are performed at plastic sheet production line. The plastic extrusion line that will be optimized is presented in Figure 1.

Figure 1. The plastic extrusion line that will be optimized

3. Results and Discussion

In the first stage; factors are determined and the observations are measured from the extrusion line for different combinations of these factors. Totally 17 different combination of factors are observed from the extrusion line. Because of having commercial confidentiality, the factor

levels are coded between -1 and +1 (where -1 means the minimum value for the related factor, while +1 is the maximum value). The coded values for the observations are given in Table 2. The coding is carried out using Equation (1):

$$X_{coded} = \frac{X_{uncoded} - ((X_{max} + X_{min})/2)}{(X_{max} - X_{min})/2} \quad (1)$$

Table 2. Experimental results

Run	Factors					Response
	Roller Rotation Speed (rpm/min) (X_1)	Extrusion Speed (rpm/min) (X_2)	Pump Speed (rpm/min) (X_3)	Pressure (bar) (X_4)	Withdrawal Speed of the Molded Product (rpm/min) (X_5)	Thickness of the Plastic Sheet (mm) (Y)
1	0.2000	-0.3333	0.7635	-1.0000	0.2121	1.21
2	0.0000	-0.3333	1.0000	-1.0000	0.2121	1.19
3	0.0000	-0.3333	0.8424	-1.0000	-0.3939	1.21
4	0.0000	0.3333	0.5468	1.0000	-0.3939	1.18
5	-1.0000	1.0000	0.3399	-1.0000	-0.0909	1.17
6	-1.0000	-0.1111	0.3498	1.0000	-1.0000	1.23
7	-1.0000	-0.1111	0.2512	1.0000	-1.0000	1.20
8	-0.7000	1.0000	0.2611	1.0000	-0.8182	1.19
9	0.0000	0.3333	-1.0000	1.0000	-0.3939	1.17
10	0.0000	0.3333	0.7833	1.0000	-0.3939	1.18
11	-0.5000	0.5556	0.5567	1.0000	0.2121	1.19
12	0.1000	-0.3333	0.7833	0.0000	0.2121	1.20
13	-1.0000	-0.3333	0.3103	1.0000	-0.6970	1.19
14	0.0000	-0.3333	0.5271	1.0000	0.2121	1.20
15	1.0000	-0.3333	0.8719	1.0000	1.0000	1.19
16	-1.0000	-0.3333	0.2414	1.0000	-0.3939	1.19
17	0.9000	-1.0000	0.8621	1.0000	0.6970	1.22

All the responses are the mean of 5 replications. For the experimental runs given in Table 2, the mathematical model between the factors (X_i) and the response (Y) is calculated using Minitab and given in Equation (2) below:

$$\begin{aligned} Y = {} & 1.52042432635235 - 0.803139610746138X_1 - 0.98863441454442X_2 \\ & - 0.718275024626442X_3 - 0.121073056266713X_4 \\ & + 0.745000089193886X_5 - 0.20179481373935X_1X_2 \\ & + 0.127855823714148X_1X_3 + 0.835595470565313X_1X_4 \\ & + 0.216613510668006X_1X_5 + 0.153789373540897X_2X_3 \\ & + 0.908741850081363X_2X_4 + 0.112013376348068X_2X_5 \\ & + 0.349433050887672X_3X_4 - 0.821330505624133X_3X_5 \\ & - 0.287224931918107X_4X_5 \end{aligned} \quad (2)$$

Coefficient of determination (R^2) is calculated as 99.98% (which means these 5 factors are sufficient to explain the variation in the plastic sheet thickness. For the ANOVA, P-value approach is used by the aid of Minitab. If P-value (=0.047) < Type-I Error Probability (α=0.05 for 95% confidence level) then we can conclude that the mathematical model is significant (Mason et al., 2003; Montgomery, 2008; Myer et al., 2008). The ANOVA results indicate that the mathematical model is significant and can be used for optimization. The "Minitab Response Optimizer" result is presented in Figure 2. The target value for the plastic sheet thickness is 1.15 mm according to the firm's quality standards. The thickness being below this value affects the strength of the product; if it is above this value, it causes excessive use of raw materials and an increase in costs. The optimum process parameters with coded factor levels (red colored values at Figure 2) are calculated as: X_1= -0.9, X_2= 1.0, X_3= 0.3005, X_4= -0.1, and X_5= -0.1515. The plastic sheet thickness is expected to be 1.1507 mm (blue colored value in Figure 2) for these optimized parameters.

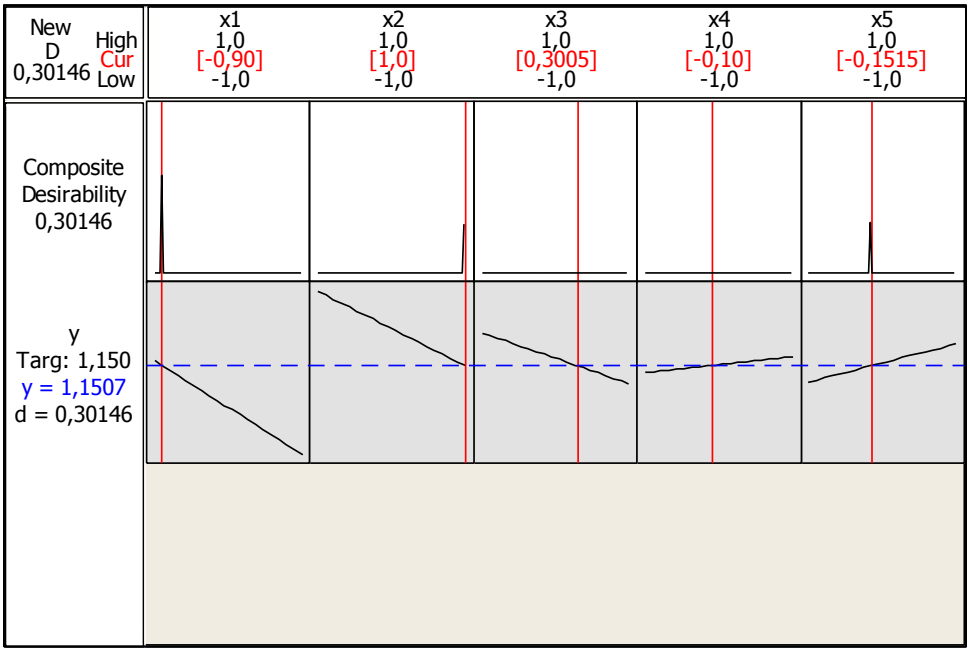


Figure 2. "Minitab Response Optimizer" result

Finally the confirmations are performed and presented in Table 3. Five replications are performed for the optimum factor levels and the mean value is calculated as 1.152 mm. According to the results it can be clearly indicated that the predicted results are very close to the observed results and the overall prediction error percentage (PE(%)) is less than 1%. This obtained thickness value is less then the observations given in Table 2 and very close to the target value of 1.15 mm thickness. This means the process parameter optimization is completed and these optimized parameter levels can be used in mass production.

Table 3. Confirmation

Response	Observed value from the production line	Minitab prediction	PE(%)
Plastic sheet thickness (mm)	1.152	1.1507	0.06

4. Conclusion

In this study the aim was to obtain 1.15 mm plastic sheet thickness for the semi-product produced by the extrusion line. Experimental results are observed from the production line and RSM is used for mathematical modeling and optimization. RSM has not been previously used for this factor combination to optimize the plastic extrusion process and this is the novelty of this study. Confirmations are performed for the optimized factor levels. The results show that the RSM provided good results for the extrusion process problem dealt with.

Acknowledgements

This research is supported by Sahlan Plastics Co. (Balıkesir –Turkey). The authors would gratefully like to thank Omer Vural and Sibel Gyunay Sali for their support.

References

Al-Refaie, A. & Musallam, A. (2019). Using mixed goal programming to optimize performance of extrusion process for multiple responses of irrigation pipes. Proceedings of the Institution of Mechanical Engineers Part E-Journal of Process Mechanical Engineering, 233, 412–424. https://doi.org/10.1177/0954408918781624

Chen, D. C., Li, Y. J. & Tzou, G. Y. (2013). An investigation into optimized ti-6al-4v titanium alloy equal channel angular extrusion process. International Applied Science on Precision Engineering Conference (Nan Tou, Taiwan), 18–22 October.

Fowler, K. R., Jenkins, E. W., & LaLonde S. M. (2010). Understanding the effects of polymer extrusion filter layering configurations using simulation-based optimization. Optimization and Engineering, 11, 339–354. https://doi.org/10.1007/s11081-009-9096-0

Iqbal, U. M., Kumar, V. S. S., & Gopalakannan, S. (2016). Application of response surface methodology in optimizing the process parameters of twist extrusion process for AA6061–T6 aluminum alloy. Measurement, 94, 126–138. https://doi.org/10.1016/j.measurement.2016.07.085

Karaoglan, A. D. (2021). Optimizing plastic extrusion process via grey wolf optimizer algorithm and regression analysis. Journal of Scientific & Industrial Research, 80 (1), 34-41.

Mason, R. L., Gunst, R. F., & Hess, J. L. (2003). Statistical design and analysis of experiments, 2nd ed., John Wiley & Sons, Hoboken, New Jersey.

Montgomery, D. C. (2008). Design and analysis of experiments, 8th Ed., John Wiley & Sons, New Jersey–Haboken.

Myer, R. H., Montgomery, D. C., Anderson-Cook, C. M. (2008). Response surface methodology: process and product optimization using designed experiments, 3rd Ed., John Wiley & Sons, Hoboken, New Jersey.

Zhou, M. & Paik, J. (2004). An application of neural network and genetic algorithm for optimizing food extrusion process parameters. International Journal of Industrial Engineering - Theory Applications and Practice, 11, 132–139.

Stabilizing the Self-Localized Solitons of the Kundu-Eckhaus Equation by Dissipation

Hazal Yurtbak[1*], Cihan Bayındır[2, 3]

[1]*Czech Technical University, Mathematics Department, Prague, Czechia*
[2]*İstanbul Technical University, Civil Engineering Department, İstanbul, Turkey*
[3]*Boğaziçi University, Civil Engineering Department, İstanbul, Turkey*
yurtbakhazal@gmail.com, cbayindir@itu.edu.tr

Abstract

The Kundu-Eckhaus equation (KEE) is a nonlinear partial differential equation in the nonlinear Schrödinger equation (NLSE) class. This equation was introduced to the scientific literature independently by Kundu (1984) and Eckhaus (1986). It is well-known that KEE admits many different analytical solutions like the NLSE. Those solutions of the KEE are widely used in fields such as nonlinear optics, fiber optical waveforms, water waves mechanics, and hydraulics, just to name a few. In this study, the effect of loss/gain on the soliton solutions of the KEE has been investigated. With this aim, we study the dissipative Kundu-Eckhaus equation (dKEE) (Bayındır, et. al, 2021, Yurtbak, 2019). We analyze the effects of dissipation in the form of a loss term on the self-localized solitons of the dKEE. For this purpose, we propose a Petviashvili's method (PM) for the numerical construction of the soliton solution of the dKEE (Petviashvili, 1976). Using PM, we first numerically compute the soliton solutions of the dKEE and discuss their properties. Then, we analyze the effects of dissipation on the dynamics and stabilities of those soliton using a split-step Fourier method (SSFM) implemented for time-stepping purposes. We show that the dKEE equation admits one and two soliton solutions for zero potential and for photorefractive potential ($V = I_0 cos^2(x)$) cases. Since the solitons under the photorefractive potentials turned out to be unstable during temporal evolution, we introduce and discuss the effects of dissipation on the dynamics and stabilization of those solitons. The effects of dissipation on soliton characteristics and power are also discussed.

Keywords: Dissipative Kundu-Eckhaus equation, Petviashvili's method, Split-step Fourier scheme, solitons.

1. Introduction

The Kundu-Eckhaus equation is a nonlinear partial differential equation and can be considered as an extended nonlinear Schrödinger equation. This equation was introduced to the scientific literature by Kundu (1984) and Eckhaus (1987), independently and used in various branches of nonlinear physics (Demiray, 2003, Bayındır, et. al, 2021). The Kundu-Eckhaus equation (KEE) admits many different types of analytical solutions like the nonlinear Schrödinger equation (NLSE). In this paper, we concentrate on the self-localized solitons of the KEE. In order to construct soliton solutions, we extend the application of Petviashvili's method (PM) (Petviashvili, 1976) to KEE and dissipative (dKEE). PM transforms the governing the nonlinear equation (Demiray, 2003, Bayındır, et. al, 2021, Yurtbak, 2019, Petviashvili, 1976) into Fourier domain and iterates until a convergence criterion is achieved. In this paper, we study the effects of dissipation on the one and two soliton solutions of the Kundu-Eckhaus obtained via PM. We first construct the soliton solutions of the KEE and under the effect of dissipation, we perform their time stepping by split-step Fourier method discussed in (Bayındır, et. al, 2021, Yurtbak, 2019). We discuss the effects on various terms of the dKEE and their effects on the soliton dynamics. We comment on our findings and discuss the implementation and limitations of our work.

2. Mathematical Formulation

The dissipative Kundu-Eckhaus equation (dKEE) can be written as

$$i\frac{\partial U}{\partial t} + \mu_1\frac{\partial^2 U}{\partial \xi^2} + \mu_2|U|^2U + i\mu_3 U + \mu_4^2|U|^4U - 2\mu_4 i(|U|^2)_\xi U = 0, \quad (1)$$

where *U* is the dependent variable, *t* is the time and ξ is the space parameter. In this equation, the parameter μ_1 is the dispersion constant, the parameter μ_2 is the cubic nonlinearity constant and the parameter μ_4 is the quintic-non linearity and Raman scattering constant. The parameter μ_3 controls the dissipation (Demiray, 2003). Solutions of the dKEE are constructed by applying SSFM and PM in our previous research (Bayındır, et. al, 2021, Yurtbak, 2019). In order to discuss the effects of potentials, we subtract photorefractive potential term *VU* form the left hand side of the dKEE use $V = I_0 cos^2(x)$ to represent a photorefractive potential. The PM iteration scheme can be obtained by transforming Eq.1 into Fourier domain, however, it is well-known that such iterations schemes diverge. Thus, to prevent such a instability we can add or subtract a $p\eta$ term with $p > 0$ to the 1D Fourier transform of Eq.(1). We accepted $p = 10$ throughout this paper. After these operations, we obtain the formula as below

$$\hat{\eta}(k) = \frac{(p+|\mu|)\hat{\eta}}{p+\mu_1|k|^2} - \frac{F[V\eta]-F[N(|\eta|^2)\eta]}{p+\mu_1|k|^2} \quad (2)$$

where *V* is the potential term. As before, the iterations of Eq. (2) may be unbounded or it may tend to zero (Bayındır, et. al, 2021). Thus, we introduce a new variable by introducing a new variable as $\eta(\xi)=\alpha\phi(\xi)$ and its Fourier transform as $\eta(k) = \alpha\phi(k)$, to come through this problem. With this change of variables, Eq. (2) can be rewritten as

$$\hat{\phi}(k) = \frac{F[N|\alpha|^2|\phi|^2\phi]}{\mu+\mu_1|k|^2} = R_\alpha[\hat{\phi}(k)] \quad (3)$$

and corresponding iteration scheme becomes

$$\hat{\phi}_{j+1}(k) = \frac{F[N(|\alpha_j|^2|\phi_j|^2)\phi_j]}{\mu+\mu_1|k|^2} \quad (4)$$

where μ is the soliton eigenvalue, α is an algebraic condition to control the convergence of the PM. This algebraic condition is basically an energy conservation principle. We multiply the both sides of Eq. (4) with $\hat{\phi}^*(k)$, where the sign * refers to complex conjugation, and the total energy can be calculated as

$$\int_{-\infty}^{+\infty}|\hat{\phi}(k)|^2\,dk = \int_{-\infty}^{+\infty}\hat{\phi}^*(k)R_\alpha[\hat{\phi}(k)]dk \quad (5)$$

The Eq. (5) is the normalization constraint of the PM algorithm which is summarize above and can be find the self-localized solitons of the dKEE with the potential term starting with single or multi-Gaussians. We present our results for the dKEE in the next section.

3. Results and Discussion

3.1. Single soliton Dynamics

In Fig.1, spatial profile of the single soliton solution for various μ_1 is given and the direction of parameters on the figures is depicted. For our simulations we used $N=1024$, $\alpha=10^{-10}$ for the construction of the single soliton solutions. After, we repeat similar steps for μ_2 and μ_4. The parameters of the potential term are selected as I_0=2.5 where the photorefractive potential terms is calculated by $V = I_0 cos^2(x)$ throughout this paper. As illustrated in Fig. 1, the parameter μ_1 controls the shape of the soliton.

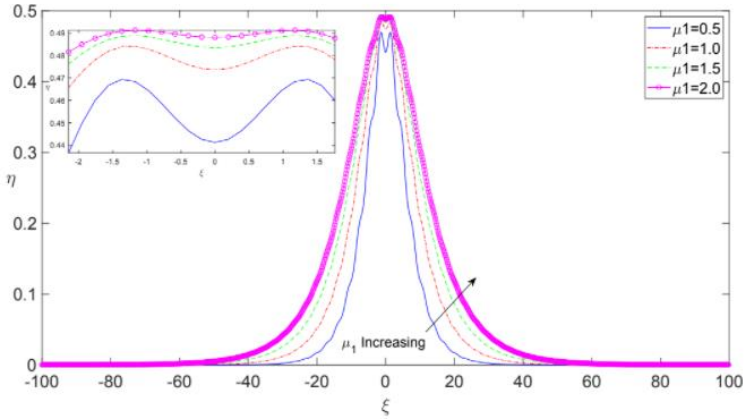


Figure 1. Self-localized single soliton solution as a function of ξ for various μ_1 values with photorefractive potential.

In Fig. 2 illustrates the power of the soliton as a function of soliton eigenvalue with photorefractive potential. The computational parameters are selected as α =10^{-7}, N =1024, μ_1= 1, μ_2=2 and μ_4= 2/3, p= 10, I_0= 2.5 as before. Clearly, solitons satisfy the Vakhitov-Kolokolov stability criteria, however, it is not a sufficient condition.

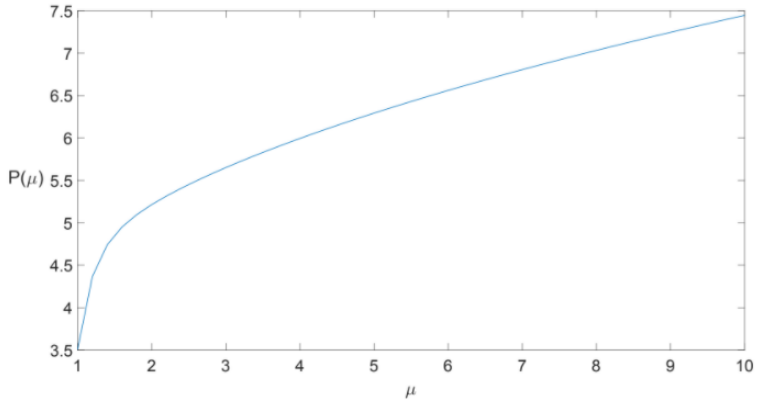


Figure 2. Self-localized single soliton solution power as a function of soliton eigenvalue, μ with photorefractive potential.

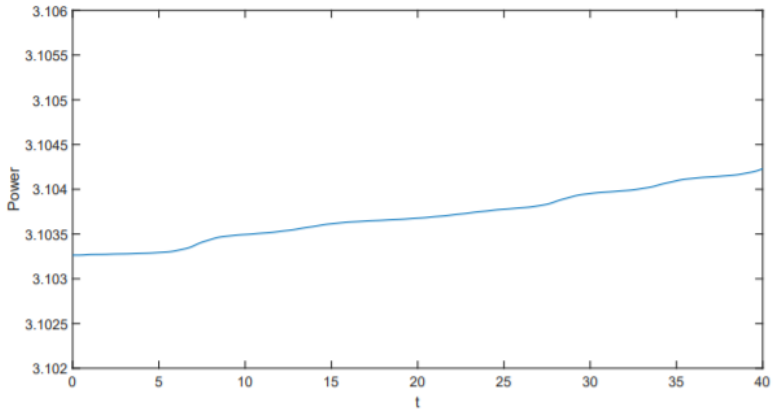


Figure 3. Self-localized single soliton solution power as a function of time under no dissipation, μ_3= 0 with photorefractive potential.

In order the assess the temporal stability, we implement a split-step Fourier scheme and starting from the soliton solutions we perform time stepping. We simulate a numerical solution for single soliton for *N=1024,* $\alpha=10^{-7}$*,* $\mu=1$. In Fig. 3, we depict single soliton power as a function of time and in Fig. 4 we depict single soliton power as a function of time for various dissipation parameters and we plot a constant power line. It is clear that, the unstable soliton of the KEE can be stabilized by dissipation. All values which are under the constant line is stabilizes the for single soliton solution for the range of time considered. We can specify the best value as μ_3=0.01 and we depict the single soliton solution at two different times (*t= 0* and *t=40*) in Fig.5 which are under the effect of the photorefractive potential. We depict the real, imaginary and absolute values of *U* for μ_3=0.01.

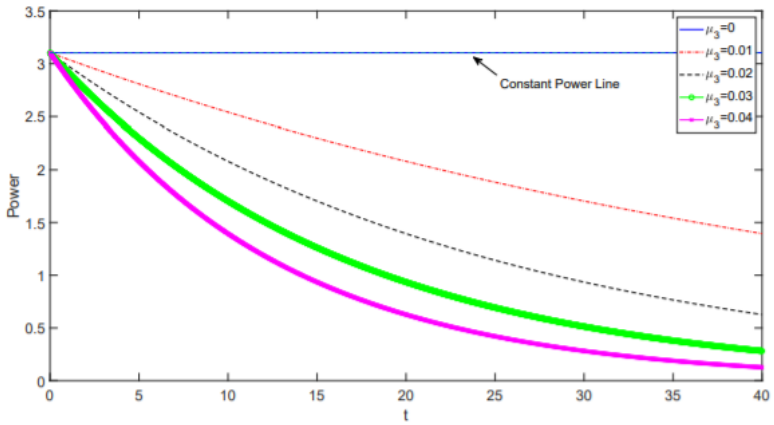


Figure 4. Self-localized single soliton solution power as a function of time for various dissipation parameter, μ_3 with photorefractive.

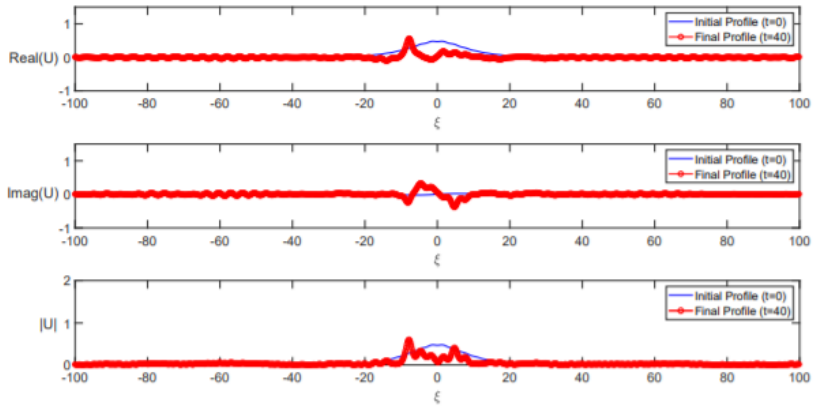


Figure 5. Self-localized single soliton solution at two different times t = 0 and t = 40 for μ_3=0.01 with photorefractive potential; a) Real part of U, b) Imaginary part of U, c) Absolute value of U.

3.2 Two soliton Dynamics

Starting from the two Gaussian initial conditions and following similar steps, we construct the two soliton solution with photorefractive potentials. In Fig. 6, the two soliton solution is depicted as a function of spatial parameter for various μ_1 and the direction of increasing μ_1 is marker. The computational parameters are selected to be *N=1024,* $\alpha=10^{-5}$ and μ_3=0. The simulations are repeated for various values of μ_2 and μ_4 and given in (Yurtbak, 2019).

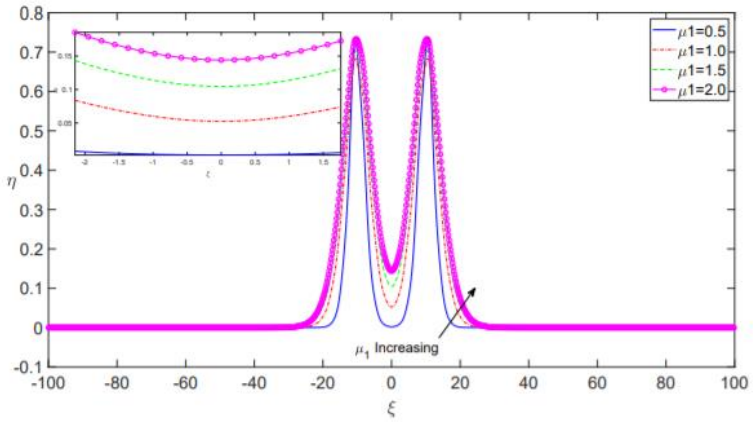


Figure 6. Self-localized two soliton solution as a function of ξ for various μ_1 values with photorefractive potential.

Trend of the power of the two soliton solutions as a function of soliton eigenvalue under the effect of photorefractive potential turned out to be similar to Fig.2 and Fig.3 of the single soliton case and for the sake of brevity of the presentation is not depicted here. For details, the reader is referred (Yurtbak, 2019).

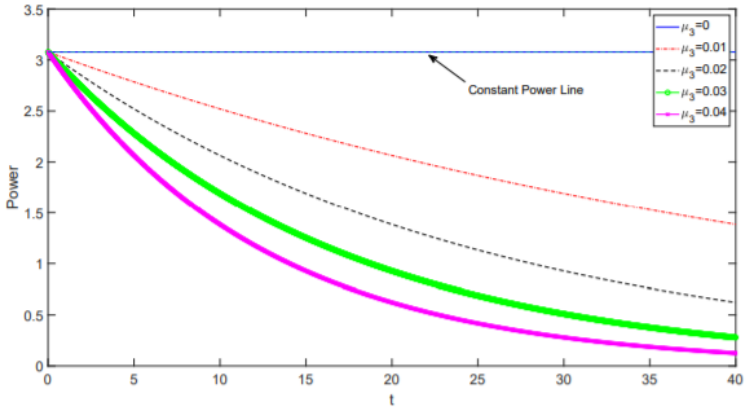


Figure 7. Self-localized two soliton solution power as a function of time for various dissipation parameter, μ_3 with photorefractive potential.

Lastly, we provide a function for various values of dissipation in Fig. 7. We obtain a time-power function of the two soliton in Fig.7 and investigate behavior of the function under dissipation. A closer inspection of Fig.4 and Fig.7 reveals that with no dissipation power grows unboundedly causing unstable behavior (Yurtbak, 2019). The values of μ_3, which lead to lines under the constant line can be used to stabilize solitons of the KEE for the temporal range considered. We specify best value for numerical solution and simulate the temporal dynamics for μ_3= 0.01. In this way, using the split-step time stepping algorithm we produce Fig. 8. This figures compares the initial and final profiles at two different times (*t=0* and *t=40*) for the dissipation parameter of μ_3=0.01. Although some splitting is observed in the crest of the soliton, the base of the soliton is well-preserved, which is beneficial for many engineering purposes.

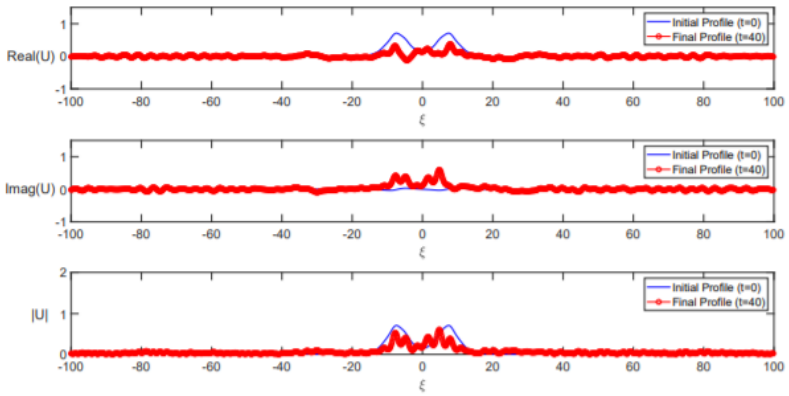


Figure 8. Self-localized two soliton solution at two different times *t=0* and *t=40* for μ_3= 0.01 with photorefractive potential; a) Real part of *U*, b) Imaginary part of *U*, c) Absolute value of *U*.

4. Conclusion

In this study, we have studied the effects of dissipation on the self-localized solitons of the KEE. We implemented a Petviashvili method to construct the single soliton and two soliton solutions of the KEE. We have showed that the unstable single and two soliton solution of the KEE with nonzero optical potentials, i.e V≠0 for $V= I_0 cos^2(x)$ type photorefractive potential can be stabilized by imposing dissipation. We discussed the effects of the strength of the dissipation term on soliton dynamics. Our results can be used in the derivation of the numerical *N*-soliton solutions of KEE under dissipation. Possible application areas include but are not limited to dissipative optical media, dissipative hydrodynamic media such as the ocean exposed to oil spill and other dissipative media in matter physics and Bose-Einstein condensation.

References

Bayındır, C. & Yurtbak, H. (2021). A split-step Fourier scheme for the dissipative Kundu-Eckhaus equation and its rogue wave dynamics. TWMS Journal of Applied and Engineering Mathematics, 11 (1), 56-65. http://jaem.isikun.edu.tr/web/images/articles/vol.11.no.1/06.pdf

Demiray, H. (2003). An analytical solution to the dissipative nonlinear Schrödinger equation. Applied Mathematics and Computation, 145 (1), 179-184. https://doi.org/10.1016/S0096-3003(02)00476-9

Eckhaus, W. (1986). The long-time behaviour for perturbed wave-equations and related problems. In: Kröner E., Kirchgässner K. (eds) Trends in Applications of Pure Mathematics to Mechanics. Lecture Notes in Physics, vol 249. Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0016391

Kundu, A. (1984). Landau & Lifshitz and higher-order nonlinear systems gauge generated from nonlinear Schrödinger-type equations. Journal of Mathematical Physics, 25, 3433-3438. https://doi.org/10.1063/1.526113

Petviashvili, V.I. (1976). Equation of an extraordinary soliton. Soviet Journal of Plasma Physics, 257 (2). https://ui.adsabs.harvard.edu/abs/1976FizPl...2..469P

Yurtbak, H. (2019). Analytical and numerical analysis of the dissipative Kundu-Eckhaus equation, MS Thesis, FMV Işık University. Available from: https://doi.org/10.13140/RG.2.2.16409.77921

Random forest regression and an alternative model selection procedure

Mehmet Ali Kaygusuz[1] and Vilda Purutçuoğlu[1,2*]

[1] *Middle East Technical University, Department of Statistics, Ankara, Turkey*
[2] *Middle East Technical University, Department of Biomedical Engineering, Ankara, Turkey*
makaygusuz1988@gmail.com, vpurutcu@metu.edu.tr

Abstract

Multiple testing procedures are very popular in recent years due to the huge amount of data in genetics, engineering and finance. In order to construct proper models for those data, different approaches are suggested. More recently, the conditional randomization tests (CRT) are proposed while there is no distributional assumption about the model. On the other hand, the hold-out-randomization test is suggested by using both bootstrap and cross validation algorithms. Moreover, the Model-X knock-off filter is applied for causal models with Genome-wide studies. Hereby, in this study, we investigate the computational cost and statistical properties of the random forest model which is one of the well- known causal regression approaches. In our analyses, we detect its accuracy by using two distinct model selection criteria, namely, Bozdogan's Consistent Akaike information criterion with Fisher information matrix (CAICF) and Information and COMPlexity (ICOMP) criterion . We evaluate the performance of the underlying extended model via protein- protein interaction networks' and simulated datasets under different dimensions and compare the results via the original random forest model.

Keywords: Causal models, multiple testing, model selection, random forest algorithm

1. Introduction

Traditionally, Causal models have a crucial role to find relationship between dependent variables. On the other hand, Candes et al.(2018) suggested that the conditional randomization test, in particular, when we do not know the conditional distribution of variables is very useful to construct the model. However, this testing procedure can be computationally demanding for high dimensional causal regression. For this reason, we use the random forest method to propose more computationally efficient selection procedure by the CRT. Moreover, in modern statistical applications, it is very important to decide on the relationship between variables and causal models are very useful models for this purpose, particularly, in Genome wide studies (GWAS). In general, these models are used in broad range application such as genetics, epistemology and time series analysis. More recently, the covariance properties of those are studied by Shah et al. (2020a). More importantly, the kernel conditional independence tests with an application causal discovery was studied by Zhang et al.(2011) and Pfister et al. (2017). On the other side, Shah and Peters (2020b) showed that the conditional independence test can be very hard without knowing the conditional distribution of (X,Y,Z). Indeed, Candes et al. (2018) already proposed the conditional randomization tests for logistic regression. The randomization test is a special case of statistical significance test where the distribution of the test statistic under the null hypothesis is calculated for all possible values of the test statistic.

On the other hand, to make randomization experiments can be challenging, in particular, when the number of random variable (p) is bigger than the number of observation (n). Thus, we use the random forest algorithm for causal models to suggest computationally accurate model selection procedure. Random forest algorithm is introduced by Breiman (2001) and widely applied for different classification

and regression purpose. Moreover, random forest methods are an ensemble method which use decision trees and it is a special type of Bagging (Bootstrap aggregating) method. Additionally, it uses a random subset of features. Thus, it is known a variance reduction technique and reduces the Mean Square Error (MSE). Recently, Wager and Athey (2018) and Athey et al. (2019) proposed a novel type of random forest algorithm for causal models.

On the other side, the selection of optimal model among alternatives is another common problem in many modelling approaches. Because of the high dimension with respect to the sample size, various model selection criteria have been developed for both parametric and non-parametric models. AIC, BIC and StaRS are some well- known alternatives. Information and complexity (ICOMP) selection criterion is an another strong candidate to select the best model in causal models since Deniz et al. (2011) introduced ICOMP selection criterion for structural equation modelling with mixed and categorical data. Furthermore, Bülbül et al. (2019) and Kaygusuz and Purutçuoğlu (2019) suggested Consistent AIC with Fisher information matrix (CAICF) and ICOMP selection criteria for Gaussian graphical models in the low and high dimensional settings. Then, Kaygusuz and Purutçuoğlu (2021) proposed the non-parametric and bayesian bootstrap procedure for Gaussian graphical models to improve CAICF and ICOMP selection criteria for high dimensional data. In this study, we compare CAICF and ICOMP selection criteria when the random forest algorithm is implemented as causal models.

Thereby, the organization of the paper is as follows: The second section includes the definition of Candes et al.(2019) conditional randomization test, the mathematical representation of causal models. Random forest algorithm and CAICF and ICOMP criterion are given in the third section. Finally, we present data analysis and conclusion in the fourth and the fifth section respectively.

2. Theory and Methods

2.1 Conditional randomization test

The conditional randomization tests (CRTs) are introduced by Candes et al. (2018) without any assumption of dependent variable Y. They merely assume that the conditional distribution of (*X*,*Y*,*Z*) is known. Whereas the detection of the conditional distribution can be challenging, in particular, for high dimensional data. Thus, it is accepted that the conditional distribution of *X*|*Z* is known and *Q*(.|*Z*) shows the distribution of *X* given *Z* = *z*, conditional on $Z_1, \ldots, Z_n$. The CRT generates the *X* values from data sets which have the following distribution : $X_1^{(1)} \sim Q(. \mid Z_i)$, *w*here it is obtained independently from every *i* = 1, . . . , *N* and independently from real values of X_i's and Y_i's. According to the null hypotheses in equation, one can write

$(X|Y=y,Z=z) \overset{d}{=} (X|Z=z) \sim Q(.|z)$. In here, $\overset{d}{=}$ indicates the equality in distribution. Hence, we can write the above equation as follows.

$$(X^{(1)}, Y, Z) = (X, Y, Z) \quad \text{for all } (X^{(1)} = X_1^{(1)}, \ldots, X_n^{(1)} \text{ by } H_0. \tag{1}$$

If we have a difference between two sides of equation, we can reject H_0. In order to obtain the test value of null hypotheses, we need to repeat this steps (k) times. On the other side, $(X_i^{(k)}|X, Y, Z) \sim Q(.|Z_i)$ is obtained independently for *i* = 1, . . ., *N* and *k* = 1, . . ., *K*. In this expression, we assume that the conditional distribution of (X, Y, Z), $(X^{(1)}, Y, Z)$, ..., $(X^{(k)}, Y, Z)$ are identically distributed and ex-changeable by the null hypotheses. For this reason, any statistics of T(X,Y,Z) is also exchangeable too as follows: $T(X, Y, Z), T(X^{(1)}, Y, Z), \ldots, T(X^{(k)}, Y, Z)$. In this representation, we can give the algorithm to compute CRT as below: The datais referred as (X_i,Y_i) where X is independent variable and Y is

dependent variable for i = 1,...,n observations. Moreover, T(X,Y,Z) is a test statistics to test conditional relationship between X and Y $k \in (1,...,K)$ copies the sample $X^{(k)}\ L(X_i,\ Y_i)$ for all i to compute $T(X^{(k)}, Y, Z)$. Hence, the p-value is computed as $\mathrm{p} = \frac{\left\{1+\sum_{k=1}^{K} 1\ \right\}\ \{T(X^{(k)},Y,Z)\geq T(X,Y,Z)\}}{1+K}$. Here, the p-values are obtained by true X vector against the values of the CRT copies.

2.2 The Mathematical Representation of Causal Models

Let us define random variables X_1, ..., X_p for graphical models as a pair G, $L(X)$ where $L(X) = L(X_1,..., X_\mathrm{p})$ is a joint distribution for a directed acyclic graph. Then, we define $S = (S_1,.., S_\mathrm{p})$ for p equations as $S_j:\ X_j = f_j\left(X_{PA_j},\ N_j\right), \qquad (j = 1, \dots, p)$,where $L(N) = L(N_1,.., N_\mathrm{p})$ are the noise variables and PA_j indicates the parents of j. We assume that the noise terms are jointly independent and L(N) is a product distribution. Moreover, causal models are obtained via directed edges from each variables X_k, for $k \in PA_\mathrm{j}$.

$$X_j = \sum_{k\in PA_j}\ \beta_{jk}\, X_k + \ N_j. \qquad (j=1,..,n)\ \ k \in PA_j \tag{2}$$

In Equation 2, N_j 's are independent identically distributed according to $N(\mu, \sigma^2)$ and for each j $\in \{1,..,n\}$ it must be $\beta_{jk} \neq 0$ for all k $\in PA_j$.

3. Proposed Method

3.1 Random Forest Algorithm

Random forest algorithm is introduced by Breiman (2001) and is widely used to solve regression or classification problems in different scientific domains such as economics, genetics, engineering. This algorithm provides a collection of decision trees, and each tree in the ensemble is comprised of a data sample obtained from a training set with replacement, called the bootstrap sample. In the bootstrap training sample, one-third of it is taken as test data, known as the out-of-bag sample. The individual decision trees can be averaged when we study a regression assignment. As an advantage of random forest algorithm, it reduces the variance and prediction error. But, it can suffer from the computational cost to compute each individual tree. Hereby, we can summarize the procedure for the random forest algorithm as follows: **Step1**:Generate pairs $(X_i\ ,\ Y_i\ )$ i= 1,...,n where $X_\mathrm{i} \in R^\mathrm{d}$ indicates independent variable and $Y_\mathrm{i} \in R$ is a response variable.Calculate $(X_1,\ Y_1)\ ,\dots,\ \ (X_n,\ Y_n)$ pairs.

Step 2:Conduct bootstrap sample $(X_1^*,\ Y_1^*)$,..., $(X_n^*,\ Y_n^*)$ which is obtained randomly drawing n times with replacement from the data $(\ X_1,\ Y_1)\ ,\dots,\ \ (X_n,\ Y_n)$.

Step 3:Repeat step 1and 2 for B times to obtain the random forest estimator as follows $\hat{\mu}_{RF}=\frac{1}{B}\ \ \sum_{j=1}^{B}\ \ (\hat{\mu}_j\,(X)\,))$.

3.2 Consistent AIC and ICOMP Selection Methods

The model selection is the problem of selecting one from among a large set of candidate computational models. The two most common likelihood model selection methods are Akaike (1973) and Schwarz (1978) criteria. Akaike information criterion (AIC) uses the Kullback-Leibler divergence and the maximum likelihood approach to find the best model and it is described basically a penalized version of the attained maximum log-likelihood for each candidate model. Schwarz criterion is known as Bayesian information criterion (BIC) and has a form of the penalized log-likelihood function too, but, it works with Bayesian prior and it has a strong penalty for additional parameters with respect to AIC. In the literature, some extensions of AIC are used to select the right model more consistently with additional

penalty terms. Three of those extended selection criterions are proposed by Bozdoğan (1987 and 2010). We call these selection methods as an extended AI's and our purpose is to propose more consistent model selection as an extension of these selection criteria. Accordingly, these AIC based criteria propose to calculate a metric distance between the model and the true distribution in a such a way that a consistent inference method can make this distance as small as possible. In the study of Bozdoğan (1987), the Kullback- Leibler divergence is applied and this criterion is named the consistent AIC (CAIC) whose expression as below :

$$CAIC(k) = -2\log L\left(\hat{\theta}_k\right) + k[logn + 1], \tag{3}$$

where the likelihood of θ is shown by log $L(\hat{\theta}_k)$ and *k* is the degrees of freedom of the distribution. We can see similarity between CAIC(*k*) and BIC of *k* log *n* and *k*[log *n* + 1], respectively by suggesting a stronger penalty term. This penalty term makes a difference while finding in a more parsimonious model than AIC and BIC. Here, log *n* + 1 provides the consistency of the model selection.

Another Bozdogan 's selection criterion is based on the Consistent AIC with Fisher information (CAICF(*k*)). The method has an increasing penalty term for the over-parametrization whose expression is presented as below.

$$\text{CAICF(k)} = -2\log \text{L}(\hat{\theta}_k) + \text{k[log n + 2]} + \log|\text{I}(\hat{\theta}_k)|. \tag{4}$$

In here, log $L(\hat{\theta}_k)$ indicates the likelihood estimation of θ, as used before, *k* is the degrees of freedom of the distribution and $I(\hat{\theta}_k)\,\hat{F}^{-1}$ denotes the Fisher information matrix. AIC and BIC have some limitations when the number of variables (*p*) is greater than the number of observations (*n*). For this reason, Bozdoğan (2010) suggest the Information and COMPlexity (ICOMP) for the covariance matrix. This new criterion can penalize the free parameters and the covariance matrix directly in the same time with a third term.This third term in the loss function can calculate the distance when parameter estimates are correlated with the model fitting stage. Below, the equation of ICOMP criterion is represented as follows.

$$ICOMP = -2\log L\left(\hat{\theta}_k\right) + 2C(\hat{\Sigma}) \tag{5}$$

in which log $L(\hat{\theta}_k)$ is the log-likelihood of E, $\hat{\theta}_k$ shows the maximum likelihood estimate of the parameter vector of θ_k , *C* indicates a real-valued complexity measure and $\hat{\Sigma} = \widehat{cov}(\hat{\theta}_k)$ denotes the estimated covariance matrix of the parameter vector of the candidate model. This covariance matrix might be obtained in different ways. Bozdoğan (2010)'s choice is the computation of the inverse of the Cramer-Rao lower bound matrix. In our paper, the estimated inverse Fisher information matrix (IFIM) of the model can make more consistent accuracy of the parameter estimation as $\hat{F}^{-1} = \left\{-E\left(\frac{\partial^2 logL(\theta)}{\partial\theta\partial\theta'}\right)\right\}^{-1}$. In this expression, the ($s \times s$)-dimensional second-order partial derivatives of the log-likelihood function of the estimated model is denoted by $\hat{F}^{-1}$ and ∂ stands for the partial denotation of the given function. Finally, a more general form of ICOMP can be denoted as the following way:

$ICOMP = -2\log L\left(\hat{\theta}_k\right) + 2C(\hat{F}^{-1})$ when $C(\hat{F}^{-1}) = \frac{s}{2}\log\left[\frac{\widehat{tr}F^{-1}}{s}\right] - \frac{1}{2}\log\left|\hat{F}^{-1}\right|$. The second term shows that the information complexity of the estimated inverse Fisher information matrix of the model and $s = \dim(\hat{F}^{-1}) = rank\left(\hat{F}^{-1}\right)$ while dim(.) shows dimension of given matrix.

4. Results and Discussion

We have two simulated datasets. In this first data, the number of observations (n) is 60 while the number of variables (p) is 100 and in the second data, the number of observations (n) is equal to 20 while the number of variables (p) is 100 as a large and moderate network structure whose topologies are scale-free. We have applied conditional randomization procedure for causal models in both data. For the first dataset, the CRT algorithm selected the 1st, 2d, 3rd, 5th, 8th, 11th, 14th, 19th, 31st, 32rd, 38th, 65th, 73rd, 78th, 84th, 98th variables, which is totally 16 variables among 100. For the second data set, the CRT algorithm chosen the 1st, 2nd, 3rd, 9th, 11th, 12nd, 14th ,17th, 21st, 26th, 28th , 30th, 41st, 42nd, 43th, 46th, 55th, 59th, 61th, 65th, 74th, 81st , 82nd, 87th, 88th, 92nd, 95th variables, which is totally 27 variables among 100. After deletion of variables obtained by CRT, we conducted random forest algorithm whose results are seen in the above tabulated values. The findings indicate that CRT is promising, specifically, under CAICF approach with significant increases in accuracy and F-score. On the other side, ICOMP is invariant to the CRT procedure.

Table 1: Accuracy and F-scores of ICOMP and CAICF w/o random forest algorithm.

Data	Measures	ICOMP without random forest	ICOMP with random forest	CAICF without random forest	CAICF with random forest
Dataset 1 n = 60, p = 100	Accuracy F-score	0.182 0.308	0.182 0.308	0.364 0.533	**0.455** **0.625**
Dataset 2 n = 20, p = 100	Accuracy F-score	0.159 0.266	0.159 0.266	0.182 0.308	**0.365** **0.534**

5. Conclusion

We have proposed an alternative model selection procedure for causal model using with Bozdoğan`s CAICF and ICOMP selection criteria. However, when the numbers of parameters (p) are more than the number of observations (n), model selection procedure can be challenging. For this reason, we have suggested random forest algorithm for causal models to propose more consistent model selection procedure. Before this algorithm, we have applied Candes et al. (2018)`s conditional randomization test for causal models. According to real datasets results, we have seen an improvement in accuracy of causal models for CAICF. On the other hand, we have to conduct more comprehensive simulation analysis with the proposed model selection procedure for the different scales, the number of observations and the number of variables. As a future work, we can also use different types of algorithms such as boosted random forest algorithm to improve the accuracy of causal models. Furthermore, causal models with an application to survival analysis are very attractive research area to investigate computational efficiency of random forest algorithm.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle , IN Petrov, BN.; Csaki, F. Editors Second International Symposium on information theory. Budepest: akademiai Kiad, 267-281. https://doi.org/10.1007/978-1-4612-1694-0_15

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forest, Annals of Statistics, 47 (2), 1148-1178. Doi: 10.1214/18-AOS1709

Bülbül, G. B., Purutçuoğlu,V., & Purutçuoğlu, E. (2019). Novel models election criteria on sparse biological networks, International Journal of Environmental Science and Technology , 16 , 1-12. https://doi.org/10.1080/ 00949655.2020.1870689

Bozdoğan, H. (1987). Model selection and AIC :the general theory and its analytical extensions, Psychometrica, 52 (3), 345-370. https://doi.org/10.1007/BF02294361

Bozdoğan, H. (2010). A new class of information complexity (ICOMP) criteria with an application to costumer profiling and segmentation, Istanbul University Journal of the School of Business Administration, 39 (2), 370-398.

Breiman, L. (2001). Random forest, Machine Learning, 45 (1), 5-32. https://doi.org/10.1023/A:1010933404324

Candes, E. J. , Fan, Y., Janson, L. & Lv, J. (2018). Planning for gold: model-X knock-offs for high dimensional controlled variable selection, The Royal Statistical Society Series B, 80 (3), 551-577. https://doi.org /10.1111/rssb.12265

Deniz, E., Bozdogan, H., & Katragadda, S. (2011). Structural equation modelling of categorical mixed-dsta using the novel gifi transformation and information complexity criterion, Istanbul University Journal of the School of Business Administration, 40 (1), 86-123.

Kaygusuz, M. A. & Purutçuoğlu, V. (2019). Model selection methods for sparse biological networks, Chapter in: Lecture Notes on Data Engineering and Communications Technologies. Editors: T. Yiğit, U. Köse, Springer.

Kaygusuz, M. A. & Purutçuoğlu,V. (2021). Model selection criteria with bootstrap algorithms: Application with biological networks, Artificial intelligence for Data Driven Techniques, Elsevier (In Press).

Pfister, N., Bühlmann, P., Schölkopf, B., & Peters, J. (2017). Kernel-based tests for joint independence. Journal of the Royal Statistical Society: Series B, 80:5–31. https://doi.org/10.1111/rssb.12235

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464.

Shah, R. D., Frol, B., Thanei, G. A., & Meinhausen, N. (2020a). Right singular vector projection graphs: fast hig dimensional covariance matrix estimation under latent confounding, Journal of The Royal Statistical Society: Series B, 82 (2), 361-389. https://doi.org/10.1111/rssb.12359

Shah, R. D. & Peters, J. (2020b). The hardness of conditional independence testing and the generalized covariance measure, The Annals of Statistics, 48 (3), 1514-1538. Doi:10.1214/19-AOS1857

Wager, S. & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests, Journal of the American Statistical Association, 113 (523), 1228-1242. https://doi.org/10.1080/01621459 .2017.1319839

Zhang, K., Peters, J., Janzing, D., & Schölkopf, B. (2011). Kernel-based conditional independence test and application in causal discovery. 27th Annual Conference on Uncertainty in Artificial Intelligence (UAI), 804–813.

Parabolic Optimal Control Problems Described by Partial Differential Inclusions

Sevilay Demir Sağlam [1*], Elimhan N. Mahmudov [2]

[1]*Istanbul University, Department of Mathematics, Istanbul, Turkey*
[2]*Istanbul Technical University, Department of Mathematics, Istanbul, Turkey*
[2]*Azerbaijan National Academy of Sciences, Institute of Control Systems, Baku, Azerbaijan*
sevilay.demir@istanbul.edu.tr, elimhan22@yahoo.com

Abstract

The paper concerns the optimization of partial differential inclusions of the parabolic type given by polyhedral set-valued mappings. We derive the optimality conditions for the problems under consideration by employing the result of the discrete approximation problem associated with the continuous problem. We formulate the sufficient conditions by passing formally to the limit as the discrete steps tend to zero in the discrete approximation problem. We consider some linear optimal control problems to demonstrate the above approach.

Keywords: Partial differential inclusions, polyhedral optimization, optimality conditions

1. Introduction

Several areas involving optimal control problems described by ordinary and partial differential equations and/or inclusions have advanced significantly in the last decade. Variational analysis of partial differential inclusions is performed using discrete approximations and advanced generalized differentiation methods to provide new optimality conditions of the Euler-Lagrange and Hamiltonian types. Several mathematical studies of parabolic equations and inclusions have been conducted, motivated by mathematical and physical problems (Cheng et al., 2011). The optimality conditions obtained by Mahmudov are more exact since they involve usable forms of the Weierstrass-Pontryagin conditions and Euler-Lagrange type of inclusions (Mahmudov, 2021a, 2021b, 2021c). Many of the necessary and sufficient conditions for optimality for partial differential inclusions presented by Mahmudov in his survey studies eventually necessitate the development of new types of equivalence results. In the paper (Mahmudov, 2008, 2009), sufficient conditions of optimality for parabolic differential inclusions are derived, and duality theorems are proven. We use difference approximations of partial derivatives and grid functions on a uniform grid in this study to approximate the problems under review. We obtain the necessary and sufficient optimality condition for a discrete-approximate problem associated with the continuous problem. Then, we formulate sufficient conditions of optimality for convex problems with partial differential inclusions by formally passing to the limit as the discrete steps tend to zero in the discrete approximation problem. Later, we look into the differential problem for several types of set-valued mapping, such as polyhedral mapping. We calculate the LAM(locally adjoint mapping) for the polyhedral set-valued mapping and prove sufficient conditions of optimality for polyhedral problems with partial differential inclusions. At the end of the section, a linear optimal control problem is discussed.

The essential properties and definitions of convex analysis structures used in this paper can be found in Mahmudov (2011) and Mordukhovich (2006). As usual, $\mathbb{R}^n$ is an n-dimensional Euclidean space,

* *Corresponding Author*

$\langle x,v\rangle$ is the inner product of elements $x,v\in\mathbb{R}^n$ and (x,v) is a pair of x,v. A set-valued mapping $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ is convex if its graph $\operatorname{gph}F=\{(x,v):\ v\in F(x)\}$ is convex in $\mathbb{R}^{2n}$. And F is convex valued if $F(x)$ is a convex set for each $x\in domF=\{x:F(x)\neq\varnothing\}$. Moreover, we introduce the Hamiltonian function and the notation of the argmaximum set for a set-valued mapping, respectively $H_F(x,v^*)=\sup_v\{\langle v,v^*\rangle:v\in F(x)\},\ v^*\in\mathbb{R}^n;\ F_A(x;v^*)=\{v\in F(x):\langle v,v^*\rangle=H_F(x,v^*)\}$. For a convex set-valued F, we let $H_F(x,v^*)=+\infty$, if $F(x)=\varnothing$.

A convex cone $K_A(z_0)$ is called the cone of tangent directions at a point $z_0=(x^0,v^0)\in A$ if from $\bar{z}=(\bar{x},\bar{v})\in K_A(z_0)$ it follows that $\bar{z}$ is a tangent vector to the set A, i.e., there exists a function $\varphi(\lambda)\in\mathbb{R}^{2n}$ satisfying $z_0+\lambda\bar{z}+\varphi(\lambda)\in A$ for sufficiently small $\lambda>0$, where $\lambda^{-1}\varphi(\lambda)\to 0$, as $\lambda\downarrow 0$. Evidently, for a convex set A at a point $(x^0,v^0)\in A$ setting $\varphi(\lambda)\equiv 0$ we have $K_A(z_0)=\{(\bar{x},\bar{v}):\bar{x}=\lambda(x-x^0),\bar{v}=\lambda(v-v^0),\lambda>0\},\forall(x,v)\in A$. As usually, $K_A^*(x^0,v^0)$ is the dual cone to a cone of tangent vectors $K_A(x^0,v^0)$. A set-valued mapping $F^*(\cdot;(x,v)):\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ defined by $F^*(v^*;(x,v))=\{x^*:(x^*,-v^*)\in K^*_{gphF}(x,v)\}$ is called the locally adjoint mapping (LAM) to F at a point (x,v) where $K^*_{gphF}(x,v)$ is the cone dual to the cone $K_{\operatorname{gph}F}(x,v)$.

This paper is about the polyhedral optimization of parabolic partial differential inclusions. In related problems, we examine the main properties of polyhedral mappings in depth (Demir Sağlam et al., 2020, 2021a, 2021b, 2021c, 2021d). A polyhedral convex set in $\mathbb{R}^n$ is a set that can be expressed as the intersection of some finite family of closed half-spaces. In particular, if the finite system of inequalities is homogeneous, the set of solutions to this finite system of inequalities is called the polyhedral cone.

2. Problem Definition

At first, we study the convex problem of partial differential inclusions of the parabolic type:

$$\text{minimize } I[u(\cdot,\cdot)]=\iint_D f\left(u(x,t),x,t\right)dxdt, \tag{1}$$

$$\frac{\partial^2u(x,t)}{\partial x^2}-\frac{\partial u(x,t)}{\partial t}\in F\left(u(x,t),x,t\right),\ (x,t)\in D, \tag{2}$$

$$u(0,t)=\alpha_0(t),\ u(x,0)=\beta_0(x),\ u(1,t)=\beta_1(t),\ D=[0,1]\times[0,1], \tag{3}$$

where $F(\cdot,x,t):\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ is a convex set-valued mapping, $f(\cdot,x,t)$ is proper convex function and $\alpha_0:[0,1]\to\mathbb{R}^n,\ \beta_i:[0,1]\to\mathbb{R}^n,(i=0,1)$ are absolutely continuous functions.

The problem is to find a classical solution $\tilde{u}(x,t)$ of the so-called first boundary value problem (1)-(3) that minimizes the functional I over a set of feasible solutions. For convex problem (1)-(3), a feasible solution is understood as an absolutely continuous function satisfying almost everywhere (a.e.) the partial differential inclusions (2) in $D=[0,1]\times[0,1]$ and the boundary conditions (3). It should be noted that the definition of a solution in one sense or another (classical, generalized, etc.) does not in any way

restrict the class of problems under consideration. Let us denote $\Omega=(0,1)\times(0,1)$, $\Gamma_0=\{(0,t):t\in(0,1)\}\cup\{(x,0):x\in(0,1)\}$, $\Gamma=\{(1,t):t\in(0,1)\}$. A function belonging to the space $C^{1,2}(\Omega)\cap C^1(\Omega\cup\Gamma_0\cup\Gamma)$ satisfying differential inclusion (2) in Ω, initial conditions $u(0,t)=\alpha_0(t),\ u(x,0)=\beta_0(x)$, and boundary conditions $u(1,t)=\beta_1(t)$ in Γ_0 and Γ respectively, we call a classical solution. Here $C^{1,2}(\Omega)$ is the space of functions $u(x,t)$ having continuous derivatives $\frac{\partial^2 u(x,t)}{\partial x^2}, \frac{\partial u(x,t)}{\partial t}$.

3. Main Results on Optimality Conditions of Partial Differential Inclusions

We use the result of the discrete approximation problem connected with the continuous problem to construct the optimality conditions for the problems under study. In the discrete approximation problem, we construct the sufficient conditions by passing formally to the limit as the discrete steps tend to zero. Since the construction of the Euler-Lagrange inclusion as well as transversality conditions are complicated by the accompaniment of discrete and discrete-approximation problems, we give only the final result.

Theorem 3.1 Suppose that $f(\cdot,x,t)$ is continuous convex proper function and $F(\cdot,x,t):\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ is a convex set-valued mapping. Then for the optimality of the solution $\tilde{u}(x,t)$ in the problem (1) –(3) it is sufficient that there exists a classical solution $u^*(x,t)$ such that the conditions $(a)-(c)$ hold almost everywhere on D:

$$(a)\quad \frac{\partial^2 u^*(x,t)}{\partial x^2}+\frac{\partial u^*(x,t)}{\partial t}\in F^*\left(u^*(x,t);\left(\tilde{u}(x,t),\frac{\partial^2\tilde{u}(x,t)}{\partial x^2}-\frac{\partial\tilde{u}(x,t)}{\partial t}\right),x,t\right)-\partial f\left(\tilde{u}(x,t),x,t\right),$$

$$(b)\quad u^*(1,t)=0,\ u^*(x,1)=0,\ u^*(0,t)=0,\ (x,t)\in D.$$

The following condition guarantees (Theorem 2.1 Mahmudov (2011)) that the LAM F^* is nonempty:

$$(c)\quad \frac{\partial^2\tilde{u}(x,t)}{\partial x^2}-\frac{\partial\tilde{u}(x,t)}{\partial t}\in F_A\left(\tilde{u}(x,t);u^*(x,t),x,t\right).$$

We used this approach to a particular case; the construction of further applications will be a topic for further research. Later on, we investigate the differential problem for the different particular cases of set-valued mapping F. Let us consider the "polyhedral" problem (1)-(3), where a polyhedral set-valued mapping F is defined as follows $F(w)=\{v:Pw-Qv\le d\}$. Here P,Q are $m\times n$ dimensional matrices, d is a m-dimensional column-vector. It follows from the inclusion (2) that

$$Pu(x,t)-Q\left(\frac{\partial^2 u(x,t)}{\partial x^2}-\frac{\partial u(x,t)}{\partial t}\right)\le d,\ (x,t)\in D. \tag{4}$$

For the polyhedral set-valued mapping, we need to calculate the LAM F^*. On the definition of cone of tangent directions $K_{gphF}(\tilde{w},\tilde{v})=\{(\bar{w},\bar{v}):(\tilde{w},\tilde{v})+\gamma(\bar{w},\bar{v})\in gphF$ for sufficiently small $\gamma>0\}$.

For a point $(\tilde{w},\tilde{v}) \in gphF$, we put $I(\tilde{w},\tilde{v}) = \{i : P^i\tilde{w} - Q^i\tilde{v} = d_i, i = 1,\ldots,m\}$ where P^i, Q^i be the i-th row of the matrices P, Q respectively and d_i be the i-th component of the vector d. By using Farkas Theorem 1.13 Mahmudov (2011), we have $(w^*, v^*) \in K_{gphF}(\tilde{w},\tilde{v})$ if and only if $w^* = -\sum_{i\in I(\tilde{w},\tilde{v})} P^{i*}\lambda_i$, $v^* = -\sum_{i\in I(\tilde{w},\tilde{v})} Q^{i*}\lambda_i$, $\lambda_i > 0$, where P^{i*}, Q^{i*} are transposed vectors of P^i, Q^i respectively. Now setting $\lambda_i = 0$, $i \notin I(\tilde{w},\tilde{v})$ and denoting by λ the vector column with components λ_i, we obtain that the polyhedral dual cone is calculated by the formula

$$K^*_{gphF}(\tilde{w},\tilde{v}) = \{(w^*, v^*) : w^* = -P^*\lambda, v^* = Q^*\lambda, \langle P\tilde{w} - Q\tilde{v} - d, \lambda\rangle = 0,\ \lambda \geq 0\}. \tag{5}$$

Moreover, for the polyhedral LAM from the definition, we have

$$F^*(v^*; (\tilde{w},\tilde{v})) = \{-P^*\lambda :, v^* = -Q^*\lambda, \langle P\tilde{w} - Q\tilde{v} - d, \lambda\rangle = 0,\ \lambda \geq 0\}. \tag{6}$$

By Theorem 3.1 and using the formula (6), we have

$$\frac{\partial^2 u^*(x,t)}{\partial x^2} + \frac{\partial u^*(x,t)}{\partial t} + P^*\lambda(x,t) \in -\partial f(\tilde{u}(x,t), x, t),\ u^*(x,t) = -Q^*\lambda(x,t), (x,t) \in D,$$

$$\langle P\tilde{u}(x,t) - Q\left(\frac{\partial^2\tilde{u}(x,t)}{\partial x^2} - \frac{\partial\tilde{u}(x,t)}{\partial t}\right) - d, \lambda(x,t)\rangle = 0,\ \lambda(x,t) \geq 0. \tag{7}$$

Theorem 3.2 Suppose that $f(\cdot,x,t)$ is continuous convex proper function and $F(\cdot,x,t) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a polyhedral set-valued mapping. Then for the optimality of the solution $\tilde{u}(x,t)$ in the polyhedral problem it is sufficient that there exists $\lambda(x,t)$ such that the conditions $(i)-(iii)$ hold almost everywhere on D:

(i) $\frac{\partial^2 Q^*\lambda(x,t)}{\partial x^2} + \frac{\partial Q^*\lambda(x,t)}{\partial t} - P^*\lambda(x,t) \in \partial f(\tilde{u}(x,t), x, t)$,

(ii) $\lambda(1,t) = 0$, $\lambda(x,1) = 0$, $\lambda(0,t) = 0$,

(iii) $\langle P\tilde{u}(x,t) - Q\left(\frac{\partial^2\tilde{u}(x,t)}{\partial x^2} - \frac{\partial\tilde{u}(x,t)}{\partial t}\right) - d, \lambda(x,t)\rangle = 0,\ \lambda(x,t) \geq 0.$

Example 3.1 We consider some linear optimal control problems to demonstrate the above approach

minimize $I[u(\cdot,\cdot)] = \iint_D f(u(x,t), x, t)\,dxdt$

$$\frac{\partial^2 u(x,t)}{\partial x^2} - \frac{\partial u(x,t)}{\partial t} = Au(x,t) + Bw(x,t),\ w(x,t) \in U,\ (x,t) \in D, \tag{8}$$

$u(0,t) = \alpha_0(t),\ u(x,0) = \beta_0(x), u(1,t) = \beta_1(t),$ where A and B are $n\times n$ and $n\times r$ matrices,

respectively, $U \subset \mathbb{R}^r$ is a convex closed set, and f is continuously differentiable function of u. It is required to find the controlling parameter $\tilde{w}(x,t) \in U,\ (x,t) \in D$ such that the corresponding solution $\tilde{u}(x,t)$ minimizes $I[u(\cdot,\cdot)]$. In this case, $F(u) = Au + BU$. By elementary computations for $v = Au + Bw$, we find that

$$F^*(v^*,(u,v)) = \begin{cases} A^*v^*, & -B^*v^* \in \left[\text{cone}(U-w)\right]^*, \\ \varnothing, & -B^*v^* \notin \left[\text{cone}(U-w)\right]^*, \end{cases} \tag{9}$$

Then, using the formula (9), we have $\dfrac{\partial^2 u^*(x,t)}{\partial x^2} + \dfrac{\partial u^*(x,t)}{\partial t} = A^*u^*(x,t) - f'\left(\tilde{u}(x,t),x,t\right)$, and taking into account Theorem 3.1, we obtain that $\left\langle w - \tilde{w}(x,t), B^*u^*(x,t)\right\rangle \le 0, \quad w \in U$, which implies a new boundary condition and the Pontryagin maximum principle $\left\langle B\tilde{w}(x,t), u^*(x,t)\right\rangle = \sup\limits_{w\in U}\left\langle Bw, u^*(x,t)\right\rangle$. Thus, we have obtained the following result.

Theorem 3.3 The solution $\tilde{u}(x,t)$ corresponding to the control $\tilde{w}(x,t)$ minimizes $I(u)$ in the Example 3.1 if there exists a function $u^*(x,t)$ satisfying the relations (d)-(f):

(d) $\dfrac{\partial^2 u^*(x,t)}{\partial x^2} + \dfrac{\partial u^*(x,t)}{\partial t} = A^*u^*(x,t) - f'\left(\tilde{u}(x,t),x,t\right)$,

(e) $u^*(1,t) = u^*(x,1) = u^*(0,t) = 0$,

(f) $\left\langle B\tilde{w}(x,t), u^*(x,t)\right\rangle = \sup\limits_{w\in U}\left\langle Bw, u^*(x,t)\right\rangle$.

5. Conclusion

In this paper, we formulate the conditions of optimality for the boundary value problems posed by parabolic partial differential inclusions. We use convex and non-smooth analysis structures to provide necessary and sufficient conditions for optimality in terms of Euler-Lagrange inclusions, as well as transversality conditions based on approximation of some subtle calculations. Furthermore, we demonstrate the basic theory's results by giving an example, the so-called linear optimal control problem.

References

Cheng, Y., Cong, F., & Xue, X. (2011). Boundary value problems of a class of nonlinear partial differential inclusions. Nonlinear Analysis: Real World Applications, 12(6), 3095-3102. https://www.sciencedirect.com/science/article/abs/pii/S1468121811001040

Demir Sağlam, S. (2021). The Optimality Principle For Second-order Discrete and Discrete-Approximate Inclusions. An International J. of Optim. and Control: Theories & Appl. 11(2), 206-215. https://doi.org/10.11121/ijocta.01.2021.001056

Demir Sağlam, S., & Mahmudov, E.N. (2020). Optimality Conditions for Higher Order Polyhedral Discrete and Differential Inclusions. Filomat, 34(13), 4533-4553. https://www.pmf.ni.ac.rs/filomat-content/2020/34-13/34-13-24-12764.pdf

Demir Sağlam, S., & Mahmudov, E.N. (2021a). Duality problems with second-order polyhedral discrete and differential inclusions, Bull. Iran. Math. Soc., Doi:10.1007/s41980-021-00531-9. https://link.springer.com/content/pdf/10.1007/s41980-021-00531-9.pdf

Demir Sağlam, S., & Mahmudov, E.N. (2021b). Polyhedral Optimization of Second-Order Discrete and Differential Inclusions with Delay. Turkish J. of Math., 45(1), 244-263. https://journals.tubitak.gov.tr/math/issues/mat-21-45-1/mat-45-1-16-2005-50.pdf

Demir Sağlam, S., & Mahmudov, E.N. (2021c). The Lagrange Problem for Differential Inclusions with Boundary Value Conditions and Duality, Pacific J. of Optim., 17(2), 209-225. http://www.yokohamapublishers.jp/online2/pjov17-2.html

Demir Sağlam, S., & Mahmudov, E.N. (2021d). Convex optimization of nonlinear inequality with higher order derivatives, Applicable Analysis, DOI: 10.1080/00036811.2021.1988578 https://www.tandfonline.com/doi/abs/10.1080/00036811.2021.1988578

Mahmudov, E.N. (2011). Approximation and Optimization of Discrete and Differential Inclusions. Elsevier, Boston, USA. https://www.sciencedirect.com/book/9780123884282/approximation-and-optimization-of-discrete-and-differential-inclusions

Mahmudov, E.N. (2008). Sufficient conditions for optimality for differential inclusions of parabolic type and duality. J. Global Optim., 41(1), 31-42. https://dl.acm.org/doi/10.1007/s10898-007-9164-y

Mahmudov, E.N. (2009). Optimal control of Cauchy problem for first-order discrete and partial differential inclusions. J. Dynam. Contr. Syst., 15(4), 587-610. https://dl.acm.org/doi/10.1007/s10883-009-9073-0

Mahmudov, E.N. (2021a). Optimal control of first order partial differential inclusions in bounded region. International J. of Control, DOI:10.1080/00207179.2021.1886328 https://www.tandfonline.com/doi/abs/10.1080/00207179.2021.1886328

Mahmudov, E.N. (2021b). Optimization of Higher-Order Differential Inclusions with Endpoint Constraints and Duality. Adv. Math. Models Appl., 6(1), 5-21. http://jomardpublishing.com/UploadFiles/Files/journals/AMMAV1N1/v6n1/MahmudovE.pdf

Mordukhovich, B.S. (2006). Variational Analysis and Generalized Differentiation, Vols.I and II. Springer, Springer-Verlag Berlin Heidelberg. https://www.springer.com/gp/book/9783540254379

Yusubov, S.S., & Mahmudov, E.N. (2021c). Optimality conditions of singular controls for systems with Caputo fractional derivatives. J. of Industrial & Management Optim., DOI:10.3934/jimo.2021182 https://www.aimsciences.org/article/doi/10.3934/jimo.2021182

Ocean Energy Conversion Analysis by Compressive Sensing

Ali Rıza Alan[1*], Cihan Bayındır[1,2]

[1]*İstanbul Technical University, Civil Engineering Department, İstanbul, Turkey*
[2]*Boğaziçi University, Civil Engineering Department, İstanbul, Turkey*
alan21@itu.edu.tr, cbayindir@itu.edu.tr

Abstract

Considering that many types of energy resources in the world are limited, it is inevitable to focus on renewable energy sources to meet the needs of the world's increasing population. Ocean energy is undoubtedly one of the most efficient resources among renewable energy sources in terms of its potential. There are many studies in the literature that have been conducted about energy conversion and which are guiding for further studies (Clément et al., 2002). Ocean wave energy conversion modeling methods are principally based on the basic wave parameters with some additional concepts. Although some concepts have been developed so far, with the advancing technology, scientists have sought superior solutions. The methods to be used to analyze ocean wave energy are as significant as obtaining the ocean energy itself, especially in the big data era for electricity generation and grid connection purposes. The compressive sensing (CS) technique, which outperforms the classical techniques since it uses a smaller number of samples (Candès et al., 2006a, Candès, 2006b), is one of the algorithms that can be used for such purposes and found come applications in coastal and ocean engineering (Bayındır, 2016, Bayındır, 2019, Bayındır et al., 2021). In this paper, we examine the utilization of the CS for the efficient analysis and assessment of ocean wave energy and ocean energy conversion in general. Constructing a time series of the wave power following Goda (2010), Bayındır (2009) and Saulnier et al. (2009), the application of the CS proves to be an advantageous tool for the measurement, analysis, and assessment of wave energy and ocean energy conversion. We discuss our findings and comment on their possible usage and applications.

Keywords: Ocean energy, wave energy, energy converter, compressive sensing

1. Introduction

The world population is increasing day by day and the global energy need is growing exponentially due to the needs of this increasing population (Clément et al., 2002). Countries that have followed a fossil fuel-based energy consumption policy for years because they are easy and sometimes cheap, despite being harmful to the environment, have started to search for new types of energy since such fuels are not sustainable and new energy needs occur. The issue of energy generation from waves is not a new issue, as it was seen in some patents towards the end of the 18th century (Clément et al., 2002). However, modern studies on this subject started to gain importance with the oil crisis in the 70s, with people turning to renewable energy sources (Drew et al., 2009). Studies on this subject leaped forward recently due to global climate change and increasing gases like carbon dioxide. Offering the highest energy density among renewable energy sources, having limited negative features compared to other sources during use, being compatible with the seasonal variation of wave energy and seasonal variation of electricity demand, and being able to travel long distances with very low energy loss are the biggest advantages of wave energy (Clément et al., 2002, Drew et al., 2009). Since it is the largest and most

* *Corresponding Author*

efficient area where wave energy can be obtained, the oceans have naturally been the focus of studies on this subject. The literature on this subject is vast and the reader is referred to Clément et al. (2002) and the references therein for a more comprehensive discussion of the subject.

2. Problem Definition

As can be understood from all these studies, the issue of wave energy conversion in the ocean is a must. Effective and fast energy conversion and their measurement and analysis at least as important as the energy conversion itself. In this paper, we investigate the possible usage of the compressive sensing (CS) technique for the analysis and measurement of the time series of wave energy and power. Although there are some studies on CS in various branches of coastal and ocean engineering (Bayındır, 2016, Bayındır, 2019, Bayındır et al., 2021), to our best knowledge, there is no direct work on wave energy conversion measurement or analysis. We discuss the implementation of our approach, its possible application areas, benefits, and usage in the coming sections of this paper.

3. Proposed Method

3.1. Generation of a Random Sea State and Harvester Dynamics

In order to simulate a realistic ocean wavefield and to simulate the ocean wave energy converter dynamics in such a field, we start with the Bretschneider-Mitsuyasu spectrum (Goda, 2010). This spectrum can be given as

$$S(f) = 0.257 H_s^2 T_s^{-4} f^{-5} \exp[-1.03(T_s f)^{-4}] \quad (1)$$

where f denotes the frequency, H_s is the significant wave height, T_s is the significant wave period. The frequency spectrum given in Eq. (1) can be converted into angular frequency spectrum by

$$S(\omega) = S(f)\frac{df}{d\omega} \quad (2)$$

where the Jacobian term is given as $df/d\omega = 1/(2\pi)$. In order the create a realistic wave field in the spatial domain, the wavenumber spectrum needs to be obtained. Using the equality of energies in the angular frequency spectrum and wavenumber spectrum, the wavenumber spectrum can be obtained via

$$S_k(k) = S(\omega)\frac{d\omega}{dk} \quad (3)$$

where S_k is the wavenumber spectrum (Bayındır, 2009). The parameters ω and k are related by the dispersion relationship

$$\omega^2 = gk\tanh(kh). \quad (4)$$

Using the dispersion relationship given by Eq. (4), the expression for group velocity can be derived as

$$\frac{d\omega}{dk} = C_g = \frac{1}{2}\left(1 + \frac{2kh}{\sinh(2kh)}\right)\frac{\omega}{k}. \quad (5)$$

Therefore, using Eqs. (1)-(5), the wavenumber spectrum can be obtained (Bayındır, 2009). After the calculation of the wavenumber spectrum, the nodal amplitudes for each dk interval can be obtained by energy equality

$$\frac{1}{2}a_r^2 = S_k(k_r)dk \quad (6)$$

where the index is $r = 0,1,\ldots,\frac{N}{2}$ and the parameters are $k_r = rdk$, $dk = 2\pi/L$ where L denotes the periodic domain length. The parameter N is the total number of spectral components and is chosen to be a power of 2 in order to use FFT routines efficiently (Bayındır, 2009). Two-sided amplitude spectrum can be constructed from the one-sided amplitude spectrum using the symmetry relation

$$a_s = a_{N-s} \quad (7)$$

where the index is $s = \frac{N}{2} + 1, \frac{N}{2} + 2, \ldots, N - 1$. In order to ensure the randomness of the wavefield, we generate uniformly distributed random phases Θ_r with values in the interval of $[0,2\pi]$ for $r = 0,1,\ldots,\frac{N}{2}$. Using these random phases and amplitudes, the complex amplitudes, A, are computed by

$$A_j = \frac{a_j \exp(i\Theta_j)}{2} \quad (8)$$

where the index is $j = -\frac{N}{2}+1, -\frac{N}{2}+2, \ldots, \frac{N}{2}$, the imaginary unity is $i = \sqrt{-1}$ and the parameter Θ_j shows the symmetric uniformly distributed random phase shifts in the interval of $[0,2\pi]$. Then, based on these complex amplitudes, initial water surface elevation can be obtained by

$$\eta_p = \sum_{j=-N/2+1}^{N/2} A_j \, exp(ik_j x_p) \quad (9)$$

where the index is $p = 0,1,\ldots, N-1$, and the discrete locations are given by $x_p = pdx$, and the other parameters are as before. Using the complex amplitudes, A, the initial water surface fluctuations are computed by using Eq. (11) where efficient IFFT routines are employed. Similar to the water surface fluctuation, η, the initial velocity potential on the sea surface, ϕ^s, can be constructed in a similar way with appropriate scaling terms (Bayındır, 2009). After the construction of the initial wave profile this way, the temporal evolution of the wavefield is modeled using the kinematic and dynamic boundary conditions given as

$$\eta_t - \phi_z^s = 0 \quad (10)$$

and

$$\phi_t^s + g\eta = 0. \quad (11)$$

These equations are solved using a 4th order Runge-Kutta time-stepping algorithm. For the details of the wavefield simulations with linear and nonlinear effects, the reader is referred to Bayındır (2009). In order to model the dynamics and construct the time series of heave velocity and converted power of a point absorber axisymmetric buoy type power take-off (PTO) device, the transfer function analysis is used. The heave velocity spectrum, $S_{\dot{z}\dot{z}}$, is then calculated from the Bretschneider-Mitsuyasu frequency spectrum (renamed as $S_{\eta\eta}(f)$) by the transfer function method as

$$S_{\dot{z}\dot{z}}(f) = S_{\eta\eta}(f)\left|H_{\dot{z}\eta}\right|^2(f) \quad (12)$$

where, $H_{\dot{z}\eta}$, is the complex, linear, frequency domain transfer function which relates the water surface fluctuation, η , to the heave velocity, $\dot{z}$, of the buoy type power take-off device. This transfer function is a device-specific function and its selection is discussed in the coming sections of this paper. The instantaneous power converted by a PTO device can be calculated as

$$P_{PTO}(t) = C_{PTO}\dot{z}^2(t) \quad (13)$$

where the parameter C_{PTO} is selected to be $C_{PTO} = 2x10^5\ kg/s$ following Mérigaud & Ringwood (2017). The total power converted over a duration of time can be calculated by

$$\hat{P}_{PTO} = C_{PTO} \sum_{n=1}^{M_w} \dot{z}_n^2 \quad (14)$$

where M_w is the number of waves in that duration.

3.2. Review of the Compressive Sensing

CS has turned out to be a revolutionary algorithm in signal processing, applied mathematics, and engineering. We try to very briefly summarize CS in this section. A *K*-sparse signal *P* is a signal which has only *K* nonzero element among its all *N* elements. Transforming *P* into an orthogonal domain such as Fourier domains with using the orthogonal transformation matrix ψ, the signal can be represented by $P = \psi\hat{P}$. Here, $\hat{P}$ shows the coefficient vector. After eliminating the zeros of the signal, its sparse representation can be given as $P_s = \psi\hat{P}_s$. Here, the parameter P_s is identified as the signal with non-zero components. CS algorithm states that the *K*-sparse signal *P* with *N* entries can be exactly regenerated using $M \geq C\mu^2(\phi,\psi)K\log(N)$ measurements. Here, *C* denotes a positive constant, ϕ symbolizes the sensing basis, ψ symbolizes the transformation basis, and $\mu^2(\phi,\psi)$ is the mutual coherence between these two (Candès et al., 2006, Candès, 2006). After the sampling using the sensing matrix ϕ and *M*

random projections the constraint of the CS optimization problem becomes $g = \phi\psi\hat{P}$. Thus, the CS problem can be summarized as

$$\min\left\|\hat{P}\right\|_{l_1} \text{ under constraint } g = \phi\psi\hat{P} \quad (15)$$

where $\left\|\hat{P}\right\|_{l_1} = \Sigma_i\left|\hat{P}_i\right|$. Among all possible solutions to this problem, its l_1 solution can be obtained via l_1 optimization technique and $P_{cs} = \psi\hat{P}$. For a more comprehensive analysis of the CS, the reader is referred to Candès et al. (2006a) and Candès (2006b), and for some applications of CS in coastal and ocean engineering, the reader is referred to Bayındır (2019) and Bayındır et al. (2021).

4. Results and Discussion

In Figure 1, we depict the Bretschneider-Mitsuyasu spectrum used for the simulation of the ocean waves. The parameters of this spectrum given in Eq. (1) is selected to be H_s=*2m* and T_s=*10s*. The linear sea state corresponding to a random realization of the spectrum depicted in Figure 1 obtained by using the Eqs. (1)-(11) is illustrated in Figure 2 after an evolution time of *t=100s.* The domain of computations for this simulation is selected as *L=1000m*.

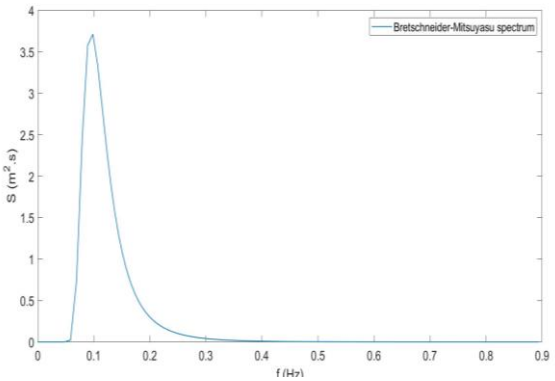


Figure 1. Bretschneider-Mitsuyasu spectrum.

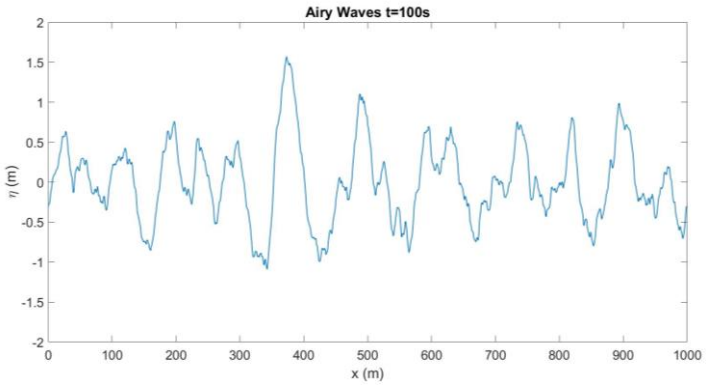


Figure 2. Simulation of the spatial variation of a linear sea.

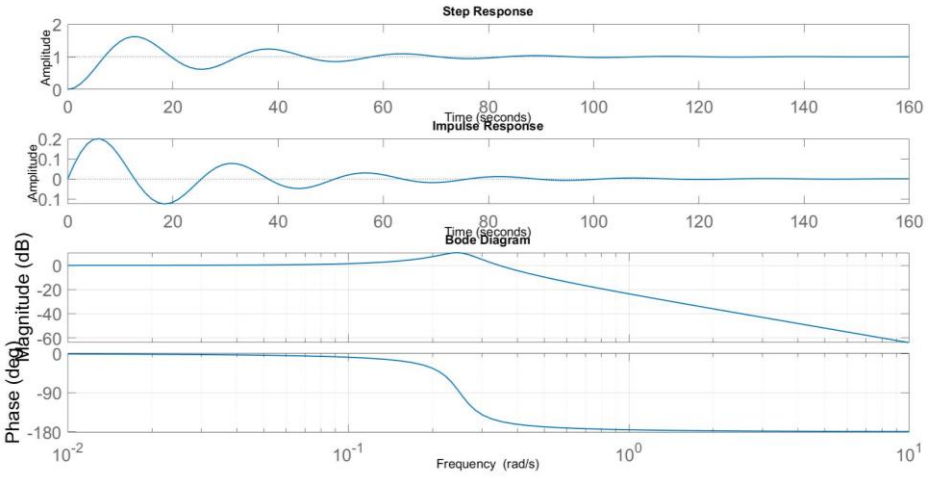


Figure 3. The transfer function $H_{\dot{z}\eta}$ a) Step response in time domain b) impulse response in time domain c) Bode diagram of frequency response in spectral domain

As discussed in the preceding section, the transfer function $H_{\dot{z}\eta}$ is used to model the response of the heave velocity, $\dot{z}$, of the buoy type PTO ocean energy converter to the water surface fluctuation, η. For this purpose, a second-order transfer function with a damping ratio of *0.15* and resonance frequency of

0.25 rad/s is considered. The step response, impulse response, and the frequency response in the form of a Bode diagram of the transfer function $H_{\dot{z}\eta}$ with these parameters are depicted in Figure 3. The buoy heave velocity time series, $\dot{z}(t)$, corresponding to a realization of the water surface fluctuation at buoy location, $\eta(t)$, is depicted in Figure 4.

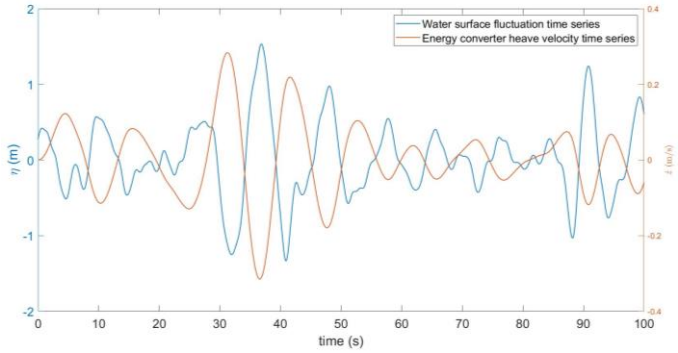


Figure 4. A time-series of the water surface fluctuation, η, and a time series of the energy converter heave velocity, $\dot{z}$

The power converted by the buoy type PTO converter is calculated from the heave velocity time series, $\dot{z}(t)$, using Eq. (13). The resulting converted power time series is depicted in Figure 5a, for a time span of *t=0-100s.*

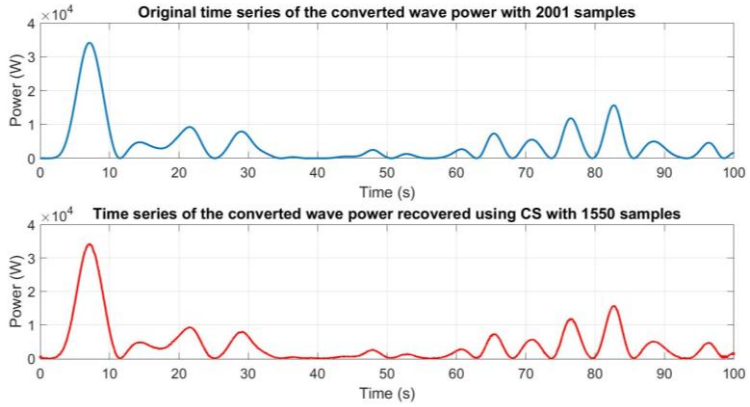


Figure 5. A time-series of the converted instantaneous wave power a) obtained by classical sampling with 2001 samples b) obtained by compressive sampling with 1550 samples

As one can realize by checking Figure 5, the CS can be effectively used to the reconstruction of the original time-series using a fewer number of components. The power time-series depicted in Figure 5a has a sparse representation in the frequency domain and thus random sampling is performed in the time domain. The exact construction of the original time-series with *N=2001* samples is achieved using *M=1550* compressive samples. This undersampling ratio brings a significant advantage for the measurement, observation, analysis, and interpolation/extrapolation of the environment of the ocean wave converters and the harvesting device parameters, statistics, and health.

5. Conclusion

In this paper, we have investigated the possible usage of compressive sampling for the measurement, observation, analysis, and interpolation/extrapolation in the ocean wave energy conversion technology. For this purpose, we modeled the time-series dynamics of a buoy type power take-off converter in a random sea environment and modeled the converted power by such a device. Then, we showed that such

a power time series can be efficiently measured and analyzed using compressive sensing. Our findings will lead to many effective measurement and analysis devices for ocean energy harvesting applications.

Acknowledgements

This work was supported by the Research Fund of the İstanbul Technical University. Project Code: MGA-2020-42544.

References

Bayındır, C. (2009). Implementation of a computational model for random directional seas and underwater acoustics, M.S. Thesis, University of Delaware. http://udspace.udel.edu/handle/19716/5425

Bayındır, C. (2016). Compressive spectral method for the simulation of nonlinear gravity waves, Scientific Reports, 22100. https://doi.org/10.1038/srep22100

Bayındır, C. (2019). Early detection of rogue waves using compressive sampling. TWMS Journal of Applied and Engineering Mathematics, 9 (2), 198-205. http://jaem.isikun.edu.tr/web/images/articles/vol.9.no.2/03.pdf

Bayındır, C., & Namlı, B. (2021). Efficient sensing of the von Karman vortices using compressive sensing, Computers & Fluids, 104975. https://doi.org/10.1016/j.compfluid.2021.104975

Candès, E. J., Romberg, J., & Tao, T. (2006a). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory, 52(2), 489-509. https://doi.org/10.1109/TIT.2005.862083

Candès, E. J. (2006b). Compressive sampling. Proceedings of the international congress of mathematicians, 3, 1433-1452. https://doi.org/10.4171/022-3/69

Clément, A., McCullen, P., Falcão, A., Fiorentino, A., Gardner, F., et.al. (2002). Wave energy in Europe: current status and perspectives. Renewable and sustainable energy reviews, 6(5), 405-431. https://doi.org/10.1016/S1364-0321(02)00009-6

Drew, B., Plummer, A. R., & Sahinkaya, M. N. (2009). A review of wave energy converter technology. Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, 223(8), 887-902. https://doi.org/10.1243/09576509JPE782

Goda, Y. (2010). Random seas and design of maritime structures. World Scientific Publishing Company. https://doi.org/10.1142/7425

Mérigaud, A., & Ringwood, J. V. (2017). Free-surface time-series generation for wave energy applications. IEEE Journal of Oceanic Engineering, 43 (1), 19-35. https://doi.org/10.1109/JOE.2017.2691199

Saulnier, J.-B., Ricci, P., Clément, A. H. & Falcão, A. F. de O. (2009). Mean power output estimation of WECs in simulated sea, Proceedings of the 8th European Wave and Tidal Energy Conference, Uppsala, Sweden. http://www.homepages.ed.ac.uk/shs/Wave%20Energy/EWTEC%202009/EWTEC%202009%20(D)/papers/165.pdf

A Bidirectional Generalized Synchronization of Nonlinear Advection-Diffusion-Reaction Processes

Shko Ali Tahir[1*], Murat Sarı[2]

[1]*University of Sulaimani, Department of Mathematics, Sulaimani, Iraq*
[2]*Yildiz Technical University, Department of Mathematics, Istanbul, Turkey*

shko.tahir@univsul.edu.iq, sarim@yildiz.edu.tr

Abstract

This research focuses on the generalized synchronization (GS) of two dependent chaotic nonlinear advection-diffusion-reaction (ADR) processes with source terms. It is based on the bidirectionally coupled driver-response concept. The approach combines the backward differentiation formula-Spline (BDFS) scheme with the Lyapunov direct method without any transformation. The sufficient condition for nonlinear coupled ADR problems on the GS is given at very low diffusion effect, without detracting from natural features. The effectiveness of bidirectional coupling for monitoring the synchronized motions of two identical ADR problems is presented for various coupling strength regions. Next, a simulation example based on the ADR equations is presented to demonstrate the proposed approach, providing both synchronization and observer conception.

Keywords: Nonlinear ADR equations, generalized synchronization, BDFS method, Lyapunov direct method

1. Introduction

The nonlinear coupled ADR system with source functions has received much attention by characterizing the reaction and diffusion or the interaction between advection and diffusion mechanisms. Researchers in this field have taken into consideration problems in various disciplines such as fluid dynamics, financial mathematics, turbulence, traffic flow, shock waves and gas dynamics. Next, we can address the coupled two identical chaotic ADR problems, one with the source function presented as the driver and the other as the response configuration, as follows:

$$\begin{cases} u_{1_t}(x,t) + \nabla.\big(u_1(x,t),u_2(x,t)\big) + \nabla\cdot\Big(-\gamma\,\nabla\Big(\big(u_1(x,t),u_2(x,t)\big)\Big)\Big) + \mathcal{F}_1(x,t) = \hbar u_{2k}(x,t)\ldots driver \\ u_{2_t}(x,t) + \nabla\cdot\big(u_1(x,t),u_2(x,t)\big) + \nabla.\Big(-\gamma\,\nabla\Big(\big(u_1(x,t),u_2(x,t)\big)\Big)\Big) + \mathcal{F}_2(x,t) = \hbar u_{1k}(x,t)\ldots response. \end{cases} \tag{1}$$

Here $(x,t) \in [a,b] \times [t_0,T]$; γ defines the diffusion coefficient; $\mathcal{F}_1$ and $\mathcal{F}_2$ are the forcing terms. u_{1k} and u_{2k} provide control functions. $\hbar$ is the coupling strength. The initial and boundary conditions are given by

$$u_i(x,t_0) = u_i(x)\ , u_i(a,t) = g_i(t),\ u_i(b,t) = g_i(t), i = 1{,}2 \tag{2}$$

where both boundary and initial functions are known. In the last decade, researchers have studied the problem of chaotic phenomena for successful applications in different dynamical systems of various origins. In particular, their behaviour is characterized by instability over time and limited predictability. One of the first studies was carried out by Huygens (1973) on the coupled pendulum. A recent research

* *Corresponding Author*

shows that the surprising synchronization phenomena generated between coupled chaotic processes (Pecora and Carroll, 1990; Pecora et al., 1997). Further research on the synchronization of coupled chaotic systems is still crucial because of its potential applications in various fields (Nguyen and Hong, 2011; Khan and Singh, 2015; Tahir et al., 2018). In the case of synchronization of driver-response configurations, the designed controller functions create trajectories of the state variables of driver-response systems to monitor the synchronized motions. For two identical chaotic problems, synchronization can be performed in two categories: unidirectional coupling synchronization and bidirectional coupling synchronization (Volos et al., 2015; Cuimei et al., 2019). Unidirectional coupling synchronization was defined as a mechanism consisting of a drive system and a response system. The two systems are synchronized by presenting the state of the driver system on the response system (Tahir et al., 2017). However, bidirectional coupling synchronization, also called mutual coupling synchronization, can be obtained by introducing the state of one system to that of another, where each system can be considered a drive system or a response system.

Chaos, due to many of its characteristics, can be associated with the properties of coupled nonlinear ADR problems with low viscosity values. Various studies have been conducted to understand the close relationship between ADR mechanisms and chaos (Brummitt et al., 2009, Garcia et al., 2009, Kawamura et al., 2017, Sari and Tahir, 2021). The close relationship between chaos and nonlinear ADR processes has been highlighted in the related studies, and these processes are represented by highly nonlinear equations as they involve the interaction between ADR mechanisms with free parameters. Therefore, a specialized technique is still needed to cope with such problems while preserving the physical characteristics of the coupled ADR processes in nature (Jima et al., 2019, Sari et al., 2019, Singh et al., 2019 and Tahir et al., 2021). Due to the state of the coupling signal, the GS can be classified as bidirectional and unidirectional (Yongguang and Suochun, 2004; Yu and Zhang, 2004; Kanako and Toshiki, 2019). In this work, we present new synchronization numerical schemes by combining the BDFS scheme with the Lyapunov direct method based on the bidirectional coupled driver-response concept. This approach, passing through the BDFS scheme, which is effective in studying the coupled ADR equations, replaces the coupled nonlinear ADR equation by an ODE system that does not require linearization. The BDFS scheme is unconditionally stable and produces highly accurate solutions in space and time. The GS has been studied to design the controller function of coupled nonlinear ADR problems without any linearization by selecting the appropriate coupling parameters. Interestingly, this work focuses on more attractive bidirectional coupling synchronization so as not to linearize natural processes and thus cause loss of real features.

The paper is organized as follows. Section 2 examines the BDFS and Lyapunov direct methods for solving the dynamics and synchronization of bidirectional coupling of behaviours of nonlinear coupled ADR processes. Then, we provide numerical examples to demonstrate the effectiveness of the proposed method in Section 3. Concluding remarks and recommendations are presented in Section 4.

2. Proposed Method

In this section, the required solutions of the driver-response ADR chaotic problem (1)-(2) are approximated by the cubic interpolating splines:

$$S_i(x,t) = \sum_{j=-1}^{m+1} \alpha_{i,j}\, B_j(x),\ \ i = 1,2 \tag{3}$$

where $\alpha_{i,j}$ is time-dependent parameter. We denote the well-known B-spline functions with $B_j(x)$. By taking into account of the interpolating conditions at points, we have

$$S_i(x_j,t) = u_i(x_j,t),\ \ i = 1,2, j = 0,\dots,m$$

$$S_i''(a,t) = S_i''(b,t).$$

By substituting the interpolating conditions at boundary points $x_0 = a,\ x_m = b$ and satisfying the natural cubic splines, one finds

$\alpha_{i,-1}(t) = 2\alpha_{i,0}(t) - \alpha_{i,1}(t),$
$\alpha_{i,m+1}(t) = 2\alpha_{i,m}(t) - \alpha_{i,m-1}(t),$
$\alpha_{i,0}(t) = u_i(x_0, t) = g_i(t),\ i = 1,2,$
$\alpha_{i,m+1}(t) = u_i(x_m, t) = g_i(t),\ \ i = 1,2.$

The approximated cubic spline S_i must also satisfy the initial conditions at points $x_0, \dots, x_m$ and then at t_0
$S_i(x_0, t_0) = u_{i,0}(x_0),$
$S_i(x_0, t_0) = u_{i,0}(x_0),\ i = 1,2,\ j = 0, \dots, m,$
$S_i(x_m, t_0) = u_{i,0}(x_m).$

Then, by using the above relations, we have
$A_1 u_i(t_0) = u_{i,0}$

where $A_1 = \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}$ with the size $2(m-1) \times 2(m-1)$ and matrix A is an $(m-1) \times (m-1)$ tridiagonal matrix and $u_{i,0} = \left[u_{i,0}(x_1) - \frac{1}{6}u_{i,0}(x_0), u_{i,0}(x_2), \dots, u_{i,0}(x_{m-1}) - \frac{1}{6}u_{i,0}(x_m)\right]$.
Now, by virtue of above relations, we reach the following ODEs:

$$A_1 \frac{dv}{dt} = D_1 v(t) + \Im\big(v(t)\big) \qquad A_1 v(t_0) = v_0. \tag{4}$$

Here, matrix D_1 is a tridiagonal matrix with the size of $2(m-1) \times 2(m-1)$. We have used the BDF to solve the tridiagonal system encountered in (4) and employed the Newton method for the approximation of v_n at p −order with the time step $\Delta t = (T - t_0)/N$ with the knots $t_n = t_0 + n\Delta t$ for $n = 0, \dots, N$ given by
$A_1 v_n - \sigma h[D_1 v_n + k_i \Im(v_n)] - \sum_{j=0}^{p} \mu_j\, v_{n-j} = 0.$
The coefficients σ, μ are known. In order to observe the GS behaviours of the nonlinear coupled ADR problems with the property that every approximate solution converges to an exact solution at the low value of the viscosity coefficient based on the Lyapunov direct method. Again system (4) is given by
$\frac{dv}{dt} = A_1^{-1}\big[D_1 v(t) + \Im\big(v(t)\big)\big]$.

Due to the fact that A_1^{-1}, D_1 is considered to be a tridiagonal matrix defining regular systems and negative definite when the eigenvalues are all different and real parts of them are negative. We can study the coupling parameters for synchronizing two chaotic ADR problems with respect to the coupling strength at the optimal value. Furthermore, the stability leads to that $t > T - t_0$ and $\hbar > 0$. Then, the GS behaviour of the nonlinear coupled ADR problem consisting of controller function $u_{k,i},\ i = 1,2$, gives the following property

$$\lim_{t \to \infty} \|e(t)\| = \lim_{t \to \infty} \|\mathbb{V}(.,t_{2n}) - \mathbb{S}(.,t_{2n})\| = 0,$$

where, $\mathbb{V}(.,t_{2n}) = \begin{bmatrix} u_1(.,t_n) \\ u_2(.,t_n) \end{bmatrix}$ and $\mathbb{S}(x_{2n},.) = \begin{bmatrix} S_1(.,t_n) \\ S_2(.,t_n) \end{bmatrix}$. In the next section, we explore the GS behaviour between two identical nonlinear coupled ADR processes without losing the inherent features of nature and with reducing computational difficulties in finding numerical solutions at a low value of the viscosity coefficient through the BDFS method.

3. Result and Discussion

In this section, we apply the techniques outlined above to the synchronization problem of identically coupled chaotic ADR problems to demonstrate the effectiveness of the proposed generalized synchronization. The globally generalized synchronization with respect to control functions is also confirmed by the simulation result at various coupling strengths. The computational domain $[a, b]$ is discretized to equally spaced elements. All computations have been carried out using MATLAB 2020 on a workstation with 16 significant decimal digits. The error norm is given to explain the synchronization of the driver and the response equations:

$$\hbar \rightarrow \|E(u_i)\|_\infty = \|\mathbb{V} - S_i\|_\infty.$$

To illustrate the efficiency of the proposed numerical scheme, we consider the coupled nonlinear Burgers equation with source functions in the following form

$$\begin{cases} u_{1_t} - 0.00001\, u_{1_{xx}} + u_1\, u_{1_x} = \mathcal{F}_1(x,t) + \hbar u_{2k}(x,t) \ldots driver \\ u_{2_t} - 0.00001\, u_{2_{xx}} + u_2\, u_{2_x} = \mathcal{F}_2(x,t) + \hbar u_{1k}(x,t) \ \ldots response. \end{cases} \tag{5}$$

The initial, boundary and source functions are taken from the exact solutions. The exact solutions of the proposed problem are:

$u_{1_t}(x,t) = e^{-t}\left(x - \frac{x^2}{2}\right)$,

$u_{2_t}(x,t) = e^{-t}\, x\,(1 - x)$.

The control functions u_{1k} and u_{2k} can be determined by the Lyapunov direct method. Full synchronization of the proposed coupled model has been observed for $\hbar \geq 0.05$. As shown in the simulation, the GS behaviour in the proposed problem (5) is observed as the viscosity decreases. The synchronization is observed in Figure 1 when $\hbar$ becomes larger with relatively small viscosity values. It is clear that the synchronization errors of bidirectional coupling converge to zero in a much shorter time than that of the unidirectional coupling.

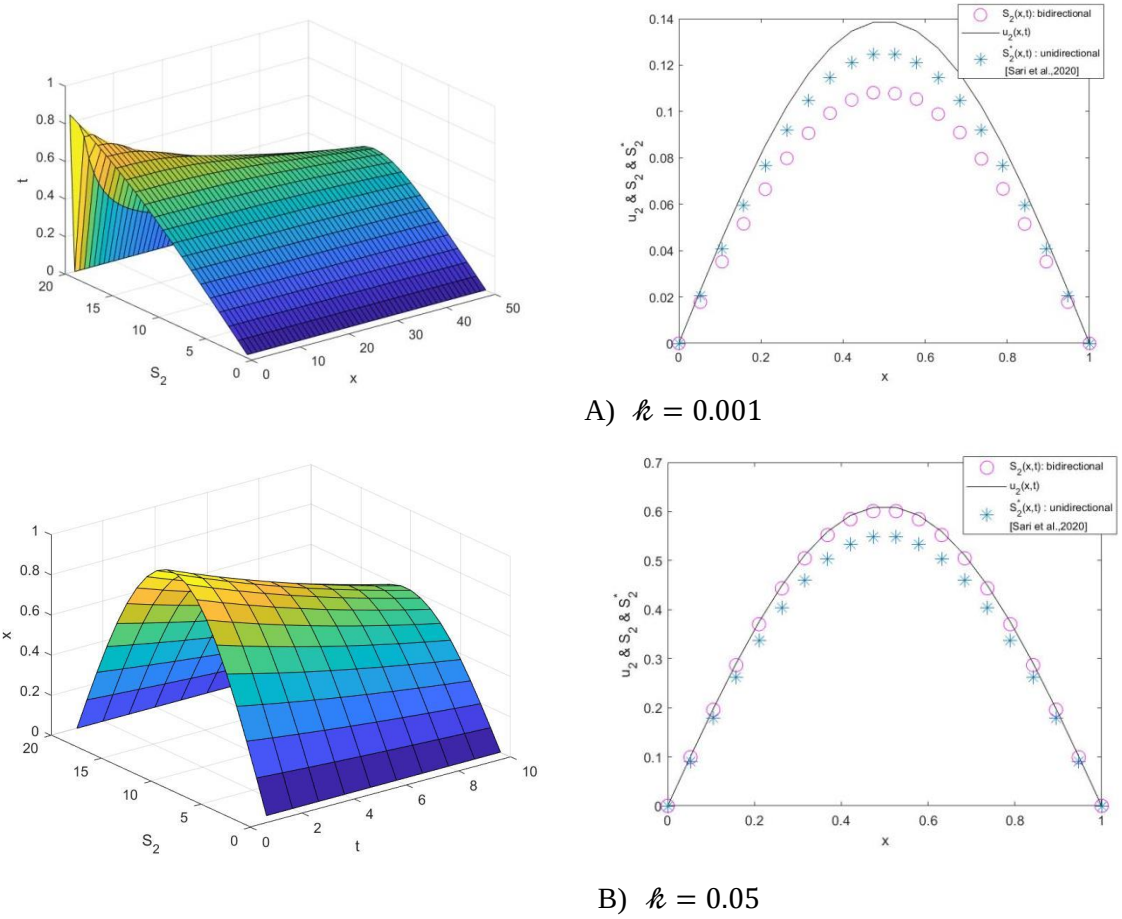


Figure 1. Synchronizational behaviour of the proposed problem at various coupling strengths $\hbar = 0.001, 0.05$.

4. Conclusions and Recommendation

In this research, the GS of spatial chaotic identical nonlinear ADR processes with the forcing term, which bidirectionally coupled, have been analysed. Based on the BDFS and Lyapunov direct method, generalized synchronization of two identical Burgers equations has been realized. From the numerical results produced, it has been notably concluded that the bidirectional coupling can be synchronized in a much shorter time than the unidirectional coupling at a very low viscosity value. For further research, synchronization theory remains a challenging problem of nonlinear behaviour in a network of the ADR processes in more realistic environments.

References

Brummitt C. D., & Sportt J. C. (2009). A search for the simplest chaotic partial differential equation, Physics Letters A, 373(1), 2717-1721. https://doi.org/10.1016/j.physleta.2009.05.050.

Cuimei J., Akbar Z., Senel T. M., & Tongxing L. (2019). Synchronization of bidirectional N-coupled fractional-order chaotic systems with ring connection based on antisymmetric structure, Advanced in Difference Equations, 6(3), 1-16. https://doi.org/10.1186/s13662-019-2380-1.

Garcia P., Acosta A., & Leiva H. (2009). Synchronization conditions for master-slave reaction diffusion systems, Europhysics Letters, 88(6), 1-4. https://doi:10.1209/0295-5075/88/60006.

Huygens C., (1973). Horoloquim oscilatoruim apud f.muquet,parisiis, The pendulum clock, Ames: Iowa State University Press, (English translation).

Jima M. K., & Shiferaw A. (2018). Numerical solution of the coupled Burgers equation using differential quadrature method based on Fourier expansion basis. Applied Mathematica, 9(2), 821-835. https://doi.org/10.4236/am.2018.97057.

Kanako S., Toshiki O. (2019). Synchronization of coupled nonlinear system with bidirectional sampled data coupling, IFAC-Papers online, 52(16), 934-939. https://doi.org/10.1016/j.ifacol.2019.12.033.

Kawamura Y., Shirasaka S., Yanagita T., & Nakao H. (2017). Optimizing mutual synchronization of rhythmic spatiotemporal patterns in reaction-diffusion systems, Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 96(1), 1-15. https://doi.org/10.1103/PhysRevE.96.012224.

Khan, A., & Singh P. (2015). Chaos synchronization in Lorenz system, Applied Mathematics, 6(1), 1864-1872. https://doi: 10.4236/am.2015.611164.

Nguyen L.H. and Hong K.S. (2011). Synchronization of coupled chaotic Fitz-Hugh-Nagumo neurons via Lyapunov functions, Mathematics and Computers in Simulation, 8(2), 590-603.

Pecora L., & Carroll T. (1990). Synchronization in chaotic systems, Physical Review Letters, 6(8), 821-824. https://doi.org/10.1103/PhysRevLett.64.821.

Pecora L., Carroll T., Johnson G., & Mar D. (1997). Fundamentals of synchronization in chaotic systems, concepts and applications, Chaos, 7(4), 520-543. https://doi.org/10.1063/1.166278.

Sari M., Tahir S.A., & Bouhamidi A. (2019). Behaviour of advection–diffusion–reaction processes with forcing terms. Carpathian Journal of Mathematics, 35(1), 233-252.

Sari M., & Tahir S.A. (2021). Synchronization of the nonlinear advection-diffusion-reaction processes, Mathematical Methods in the Applied Sciences, 44(15), 11970-11984. http://doi.org/10.1002/mma.6984.

Singh A, Das S., Ong S.H., & Jafari H. (2019). Numerical solution of nonlinear reaction–advection–diffusion equation, Journal of Computational and Nonlinear Dynamics, 14(1):41003. https://doi.org/10.1115/1.4042687.

Tahir S. A., Bouhamidi A., & Sari M. (2018). Designing a response approach in chaotic systems, Sigma Journal of Engineering and Natural Sciences, 36(1), 441-457.

Tahir S. A., Sari M. (2021). Simulations of nonlinear parabolic PDEs with forcing function without linearization, Mathematica Slovaca, 71(4), 1005-1018. https://doi.org/10.1515/ms-2021-0035.

Tahir S. A., Sari M., & Bouhamidi A. (2017). Generalized synchronization of identical and non-identical chaotic dynamical systems via master approaches, An International Journal of Optimization and Control: Theories & Applications, 7(3), 248-254. http://doi.org/10.11121/ijocta.01.2017.00509

Volos K., Kyprianidis M., Stouboulos N., MuozPacheco M., & Pham V. (2015) Synchronization of chaotic nonlinear circuits via a Memristor, Journal of Engineering Science and Technology Review, 8(2) 44-51. https://doi.org/10.1063/1.4917383.

Yongguang Y., Suochun Z. (2004). The synchronization of linearly bidirectional coupled chaotic systems, Chaos, Solitons and Fractals, 22(1), 189-197. https://doi.org/10.1016/j.chaos.2003.12.088.

Yu Y., & Zhang S. (2004). The synchronization of linearly bidirectional coupled chaotic systems, Chaos, Solitons and Fractals, 22(1), 189-97, 2004. https://doi.org/10.1016/j.chaos.2003.12.088.

An Application of Statistical Design and Analysis of Experiments for System Performance Evaluation

Levent Eriskin*, Mumtaz Karatas

Department of Industrial Engineering, National Defence University, Turkish Naval Academy
34640, Istanbul, Turkey
leriskin@dho.edu.tr, mkaratas@dho.edu.tr

Abstract

In this paper, we consider evaluating and comparing candidate systems for acquisition using statistical design and analysis of experiments techniques. Due to numerous decision variables inherent in acquisition decisions and the high technology of the candidate systems, enterprises need decision support for making these decisions. In this regard, utilizing quantitative methods is of utmost importance. To address this issue, statistical design and analysis of experiments field provides quantitative techniques for designing appropriate experiments and analyzing the experimental data. These techniques are particularly useful for enterprises that procure systems as Commercial-Off-The-Shelf (COTS). Even though procuring a COTS product has advantages for ensuring a shorter acquisition lead time, it inherently involves risks since design specifications declared by the provider may not reflect the real performance of the system under various operating conditions. Consequently, these risks oblige enterprises to conduct a rigorous evaluation. In order to show how these quantitative techniques can be applied to real-life problems, we present an application dealing with the acquisition of a naval gun system as COTS. The case study shows that the analysis provides adequate decision support for evaluating and comparing candidate system alternatives.

Keywords: Design of experiments, nested design, system performance evaluation, case study.

1. Introduction

Acquisition decisions are inherently complex for enterprises when the system of interest is a complex one involving high technology. Moreover, the high costs of such systems make Decision Makers (DMs) aspire for better value for money. These compelling factors and numerous decision variables make it impossible for the DM to reach a solution with a holistic judgement. To address these decision problems, the field of Operations Research (OR) offers various techniques for providing decision support to the DM. The OR field utilizes mathematical modeling, statistical analysis, and optimization techniques for partitioning the problem into manageable pieces and solving each of these with analytical techniques.

Analytical techniques heavily rely on quantitative data. In this regard, measuring the performance of candidate systems in quantitative terms has become an important step in system acquisition processes. In order to obtain these data, a thorough test and evaluation planning is required. A combined test and evaluation process can be defined as a process where a system is tested and related data are analyzed for comparing them against requirements or alternative system solutions (Eriskin and Gunal, 2019). Test and evaluation of systems can be partitioned into two main categories, namely, (*i*) Developmental Test and Evaluation (DTE) and (*ii*) Operational Test and Evaluation (OTE) (Eriskin and Gunal, 2019). DTE aims at promoting engineering processes where the developed system is tested and evaluated against predefined technical and performance specifications. OTE, on the other hand, focuses on testing and

* *Corresponding Author*

evaluating the system under real-life conditions where various factors including uncontrollable ones are considered. For evaluating and comparing the performance of system alternatives that are under consideration for the acquisition, OTE is conducted.

Designing an experiment statistically is very important because with such experimentation appropriate data can be collected and analyzed with relevant statistical methods. Hence, objective and meaningful conclusions can be drawn from the analysis. In general, there are two main pillars of experimentation: experimental design and statistical analysis (Montgomery, 2017). These pillars are closely related and should be planned together since experimental analysis dictates what kind of statistical analyses can be conducted. Being a branch of applied statistics, Design and Analysis of Experiments (DOE) provides techniques for designing the appropriate experiments and analyzing the experimental data. DOE allows for multiple factors, both controllable and uncontrollable, to be included in the design simultaneously and identifies important interactions among them. DOE mainly aims at quantifying and explaining the variation in the response (dependent) variable (i.e., performance measure in the system evaluation context) caused by different levels of experimental factors (independent variables).

There exist studies in the literature dealing with evaluation and selection of various systems, such as attack helicopter (Cheng et al., 1999), the surface to air missile (Lee et al., 2010; Wang et al., 2014), tactical missile system (Chen, 1996; Cheng, 1999; Chen and Shyu, 2006), and space system (Burk et al., 1997; Rayno et al., 1997; Parnell et al., 1998), however, only Wang et al. (2014) applies DOE for performance measurement of alternative systems. In this context, we consider the problem of selecting the best Naval Gun System (NGS) among several candidates by conducting an OTE. Being one of the decision-making problematics, selection or choice problem is frequently encountered in real-life applications and usually involves multiple criteria to be evaluated (Eriskin, 2021). Consequently, this study aims to provide decision support to the DMs for acquisition decisions. We utilize DOE techniques for the experimental design and statistical analysis.

The paper is organized as follows: In Section 2 we define the problem. Section 3 provides the details of the design of experiment and statistical analysis. Finally, we conclude in Section 4.

2. Problem Definition

We consider an illustrative NGS acquisition process as Commercial-Off-The-Shelf (COTS). COTS systems are already developed and produced systems that are available for acquisition. Even though procuring a COTS has advantages for ensuring a shorter acquisition lead time, it inherently involves risks. For instance, COTS are developed for the common user, and may not be compatible and interoperable with buyer's systems that are already in the inventory (Eriskin and Gunal, 2019). Moreover, design specifications declared by the provider may not reflect the real performance of the system under various operating conditions. Consequently, these risks oblige the buyers to conduct a rigorous evaluation.

A Navy announces that it will procure 10 modern NGSs as COTS for its newly produced frigates. An NGS comprises a naval gun and two (main and secondary) Fire Control Radars (FCRs). Three companies (i.e., Alfa, Bravo, Charlie) are competing in order to get the deal. The Navy stipulates that these companies participate in the OTE that will be conducted by the Navy itself. In this OTE, the Navy wants to evaluate the performance of the candidate systems in terms of precision of hitting surface targets. The performance measure in this OTE is the deviation of shells from the target in yards. The Navy assigns two ships for the OTE. The idea is to install portable versions of the candidate NGSs on board these ships and perform a live firing experiment with its crews. Hence, it will be possible to experiment these three systems with two different crews at sea under realistic conditions. A group of experts is gathered for conducting and analyzing the OTE.

3. DOE Procedure

In this section, we first explain the design of the experiment and then provide the details of the pertaining statistical analysis.

3.1. Design of Experiment

In order to design proper experimentation, system dynamics should be clearly understood. The naval guns are directed to surface targets with FCRs. There are two FCRs within each NGS. The FCR used in firing affects the performance of a NGS and both FCRs are used interchangeably in surface warfare. Field experts estimate that the mean range of the first contact from a surface target is 15,000 yards, hence, the firing range in the OTE should be this distance. Even though both ships participating in the OTE have well-trained personnel, it is doubted that their performance in naval gun firing may differ, hence, should be taken into account. The last factor that is believed to have an impact on the performance of a NGS is the sea state. Sea state causes pitch and roll, which affects the stability of the ship and the NGS. The sea state is measurable but uncontrollable.

Consequently, the OTE team determines a total of 4 factors that should be considered in the experiment. These design factors and pertaining levels are given in Table 1.

Table 1. Design factors and their levels

Factor	Levels
Naval Gun System	Alfa, Bravo, Charlie
Ship	S1, S2
Fire Control Radar	R1, R2
Sea State	Measurable but uncontrollable

In certain multifactor experiments, the levels of one factor (e.g., factor B) may seem quite similar for different levels of another factor (e.g., factor A). However, this similarity may mislead us in designing an experiment in the sense that they are identical but in fact, they are not. In such cases, we say the levels of factor B are nested under the levels of factor A. Since FCRs are specific to each NGS, levels of FCR factor are nested under the levels of NGS factor. Therefore, we need a design with both nested and factorial factors. For detailed information regarding nested designs and their statistical analysis, the interested reader is referred to (Montgomery, 2017). In this experiment, the FCR factor is the nested one (nested under the NGS factor) while the factors NGS and Ship are the factorial factors. In order to calculate the error term, each factor combination is replicated twice, hence the OTE team decides to perform a total of 24 runs for the experiment. Aiming to hedge against nuisance factors, all 24 runs are randomized. The resulting nested design is illustrated in Figure 1.

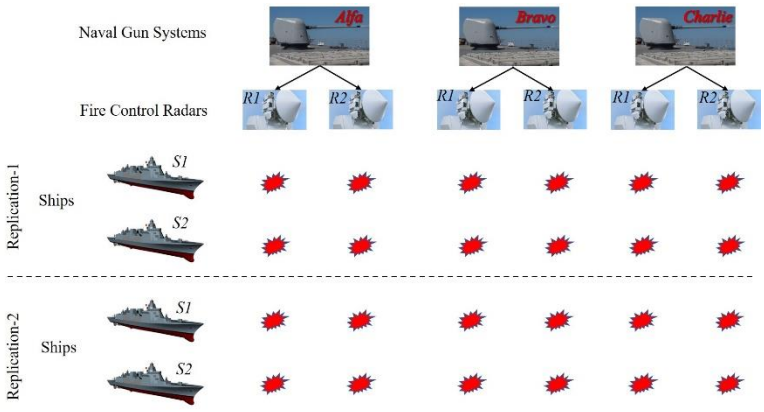


Figure 1. Nested design for the NGS OTE

3.2. Statistical Analysis

The nested design for the NGS OTE and the deviations measured for each firing are given in Table 2.

Table 2. Nested design and results of the experiment (deviations in yards)

Std Order	Run Order	PtType	Blocks	NGS	FCR	Ship	**Dev**	Std Order	Run Order	PtType	Blocks	NGS	FCR	Ship	**Dev**
3	1	1	1	1	2	1	**80**	18	13	1	1	2	1	2	**96**
16	2	1	1	1	2	2	**62**	9	14	1	1	3	1	1	**135**
14	3	1	1	1	1	2	**74**	23	15	1	1	3	2	1	**136**
12	4	1	1	3	2	2	**107**	19	16	1	1	2	2	1	**103**
6	5	1	1	2	1	2	**89**	21	17	1	1	3	1	1	**118**
20	6	1	1	2	2	2	**95**	4	18	1	1	1	2	2	**87**
1	7	1	1	1	1	1	**77**	7	19	1	1	2	2	1	**92**
13	8	1	1	1	1	1	**87**	8	20	1	1	2	2	2	**93**
17	9	1	1	2	1	1	**104**	10	21	1	1	3	1	2	**113**
24	10	1	1	3	2	2	**128**	22	22	1	1	3	1	2	**113**
11	11	1	1	3	2	1	**119**	2	23	1	1	1	1	2	**83**
15	12	1	1	1	2	1	**80**	5	24	1	1	2	1	1	**114**

To that end, a linear model with both nested and factorial factors is constructed for the design. The fixed effect linear model is shown in Equation (1).

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_{k(i)} + (\tau\beta)_{ij} + (\beta\gamma)_{jk(i)} + \epsilon_{ijkl} \qquad \begin{cases} i = 1,2,3 \\ j = 1,2 \\ k = 1,2 \\ l = 1,2 \end{cases} \tag{1}$$

In this equation τ_i represents the NGS effect, β_j represents the Ship effect, $\gamma_{k(i)}$ represents the FCR effect, $(\tau\beta)_{ij}$ is the effect of NGS*Ship interaction, and $(\beta\gamma)_{jk(i)}$ is the effect of Ship*FCR interaction. Note that since levels of FCR factor are nested under the levels of NGS, not every level of FCR factor appears with every level of NGS factor. Thus, there are no NGS*FCR interactions and 3-way interactions of factors.

The statistical analyses are conducted with Minitab 19 Statistical Software. Firstly, we examine the residual plots for adequacy checking of the Analysis of Variance (ANOVA) model. We generated the "four in one plot" with Minitab 19 and observed that residual plots are satisfactory in terms of (*i*) Normality, (*ii*) Homoscedasticity, and (*iii)* Auto-correlation.

ANOVA results of the model are given in Table 3. The R^2 figure of the ANOVA is 88.22%, hence, 88.22% of the variation in the response variable (deviations) is explained by the model. ANOVA reveals that the NGS effect is significant. Therefore, we reject the null hypothesis suggesting that the mean deviations of all NGSs are equal. Consequently, we conclude that there is a statistical difference between NGSs' hitting performances. Also, we can reject the null hypothesis claiming that there is no difference between Ships' performances at $\alpha = 0.05$ since the *p*-value is 0.043 for the Ship effect. None of the interaction terms is statistically significant in the analysis.

Table 3. ANOVA results

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
NGS	2	7149,07	3574,53	41,04	0,000
Ship	1	444,93	444,93	5,11	0,043
FCR(NGS)	3	79,22	26,41	0,30	0,823
NGS*Ship	2	56,27	28,13	0,32	0,730
FCR(NGS)*Ship	3	96,14	32,05	0,37	0,777
Error	12	1045,15	87,10		
Total	23	8870,78			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
9,33252	88,22%	77,42%	52,87%

Since the NGS effect is statistically significant, we examine the Tukey pairwise comparisons of NGS performances with a 95% confidence level. Looking at the pairwise comparisons presented in Table 4 and Figure 2, we observe that NGS-1 (Alfa) has the best (minimum) deviation performance among all NGSs. Consequently, we can conclude that the Alfa NGS has the best performance figures in terms of "Precision". %95 confidence intervals for the pairwise differences in means also indicate that the differences are statistically significant.

Table 4. Grouping information using the Tukey method and 95% confidence

NGS	N	Mean	Grouping		
3	8	121,017	A		
2	8	98,130		B	
1	8	78,791			C

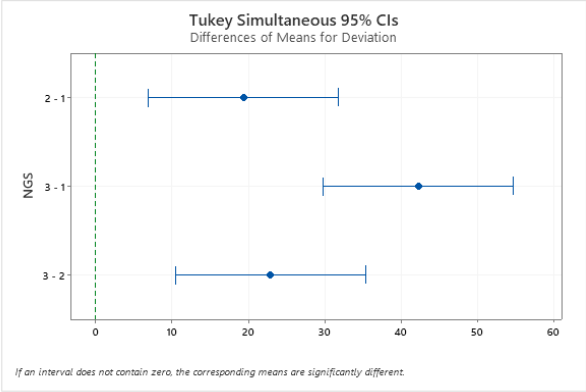


Figure 2. Tukey pairwise comparisons

Figure 3 shows the main effects plot for NGS and Ship factors. This figure verifies the superiority of NGS-1 (Alfa) in terms of precision. We also observe that Ship-2 performs better in the live firing experiments. In Figure 4, the interaction plot between these two main factors is presented. As seen from the figure, there is no interaction between these two factors, which verifies the ANOVA results given in Table 3.

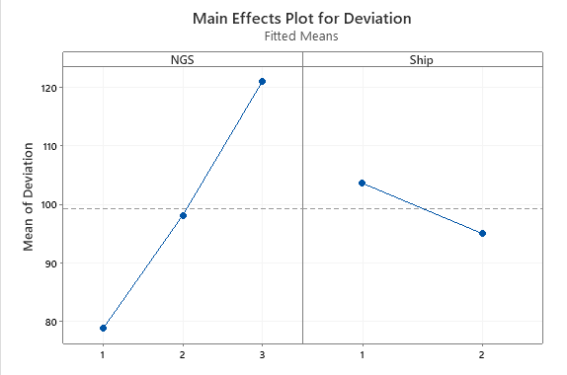


Figure 3. Main effects plot for NGS and Ship factors

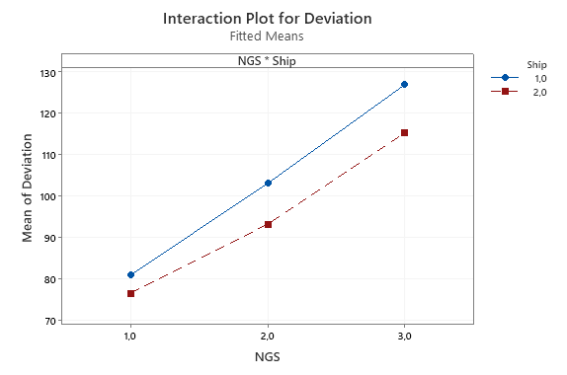


Figure 4. Interaction plot for NGS*Ship interaction

4. Conclusion

We considered evaluating and comparing candidate systems for acquisition using DOE techniques. Due to the complexity of the acquisition decisions and costs of the candidate systems, enterprises require decision support for making these decisions. Hence, utilizing quantitative methods is of utmost

importance. Being a branch of the applied statistics, the DOE provides quantitative techniques for designing the appropriate experiments and analyzing the experimental data.

One issue to keep in mind that the DOE process should be considered holistically. Therefore, all stakeholders and participants should be involved in the process from the beginning. For instance, analysts that will plan the experiment and conduct the statistical analysis should be part of the process from the beginning and evaluate what can be measured quantitatively and what cannot. Moreover, for a proper design and statistical analysis, system dynamics should be clearly understood by the analyst.

In this paper, we presented an illustrative case study where a nested design including both nested and factorial factors is performed. As future work, various applications utilizing other experimental designs and combining the resulting analyses with multi-criteria decision aid techniques can be presented.

References

Burk, R. C., & Parnell, G. S. (1997). Evaluating future military space technologies. Interfaces, 27(3), 60-73.

Chen, S. M. (1996). Evaluating weapon systems using fuzzy arithmetic operations. Fuzzy Sets and Systems, 77(3), 265-276.

Chen, R. S., & Shyu, J. Z. (2006). Selecting a weapon system using zero-one goal programming and analytic network process. Journal of Information and Optimization Sciences, 27(2), 379-399.

Cheng, C. H. (1999). Evaluating weapon systems using ranking fuzzy numbers. Fuzzy Sets and Systems, 107(1), 25-35.

Cheng, C. H., Yang, K. L., & Hwang, C. L. (1999). Evaluating attack helicopters by AHP based on linguistic variable weight. European Journal of Operational Research, 116(2), 423-435.

Erişkin, L. (2021) Preference modelling in sorting problems: Multiple criteria decision aid and statistical learning perspectives. Journal of Multi-Criteria Decision Analysis; 1-17.

Eriskin, L., & Gunal, M. M. (2019). Test and evaluation for weapon systems: concepts and processes. Operations Research for Military Organizations, 98-110.

Lee, J., Kang, S. H., Rosenberger, J., & Kim, S. B. (2010). A hybrid approach of goal programming for weapon systems selection. Computers & Industrial Engineering, 58(3), 521-527.

Montgomery, D. C. (2017). Design and analysis of experiments. John Wiley & Sons, Hoboken, NJ.

Parnell, G. S., Conley, H. W., Jackson, J. A., Lehmkuhl, L. J., & Andrew, J. M. (1998). Foundations 2025: A value model for evaluating future air and space forces. Management Science, 44(10), 1336-1350.

Rayno, B., Parnell, G. S., Burk, R. C., & Woodruff, B. W. (1997). A methodology to assess the utility of future space systems. Journal of Multi-Criteria Decision Analysis, 6(6), 344-354.

Wang, P., Meng, P., & Song, B. (2014). Response surface method using grey relational analysis for decision making in weapon system selection. Journal of Systems Engineering and Electronics, 25(2), 265-272.

Optimization Versus Metaheuristics in Forecasting: A Comparative Study for Energy Demand Forecast of Turkey

Alper Döyen*, Yakup Atasagun

Konya Technical University, Industrial Engineering Department, Konya, Turkey
alperdoyen@ktun.edu.tr, yatasagun@ktun.edu.tr

Abstract

Turkey is a developing country and has a high import energy dependency rate to meet its energy demand, which makes it more critical to forecast the energy demand accurately. Fitting a regression model is a well-accepted method to estimate energy demand of a country. The prediction quality of a regression model is directly related to the parameters of predictor variables. Metaheuristic algorithms and artificial neural networks (ANN) are widely employed in the literature to determine the values of these parameters. However, instead of employing these algorithms, optimization can be used to determine regression parameters in a more intelligent manner. For that purpose, in this study we use goal programming (GP), which is a well-known multi-objective mathematical programming technique, and well-known least squares estimation (LSE). The performance of GP and LSE are compared to the performance of some metaheuristic and ANN methods proposed in the literature for the energy demand prediction of Turkey. It is observed that GP and LSE outperform all the metaheuristics and the ANN in different performance measures with a much less effort in developing and running.

Keywords: Energy demand forecasting, goal programming, multiple linear regression

1. Introduction

Energy is important for all countries, but it is especially important for developing countries. The energy demand of those countries increases dramatically due to the rapid rise in per capita income and population. Turkey has one of the fastest growing energy demands among the Organization for Economic Co-operation and Development (OECD) countries in the past two decades. Regarding to that, Turkey's reliance on imported energy increases. Energy imports are expensive, and energy storage is neither simple nor cost-effective, therefore maintaining the energy supply-demand balance is indispensable. This balance is provided by efficient energy planning. The basic step in efficient planning is the accurate estimation of energy demand.

Energy demand of a country is related to various factors such as, gross national product, population, industrial production index, import, export, weather temperatures, electricity prices, oil prices, number of vehicles, inflation percentages etc. Considering mentioned factors above, many different methodologies were applied to forecast the energy demand of Turkey. Some of the important studies are listed in Table 1, briefly.

Most of the above studies try to find a good regression model to predict the energy demand of Turkey. Regression analysis is a methodology that allows finding a functional relationship (i.e., model/equation) among response (or dependent) variables and predictor (or independent) variables. If the functional relationship is linear and there exists only one predictor variable, then it is called simple linear regression. If there are more than one independent variable that affect the value of a dependent variable, then the regression model is a multiple linear regression model. Since the amount of energy demand

* *Corresponding Author*

depends on numerous predictor variables, fitting a multiple linear regression model is quite appropriate for energy demand prediction. Multiple linear regression is explained in the next section.

Table 1. Some studies aiming to forecast Turkey's energy/electricity demand.

Method Used	**Type of method**	**Author(s)**	**Forecasted variable**
Statistical approaches	ARIMA	Erdogdu (2007)	Electricity demand
	Ridge Regression and Partial LSE	Bulut and Yıldız (2016)	Energy demand
	Linear Regression	Aydın (2014)	Energy demand
Artificial Neural Network		Günay (2016)	Electricity demand
		Kankal et al. (2011)	Energy demand
Metaheuristics	Genetic Algorithm	Ceylan and Öztürk (2004)	Energy demand
	Particle Swarm Opt. (PSO)	Ünler (2008)	Energy demand
	PSO and ACO	Kıran et al. (2012a)	Electricity demand
	Ant Colony Opt. (ACO)	Toksarı (2007)	Electricity demand
	Differential Evolution	Beskirli et al. (2017)	Energy demand
Hybrid methods	Hybridized PSO and ACO	Kıran et al. (2012b)	Energy demand
Grey Theory	Grey Prediction	Akay and Atak (2007)	Electricity demand

Fitting a regression model means determining the coefficients (parameters) of the model. The literature essentially stands on proposing new metaheuristics and artificial neural networks to determine the values of those coefficients, effectively (Kıran et al., 2012a). However, both developing and running processes of these types of algorithms may be grueling.

In this study, we fit multiple linear regression models by using least square estimation (LSE) and goal programming (GP). Applying LSE and GP methods need little effort. Therefore, if it can be shown that the estimation quality of LSE and GP is comparable to that of metaheuristics or ANNs, then the prior methods may be preferred. The study mainly aims to evaluate the performance of LSE and GP in comparison to the performances of some metaheuristics and artificial neural networks existed in the literature. Moreover, due to our knowledge, goal programming has never been applied for the energy demand estimation of Turkey. Therefore, this study also contributes to the literature in this perspective.

2. Multiple Linear Regression

In this study, Turkey's energy demand is modeled by multiple linear regression. In multiple linear regression, the response variable Y is related to k independent (predictor) variables ($X_i, i = 1, \cdots k$) as shown in equation (1). The $\beta_i, i = 0,1, \cdots k$ parameters are called regression coefficients.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + \varepsilon \quad (1)$$

Consider the following model, which is a second order response surface model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon \quad (2)$$

Here, we can let $X_3 = X_1^2,\ X_4 = X_2^2,\ X_5 = X_1 X_2,\ \beta_3 = \beta_{11},\ \beta_4 = \beta_{22}\ and\ \beta_5 = \beta_{12}$. Then this model is also a multiple linear model as shown in equation (3). Computational results for this type of a model are also reported in the computational results section.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (3)$$

To estimate the Y values a fitted model is developed as follows:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots \hat{\beta}_k X_k \quad (4)$$

where $\hat{Y}$ is the predicted value and $\hat{\beta}_i$ are the estimates of the regression coefficients. The estimation errors for each data point are called as residuals and calculated by $e_i = y_i - \hat{y}_i$. Different functions of these residuals lead to different performance metrics which are used to evaluate the quality of the fitted model. Sum of square errors (SSE), mean absolute error (MAE) and root of mean square error (RMSE) are the performance metrics applied in this study. Their formulations are given below.

$$\text{SSE}=\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (5)$$

$$\text{MAE}= \frac{1}{}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \quad (6)$$

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (7)$$

3. Least Squares Estimation

Least squares estimation (LSE) is a kind of optimization method, which was first introduced by Legendre (1805). The LSE method chooses the model parameters (the *β*s) so that the sum of squares of the errors (SSE) is minimized.

$$L = e_1^2 + e_2^2 + \cdots e_n^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2 \quad (8)$$

$$L = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ji}\right)^2 \quad (9)$$

The least squares function *L* defines the sum of squares of the errors (SSE) in equations (8-9). The *L* function is differentiated with respect to all β coefficients and the partial derivatives are all equated to zero. The resulting system of equations is solved, thus the $\hat{\beta}$ values are obtained.

4. Goal Programming

Goal Programming (GP) is a multi-objective programming technique, which was first introduced by Charnes et al. (1955). The essence of GP is to try to achieve a set of goals (targets) as closely as possible. Unwanted deviations from the goals are minimized in the objective function. The mathematical model of a general GP is given below in equations (9-11). Here, the model has two sets of constraints: goal constraints and system (hard) constraints. Q is the set of hard constraints. The $f_i(\boldsymbol{x})$ is a linear function of $\boldsymbol{x}$, and b_i is the target value for that objective, d_i^- and d_i^+ determine the negative and positive deviations from b_i. The w_1 and w_2 are the respective weights related to the negative and positive deviations.

$$\min Z = \sum_{i=1}^{n} w_1 d_i^- + w_2 d_i^+ \quad (9)$$

$s.t.$

$$f_i(\boldsymbol{x}) + d_i^- - d_i^+ = b_i, \qquad i = 1, \cdots, n \quad (10)$$

$$\boldsymbol{x} \in Q \quad (11)$$

We know that $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots \hat{\beta}_k X_k$ is the fitted multiple linear regression equation. Since the residual is $e_i = y_i - \hat{y}_i$, we can write the equation (12).

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + e_i, \quad \forall i \quad (12)$$

When we consider $e_i = d_i^- - d_i^+$, the GP can be applied to obtain good estimations of β coefficients as following:

GP1:

$$\min Z = \frac{1}{|N|}\sum_{i=1}^{n}(d_i^- + d_i^+) \tag{13}$$

$s.t.$

$$\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + d_i^- - d_i^+ = Y_i, \qquad i = 1, \cdots, n \tag{14}$$

$$\hat{\beta}_0, \hat{\beta}_1, \cdots \hat{\beta}_k, d_i^-, d_i^+ \geq 0 \tag{15}$$

Here in this model (GP1), the objective function finds the mean absolute error (MAE) metric for the regression. To get sum of squares error (SSE) metric, only the objective function of the above model needs to be changed, while all the constraints remain the same. The objective function of this new model (GP2) would be as in equation (16) below.

$$\textbf{GP2}: \min Z = \sum_{i=1}^{n}(d_i^- + d_i^+)^2 \tag{16}$$

5. Computational Results

In this section, we compared the performance of LSE and Goal Programming (GP1 and GP2) with some metaheuristics and ANN algorithms found in the literature. For this purpose, we have made several analyses. The LSE is executed on MS Excel. The GP1 and GP2 models are implemented in the General Algebraic Modelling System (GAMS) and use the CPLEX and CONOPT solvers, respectively. The running times for all the analyses are just around one CPU second.

In the first analysis, we considered a model that includes four indicators (GDP, population, import and export) to estimate Turkey's energy demand by using a benchmark data given for 1979-2005 years. That data set was used by different researchers (for different methods) to evaluate the performance of their multiple linear regression (Table 2) and second-order response surface models (Table 3). Please bear in mind that GP1 results a model for the MAE metric and GP2 results a model for the MSE (or RMSE) metric, therefore SSE and MAE values are not available (n/a) for GP1 and GP2, respectively.

Both of the Tables 2 and 3 show that the LSE and the GPs (GP1 and GP2) result better regression models than the well-known metaheuristics since the related values are smaller. Best values obtained are shown in italics. The last column, namely *Imp (%)*, indicates how much improvement is obtained as a percentage by the proposed methods regarding to the best value given in the literature. For example, in Table 2 the best MAE value given by the previous methods is 1.10 (found by DE and HAPE algorithms). We have obtained 1.02 by GP1. Therefore, our study improved this value by 7.2 %.

Table 2. Results for the first data set (multiple linear regression model) analysis

Performance Metric	ACO (Toksari, 2007)	PSO (Ünler, 2008)	HAPE (Kiran etal, 2012)	DE (Beskirli etal, 2017)	LSE	GP1	GP2	Imp (%)
SSE	45.724	42.614	41.7121	41.712	*41.712*	n/a	*41.712*	0
MAE	1.133	1.110	1.10	1.10	1.10	*1.02*	n/a	7.2

Table 3. Results for the first data set (second-order response surface model) analysis

Performance Metric	ACO (Toksari, 2007)	PSO (Ünler, 2008)	HAPE (Kiran etal, 2012)	DE (Beskirli etal, 2017)	LSE	GP1	GP2	Imp (%)
SSE	27.947	27.664	20.567	19.757	*17.652*	n/a	*17.652*	10.7
MAE	n/a	n/a	0.659	0.666	0.622	*0.494*	n/a	25

The second comparison is done by artificial neural network (ANN) prediction of Günay (2016), which uses a data set consisting of six predictor variables (population, GDP per capita, inflation rate, unemployment rate, average summer temperature and average winter temperature) to predict gross electricity demand. The data ranges from 1975 to 2013 years. The comparison of our models with ANN of Günay (2016) is given in Table 4. Both the LSE and the GPs (GP1 or GP2) give better results than the ANN. In addition, LSE and GP1 are the best methods for the SSE and MAE metrics, respectively.

Table 4. Results for the second data set analysis

Performance Metric	ANN (Günay, 2016)	LSE	GP1	GP2	Imp (%)
SSE	1758.2	*1654.8*	n/a	1655.6	5.9
MAE	5.50	5.38	*5.12*	n/a	6.9

In the paper, Günay (2016) also presents a reduced model excluding some of the insignificant variables. This model includes four variables (population, GDP per capita, inflation rate and avg. summer temperature). Regarding to this model, data of 2007-2013 period was used to show that ANN of Günay (2016) had superior performance than the official predictions of Ministry of Energy and Natural Resources of Turkey (TEIAS, 2017), ARIMA of Erdogdu (2007) and grey prediction of Akay and Atak (2007). Therefore, we just compared our models by the ANN of Günay (2016) on this data set. The results are given in Table 5, where RMSE is used as a performance metric instead of SSE to stick to the metric of Günay (2016). Note that RMSE can be calculated directly from SSE, thus LSE and GP2 can be used to find it. Due to the results reported in Table 5, we can say that LSE and GP models are significantly better than the ANN model. In addition, since the superiority of the ANN used in comparison has been shown over some statistical methods and grey prediction, we can say that also LSE and GP perform better than those methods.

Table 5. Results for the third data set analysis

Performance Metric	ANN (Günay, 2016)	LSE	GP1	GP2	Imp (%)
RMSE	5.7	1.03	n/a	*1.00*	*82.5*
MAE	37.14	5.98	*3.96*	n/a	89.3

6. Conclusion and Future Work

Metaheuristics and ANNs are well accepted methods in fitting multiple linear equations to forecast energy demand of Turkey. In this study, it is shown that GP and LSE methods outperform metaheuristics and ANNs in terms of prediction quality for the considered forecasting problem. The methods we implemented improved the forecasting errors for all data sets used in earlier studies. This improvement could be as high as 89 percent, according to our findings. Moreover, the presented methods are easy in implementing and fast in running times. Considering the promising results found, as a future work, we plan to use these optimization methods with up-to-date data to forecast Turkey's energy demand for the following years. We intend to apply stepwise linear regression to determine the predictor variables. In stepwise regression, variables are added or removed in succession, depending on whether their adding or removing results statistically significant contribution or not.

References

Akay, D., & Atak, M. (2007). Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. Energy, 32(9), 1670-1675.

Aydin, G. (2014). Modeling of energy consumption based on economic and demographic factors: The case of Turkey with projections. Renewable and Sustainable Energy Reviews, 35, 382-389.

Beskirli, M., Hakli, H., & Kodaz, H. (2017). The energy demand estimation for Turkey using differential evolution algorithm. Sādhanā, 42(10), 1705-1715.

Bulut, Y. M., & Yıldız, Z. (2016). Comparing energy demand estimation using various statistical methods: the case of Turkey. Gazi University Journal of Science, 29(2), 237-244.

Ceylan, H., & Ozturk, H. K. (2004). Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach. Energy Conversion and Management, 45(15-16), 2525-2537.

Charnes, A., Cooper, W. W., & Ferguson, R. O. (1955). Optimal estimation of executive compensation by linear programming. Management science, 1(2), 138-151.

Erdogdu, E. (2007). Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey. Energy policy, 35(2), 1129-1146.

Günay, M. E. (2016). Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey. Energy Policy, 90, 92-101.

Kankal, M., Akpınar, A., Kömürcü, M. İ., & Özşahin, T. Ş. (2011). Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. Applied Energy, 88(5), 1927-1939.

Kıran, M. S., Özceylan, E., Gündüz, M., & Paksoy, T. (2012a). Swarm intelligence approaches to estimate electricity energy demand in Turkey. Knowledge-Based Systems, 36, 93-103.

Kıran, M. S., Özceylan, E., Gündüz, M., & Paksoy, T. (2012b). A novel hybrid approach based on particle swarm optimization and ant colony algorithm to forecast energy demand of Turkey. Energy conversion and management, 53(1), 75-83.

Legendre, A. M. (1805). Nouvelles méthodes pour la détermination des orbites des comètes: avec un supplément contenant divers perfectionnemens de ces méthodes et leur application aux deux comètes de 1805. Courcier.

TEIAS (2007). Turkish Electricity Energy 10-year Capacity Projection Report (in Turkish).

Toksarı, M. D. (2007). Ant colony optimization approach to estimate energy demand of Turkey. Energy Policy, 35(8), 3984-3990.

Ünler, A. (2008). Improvement of energy demand forecasts using swarm intelligence: The case of Turkey with projections to 2025. Energy policy, 36(6), 1937-1944.

The Effects of Turbulent Fluctuations on Nonlinear von Kármán Vortex Shedding

Kubilay Ateş[1], Cihan Bayındır[2,3*]

[1]*Yıldız Technical University, Civil Engineering Department, İstanbul, Turkey.*
[2]*İstanbul Technical University, Civil Engineering Department, İstanbul, Turkey.*
[3]*Boğaziçi University, Civil Engineering Department, İstanbul, Turkey.*
l4918015@std.yildiz.edu.tr, cbayindir@itu.edu.tr

Abstract

Flows around bluff bodies generate wakes and vortices downstream of flow. This phenomenon is known as vortex shedding and such vortices are generally named as von Kármán vortices after Theodore von Kármán. Although such a phenomenon is introduced to the scientific literature by the study of fluid flows, it is also observed in other fields such as Bose-Einstein condensation. Due to the complexity of the governing equations and involved complex geometries, such phenomena are generally studied numerically using different software and various turbulent modelling techniques. One of the other commonly utilized models for the study of nonlinear vortex shedding is the complex Ginzburg-Landau (GL) equation (Cohen et. al, 2003, Gillies, 1998, 2001, Roussopoulos et al., 1996). This dynamic equation is an equation in the nonlinear Schrödinger class and also appears in various other branches of science. In this paper, we investigate the effects of turbulent fluctuations on the vortex shedding in the frame of the GL equation. With this aim, we solve the GL equation using a spectral scheme with a 4th order Runge-Kutta time integrator. For the spectral solution, efficient FFT routines are employed. We analyze the possible modulation instabilities causes by turbulent fluctuations, their effects on the regular stable vortices, and possible rogue vortex formation (Bayındır, 2006a, 2006b). We also study the dynamics and statistics of such vortices under the effect of turbulent fluctuations. Our findings can be used for controlling, mitigating, or resonating the vortices and wake for many different engineering purposes including but are not limited to structural safety and serviceability considerations, noise reduction, energy harvesting, just to name a few.

Keywords: Vortex shedding, Ginzburg-Landau equation, turbulent fluctuations, spectral method

1. Introduction

One of the most commonly studied phenomena in fluid dynamics is the vortex shedding and wake formation around bluff bodies. Although the presence of such vortices are introduced by von Kármán who investigated fluid flows, similar vortex and wake structures are also observed in some other branches of physics such as Bose-Einstein condensation. For the investigation of such vortices the commonly used approach is the simulation of the Navier-Stokes equation for complex geometries. Such studies generally utilize commercial fluid dynamics software. Another approach for similar studies is to use dynamic equations which are derivable from the Navier-Stokes equations. One of these equations is the complex GL equation.

In this study, we investigate the effects of turbulent fluctuations on the dynamics and characteristics of the vortex shedding and wakefields modeled in the frame of the complex GL equation. With this motivation, we solve the complex GL equation using a spectral scheme with a 4th order Runge-Kutta integrator. In order to simulate a flow field around a circular cylinder located at the origin, the periodic

* *Corresponding Author*

vanishing amplitude conditions are imposed. We show that depending on the strength of the turbulent fluctuations, the vortex wavelength and their reach downstream of the flow are significantly affected. The range of the complex amplitude is found out to be unaffected. We discuss our findings and their applicability in a real setting.

2. Mathematical Formulation

Flows around bluff bodies generate wakes and vortices downstream of flow. This phenomenon is known as vortex shedding and such vortices are generally named as von Kármán vortices after Theodore von Kármán. A dimensionless form of the Ginzburg-Landau equation (GLE) is given by

$$\frac{\partial A}{\partial t}+U\frac{\partial A}{\partial x}=\mu(x)A+(1+iC_D)\frac{\partial^2 A}{\partial x^2}-(1+iC_N)|A|^2 A+F(x,t) \tag{1}$$

Here $A(x,t)$ is the complex amplitude, U, C_D, C_N and are real constants that show the advection speed, diffusion, and nonlinearity constants (Cohen et al., 2003). The $F(x,t)$ is the forcing function and throughout this study unforced vortices are considered thus this function is selected as $F(x,t)=0$. The wake growth parameter $\mu(x)$ is defined by

$$\mu(x)=\mu_o+\mu' x \tag{2}$$

In order to analyze the effects of noisy perturbations of the characteristics and stability of von Kármán vortices we solve the GLE using a 4th order Runge-Kutta time-stepping algorithm and Fourier spectral method. This method is summarized below. Rewriting Eq. (1) as

$$\frac{\partial A}{\partial t}=-U\frac{\partial A}{\partial x}+\mu(x)A+(1+iC_D)\frac{\partial^2 A}{\partial x^2}-(1+iC_N)|A|^2 A+F(x,t)=g(A,x,t) \tag{3}$$

Four slopes of the Runge-Kutta time-stepping algorithm can be calculated using

$$\begin{aligned} m_1&=g(A_n,t_n,x)\\ m_2&=g(A_n+0.5m_1dt,t_n+0.5dt,x)\\ m_3&=g(A_n+0.5m_2dt,t_n+0.5dt,x)\\ m_4&=g(A_n+m_3dt,t_n+dt,x) \end{aligned} \tag{4}$$

Then the value of the complex amplitude and time at the next time step can be evaluated using

$$\begin{aligned} A_{n+1}&=A_n+dt(m_1+2m_2+2m_3+m_4)/6\\ t_n&=t_{n+1}+dt \end{aligned} \tag{5}$$

In order to prevent temporal instability, the time step is selected as $dt=0.005$. The spatial derivatives in Eq. (3) can be computed spectrally using Fourier series by

$$\frac{\partial A}{\partial x}=F^{-1}[ikF[A]] \quad \text{and} \quad \frac{\partial^2 A}{\partial x^2}=F^{-1}[-k^2F[A]] \tag{6}$$

Here, F denote the Fourier and F^{-1} denote the inverse Fourier transform operations, and k denotes the wavenumber vector which includes $N=512$ multiples (spectral components) of the fundamental wavenumber vector, k_0. The domain length for our simulations is selected to be $L=120$ and the corresponding fundamental wavenumber, $k_0=2\pi/L$, is used.

3. Results and Discussion

In order to illustrate the effects of turbulent fluctuation we first simulate the numerical simulation of the Ginzburg-Landau equation for the turbulent fluctuation (noise) free case. For this simulation, the

computational parameters are selected as $U=5,\ C_D=1,\ C_N=0, \mu_o=3.57,\ \mu'=-0.0434$. The initial condition is selected as $A(x,0)=10^{-4}$.

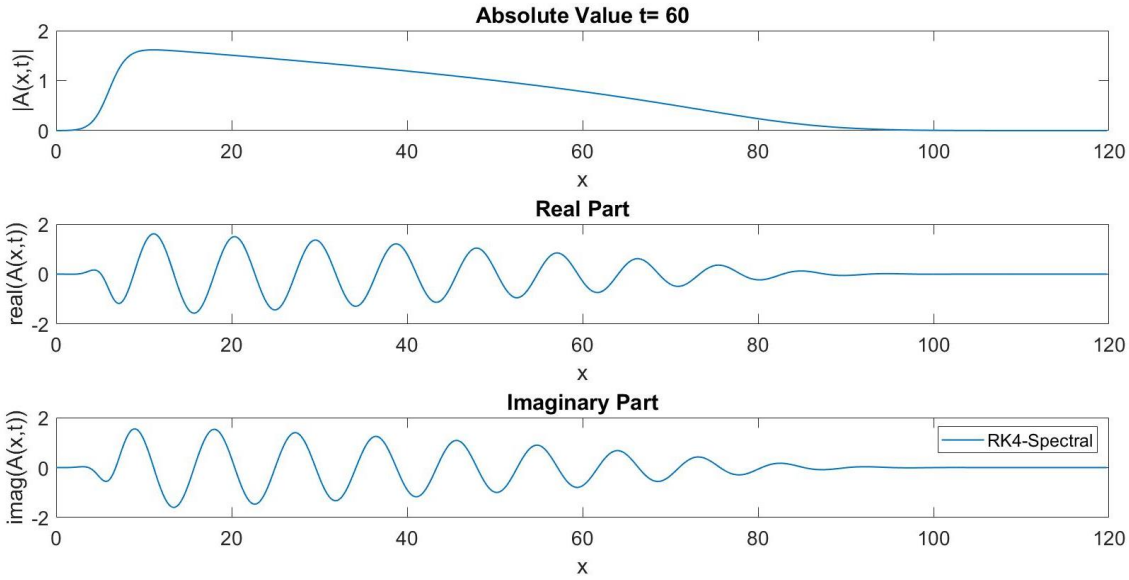


Figure 1. Simulation of vortices downstream of a circular cylinder at $x = 0$ for turbulent fluctuation free case.

It is known that this selection of the parameters will lead to a growth parameter slightly larger than the critical growth parameter, thus growing wake oscillations will be observed downstream of the flow around a circular cylinder located at $x=0$. We present the vortex street obtained downstream of such a flow in Fig. 1. Our results are in good agreement with finite element solutions presented by Cohen et al. (2003). The results illustrated in Fig. 1 clearly indicate that the wake structures have emerged downstream of a circular cylinder located at the origin. The vortex street and shedding are observable up to the location $x \approx 90$ for a simulation time of $t \approx 60$.

Next, we investigate the effect of turbulent fluctuations on the dynamics and characteristics of the vortex street presented in Fig. 1. With this motivation, we impose uniformly distributed noise on the complex amplitude modeled in the frame of the GL equation. For this purpose, we impose noise on the complex amplitude of the

$$A(x,t) = A(x,t)_{noisefree} + a \times rand \tag{2}$$

where a show a set of uniformly distributed random numbers in the interval of $rand \in [-1,1]$. In a real setting, the noise refers to the turbulent fluctuations which can be realized by altering the flow field properties or flow geometry. In our simulations, we observe that when such a noise is imposed on the initial condition, the behavior of the wakefields remains unaffected. However, when such a noise is imposed on the complex amplitude at every time step of temporal evolution, the dynamics and characteristics of the vortex street are significantly affected. We present our findings in Fig. 2 for the vortex street depicted in Fig. 1 under the effect of a noise term with an amplitude of $a = 0.05$.

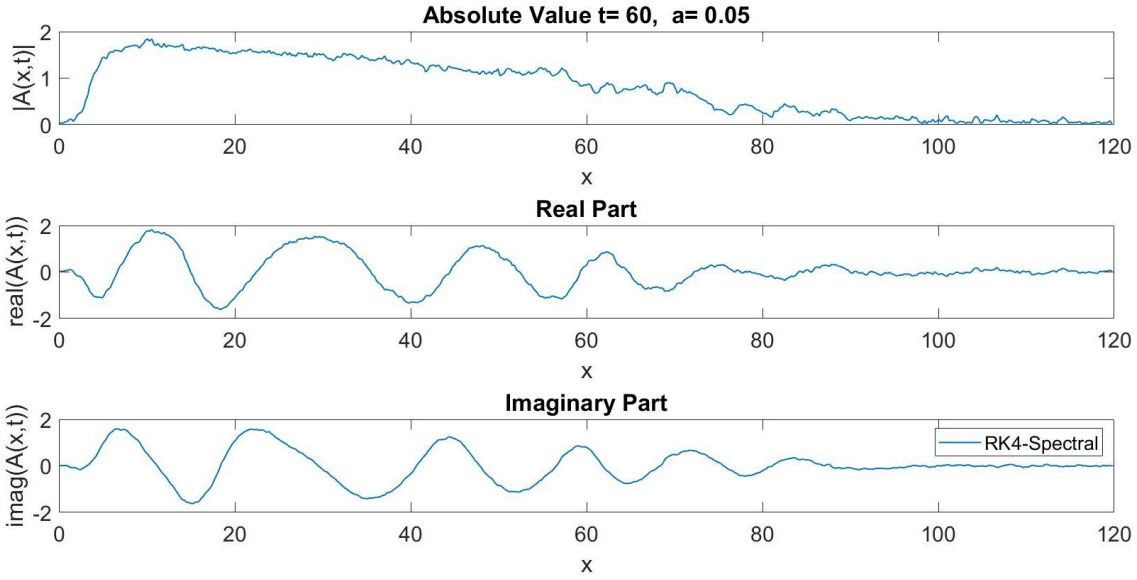


Figure 2. Simulation of vortices downstream of a circular cylinder at $x = 0$ under the effect of turbulent fluctuations with an amplitude of $a = 0.05$.

A comparison of Fig. 1 and Fig. 2 reveals that the wake and vortex structure downstream of a circular cylinder located at origin are significantly affected the turbulent fluctuations. The complex amplitudes of the wakes are not significantly affected, however, the lengths of vortices are increased which causes a reduction in the vortex wavenumber. Also, the extend on the vortices reduces to the downstream location of approximately $x \approx 82$.

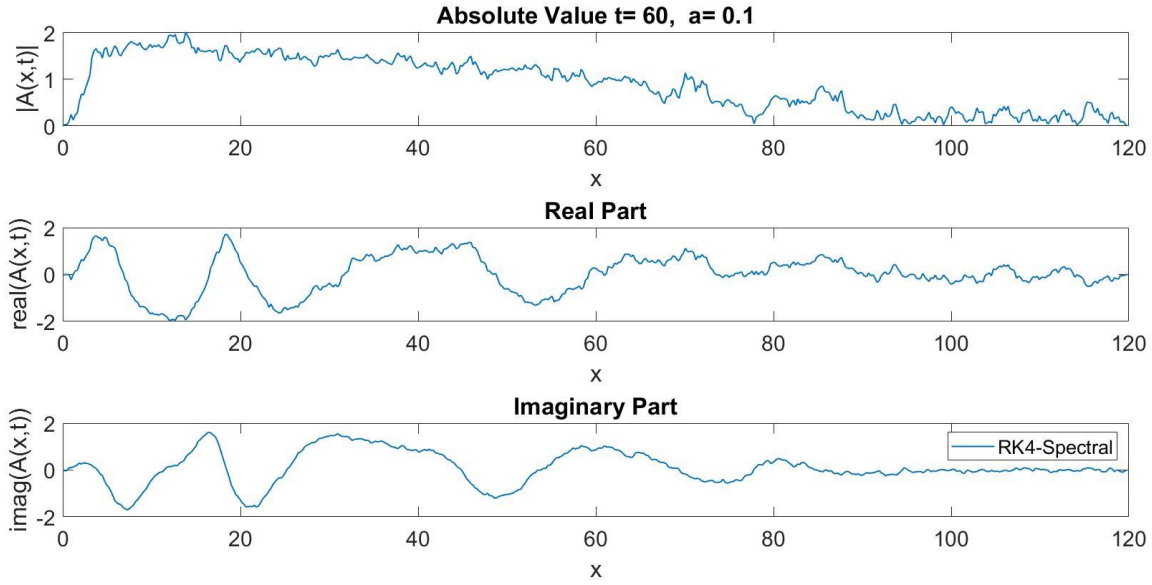


Figure 3. Simulation of vortices downstream of a circular cylinder at $x = 0$ under the effect of turbulent fluctuations with an amplitude of $a = 0.1$.

Next, we investigate the effect of a stronger turbulent fluctuation of the dynamics and characteristics of the vortex street presented in Fig. 1. With this aim, we impose a noise with an amplitude of $a = 0.1$ at every time step of temporal evolution. The results depicted in Fig. 3 confirms that stronger fluctuations lead to further suppression of the vortices. Even under such a strong fluctuation large amplitude or rogue

vortices are not observed (Bayındır, 2006a, 2006b), however, the vortex width after the location $x \approx 21$ are significantly increased leading to smaller vortex wavenumbers. For engineering purposes, such results are very promising for the wake, vortex, and flow-induced vibration control. With the help of controlled turbulent fluctuations imposed on the flow field or flow geometry, the dynamics of the wavefields can be altered and controlled. Such a wake control strategy can be used to limit the flow-induced vibration or when desired, to enhance them. Our findings can be used for avoiding the vortex shedding induced resonance or fatigue or can be used as a remedy to vortex control strategies involving the use of components such as helical strakes. Our results can also be used to enhance piezoelectric energy conversion of flow-induced wakes and vortices and their efficient measurement (Bayındır et al., 2021), just to name a few.

4. Conclusion

In this paper, we have investigated the effects of turbulent fluctuations on the dynamics and characteristics of the vortex streets and shedding in the frame of the complex Ginzburg-Landau equation. We showed that, depending on the amplitude of the turbulent fluctuations the dynamics and characteristics of vortex streets can be significantly affected. In near future, we aim to extend our analysis to higher dimensions to investigate the effect of turbulent fluctuations on multi-dimensional vortices and we aim to perform spectral analysis using Fourier and wavelet analysis to analyze the frequency and space-frequency and time-frequency domain characteristics of vortex shedding control via turbulent fluctuation injections.

Acknowledgements

The author gratefully acknowledges the support of the İstanbul Technical University. This work was supported by the Research Fund of the İstanbul Technical University. Project Code: MGA-2020-42544. Project Number: 42544.

References

Gillies, E. A. (1998). Low-dimensional control of the circular cylinder wake. Journal of Fluid Mechanics, 371, 157-178. https://doi.org/10.1017/S0022112098002122

Roussopoulos, K., & Monkewitz, P.A. (1996). Nonlinear Modeling of Vortex Shedding Control in Cylinder Wakes. Physica D, 97, 264-273. https://doi.org/10.1016/0167-2789(96)00151-0

Gillies, E.A. (2001). Multiple Sensor Control of Vortex Shedding, AIAA Journal, 39 (4), 748-750. https://doi.org/10.2514/2.1374

Cohen, K., Siegel, S., McLaughling, T. & Myatt, J. (2003). Fuzzy logic control of circular cylinder vortex shedding model, 41st Aerospace Sciences Meeting and Exhibit AIAA 2003-1290. https://doi.org/10.2514/6.2003-1290

Bayındır, C. (2016a). Rogue waves of the Kundu-Eckhaus equation in a chaotic wavefield, Phys. Rev. E., 93, 032201. https://doi.org/10.1103/PhysRevE.93.032201

Bayındır, C. (2016b). Rogue wave spectra of the Kundu-Eckhaus equation, Phys. Rev. E., 93, 062215. https://doi.org/10.1103/PhysRevE.93.062215

Bayındır, C., & Namlı, B. (2021). Efficient sensing of the von Karman vortices using compressive sensing, Computers & Fluids, 104975. https://doi.org/10.1016/j.compfluid.2021.104975

A Neural Network Learning Approach for Solving the Knapsack Problem

Ertan Yakıcı[1*], Tolga Önel[2]

[1] *National Defence University, Barbaros Naval Sciences and Engineering Institute, İstanbul, Turkiye*
[2] *HAVELSAN Inc., İstanbul, Turkiye*
eyakici@dho.edu.tr, tonel@havelsan.com.tr

Abstract

Widely known knapsack problem is an optimization problem which is in the non-deterministic polynomial-time complete (NP-C) class of problems. Hence, there is no known polynomial-time exact algorithm to guarantee an optimal solution for the Knapsack problem. A brute force search for the decision of choosing the most valuable items while remaining within the weight capacity limit has 2^n possibilities for n items. Going through all of the possibilities requires $O(2^n)$ time complexity. With the heuristic improvements to the brute force search, better time complexities can be achieved. However, optimal solutions cannot be guaranteed by these heuristic improvements. Although it is not an exact solution technique, the neural network has the strength of learning from known samples and then solving similar problems in very short periods. In this study, we train a neural network with the previously solved knapsack problem instances. After the training phase, we experiment with the method using 115 problem instances. 72% accuracy with the test data set is achieved in this experiment.

Keywords: Knapsack problem, neural networks, optimization

1. Introduction

Among a finite number of items, whose weights and the benefits are known, choosing a subset with the maximum benefit without exceeding the maximum weight that can be carried is named as the 0-1 Knapsack problem (KP01) (Lv et al., 2016). KP01 is known as a non-deterministic polynomial-time complete (NP-complete) problem (Basset et al., 2017). Dynamic programming, branch and bound, brute force, Lagrangian decomposition-based, genetic and other heuristic algorithms are proposed in the literature to solve the KP01 (Pan and Zhang, 2018). The algorithms proposed in the literature cannot guarantee to solve the KP01 to optimality in polynomial-time. Among the known heuristic methods, genetic algorithm with O(n) time complexity is the fastest of all (Shaheen and Sleit, 2016).

Supervised learning algorithms are used for learning the function related to an input-output pair. With the learned function, the output of an input that is not seen before can be approximately predictable. Neural networks are widely used as a supervised learning mechanism to extract information from the data. A neural network can be trained using the known inputs and the outputs related to the phenomenon. With an unprecedented example of the phenomenon, the output corresponding to the input can be determined with the trained neural network in O(n) time.

In this study, the use of neural networks is tested for solving the KP01 problem. We train a neural network using optimally solved KP01 examples. Using the trained neural network, we manage to achieve 72% similarity (in the optimal set of items) with the pre-solved KP01 test set examples. Hence, the proposed method may be suitable for the applications, where very fast but approximate solutions are required. Our findings also show that the use of neural networks can be a good candidate mechanism for solving hard optimization problems with similar input quickly after the neural network is trained.

[*] *Corresponding Author*

To the best of our knowledge, there is no similar study reported before. Therefore, this study serves as preliminary work to explore whether there is a satisfactory reason for further studies in this direction.
The paper is organized as follows: In Section 2, we describe the KP01 problem. In Section 3, we describe the proposed neural network technique for the KP01 problem. We discuss our experimental results in Section 4. Finally, we conclude and give some future research directions in Section 5.

2. Problem Definition

The decision version of the 0-1 knapsack problem is an NP-Complete problem. For a finite set of items *U*, the knapsack problem chooses a subset *U'* from *U* (*U'* ⊆ *U*), such that; $\sum_{u\in U'} w(u) \leq W$ and $\sum_{u\in U'} v(u) \geq V$, where *W* is the total maximum capacity of the knapsack and *V* is the minimum desired benefit (value) (Salkin and De Kluyver, 1975). n item has 2^n possible combinations according to their presence in the knapsack. A brute force algorithm, an exhaustive search that consists of enumerating all possible solution candidates and checking each candidate with respect to the problem's objective, solves the optimization version of the 0-1 knapsack problem in $O(2^n)$ time by examining all the possibilities for satisfying the capacity requirement and comparing values of combinations to find the maximum. There are also heuristic techniques with better time complexities that cannot guarantee to give the optimal solution for KP01 (Wilbaut et al., 2008).

3. Proposed Method

A Neural network-based approach is proposed to solve the KP01 problem. Neural networks are formed by neurons that perform a simple operation on their inputs. Generally, a neuron holds a number between 0 and 1. This number is called the activation of that neuron. Data is supplied to the network at the input layer. At the output layer, neurons' activation levels are determined according to the input data. In other words, neural networks learn the function that maps the input data to the output data. They learn the pattern in the training samples. The learning process is the adjustments of the numerical weights on the synapses that connect the neurons to each other. Once the weights are learned, a neural network can map the input data to the output data according to the function that is learned. A simple neural network is shown in Figure 1. If the neural network is a classifier that discriminates the samples of two classes, the output layer may have two neurons as shown in Figure 1. Each neuron of the output layer can correspond to one class. Related output layer neurons with the sample input data can have higher activation level than that of the other output layer neurons. Hence, the classification of the input sample data can be achieved (Alpaydın, 2020).

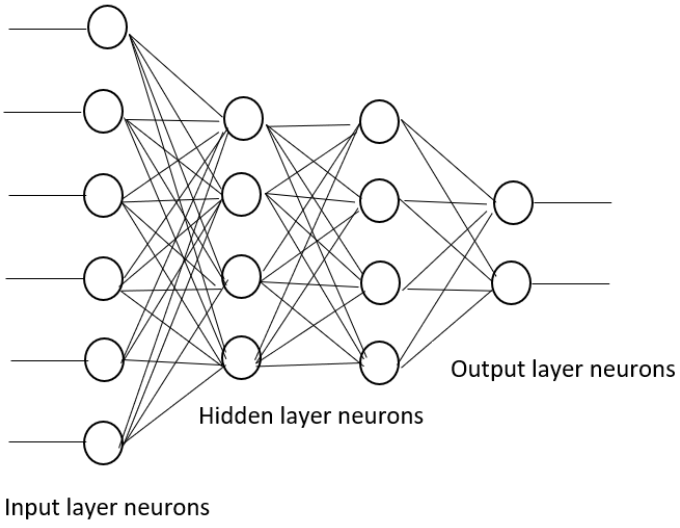


Figure 1. A representation of a simple neural network with two hidden layers.

The activation of each neuron can be determined by Equation 1.

$$a_{l+1} = \sigma(W_l a_l + b_l) \tag{1}$$

Where a_l is the activation level of neuron *l*, W_l is the weight of the synapse connecting the neuron *l* to neuron *l*+1 and b_l is the bias value of the neuron *l*. Finally, σ is the function (i.e., sigmoid) applied at the neuron *l*. In this way, activations of neurons propagate from the input layer through the output layer. Equation 1 is visualized in Figure 2. The role of the σ function is to map the activation of the neuron to a number between 0 and 1. Sigmoid function (σ) is given in Equation 2.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

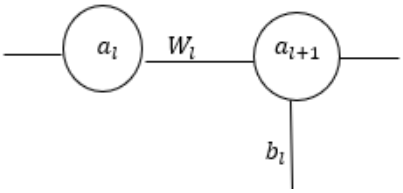


Figure 2. Activation of neurons.

When the single neuron activation in Figure 2 is generalized into the neural network as shown in Figure 1, the activations of neurons are given in Equation 3.

$$\begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \dots \\ a_n^{(1)} \end{bmatrix} = \sigma\left(\begin{bmatrix} w_{0,0}\ w_{0,1} & \dots & w_{0,n} \\ w_{1,0}\ w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{k,0}\ w_{k,1} & \cdots & w_{k,n} \end{bmatrix} \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \dots \\ a_n^{(0)} \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_n \end{bmatrix} \right) \tag{3}$$

Hence, the activation $a_0^{(1)}$ is given in Equation 4.

$$a_0^{(1)} = \sigma\left(w_{0,0}a_0^{(0)} + w_{0,1}a_1^{(0)} + \cdots + w_{0,n}a_n^{(0)} + b_0\right) \tag{4}$$

And, they can be calculated similarly for $a_1^{(1)}$ to $a_n^{(1)}$. Therefore, in general activations of neurons are given as shown in Equation (5).

$$a^{(1)} = \sigma\left(W a^{(0)} + b\right) \tag{5}$$

Learning of the neural network can be described as finding the correct weight (*W*) and bias (*b*) values. While training the network, initially these weights and bias take random values. The aim of a cost function is to minimize the gap between the desired output layer activations and the achieved output layer activations. This minimization approach is applied for all the training samples. This is done by going through the opposite direction of the gradient of the cost vector. The cost of a single neuron output at layer *L* (i.e., output layer) is the difference between the activation of the neuron ($a^{(L)}$) and the desired activation of the neuron (*y*) (Zhang et al., 2019). The desired output of the neuron is known through the labeled training data. For a single training example, the cost function of a single neuron at layer *L* is given in Equation 6.

$$C_0(\dots) = \left(a^{(L)} - y\right)^2 \tag{6}$$

where $a^{(L)}$ is given in Equation 7.

$$a^{(L)} = \sigma\left(w^{(L)}a^{(L-1)} + b^{(L)}\right) \tag{7}$$

We can define the input of the sigmoid function in Equation 7 as shown in Equation 8.

$$z^{(L)} = w^{(L)}a^{(L-1)} + b^{(L)} \tag{8}$$

Hence, the result given in Equation 9 is obtained.

$$a^{(L)} = \sigma\left(z^{(L)}\right) \tag{9}$$

The cost function in Equation 6 depends on the changes of $a^{(L)}$ and *y*. Consequently, changes in $a^{(L)}$ depend on the changes in $z^{(L)}$ as claimed in Equation 9. Finally, changes in $z^{(L)}$ depend on the changes in $w^{(L)}$, $b^{(L)}$, and $a^{(L-1)}$ as shown in Equation 8. The activation of the previous neuron (i.e. $a^{(L-1)}$) cannot be controlled directly. Likewise, a back propagation of the cost function to layer $a^{(L-2)}$ is performed (Kolbusz et al., 2019). To investigate how $w^{(L)}$ affects C_0, $\frac{\partial C_0}{\partial w^{(L)}}$ is considered. With this aim, the chain rule can be applied as given in Equation 10.

$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial z^{(L)}}{\partial w^{(L)}}\frac{\partial a^{(L)}}{\partial z^{(L)}}\frac{\partial C_0}{\partial a^{(L)}} = a^{(L-1)}\sigma'\left(z^{(L)}\right)2\left(a^{(L)} - y\right) \tag{10}$$

Since Equation 10 is related to the cost of a single training example, we must take the average of the costs for all *n* training examples as shown in Equation 11.

$$\frac{\partial C}{\partial w^{(L)}} = \frac{1}{n}\sum_{k=0}^{n-1}\frac{\partial C_k}{\partial w^{(L)}} \tag{11}$$

Note that, Equation 11 gives only one component of the gradient vector which includes all the weights and biases in the neural network. Hence, Equation 12 gives the gradient vector of the cost function

$$\nabla C = \begin{bmatrix} \frac{\partial C}{\partial w^{(1)}} \\ \frac{\partial C}{\partial b^{(1)}} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial C}{\partial w^{(L)}} \\ \frac{\partial C}{\partial b^{(L)}} \end{bmatrix} \tag{12}$$

Similarly, for the sensitivity of the bias term to the cost function, we apply Equation 13.

$$\frac{\partial C_0}{\partial b^{(L)}} = \frac{\partial z^{(L)}}{\partial b^{(L)}}\frac{\partial a^{(L)}}{\partial z^{(L)}}\frac{\partial C_0}{\partial a^{(L)}} = \sigma'\left(z^{(L)}\right)2\left(a^{(L)} - y\right) \tag{13}$$

We should also investigate the sensitivity of the previous neuron activation ($a^{(L-1)}$) to the cost function as given in Equation 14.

$$\frac{\partial C_0}{\partial a^{(L-1)}} = \frac{\partial z^{(L)}}{\partial a^{(L-1)}}\frac{\partial a^{(L)}}{\partial z^{(L)}}\frac{\partial C_0}{\partial a^{(L)}} = w^{(L)}\sigma'\left(z^{(L)}\right)2\left(a^{(L)} - y\right) \tag{14}$$

When the neurons at all layers are considered, Equation 10 can be rearranged as given in Equation (15).

$$\frac{\partial C_0}{\partial w_{jk}^{(L)}} = \frac{\partial z_j^{(L)}}{\partial w_{jk}^{(L)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}} = a^{(L-1)} \sigma'\left(z^{(L)}\right) 2\left(a^{(L)} - y\right) \quad (15)$$

where, the indices *j* and *k* represent the connection of the neuron *j* at layer *L* to the neuron *k* at layer (*L*-1).

Similarly, for the effect of $a_k^{(L-1)}$ to the C_0, Equation 16 is applied.

$$\frac{\partial C_0}{\partial a_k^{(L-1)}} = \sum_{j=0}^{n_L-1} \frac{\partial z_j^{(L)}}{\partial a_k^{(L-1)}} \frac{\partial a_j^{(L)}}{\partial z_j^{(L)}} \frac{\partial C_0}{\partial a_j^{(L)}} = a^{L-1} \sigma'\left(z^{(L)}\right) 2\left(a^{(L)} - y\right) \quad (16)$$

Considering Equation 14, the elements of the gradient vector can be found as given in Equation 17.

$$\frac{\partial C}{\partial w_{jk}^{(l)}} = a_k^{(l-1)} \sigma'\left(z_j^{(l)}\right) \frac{\partial C}{\partial a_j^{(l)}} \quad (17)$$

4. Results and Discussion

For the experiments, a neural network with 100 input layer neurons that correspond to the benefit/weight (benefit per weight) values of the KP01 samples is used. Input layer values are normalized between 0 and 1. Three hidden layers, each with 500 neurons and 100 output layer neurons, whose output values are between 0 and 1, are employed in the architecture of the neural network. Each neuron at the output layer takes the role to decide whether an item is taken to the knapsack, or not. If the output layer neuron activation is greater than ½, the corresponding item is taken to the knapsack. Otherwise, if the output value is less than or equal to ½, the corresponding item is not taken. 'relu' function and 'sigmoid' function are used for hidden layer neuron activation and output layer activation, respectively. 'adam' optimizer is employed for the gradient descent learning of the neural network parameters. Binary cross-entropy loss function and 115 previously solved KP01 samples (100 for training, 15 for testing) are used. Each KP01 sample contains 100 items whose weight and benefit values are known.

Each data point on the graphs given in Figure 3 represents the mean of corresponding 25 runs. At each run, training and test samples are chosen randomly among 115 total samples. For the implementation, the Python Tensorflow machine learning library is employed. The accuracies achieved in training and test sets versus the training epochs are given in Figure 3.

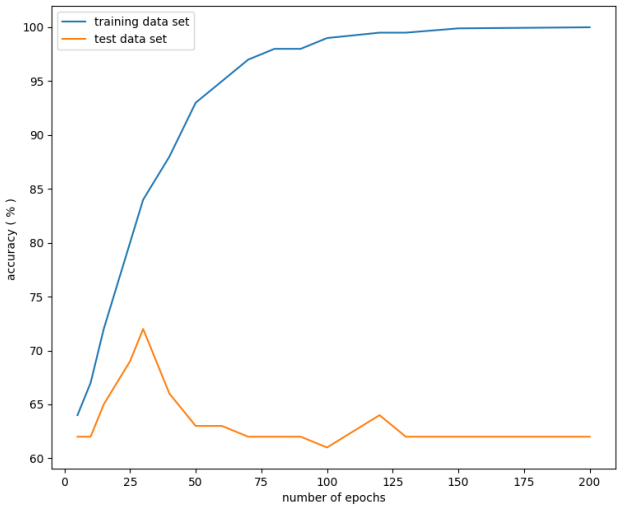


Figure 3. Training and test accuracies versus training epochs.

After 27 epochs, as the accuracy of the training data set continues to increase, the accuracy of the test data set starts to decrease. This is caused by the overfitting with the training data set. At the 27th epoch, 84% accuracy with the training data set and 72% accuracy with the test data set is achieved. At the 200th epoch, the accuracy with the training data set is 100% and the accuracy with the test data set is 63%.
In this preliminary work, although there is no comparison with other references, we believe that 72% accuracy for the test data provides a satisfactory reason for conducting further studies with large instances and different hard optimization problems.

5. Conclusion

A neural network learning approach for solving KP01 problem is tested. It is observed that the proposed method may be suitable for the applications where fast but approximate solutions are required for the optimization problems with similar training and test data sets. Although the experiments conducted in this work do not include large-sized problems, this study serves as a preliminary claim for the benefit that neural networks can provide to solution of the optimization problems. Therefore, in the future, the approach must be tested with large instances of different hard optimization problems.

References

Alpaydın E. (2020). Introduction to machine learning. MIT Press.

Basset, M.,A., Shahat, D.E., & Henawy, I.,E. (2017). Solving 0-1 knapsack problem by binary flower pollination algorithm. Neural Computing and Applications, 31, 5477-5495.

Kolbusz, J., Rozycki, P., Lysenko, O., & Wilamowski, B.N. (2019). Error back propagation algorithm with adaptive learning rate. International Conference on Information and Digital Technologies (IDT), 216-222. IEEE.

Lv, J., Wang, X., Huang, M., Cheng, H., & Li, F. (2016). Solving 0-1 knapsack problem by greedy degree and expectation efficiency. Applied Soft Computing, 41, 94-103.

Pan, X., & Zhang, T. (2018). Comparison and analysis of algorithms for the 0/1 knapsack problem. Journal of Physics: Conference Series 1069(1), 012024, 1-7, IOP Publishing.

Salkin, H. M., & De Kluyver, C. A. (1975). The knapsack problem: a survey. Naval Research Logistics Quarterly, 22(1), 127-144.

Shaheen, A., & Sleit, A. (2016). Comparing between different approaches to solve the 0/1 knapsack problem. International Journal of Computer Science and Network Security, 16(7), 1.

Wilbaut, C., Hanafi, S., & Salhi, S. (2008). A survey of effective heuristics and their application to a variety of knapsack problems. IMA journal of management Mathematics, 19(3), 227-244.

Zhang, N., Shen, S., L, Zhou, A., & Xu, Y., S. (2019). Investigation on performance of neural networks using quadratic relative error cost function. IEEE Access, 7, 106642-106652.

The Interaction of Von Kármán Vortices with the Solitons of the Complex Ginzburg-Landau Equation

Sofi Farazande[1*], Cihan Bayındır[1,2]

[1]*Boğaziçi University, Civil Engineering Department, İstanbul, Turkey*
[2]*İstanbul Technical University, Civil Engineering Department, İstanbul, Turkey*
sofi.farazande@boun.edu.tr, cbayindir@itu.edu.tr

Abstract

The complex Ginzburg-Landau (GL) equation is a well-known equation in various areas of physics that is also widely used to model vortex shedding phenomena occurring around a bluff body in a flow field, which is named as von Kármán vortex street (Roussopoulos and Monkewitz, 1996, Cohen et al., 2003). In addition, it describes nonlinear waves, second-order phase transitions, superconductivity, superfluidity, Bose-Einstein condensation, liquid crystals, and strings in field theory, etc. (Aranson and Kramer, 2002). Moreover, the GL equation can be utilized to find soliton solutions of many nonlinear systems (Akhmediev and Ankiewicz, 2001). Solitons are self-localized, solitary, nonlinear waves that emerge from a collision with a similar pulse having an unchanged shape and speed (Scott et al., 1973). Most of its applications lie in the domains of optics and fluid mechanics, which are attained by solutions of some familiar partial differential equations as Korteweg-de-Vries, modified Korteweg-de-Vries, Sine-Gordon, and nonlinear Schrödinger equations (Helal, 2002), apart from GL equation. In the present study, we aim to analyze the interaction of the soliton solutions of the GL equation with the von Kármán vortex street. For this purpose, we solve the GL equation via a spectral scheme that uses FFT routines for the space derivative and a 4th order Runge-Kutta time-stepping method to simulate vortices. Subsequently, we use the soliton solutions of GL constructed using analytical techniques and investigate their effects on Von Kármán vortices. We investigate how the vortex structure and stability are affected and whether the vortex fluctuations are reduced by the solitons. We discuss our findings and their possible usage in controlling the vortices by solitons for structural damage prevention and resonating for energy harvesting.

Keywords: von Kármán vortex, solitons of Ginzburg-Landau equation, numerical model, soliton-vortex interaction

1. Introduction

Vortices are subjects of a wide range of disciplines in fluid mechanics. They appear in the analysis of the lifting force of aircraft wings, whirlpools occurring in the ocean, the atmosphere of the planets, and are used even in the electromagnetic fields. Von Kármán vortex street is a pattern of repetitive vortices observed behind a blunt body in a flow field. In this study, the vortices are studied in the frame of the complex GL equation. Other than vortex street, GL equation is used to study nonlinear waves, second-order phase transitions, superconductivity, superfluidity, Bose-Einstein condensation, liquid crystals, and strings in field theory, etc. (Aranson and Kramer, 2002). GL equation is also used to model nonlinear phenomena in the form of solitons, which are the self-localized solutions of nonlinear systems with an infinite number of degrees of freedom. They can be formed by collision with a similar pulse having unchanged shape and speed (Scott et al., 1973). Apart from the GL equation, solitons can be obtained from the nonlinear Schrödinger and Schrödinger-KdV equations as well. Fractional order forms of such

* *Corresponding Author*

nonlinear wave equations have been solved by several numerical methods in previous studies (Yavuz et al., 2020, Yavuz et al., 2021).

In the present study, we investigate the interaction of the von Kármán vortex street and soliton solutions of the complex GL equation. For this purpose, we solve the complex GL equation numerically using the Fourier spectral method with a 4th order Runge-Kutta time integrator. We impose three soliton solutions to the equation as an external forcing term and show that hereby a significant control over the von Kármán vortex street can be provided. Finally, we discuss our findings and their possible usage in controlling the vortices by solitons.

2. Mathematical Formulation

Von Kármán vortices are wakes and vortices downstream of flow formed by the flow around a bluff body, named after Theodore von Kármán. A model for their study is the Ginzburg-Landau equation (GLE). A dimensionless form of the GLE is given by

$$\frac{\partial A}{\partial t} + U\frac{\partial A}{\partial x} = \mu(x)A + (1 + iC_D)\frac{\partial^2 A}{\partial x^2} - (1 - iC_N)|A|^2 A + F(x,t). \quad (1)$$

Here $A(x,t)$ shows the complex amplitude. The parameters U, C_D, and C_N are real constants that show the advection speed, diffusion, and nonlinearity. $\mu(x)$ is the wake growth parameter and is defined by

$$\mu(x) = \mu_0 + \mu' x. \quad (2)$$

The term $F(x,t)$ is the forcing function, for which the soliton solutions of the GLE and other nonlinear systems are used in this study (Akhmediev and Ankiewicz, 2001, Millot and Tchofo-Dinda, 2005). For the unforced simulations, $F(x,t) = 0$ is set. To simulate the forced cases, the equation is solved by substituting one bright soliton and two dark soliton solutions into the $F(x,t)$ term. The bright soliton solution of the complex GL equation used is (Akhmediev and Ankiewicz, 2001)

$$F(x,t) = a(x)\exp(i\phi(x))\exp(-i\omega t) \quad (3)$$

where $\phi(x) = \phi_0 + d\ln[a(x)]$ and $a(x)$ is defined as

$$a(x) = BC\,\mathrm{sech}(Bx) \quad (4)$$

where

$$C = \sqrt{\frac{3d\lambda^2}{2(2\beta-\epsilon)}}$$

$$B = \sqrt{\frac{\delta}{d-\beta+\beta d^2}} \quad (5)$$

$$d = d_{\pm} = \frac{3(1+2\epsilon\beta) \pm \sqrt{9(1+2\epsilon\beta)^2 + 8(\epsilon-2\beta)^2}}{2(\epsilon-2\beta)}$$

$$\omega = -\frac{\delta(1-d^2+4\beta d)}{2(d-\beta+\beta d^2)}$$

and δ, β and ϵ are real constants. δ is the linear gain (or loss) at the (spatial or temporal) central frequency, β is the spatial or temporal spectral filtering which is also called diffusion coefficient and has a condition of $\beta > 0$, and ϵ is related to the nonlinear gain or absorption process (Akhmediev and Ankiewicz, 2001).

Our first dark soliton solution has a similar structure with Eq. (3). For the case where the transverse velocity of the solution does not exist, its formula reduces to the following:

$$F(x,t) = \left[\kappa u \tanh(\kappa x) + (d + 2i)\frac{w}{d}\right]\exp[i\phi(x)]\exp(iKx - i\Omega t) \quad (6)$$

where

$$d = \frac{3(1+2\epsilon\beta) \pm \sqrt{9(1+2\epsilon\beta)^2 + 8(\epsilon-2\beta)^2}}{2(\epsilon-2\beta)}$$

$$u^2 = \frac{3d\lambda^2}{2(2\beta-\epsilon)}$$

$$\phi(x) = d\ln[\cosh(\kappa x)] \tag{7}$$
$$\Omega = \frac{\delta}{2\beta} + \frac{2\beta-\epsilon}{2\epsilon}\left(K^2 - \kappa^2 d^2 + \frac{\delta}{\beta}\right)$$
$$K = (\epsilon u)(w/d)$$

and again δ, β, ϵ, K and w are real constants. The definition of λ is the same as the bright soliton case. The other dark soliton solution is a soliton of the nonlinear Schrödinger equation, which is a reduced form of the complex GL equation. The nonlinear Schrödinger equation has dissipative soliton solutions as well (Bayındır, 2016). The formulation that we use for dark soliton is as follows (Millot and Tchofo-Dinda, 2005)

$$F(x,t) = \pm\sqrt{P_0}\tanh(t/\delta_0)\exp(i\gamma P_0 x) \tag{8}$$

where P_0 is the power required to create a soliton, and δ_0 is the initial pulse width. In our study, we calculate P_0 as

$$P_0 = \int_0^L |A|^2 dx \tag{9}$$

where L is the length of the domain. In order to analyze the effect of the presence of the soliton forcing on the von Kármán vortices, we solve the GLE using a 4th order Runge-Kutta (RK4) time-stepping algorithm and Fourier spectral method. This method is briefly summarized here. We start by rewriting Eq. (1) as

$$\frac{\partial A}{\partial t} = -U\frac{\partial A}{\partial x} + \mu(x)A + (1+iC_D)\frac{\partial^2 A}{\partial x^2} - (1+iC_N)|A|^2 A + F(x,t) = g(x,t). \tag{10}$$

Then the four slopes of the RK4 algorithm are calculated by

$$m_1 = g(A_n, F_n(x,t_n), t_n, x)$$
$$m_2 = g(A_n + 0.5m_1 dt, F_n(x, t_n + 0.5dt), t_n + 0.5dt, x)$$
$$m_3 = g(A_n + 0.5m_2 dt, F_n(x, t_n + 0.5dt), t_n + 0.5dt, x) \tag{11}$$
$$m_4 = g(A_n + m_3 dt, F_n(x, t_n + dt), t_n + dt, x)$$

Starting from the initial conditions, the value of the complex amplitude and time at the next time steps are evaluated by

$$A_{n+1} = A_n + dt(m_1 + 2m_2 + 2m_3 + m_4)/6 \tag{12}$$
$$t_{n+1} = t_n + dt$$

The time step is selected as $dt = 0.005$ which does not cause any forms of instability. The spatial derivatives in Eq. (3) can be computed spectrally using Fourier series by

$$\frac{\partial A}{\partial x} = FFT^{-1}[ikFFT[A]] \quad \text{and} \quad \frac{\partial^2 A}{\partial x^2} = FFT^{-1}[k^2 FFT[A]]. \tag{13}$$

Here, FFT denotes the Fourier and FFT^{-1} denotes the inverse fast Fourier transform operations. The parameter k is the wavenumber vector which has entries of $N = 512$ multiples of the fundamental wavenumber vector, k_0. In our simulations, the length of domain is selected to be $L = 120$ and the corresponding fundamental wavenumber, $k_0 = 2\pi/L$, is used.

3. Results and Discussion

In order to illustrate the effects of soliton forcing on the dynamics and characteristics of the vortices, we first perform the numerical simulation of the GLE for the unforced case. With this motivation, the computational parameters are selected as $U = 5, C_D = 1, C_N = 0, \mu_0 = 3.57, \mu' = -0.0434$ (Cohen et. al, 2003). The vortices found this way are depicted in blue in Figs.1-3. In order to investigate the effects of soliton forcing, we use the analytical soliton solutions given by Eqs. (3), (6), and (8). For the bright soliton solution, the computational parameters are selected as $\delta = -0.03$, $\beta = 0.9$, and $\epsilon = 0.8$. For the first dark soliton given by Eq. (6), the computational parameters are set to be $\delta = -0.01$, $\beta = 0.1$, $\epsilon = 1.5$, $K = 0$ and $w = 0$. For the second dark soliton solution given by Eq. (8) the value of $\delta_0 = 0.1$ is used. The initial condition for all cases is selected as $A(x,0) = 10^{-4}$. We selected the values of δ, β, ϵ and δ_0 with trial-and-error procedure so that the control over the vortex and their suppression are

achieved as much as possible and their stabilities are not adversely affected. We depict our findings for the soliton-controlled case in Figs. 1-3 below in red. For all the simulations, a simulation time of $t \approx 60$ is used.

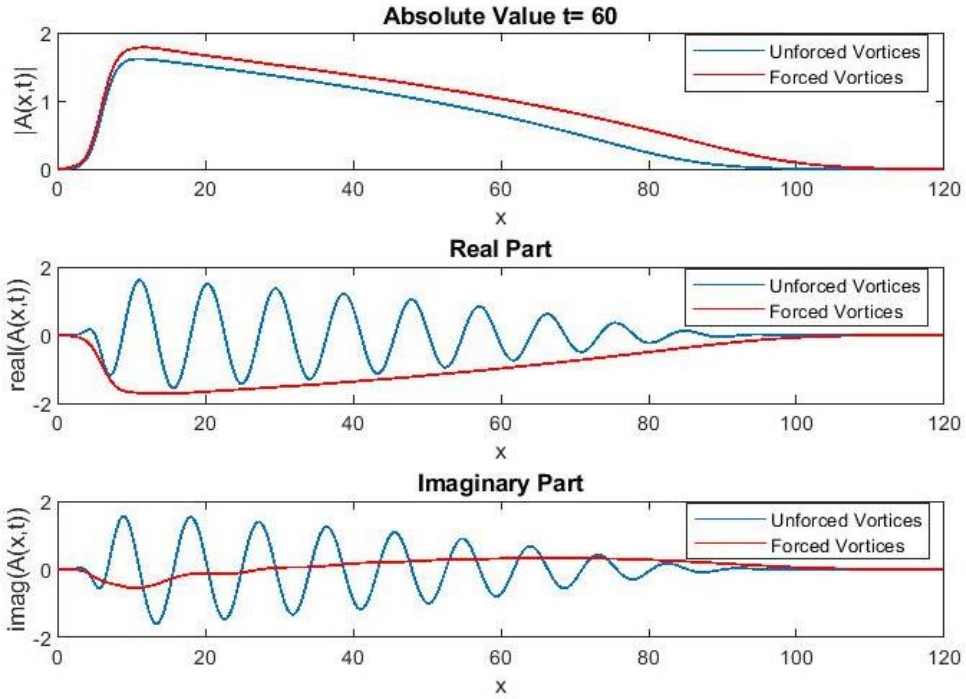


Figure 1. Simulation of vortices downstream of a circular cylinder at $x = 0$ and their interaction with a bright soliton at $x = 10$.

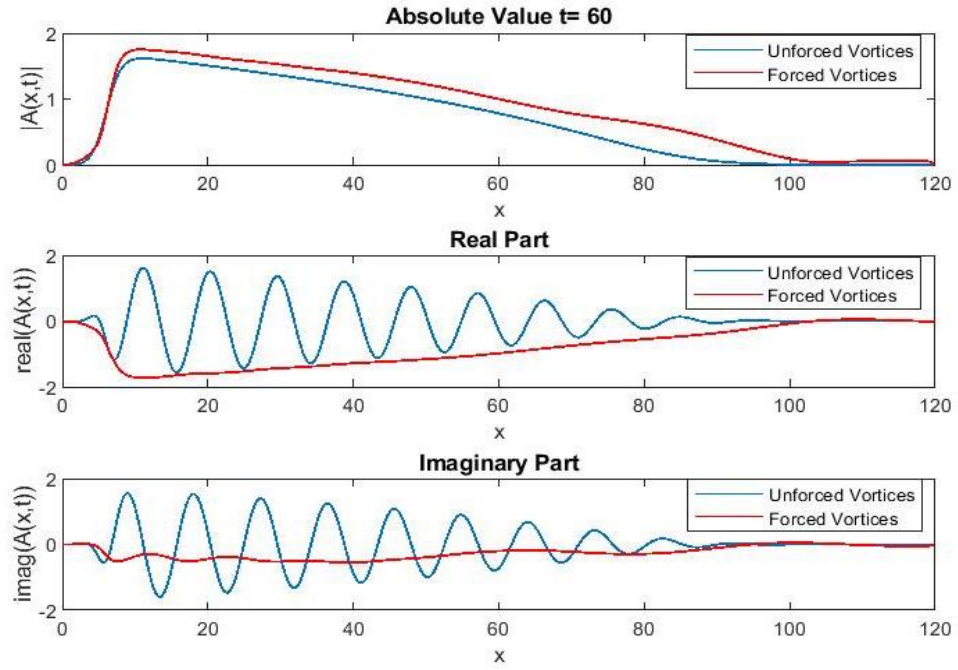


Figure 2. Simulation of vortices downstream of a circular cylinder at $x = 0$ and their interaction with a dark soliton at $x = 10$.

In all three figures, the unforced vortex street has the same parameters and therefore the same pattern. It represents the growing wake oscillations will be observed downstream of the flow around a circular

cylinder located at $x = 0$. The vortex shedding is detectable up to $x \approx 90$ for a simulation time of $t \approx 60$ There is a decay in the amplitude as we move away from the location of the circular cylinder. For the unforced case, $F(x,t) = 0$ is used in Eq. (1).

The forced cases represent the interaction of the von Kármán vortex with the soliton solutions. In this case, the forcing term $F(x,t)$ in Fig. 1, Fig. 2 and Fig. 3 are given by Eqs. (3), (6) and (8), respectively. In line with our purpose, visible control over the vortices and their suppression is provided. As illustrated in Fig. 1 and Fig. 2, the bright soliton and the dark soliton reduce the number of vortex cycles after the bluff body to nearly null in the real part and only a slight increase in the amplitude is observed. The dark soliton forcing with a power P_0 plotted in Fig. 3 similarly provides a reduction in the number of cycles. Moreover, the first wake in the vortex street is moved closer to the bluff body located at $x = 0$. The stability is established more resiliently in forced solutions of each case.

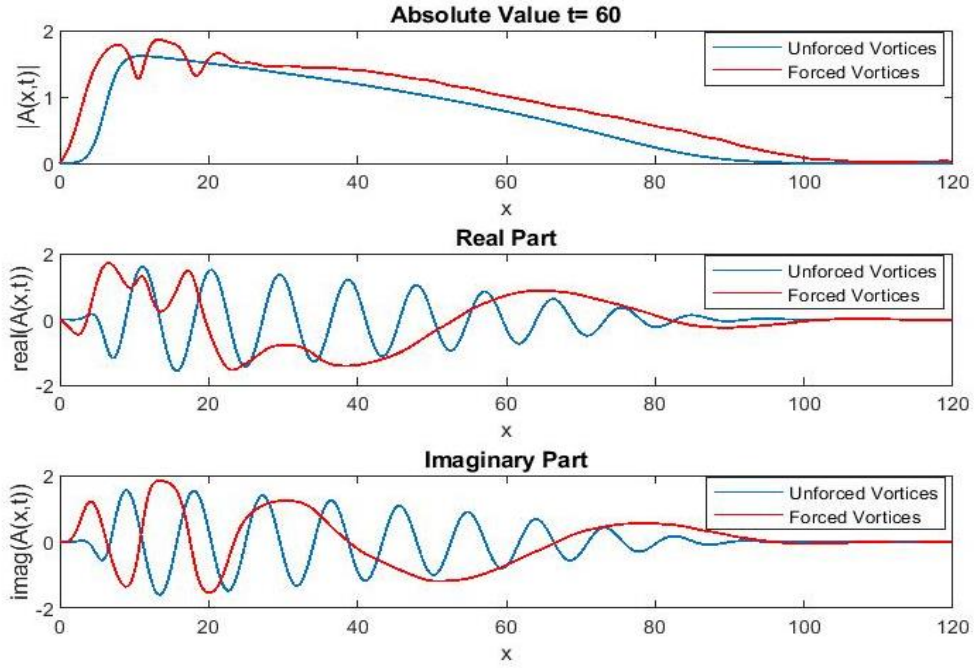


Figure 3. Simulation of vortices downstream of a circular cylinder at $x = 0$ and their interaction with a dark soliton at $x = 10$ with a power P_0.

These findings can be used to reduce the effects of vortex shedding on natural or artificial structures, such as vibration-induced resonance and fatigue. Besides, we showed that with soliton-formed feedback forcing, we can control the amount and decrease/control the frequency of the vortex shedding. These results can shed light upon many nonlinear vortex-induced vibration studies and their control via nonlinear mechanics. In the future, we aim to extend the study by applying FFT analysis to investigate the vortex dynamics in the spectral domain. Furthermore, a wavelet analysis will give clues about the location of the change of spectral properties, their possible benefits for many engineering applications.

4. Conclusion

In this paper, we have investigated the effects of the interaction of solitons of the complex Ginzburg-Landau equation with the von Kármán vortex street. We showed that applying a soliton forcing upstream of the flow close to the bluff body, causes the vortex shedding dynamics and characteristics to remarkably change. Those changes appear to be a significant reduction in the vortex cycles and even their suppression. An insignificant amount of change is also observed in the amplitude of oscillations for many engineering purposes. Our results can bring many challenges and lead to new mechanisms for

the optimal control of the vortices for engineering applications such as reducing flow-induced vibrations and resonances. The feedback forces in the soliton form or interaction of solitons with vortices can be used for this purpose. As a further investigation, we aim to analyze the spectral effects of the soliton on the dynamics of vortex streets by performing an FFT or wavelet analysis.

Acknowledgements

The authors gratefully acknowledge the support of the İstanbul Technical University. This work was supported by the Research Fund of the İstanbul Technical University. Project Code: MGA-2020-42544. Project Number: 42544.

References

Akhmediev, N. & Ankiewicz, A. (2001). Solitons of the complex Ginzburg-Landau equation. Spatial Solutions, 311-341. https://doi.org/10.1007/978-3-540-44582-1_12.

Aranson, I. S. & Kramer, L. (2002). The world of the complex Ginzburg-Landau equation. Rev. Mod. Phys. 74, 99. https://doi.org/10.1103/RevModPhys.74.99.

Bayındır, C. (2016). Analytical and numerical aspects of the dissipative nonlinear Schrödinger equation. TWMS Journal of Applied and Engineering Mathematics, 6, 135-142.

Cohen, K., Siegel, S., McLaughling, T. & Myatt, J. (2003). Fuzzy logic control of circular cylinder vortex shedding model, 41st Aerospace Sciences Meeting and Exhibit AIAA 2003-1290. https://doi.org/10.2514/6.2003-1290.

Helal, M. A. (2002). Solution solutions of some nonlinear partial differential equations and its applications in fluid mechanics. Chaos, Solitons & Fractals, 13, 1917-1929. https://doi.org/10.1016/S0960-0779(01)00189-8.

Millot, G. & Tchofo-Dinda, P. (2005). Solitons/Optical Fiber Solitons, Physical Origin and Properties. Encyclopedia of Modern Optics, 56-65. https://doi.org/10.1016/B0-12-369395-0/01246-X

Roussopoulos, K., & Monkewitz, P. A. (1996). Nonlinear Modeling of Vortex Shedding Control in Cylinder Wakes. Physica D, 97, 264-273. https://doi.org/10.1016/0167-2789(96)00151-0

Scott, A. C., Chu, F. Y. F. & Mc Laughlin, D. W. (1973). The soliton: A new concept in applied science. Proceeding of the IEEE, 61, 1443-1483. https://doi.org/10.1109/PROC.1973.9296.

Yavuz, M., Sulaiman, T. A., Usta, F., Bulut, H. (2020). Analysis and numerical computations of the fractional regularized long-wave equation with damping term. Mathematical Methods in Applied Sciences, 44, 7538-7555. https://doi.org/10.1002/mma.6343.

Yavuz, M., Sulaiman, T. A., Yusuf, A., Abdeljawad, T. (2021). The Schrödinger-KdV equation of fractional order with Mittag-Leffler nonsingular kernel. Alexandria Engineering Journal, 60, 2715-2724. https://doi.org/10.1016/j.aej.2021.01.009.

Multi-derivative, Multi-stage and Multi-step Time Integration Methods

Huseyin Tunc[1*], Murat Sarı[2]

Yildiz Technical University, Mathematics Department, Istanbul, Turkey
[1]tnchsyn@gmail.com, [2]sarim@yildiz.edu.tr

Abstract

Implicit methods can be considered as the best options for solving stiff initial value problems. The implicit time-integration algorithms are computationally costly, especially for solving large nonlinear systems. Then, an implicit approach must be optimized in terms of local degrees of freedom of algebraic systems of equations. The multi-stage implicit algorithms such as the implicit Runge-Kutta methods (IRKM), implicit Lobatto methods or implicit Radau methods have more degrees of freedom than the linear multi-step methods. The linear multi-step methods such as the backward differentiation formulations (BDFs), the Adam-Bashford methods (ABMs) and the Adam-Moulton methods (AMMs) take advantage of the optimized degrees of freedom. However, this set of algorithms has drawbacks of storage and order-preservation, especially for stiff problems. Here we show that multi-step methods generally require fewer time steps than the multi-stage methods to achieve the same accuracy for solving stiff problems. An inevitable question arises here: Could a stiff problem solver have both optimized degrees of freedom and order preservation? It has been proven here that the implicit-explicit local differential transformation method (IELDTM) as a multi-derivative method has both of the vital features, by eliminating the existing disadvantages of explicit approaches originating from the differential transformation method (DTM).

Keywords: Time integration, implicit algorithm, stiff problem, initial value problem, Taylor series

1. Introduction

Differential equations arise in modelling various natural phenomena, and their solutions provide valuable information about the related dynamical systems. In general, exact solutions of the equations cannot be derived all the time easily, or analytical expressions are very complicated to observe the behaviors of physical systems. Even for the large linear equation systems, analytical evaluations are not easy to implement, and symbolic calculations lead to enormous computational time. These drawbacks of analytical approaches can be handled by considering accurate and economic numerical methods. With the appropriate selection of parameters used in a numerical method, the convergence of the method can generally be controlled. However, stiff differential equations are not easy to handle and find reliable numerical solutions in the entire domain. Stiff behaviors are modelled by not only ordinary differential equations (ODEs) but also partial differential equations (PDEs) (Hairer and Wanner, 1996). Numerical techniques are also classified according to stiff and non-stiff problems in the literature (Hairer et al., 1993-1996). The methods constructed to solve stiff problems are of particular interest due to the instability of classical methods. Unwanted oscillations, divergence or slow convergence are commonly faced in solving stiff problems under the consideration of some inappropriate numerical techniques.

Stiff initial value problems (IVPs) are encountered in chemical kinetics, nonlinear mechanics, fluid dynamics, biochemistry and so on (Hairer et al., 1996). There are three main groups of time integration methods for IVPs: multi-step, multi-stage and multi-derivative (Hairer et al., 1993; Hairer and Wanner,

* *Corresponding Author*

1996). The multi-stage implicit algorithms such as the Norsett's diagonally implicit Runge-Kutta method (NDIRKM) (Al-Rabeh, 1993), implicit Lobatto methods (LBM) (Pinto, 1997) or implicit Radau methods (RDM) (Ding and Tan, 2009) have more local degrees of freedom than the implicit multi-step and multi-stage methods. On the other hand, the implicit linear multi-step methods such as the backward differentiation formulae (BDF) (Akinfenwa et al., 2013) and the Adam-Moulton methods (AMM) (Han and Han, 2002) take advantage of the optimized degrees of freedom. Nevertheless, the linear multi-step methods have two crucial drawbacks: order reduction issues when the higher-order methods are hybridized with low order ones for bootstrapping and storage issues. Thirdly, the multi derivative time-integration methods such as the local differential transform method (Tunc and Sari, 2019), automatic differentiation method (Abad et al., 2015), Taylor series methods (Ernsthausen and Nedialkov, 2020) and implicit-explicit local differential transform method (IELDTM) (Tunc and Sari, 2021a,b) provide optimized degrees of freedom and highly accurate results for solving IVPs. The IELDTM as a multi-derivative time integration method (Tunc and Sari, 2021a,b) was seen to have distinct numerical characteristics and eliminates the well-known stability disadvantages of the explicit ones. The IELDTM is a direction free, arbitrarily high order and stability preserved time integration method for solving IVPs for ODEs (Tunc and Sari, 2021a) and IVPs reduced from parabolic PDEs (Tunc and Sari, 2021b).
This study aims to compare the multi-derivative, multi-stage and multi-step time integration methods for solving stiff nonlinear advection-diffusion equation (the Burgers equation). The Burgers equation is reduced to a system of ODEs utilizing the Chebyshev spectral collocation method (ChSCM) in space. The performance results of all implicit time integration algorithms are extensively illustrated qualitatively and quantitatively. Throughout the study, we discuss both the advantages and disadvantages of each group of time-integration techniques.

2. Model Equation and Spatial Discretization

Behaviors of many physical processes encountered in models of advection mechanisms and diffusion transports lead to the Burgers equation. Thus, the equation arising in various physical areas of science are considered,

$$u_t + uu_x = \varepsilon u_x x,\ a \le x \le b \tag{1}$$

with the boundary conditions

$$u(a,t) = f_1(t) \quad \text{and } u(b,t) = f_2(t), \quad t > 0 \tag{2}$$

and initial condition

$$u(x,0) = g(x),\ a < x < b \tag{3}$$

where $\varepsilon > 0$ is the kinematic viscosity constant and f_1, f_2 and g are known functions. Applying the ChSCM to equation (1) (Tunc and Sari, 2021b) and imposing the boundary conditions leads to the following nonlinear IVP,

$$\frac{d\boldsymbol{c}}{dt} = f(\boldsymbol{c},t) = \varepsilon\bar{B}\boldsymbol{c} - \langle \boldsymbol{c}, \bar{A}\boldsymbol{c}\rangle + \boldsymbol{F}(\boldsymbol{c},t), \tag{4}$$
$$\boldsymbol{c}(0) = g(\boldsymbol{x})$$

where $\boldsymbol{c}(t) = [u(x_0,t), u(x_1,t), \dots, u(x_N,t)]^T, \boldsymbol{x} = [x_0, x_1, \dots, x_N]^T$, x_i are the Chebyshev-Gauss-Lobatto (CGL) collocation points, N is the order of polynomial approximation, $\langle .,. \rangle$ is the elementwise product operator**,** $\bar{A}$ is the reduced form of the Chebyshev differentiation matrix (CDM) A and $\bar{B} = \bar{A}^2$ (Tunc and Sari, 2021b). Imposing boundary conditions lead to reduction of the CDM A to $\bar{A}$ by removing the fist and the last rows and columns of A. Additionally, imposing boundary conditions yield the column vector $\boldsymbol{F}(\boldsymbol{c},t)$ defined as

$$\boldsymbol{F}(\boldsymbol{c},t) = \varepsilon\left(B_{:1}f_1(t) + B_{:(N+1)}f_2(t)\right) - \langle \boldsymbol{c}, A_{:1}f_1(t) + A_{:(N+1)}f_2(t)\rangle$$

where $B_{:1}$, $B_{:(N+1)}$, $A_{:1}$ and $A_{:(N+1)}$ denote first and last columns of the matrices A and B expect that first and last elements. Since $u(x,0) = g(x)$, the initial conditions at the CGL collocation points yield the initial condition of the IVP as $\boldsymbol{c}(0) = g(\boldsymbol{x})$.

3. Time Integration Methods

This section introduces some implicit multi-step, multi-stage and multi-derivative time-integration methods for solving IVP (4).

3.1 Multi-step Methods

The general k −step multi-step method for solving equation (4) can be expressed as

$$\sum_{j=0}^{k} \alpha_j \boldsymbol{c}_{n+j} = \Delta t \sum_{j=0}^{k} \beta_j \boldsymbol{f}_{n+j} \tag{5}$$

where $\alpha_k \neq 0$ and $\Delta t = \frac{t_f}{M}$. By defining k, α_j and β_j various multi-step algorithms can be derived. Implicit-explicit forms are available depending on the selection of β_k as follows:

- If $\beta_k = 0$, the method is explicit.
- If $\beta_k \neq 0$, the method is implicit.

Implicit linear multi-step methods yield optimized degrees of freedom irrespective of k, i.e. $degrees\ of\ freedom\ (dof) = N - 1$. In this study, we consider the backward differentiation methods (BDM) (Akinfenwa et al., 2013) and the Adam-Moulton methods (AMM) (Han and Han, 2002) illustrated in Table 1.

Table 1. Iteration formulae of the backward differentiation and Adam-Bashford methods with various orders

Group	Order	Formulae
BDM	2	$c_{n+2} - \frac{4}{3}c_{n+1} + \frac{1}{3}c_n = \frac{2}{3}\Delta t f_{n+2}, n = 0,1,..,M-2$
	3	$\boldsymbol{c}_{n+3} - \frac{18}{11}\boldsymbol{c}_{n+2} + \frac{9}{11}\boldsymbol{c}_{n+1} - \frac{2}{11}\boldsymbol{c}_n = \frac{6}{11}\Delta t f_{n+3}, n = 0,1,..,M-3$
	4	$c_{n+4} - \frac{48}{25}c_{n+3} + \frac{36}{25}c_{n+2} - \frac{16}{25}c_{n+1} + \frac{3}{25}c_n = \frac{12}{25}\Delta t f_{n+4}, n = 0,1,..,M-4$
AMM	2	$c_{n+1} = c_n + \frac{1}{2}\Delta t(f_{n+1} + f_n), n = 0,1,..,M-1$
	3	$\boldsymbol{c}_{n+2} = \boldsymbol{c}_{n+1} + \Delta t\left(\frac{5}{12}f_{n+2} + \frac{2}{3}f_{n+1} - \frac{1}{12}f_n\right), n = 0,1,..,M-2$
	4	$c_{n+3} = c_{n+2} + \Delta t\left(\frac{9}{24}f_{n+3} + \frac{19}{24}f_{n+2} - \frac{5}{24}f_{n+1} + \frac{1}{24}f_n\right), n = 0,1,..,M-3$

3.2 The Multi-stage Methods

The generalized implicit-explicit Runge–Kutta type multi-stage methods can be stated as follows:

$$k_i = f\left(t_n + c_i\Delta t, y_n + \Delta t \sum_{j=1}^{s} a_{ij}k_j\right) \tag{6}$$

$$\boldsymbol{c}_{n+1} = \boldsymbol{c}_n + \Delta t \sum_{i=1}^{s} b_i k_i$$

where s is the stage number, and the corresponding Butcher table takes the form

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \dots & a_{1s} \\ c_2 & a_{21} & a_{22} & \dots & a_{2s} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_s & a_{s1} & a_{s2} & \dots & a_{ss} \\ \hline & b_1 & b_2 & \dots & b_s \end{array}$$

Implicit and explicit multi-stage algorithms can be given by the following inferences:

- $a_{ij} = 0$ for $j \geq i$ yields explicit multi-stage methods,
- $a_{ij} \neq 0$ for any $j \geq i$ yields implicit RK methods.

In this study, we consider three widely used implicit multi-stage methods: Nørsett's three-stage, fourth-order diagonally implicit Runge–Kutta method (DIRKM), the fourth-order Lobatto IIIB method (LBM4) and the third-order Radau IIA method (RDM3). The Butcher tables of these three multi-stage methods can be found in the literature (Hairer and Wanner, 1996).

3.3 Multi-derivative Methods

The general Taylor series based multi-derivative methods yield the following type of local approximate solutions

$$\boldsymbol{c}_i(t) = \sum_{k=0}^{K} C_i(k)(t - t_i)^k + O((t - t_i)^{K+1}),\ t_i - \rho^i \leq t \leq t_i + \rho^i \quad (7)$$

where $C_i(k) = \frac{1}{k!}\frac{d^k \boldsymbol{c}}{dt^k}$ is the differential transform of function $\boldsymbol{c}(t)$ about $t = t_i$. Taking the differential transform of main equation (4) provides a recursive relation and $C_i(k)$ can be written in terms of $C_i(0)$ for all k values. By assuming local representation (7), various versions of the multi derivative time-integration methods such as the local differential transform method (Tunc and Sari, 2019), automatic differentiation method (Abad et al., 2015), Taylor series methods (Ernsthausen and Nedialkov, 2020) and implicit-explicit local differential transform method (IELDTM) has been studied for solving IVPs in the literature (Tunc and Sari, 2021a,b).

In this study, we consider the recently derived IELDTM (Tunc and Sari, 2021a,b) for the calculation of $C_i(0)$ using the following continuity condition:

$$\boldsymbol{c}_{i+1}(t_i + (1-\theta)\Delta t_i) = \boldsymbol{c}_i(t_i + (1-\theta)\Delta t_i), \quad (8)$$

$$\sum_{k=0}^{K} C_{i+1}(k)(-\theta\Delta t_i)^k = \sum_{k=0}^{K} C_i(k)\big((1-\theta)\Delta t_i\big)^k + O((\Delta t_i)^{K+1}, \theta) \quad (9)$$

where $\theta \in [0,1]$. Depending on the choice of θ, various $A-$ and $L-$ stable cases of the IELDTM can be obtained with the optimized degrees of freedom (Tunc and Sari, 2021a,b).

4. Numerical Experiments

This section provides some numerical illustrations for solving two challenging test problems, and the described methods are compared through their computational mechanisms.

Problem 1 (Tunc and Sari, 2021b)

Consider the Burgers equation (1) with initial condition

$$u(x,0) = sin\,\pi x,\ 0 < x < 1 \quad (10)$$

and the homogenous boundary conditions

$$u(0,t) = u(1,t) = 0,\ t > 0. \quad (11)$$

The exact solution of equation (1) under consideration of conditions (10)-(11) can be seen in the literature (Tunc and Sari, 2021b).

Problem 2 (Seydaoglu, 2018)

Consider the Burgers equation (1) with the polynomial type of initial condition

$$u(x,0) = 4x(1-x),\ 0 < x < 1 \quad (12)$$

and the homogenous boundary conditions

$$u(0,t) = u(1,t) = 0,\ t > 0. \quad (13)$$

The literature (Seydaoglu, 2018) covers the exact solution of equation (1) with initial and boundary conditions (12)-(13).

In Figures 1A-B and 2A-B, the multi-derivative (CIELDTM4, CIELDTM3, CIELDTM2, CIELDTM1, BIELDTM4, BIELDTM3, BIELDTM2), multi-stage (NDIRKM4, RDM4, LBM4) and multi-step (BDF4, BDF3, BDF2 and AMM4) time integration methods are compared considering the step-sizes that are the threshold values to get $\|E\|_\infty = 10^{-5}$ error for solving Problems 1-2 with the parameter values $\varepsilon = 0.001$ and $N = 40$. The corresponding CPU times to produce the results illustrated in Figures 1A and 2A are comparatively shown in Figures 1B and 2B, respectively. According to Figures 1A-B and 2A-B, the following observations can be stated:

- Order reductions of the implicit multi-step algorithms BDM and AMM are seen from the figures for solving the stiff ODE system as opposed to the IELDTMs and the implicit multi-stage methods.
- Although the implicit multi-stage methods NDIRKM4, RDM4 and LBM4 satisfy order-preservation, these algorithms yield much more local degrees of freedom than the rival ones.
- The IELDTMs have better numerical characteristics by providing both optimized degrees of freedom and order-preservation property.
- In terms of the required CPU times, the CIELDTM3 is the winner of the comparison. The RDM3, BIELDTM3 and CIELDTM4 follow the CIELDTM3, respectively for Problem 1. The RDM3, CIELDTM4 and BIELDTM3 follow the CIELDTM3, respectively for Problem 2. The lowest performances belong to the AMM4 for both problems due to the stability drawback and order reduction.
- Overall, the IELDTMs are proven to solve ODEs reduced from singularly perturbed PDEs with optimal computational cost.

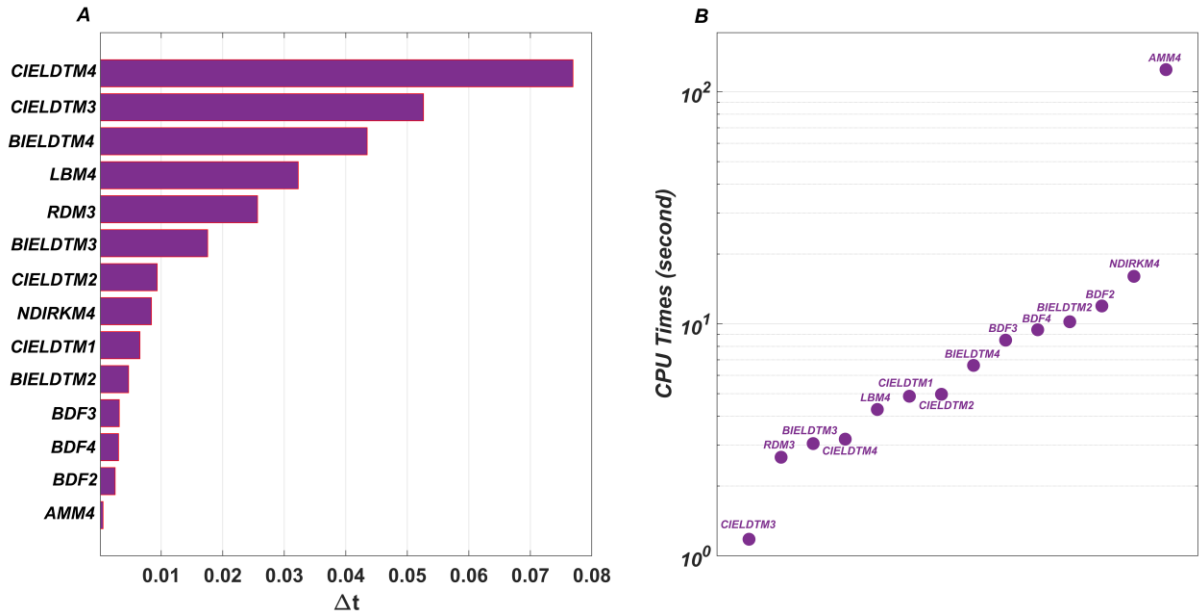


Figure 1. A) Comparison of the various time-integration methods in terms of the required step sizes to get $\|E\|_\infty = 10^{-5}$ for solving Problem 1 with $\varepsilon = 0.01$ and $N = 40$, B) Comparison of the various

time-integration methods in terms of the CPU times to get $\|E\|_\infty = 10^{-5}$ for solving Problem 1 with $\varepsilon = 0.01$ and $N = 40$.

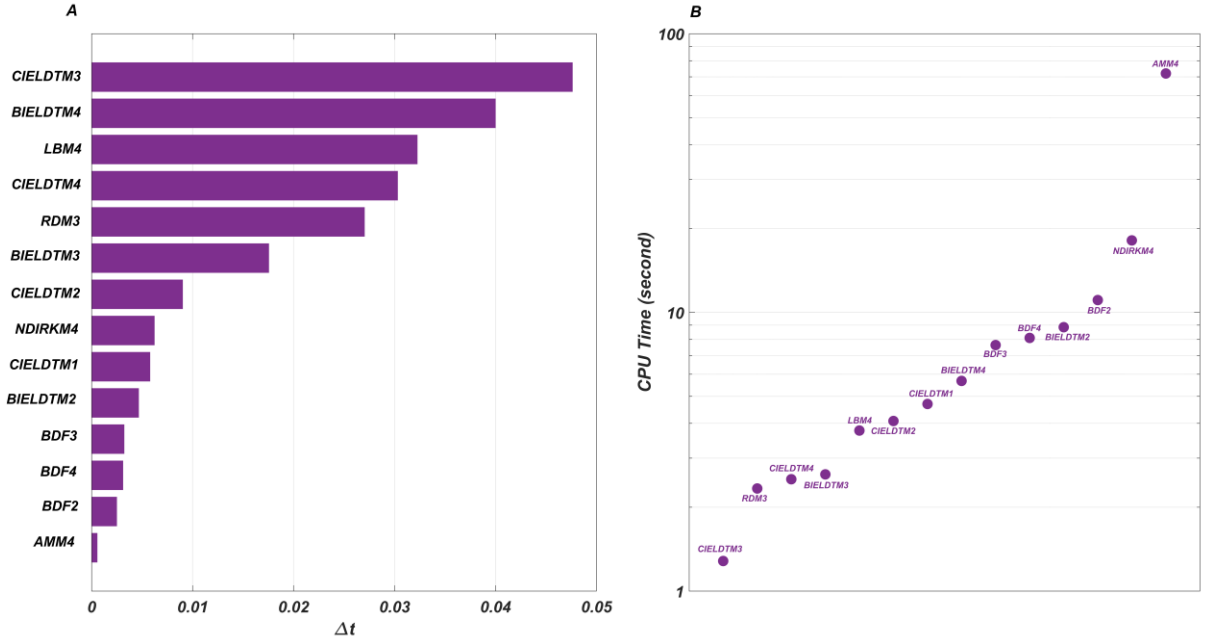


Figure 2. A) Comparison of the various time-integration methods in terms of the required step sizes to get $\|E\|_\infty = 10^{-5}$ for solving Problem 2 with $\varepsilon = 0.01$ and $N = 40$, B) Comparison of the various time-integration methods in terms of the CPU times to get $\|E\|_\infty = 10^{-5}$ for solving Problem 2 with $\varepsilon = 0.01$ and $N = 40$.

5. Conclusion

In this study, various implicit time-integration techniques have been compared for solving stiff advection-diffusion equations in terms of accuracy and computational cost. The IELDTM as a multi-derivative time integration technique has been the best option for solving the current system of IVPs reduced from the Burgers equation. The IELDTM has been seen to be a robust and versatile time-integration technique for solving stiff IVPs by providing both order-preserving property and optimized *dof*. According to our numerical experiments, the multi-stage methods have been observed to preserve the numerical orders and yield large local *dof*. Additionally, the multi-step methods have been proven to need smaller time increments and larger CPU times than the rival methods to get the same accuracy.

References

Abad, A., Barrio, R., Marco-Buzunariz, M. & Rodríguez, M. (2015). Automatic implementation of the numerical Taylor series method: A MATHEMATICA and SAGE approach. Applied Mathematics and Computation, 268, 227–245. https://doi.org/10.1016/j.amc.2015.06.042

Akinfenwa, O.A., Jator, S.N. & Yao, N.M. (2013). Continuous block backward differentiation formula for solving stiff ordinary differential equations. Computers & Mathematics with Applications, 65(7), 996-1005. https://www.sciencedirect.com/science/article/pii/S0898122113000564

Al-Rabeh, A.H. (1993). Optimal order diagonally implicit Runge-Kutta methods. BIT Numerical Mathematics, 33, 619–633. https://link.springer.com/article/10.1007%252FBF01990538

Curtiss, C. & Hirschfelder, J. (1952). Integration of stiff equations. Proceedings of the National Academy of Sciences, 38, 235–243. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1063538/

Ding, X. & Tan, J. (2009). Implicit Runge–Kutta methods based on Radau quadrature formula. International Journal of Computer Mathematics, 86(8), 1394-1404. https://doi.org/10.1080/00207160701870878

Ernsthausen, J. & Nedialkov, N. (2020). Stepsize selection in the rigorous defect control of Taylor series methods. Journal of Computational and Applied Mathematics, 368, 112483. https://doi.org/10.1016/j.cam.2019.112483

Hairer, E., Nørsett, S. & Wanner, G. (1993). Solving ordinary differential equations I -nonstiff problems. Springer.

Hairer, E. & Wanner, G. (1996). Solving ordinary differential equations II: stiff and differential-algebraic equations. Springer. https://www.springer.com/gp/book/9783540604525

Han, T.M. & Han, Y. (2002). Solving implicit equations arising from Adams-Moulton methods. BIT Numerical Mathematics, 42, 336–350. https://link.springer.com/article/10.1023/A:1021951025649

Pinto, S.G., Rodriguez, S.P.B& Torcal, J.I.M. (1997). On the numerical solution of stiff IVPs by Lobatto IIIA Runge-Kutta methods. Journal of Computational and Applied Mathematics, 82, 129-148. https://www.sciencedirect.com/science/article/pii/S0377042797000861

Seydaoglu, M. (2018). An accurate approximation algorithm for Burgers' equation in the presence of small viscosity. Journal of Computational and Applied Mathematics, 344, 2018, 473-481. https://www.sciencedirect.com/science/article/pii/S0377042718303467

Tunc, H. & Sari, M. (2019). A local differential transform approach for the cubic nonlinear Duffing oscillator with damping term. Scientia Iranica, 26(2), 879–886. http://scientiairanica.sharif.edu/article_20142.html

Tunc, H. & Sari, M. (2021a). A new implicit-explicit local method to capture stiff behavior with COVID-19 outbreak application. arXiv preprint (math.NA), 2104.05817. https://arxiv.org/abs/2104.05817

Tunc, H. & Sari, M. (2021b). A stability preserved time-integration method for nonlinear advection-diffusion-reaction processes. Journal of Mathematical Chemistry, 59, 1917–1937. https://link.springer.com/article/10.1007/s10910-021-01271-1

Analysis of Wave Runup, Overtopping and Overwash Parameters via Compressive Sensing

Ali Rıza Alan[1*], Cihan Bayındır[1, 2]

[1]*İstanbul Technical University, Civil Engineering Department, İstanbul, Turkey*
[2]*Boğaziçi University, Civil Engineering Department, İstanbul, Turkey*
alan21@itu.edu.tr, cbayindir@itu.edu.tr

Abstract

The analysis of wave overtopping and overwash is fundamental to prevent damage to coastal structures and zones. There are many studies in the literature on this subject that shed light on today's research (Kobayashi et al., 1996, van der Meer, 1955). Wave overwash modeling methods are principally based on the prediction and generation of overtopping parameters as the essential inputs. Currently, available methods are inefficient for the evaluation of big field data. Recording and analyzing these data with efficient sensing are fundamentally significant for the observation, appraisal, and prevention of catastrophic results of coastal hazards. For this purpose, new algorithms should be developed, implemented, and tested. Compressive sensing technique (CS) is one of the most efficient algorithms that can beat old-style sensing approaches by utilizing far fewer samples while accomplishing accurate recovery (Candès et al., 2006a, Candès, 2006b). In this paper, we investigate the possible usage of the CS for the viable estimation and analysis of wave runup, overtopping, and overwash for coastal areas. Using the time-series data sets of wave overtopping and overwash constructed by empirical formulas proposed in (Hughes & Thornton, 2016), we show that CS may be utilized as a powerful instrument for the estimation, investigation, and analysis of wave overtopping and overwash in coastal areas and structural health monitoring. We discuss our results and remark on their importance and possible usage areas. The results of this study will be useful for the coastal engineering community in implementing wave runup, overtopping, and overwash reduction strategies to mitigate coastal hazards and the associated human and economic losses.

Keywords: Wave runup, wave overtopping, wave overwash, compressive sensing

1. Introduction

Currently, the majority of the human population lives in coastal areas for some economic and humanitarian reasons. Hereby, the management, regulation, and protection of coastal areas are extremely necessary for people to lead their lives in an ideal way. In some cases where the structures protecting the coastal areas are insufficient to fulfill their function, some negative situations arise both for the coastal areas and for the people living there. Wave overtopping and wave overwash are among the most important of these cases. The event of water flowing over the crest of a coastal structure such as a seawall, a dike, a breakwater, etc., due to wave runup is called wave overtopping and the transport of water and sediment over the crest of a beach by water flowing is defined as wave overwash (Kobayashi et al., 1996). Wave overtopping not only has catastrophic consequences where small buildings collapse and people drown in floodwaters, but they can also cause serious hazards to the highway and even the railway. Wave overwash is dangerous in developed areas, as erosion of the beach face will leave these areas vulnerable after the storm and properties and infrastructures in these areas

* *Corresponding Author*

will be damaged (Kobayashi et al., 1996). Due to their importance, many studies have been carried out on these issues so far. Dune profile formation and overwash transport rates for different dune geometries are measured by Kobayashi et al., (1996), Hancock et al., (1995). Instantaneous discharge time series of overtopping waves by creating laboratory models with different slopes are analyzed by Hughes et al. (2016). A statistical model which identifies the exceedance probability of extreme waves' overtopping volume is discussed by Mori et al. (2003). This brief list and the references therein can give the reader an idea about the vastness of the subject.

2. Problem Definition

As can be understood from all these studies, analysis of wave runup, overtopping, and overwash is a must. Performing these analyzes effectively and quickly is extremely important in terms of efficiency in the results of the studies. One of the very useful algorithms that can be used for this purpose is compressive sensing (CS) (Candès et al., 2006a, Candès, 2006b). Although we see some applied examples of the CS algorithm, which has the main advantage of saving time and storage by achieving the same result with less data, in some coastal engineering problems (Malara et al., 2018, Bayındır, 2019, Bayındır & Namlı, 2021, Bayındır, 2016), there is currently no example related to the field of wave runup, overtopping, and overwash, which is also the subject of this paper. The purpose of this paper is to be a pioneering paper and to guide future studies on this subject since there is no example in the current literature to our best knowledge.

3. Proposed Method

3.1. Methodology for the Probabilistic Time Series Analysis of Wave Overtopping

There are numerous techniques and formulas in the literature to analyze the joint statistics of wave height and period distributions. One of such formula is given as (Goda, 2010)

$$P(\tau|x) = \frac{P(x|\tau)}{P(x)} = \frac{ax}{\sqrt{\pi}\nu}\exp\left[-\frac{a^2x^2}{\nu^2}(\tau-1)^2\right] \quad (1)$$

where *P(x)* denotes the probability distribution of the dimensionless wave height parameter $x = H/H'$. Here H' is an arbitrary reference wave height and the parameter *a* is

$$a = \frac{H'}{(8m_0)^{1/2}} = \begin{cases} 1/2\sqrt{2}, if\ \ H' = m_0^{1/2} = \eta_{rms} \\ \sqrt{\pi}/2,\ if\ \ H' = \bar{H} \\ 1,\ if\ \ H' = H_{rms} \end{cases} \quad (2)$$

It is well-known that for a linear, narrow-banded sea state *P(x)* is in Rayleigh form (Goda, 2010). In Equation (1), $P(\tau|x)$ is the conditional probability of the dimensionless period parameter $\tau = T/\bar{T}$ for a given *P(x)* (Goda, 2010). $\bar{T}$ shows the mean wave period and the parameter ν shows the narrowness of the spectral bandwidth and defined as

$$\nu = \left[\frac{m_o m_2}{m_1^2} - 1\right]^2 \quad (3)$$

where m_o, m_1, m_2 are the zeroth, first, and second moments of the spectrum. Following (Goda, 2010), the value of $\nu = 0.26$ is used throughout this study. In order to generate random wave height and wave period series we follow the approach summarized below. We use the value $H' = H_{rms} = 2m$. Thus, the parameter *a* becomes *a=1*. We generate Rayleigh distributed random number using the MATLAB random number generator to represent the statistical distributions of wave heights. A histogram including 100 waves with $H_{rms} = 2m$ generated this way is depicted in Figure 1. Then we compute the dimensionless wave height parameter *x* and calculate the conditional pdf given by Equation 1 for the typical range of $0 \leq \tau \leq 3$. After the calculation of pdf, we construct the cumulative pdf (cpdf) by simple summation formulas. Then, by means of the random inversion technique for this custom but not-built-in cpdf, we obtain the random τ values corresponding to the random x values. The pdf of τ

calculated in this fashion numerically and its theoretical distribution calculated by integrating Equation 1 numerically over the *x* parameter is depicted in Figure 2. After the random τ values are obtained, the random wave periods are calculated by $T = \tau\bar{T}$ formula. For this calculation a wave field with $\bar{T} = 8s$ is considered and the resulting wave period histogram is depicted in Figure 3.

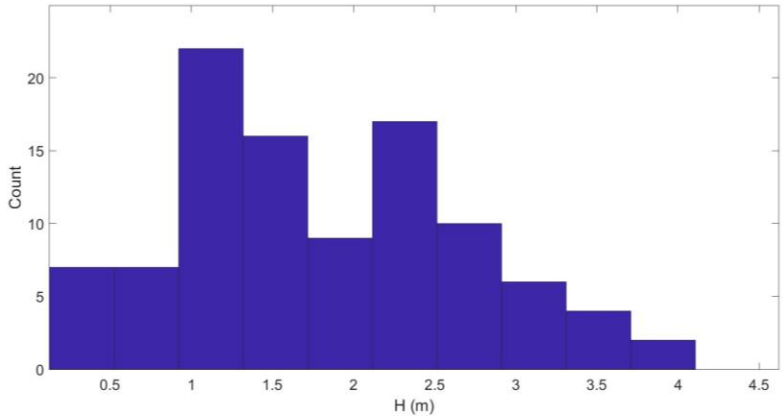


Figure 1. A histogram of wave heights of 100 waves with $H_{rms} = 2m$

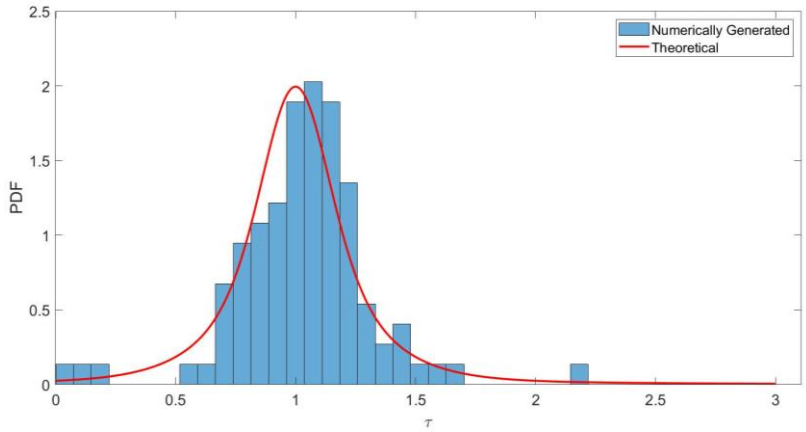


Figure 2. Probability distribution function for dimensionless time parameter $\tau = T/\bar{T}$, numerically obtained by the random inversion technique vs obtained by integration of Eq.1 over *x*

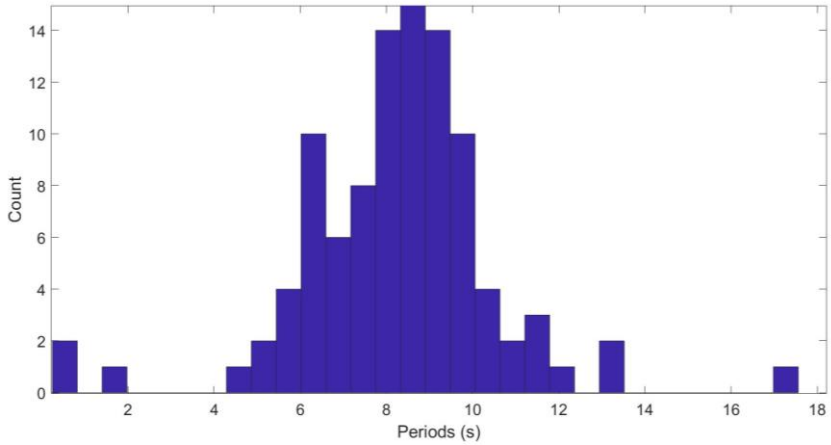


Figure 3. A histogram of wave period of the 100 waves with $\bar{T} = 8s$

The random wave height and the corresponding random wave period distributions are obtained using the approach summarized above. In Europe and at some other places on dikes and some other coastal protection structures are designed according to 2% exceedance criteria, that is the highest 2% of runup exceeds freeboard and causes overtopping. Thus, use this probabilistic thresholds criterion for simulating the time series of random wave overtopping. During the overtopping period of $0 \leq t \leq T$, where T denotes the individual wave period, the overtopping volume time series can be given by the Weibull formula (Hughes et al., 2016)

$$q(t) = \frac{V_T b}{\hat{a}} \left(\frac{t}{\hat{a}}\right)^{(b-1)} \exp\left[-\left(\frac{t}{\hat{a}}\right)^{b}\right] \quad (4)$$

where a shows the scale factor, b shows the shape factor and V_T shows the volume of the overtopping. In this study the typical values of $\hat{a}=0.4$ and $b=2$ are used following (Hughes & Thornton, 2016). For the smallest wave causing overtopping, the overtopping volume is selected as V_T=5.5 m^3/m which is also the largest volume for Dutch overtopping simulator. It is known that the overtopping volume is nonlinearly related to the wave height (Mori et al., 2003), however, it can be linearized for small values of freeboard-to-wave amplitude ratio (Mori et al., 2003). Thus, for waves higher than the smallest causing overtopping, the overtopping volume is linearly scaled. One of the overtopping time series obtained by this methodology is depicted in Figure 4.

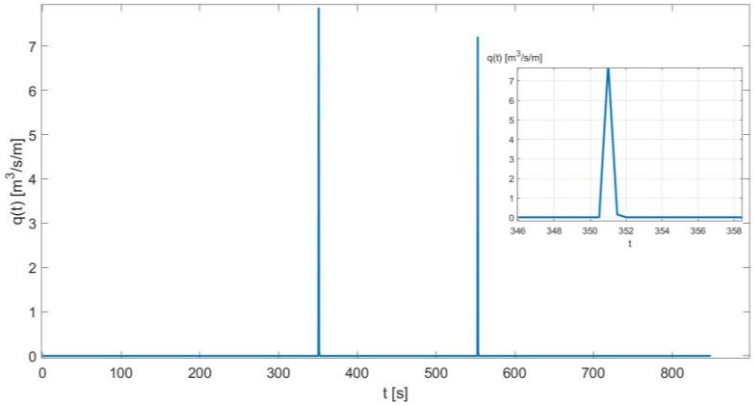


Figure 4. A time series of random wave overtopping volume with 2% overtopping probability

3.2. Review of the Compressive Sensing

Compressive sensing (CS) which has attracted the attention of researchers from many branches of science from engineering to medicine and used by these researchers in their studies, achieved a significant breakthrough in signal processing. CS stated that a K-sparse signal q, with only K out of its N elements are nonzero, can be transformed into an orthogonal domain (i.e. Fourier) by the matrix ψ (orthogonal transformation matrix). Thus, the representation of the signal becomes $q = \psi\hat{q}$. In this formula $\hat{q}$ shows the coefficient vector of transformation. After eliminating the zeros from the inputs, we can obtain $q_s = \psi\hat{q}_s$, where q_s identifies as the signal with non-zero components. A K-sparse signal q with N elements can be exactly recovered by $M \geq C\mu^2(\phi,\psi)K\log(N)$ measurements using the CS algorithm. In this formula, C symbolizes a positive constant, ϕ symbolizes the sensing basis and $\mu^2(\phi,\psi)$ is the mutual coherence between the sensing and transformation basis (Candès et al., 2006a, Candès, 2006b). After the random sampling using M random samples, one gets $g = \phi\psi\hat{q}$. Thus the CS problems becomes:

$$\min\|\hat{q}\|_{l_1} \text{ subjected to } g = \phi\psi\hat{q} \quad (5)$$

where $\|\hat{q}\|_{l_1} = \Sigma_i|\hat{q}_i|$. Among all possible solutions of this optimization problem, its l_1 solution becomes $q_{cs} = \psi\hat{q}$. For a more comprehensive discussion of CS the reader is referred to Candès et al., (2006a),

Candès, (2006b), and its applications in coastal hydrodynamics to Bayındır, (2016, 2019) and Malara et al., (2018).

4. Results and Discussion

The aim of this paper is to propose and examine the applicability of the CS for the efficient sensing of the wave overtopping time series. The typical overtopping volume time series have a sparse behavior in the time domain as depicted in Figure 4. Thus, random compressive sampling is performed in the Fourier domain and the l1 minimization problem is solved in the temporal domain. In Figure 5, we depict the random wave overtopping time series which is also depicted in Figure 4, and its CS reconstruction.

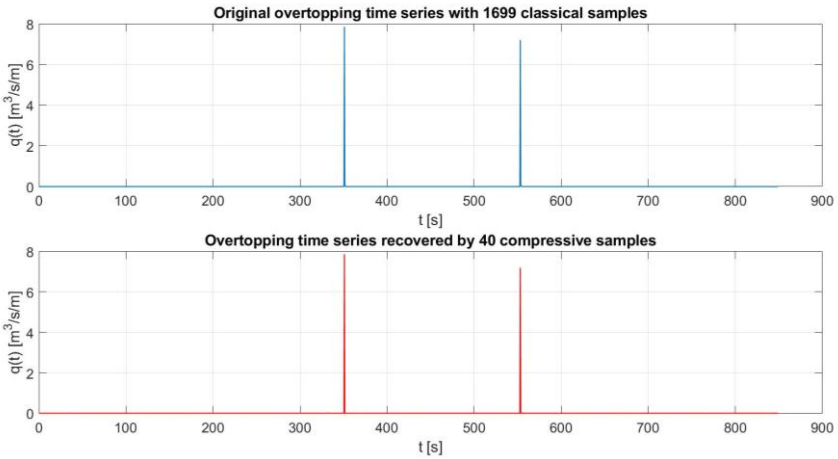


Figure 5. A time series of random wave overtopping volume with 2% overtopping probability. a) classical sampling b) compressive sampling reconstruction.

In Figure 5, one can realize that the exact reconstruction of the random wave overtopping time series is exactly reconstructed by CS. The main advantage of the CS-based approach is that, although the classical observation system uses N=1699 classical samples, the CS-based observation system uses only K=40 random samples. This significant undersampling ratio brings a major advantage for coastal measurement and observation systems in terms of memory, cost, and time of measurement. Additionally, the CS-based approach is also beneficial for the interpolation and extrapolation of the missing data. Although this result is depicted for overtopping and overwash, it is straightforward to extend our finding to random wave runup. Such runup, if it exceeds the freeboard of the coastal structure or not, has a random multifrequency behavior in the time domain, and has a sparse representation in the spectral domain. Thus, random sampling can be performed in the temporal domain or a hybrid approach may be followed for the efficient analysis of typical random run-up time series.

5. Conclusion

In this study, we investigated the efficient measurement of the wave overtopping, overwash, and runup parameters by compressive sensing. Using a joint probability distribution of wave heights and periods and their random inversion, we constructed a wave overtopping time series with and freeboard exceedance probability of 2%. The resulting time series can be treated as a sparse signal in the time domain. We showed that such sparse data can be effectively reconstructed by compressive sampling using a far fewer number of samples than Shannon's classical sampling theory states. Our findings will help coastal and ocean engineering communities to develop cheaper and faster overtopping, overwash, and runup measurement systems with lower cost and data storage requirements. Additionally, our findings can also be used to interpolate or extrapolate missing data.

Acknowledgements

The authors gratefully acknowledge the support of the İstanbul Technical University. This work was supported by the Research Fund of the İstanbul Technical University. Project Code: MGA-2020-42544. Project Number: 42544.

References

Bayındır, C. (2016). Compressive spectral method for the simulation of nonlinear gravity waves, Scientific Reports, 22100. https://doi.org/10.1038/srep22100

Bayındır, C. (2019). Early detection of rogue waves using compressive sampling. TWMS Journal of Applied and Engineering Mathematics, 9 (2), 198-205. http://jaem.isikun.edu.tr/web/images/articles/vol.9.no.2/03.pdf

Bayındır, C., & Namlı, B. (2021). Efficient sensing of the von Karman vortices using compressive sensing, Computers & Fluids, 104975. https://doi.org/10.1016/j.compfluid.2021.104975

Candès, E. J., Romberg, J., & Tao, T. (2006a). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory, 52 (2), 489-509. https://doi.org/10.1109/TIT.2005.862083

Candès, E. J. (2006b). Compressive sampling. Proceedings of the international congress of mathematicians, 3, 1433-1452. https://doi.org/10.4171/022-3/69

Goda, Y. (2010). Random seas and design of maritime structures. World Scientific Publishing Company. https://doi.org/10.1142/7425

Hancock, M. W., & Kobayashi, N. (1995). Wave overtopping and sediment transport over dunes. In Coastal Engineering 1994 (pp. 2028-2042). https://doi.org/10.1061/9780784400890.148

Hughes, S. A., & Thornton, C. I. (2016). Estimation of time-varying discharge and cumulative volume in individual overtopping waves. Coastal Engineering, 117, 191-204. https://doi.org/10.1016/j.coastaleng.2016.08.006

Kobayashi, N., Tega, Y., & Hancock, M. W. (1996). Wave reflection and overwash of dunes. Journal of Waterway, Port, Coastal, and Ocean Engineering, 122 (3), 150-153. https://doi.org/10.1061/(ASCE)0733-950X(1996)122:3(150)

Malara, G., Kougioumtzoglou, I. A., & Arena, F. (2018). Extrapolation of random wave field data via compressive sampling. Ocean Engineering, 157, 87-95. https://doi.org/10.1016/j.oceaneng.2018.03.044

Mori, N., & Cox, D. T. (2003). Statistical modeling of overtopping for extreme waves on fixed deck. Journal of waterway, port, coastal, and ocean engineering, 129 (4), 165-173. https://doi.org/10.1061/(ASCE)0733-950X(2003)129:4(165)

Van der Meer, J. W. (1995). Wave run-up and wave overtopping at dikes. Wave forces on inclined and vertical structures, ASCE. https://ci.nii.ac.jp/naid/10018654338/en/

Artificial Neural Network Solution of Advection-Diffusion Equation with Source Effects

Pelin Celenk[1], Seda Gulen[2], Murat Sarı[3*]

[1]*Yildiz Technical University, Graduate School of Science and Engineering, Mathematics, Istanbul, Turkey*
[2]*Namik Kemal University, Faculty of Arts and Science, Department of Mathematics, Tekirdag, Turkey*
[3]*Yildiz Technical University, Faculty of Arts and Science, Department of Mathematics, Istanbul, Turkey*
f2519010@std.yildiz.edu.tr, sgulen@nku.edu.tr, sarim@yildiz.edu.tr

Abstract

In this study, an artificial neural network (ANN) method is proposed to capture the behaviours represented by various versions of the advection-diffusion equations. This technique uses a trial function that depends on a neural network and satisfies the initial and boundary conditions to approach the solution of the problem. The unknown parameters of the neural network are adjusted with the gradient-descent and particle swarm optimization methods to minimize the cost function. The produced results are seen to be in good agreement with the exact solution and needed less computer storage.

Keywords: Artificial Neural Networks (ANNs), Gradient Descent (GD), Particle Swarm Optimization (PSO), Partial differential equations (PDEs)

1. Introduction

ANNs that are the most important topic of machine learning methods simulate the human brain and neural systems and attract increasing attention in solving many problems arising in engineering and science. Therefore, various researchers have paid their attention to the ANN methods due to their adaptability and resistance to errors in data (Lagaris et al., 1998). Recently, these methods have been used by researchers to obtain predictive solutions of partial differential equations that represent most real-world problems. Since solving these problems is often difficult and needed computationally intensive numerical techniques, various numerical techniques have been employed up to now. Several methods using the ANNs provide a closed-form solution that is infinitely differentiable. In addition, these methods require less computer storage than classical methods such as finite difference or finite element methods. However, since the topic is quite new, the studies are very limited. One of these studies is that Lagaris et al. (1998) presents an ANN method to solve initial and boundary value problems. After this study, various authors (Rassi, 2018; *Baymani* et al., 2010; Hayati and Karami, 2007; Eskiizmirliler et al., 2020) continued to be interested in ANN solutions of partial differential equations.

In this work, predictive solutions of some linear and nonlinear partial differential equations are obtained by an artificial neural network method. In this technique, the dependent variable is approximated by a trial function that depends on a neural network solution and satisfies the boundary condition of the problem. The gradient descent and particle swarm optimization algorithms are successfully used to adjust the unknown parameters. The computed results revealed that this technique presents accurate and reliable solution with minimal computational effort.

2. Artificial Neural Networks

As indicated in the previous section, the ANNs are modelled by simulating the human brain and neural systems. In the mathematical model, an artificial neuron is called *perceptron*. Each neuron has a real valued input and each input is multiplied by an associate weight. The sum of these products are

* *Corresponding Author*

associated to a bias. These computed values by passing through the activation function, the output of the neuron is obtained as feed-forward (FF).

In this section, we investigate how the feed-forward neural network (FFNN) method can be adapted to capture the behaviours of the partial differential equation given by

$$u_t = N(t, x, u_x, u_{xx}, \dots), \quad (1)$$

subject to certain boundary conditions. Here u is a solution to be computed and also represented by a neural network.

In the proposed technique, the trial function that satisfies the boundary conditions is expressed by

$$\Psi_t(x,t) = A(x,t) + F\big(x, t, Net(x,t,p)\big), \quad (2)$$

where $A(x,t)$ satisfies the boundary conditions and $Net(x,t,p)$ is the output of the FFNN with the adjustable parameters $p = (\alpha, \beta, \omega, \Omega)$ in the hidden layer of neural networks and inputs x and t. The second term F is constructed by employing the FFNN whose weights and bias are adjusted to solve the minimization problem

$$p = minE(x,t;p), \quad (3)$$

$$E(p) = \frac{1}{2}\sum_{i=0}^{N}\sum_{j=1}^{M}\{e_{ij}\}^2, \quad (4)$$

$$e_{ij} = \left(\frac{\partial \Psi_t}{\partial t_j} - F(t, x_1, \dots, x_n, \frac{\partial \Psi_t}{\partial x_i}\right). \quad (5)$$

To optimize the minimization problem (3), two different optimization methods, a gradient descent algorithm and particle swarm optimization, are performed comparatively.

3. Solution of the Minimization Problem

3.1 The Gradient Descent Method

The gradient descent method is one of the most used optimization methods in neural networks studies. In predictive solution of a mathematical model, an error called cost function arises between the exact solution and computed solution,

$$E(p) = \frac{1}{2}\sum_{i=0}^{N}\sum_{j=1}^{M}\{e_{ij}\}^2, \quad (6)$$

where $Net(x,t;p) = \sum_{k=1}^{m} p_k = (\alpha_k, f(\square_k t + \Omega_k\, x + \beta_k)$. (7)

The gradient descent method is constructed to optimize the cost function $E(p)$. Computations of this error includes the network outputs and the derivatives of the outputs with respect to any of its inputs.

The algorithm of the gradient descent method can be given as follows:

---

1. **Input:** λ – Learning rate
2. **Input:** $E(x_i, t_j)$ – Cost function to be minimized
3. **Input:** $p_k = (\alpha_k, f(\square_k t + \Omega_k\, x + \beta_k)$ - initial parameter vector
4. $n \leftarrow 0$ – initialized step counts
   $p_0 = (\alpha_0, f(\square_0 t + \Omega_0\, x + \beta_0)$
5. **While** p_n not converged **do**
   $n \leftarrow n + 1$ – update step count
   $p[n+1] = p[n] - \lambda[n]\nabla F(p[n])$ – update parameters
6. until convergence
7. **End**

Here, $p_n = (\alpha_n, \beta_n, \square_n, \Omega_n) \in \mathrm{R}^m$, m and n denote, the number of neurons and the iteration number, respectively.

In the gradient descent method, the selection of the parameter λ is of great importance for the convergence of the solution. Selecting a very small parameter causes time-consuming computations, while large values of the parameter produce divergent solutions.

3.2 The Particle Swarm Optimization

The particle swarm optimization (PSO) that inspires by the behaviour of swarms, especially birds, fishes and bees is a population-based stochastic optimization method. Every bird, fish or bee is called as a particle that moves with a certain velocity and searches for the global best position after some iteration in searching space.

The particle i is defined by position vector x_i and its velocity vector v_i as follows (Eskiizmirliler et al., 2020):

$v_i[n+1] = wv_i[n]$ (Inertia term)
$+c_1 r_1\left(x_{Best_i}[n] - x_i[n]\right)$ (Cognitive component)
$+c_2 r_2(g_{Best}[n] - x_i[n]$ (Social component)
$x_i(t+1) = x_i(t) + v_i(t+1)$ (New location),

where, w represents the inertial weight and controls the influence of the previous velocity c_1 and c_2 defines the coefficient of the self-recognition and r_1 and r_2 are the coefficient of social component chosen randomly in the interval $[0,1]$. The PSO algorithm can be given as follows:

Initialize all particles of the swarm with randomly generated position and velocity
Repeat
  For each particle in the swarm
  Calculated the Cost function
  Update the local best position of the particle
  Update the global best position of the swarm
  End for
  For each particle in the swarm
    Update the velocity and position of the particle according to equation:
    $v_i[n+1] = wv_i[n] + c_1 r_1\left(x_{Best_i}[n] - x_i[n]\right) + c_2 r_2(g_{Best}[n] - x_i[n]$
    $x_i(t+1) = x_i(t) + v_i(t+1)$
  End for
Until (Stopping criteria)

---

4. Illustrative Examples

In this section, the proposed method is performed for linear advection-diffusion and the inviscid Burgers equations, respectively. In order to apply the proposed technique, it is started by defining uniform space and time grid consisting of N_x and N_t points satisfying $x_i = 0,1,2,\ldots,N_x$ where $h_x = ih_x$ and $t_j = 0,1,2,\ldots,N_t$ where $h_t = T/N_t$, respectively.

The efficiency and adaptability of the approach are tested for different values of the model parameters. To evaluate the flexibility of the proposed techniques, their accuracy and efficiency with respect to various parameter values are estimated by the mean absolute error norm (MAE), the mean squared error (MSE) and the mean squared relative error (MSRE) are formulated as

$$\|u - \Psi\|_{MAE} \coloneqq \frac{1}{N_x \times N_t} \sum_{i=1}^{N_x} \sum_{j=1}^{N_t} \left|u(x_i, t_j) - \Psi(x_i, t_j)\right|, \tag{8}$$

$$\|u - \Psi\|_{MSE} \coloneqq \left(\frac{1}{N_x \times N_t} \sum_{i=1}^{N_x} \sum_{j=1}^{N_t} \left|u(x_i, t_j) - \Psi(x_i, t_j)\right|^2\right)^{1/2}, \tag{9}$$

$$MSRE \coloneqq \frac{\|u-\Psi\|_{MSE}}{\|u\|_{MSE}}. \tag{10}$$

In all cases, it is used a three-layer feed-forward neural networks having two inputs, one hidden layer with five neurons and one output and sigmoid activation function, that is

$$f(z) = 1/\left(1 + exp(-z)\right). \tag{11}$$

Example (4.1): (Mohebbi and Dehghan, 2010) Consider the advection-diffusion equation,

$$u_x + \beta u_t = \alpha u_{xx}\ ,\ (x,t) \in [0,2] \times [0,1] \tag{12}$$

with the exact solution

$$u(x,t) = exp(-\beta t)sin(x - \alpha t). \tag{13}$$

The trial function can be written as follows:

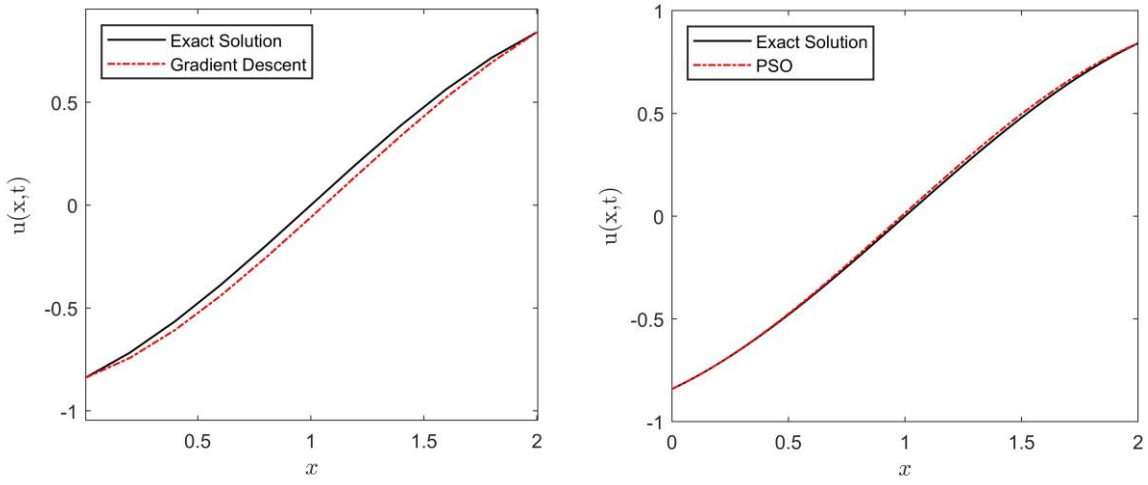
$$\Psi(x,t) = \big((2-x)/2\big)u(0,t) + (x/2)u(2,t) + (1-t)u(x,0) - \big((2-x)/2\big)(1-t)u(0,0) - (x/2)(1-t)u(2,0). \tag{14}$$

To train the network, for the Peclet numbers 1000 and 10000, the quadrature nodes are generated by discretization of a $51x51$ mesh, $N_x = 51$ and $N_t = 51$. Whereas for the Peclet numbers 20000, the nodes are generated by discretization of a $21x21$ mesh, $N_x = 21$ and $N_t = 21$. In both cases, the network is trained 100 times.

Table 1. Comparison of the MAEs of Example (4.1) with different values of Pe, α and β

x	$Pe = 1000, \alpha = 0.001$ and $\beta = 1$			$Pe = 10000, \alpha = 0.0001$ and $\beta = 1$			$Pe = 20000, \alpha = 0.0005$ and $\beta = 1$		
	Mohebbi & Dehghan (2010)	ANN-GD	ANN-PSO	Mohebbi & Dehghan (2010)	ANN-GD	ANN-PSO	Mohebbi & Dehghan (2010)	ANN-GD	ANN-PSO
0.25	1.4523e-06	2.5526e-04	1.3328e-06	1.0552e-06	2.5092e-05	1.8221e-07	1.0227e-04	2.9131e-05	7.2540e-07
0.50	1.7519e-06	1.1410e-04	1.3751e-06	4.9081e-06	4.6371e-05	6.0528e-07	2.1975e-04	2.8871e-05	4.9164e-06
0.75	1.8969e-06	1.2878e-04	1.0972e-06	6.9429e-06	6.4083e-05	9.1696e-07	3.4696e-04	8.2184e-06	8.5401e-06
1.00	9.7462e-07	1.2181e-04	9.3018e-07	2.5616e-05	7.6229e-05	1.0364e-07	4.7082e-04	2.1842e-05	9.8358e-06
1.25	1.0998e-07	9.9214e-05	9.0177e-07	7.0254e-05	7.9591e-05	9.7891e-07	6.5159e-04	4.9435e-05	8.5637e-06
1.50	2.0490e-07	6.7306e-05	8.2894e-07	1.5840e-04	7.0356e-05	7.9764e-07	8.0199e-04	6.2472e-05	5.5743e-06
1.75	2.8775e-07	3.3387e-05	5.2689e-07	2.6866e-04	4.4737e-05	5.2539e-07	7.4040e-04	4.9332e-05	2.3515e-06

The efficiency of the present methods for different values of Peclet number $(Pe), \alpha$ and β are presented in Table 1 in comparison with the work of Mohebbi and Dehghan (2010). As seen from the table, by increasing the Peclet number, the error of the ANN-PSO solution decreases. Figure 1 illustrates the behavioural solution of Example (4.1) obtained by the gradient descent and PSO algorithms.



(a) (b)

Figure 1. Solutions of the Example (4.1) obtained by (a) the gradient descent method and (b) the PSO

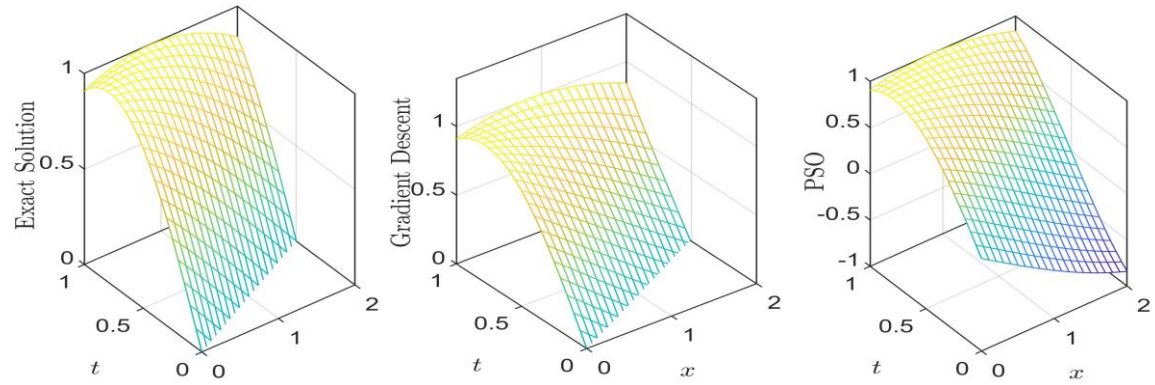


Figure 2. The exact solution, the gradient descent and PSO for Example (4.1) with 100 times training and $Pe = 1000, \alpha = 0.001$ and $\beta = 1$

Example (4.2): (Dehghan, 2005)

We next focus on the advection-diffusion equation with $\alpha = 0.1$, $\beta = 0.8$ and $(x,t) \in [0,1] \times [0,1]$.

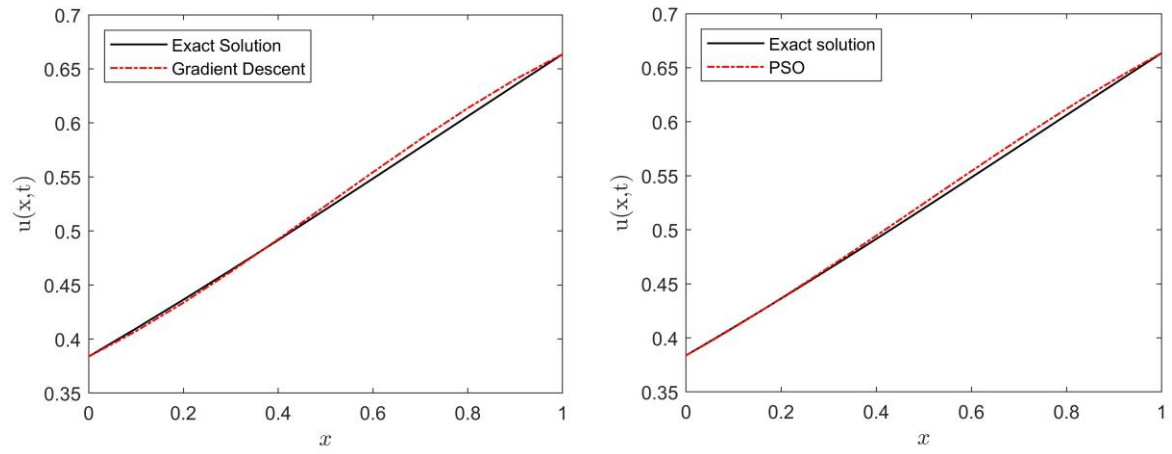
The exact solution is $u(x,t) = \sqrt{\frac{20}{20+t}} exp\left[-\frac{(x-2-0.8t)^2}{5(t+20)}\right]$. (15)

To train the network, the quadrature nodes are generated by discretization of $N_x = 101$ and $N_t = 101$ are trained 125 times.

As seen in Table 2, the MAEs of the solutions for both the gradient descent and PSO are compared in Example (4.2) with the literature (Dehghan, 2005). Behaviours of the solutions obtained by these methods have been shown in Figure 2. The computed results revealed that the proposed technique approximates better than the other methods and is applicable, effective and easy to use.

Table 2. Comparison of the MAEs of Example (4.1) with the GD and PSO

	Exact Value	Dehghan (2004) (Second Order)	Dehghan (2004) (Third order)	Dehghan (2004) (Fourth order)	ANN-GD	ANN-PSO
x		MAE	MAE	MAE	MAE	MAE
0.1	0.4097319	1.3e-03	2.7e-03	3.4e-05	4.7220e-06	4.2158e-09
0.2	0.4364170	1.1e-03	2.7e-03	3.2e-05	8.0930e-06	1.0697e-07
0.3	0.4637347	1.2e-03	2.6e-03	3.1e-05	1.0227e-05	2.6273e-07
0.4	0.4915904	1.4e-03	2.6e-03	2.9e-05	1.1239e-05	4.2864e-07
0.5	0.5198801	1.3e-03	2.7e-03	2.7e-05	1.1241e-05	5.6642e-07
0.6	0.5484904	1.1e-03	2.4e-03	2.7e-05	1.0347e-05	6.4404e-07
0.7	0.5772989	1.4e-03	2.0e-03	2.5e-05	8.6680e-06	6.3715e-07
0.8	0.6061756	1.5e-03	2.3e-03	2.2e-05	6.3132e-06	5.3007e-07
0.9	0.6349830	1.7e-03	2.5e-03	2.0e-05	3.3893e-06	3.1655e-07



(a) (b)

Figure 3. Solutions of Example (4.2) obtained by (a) GD and (b) PSO

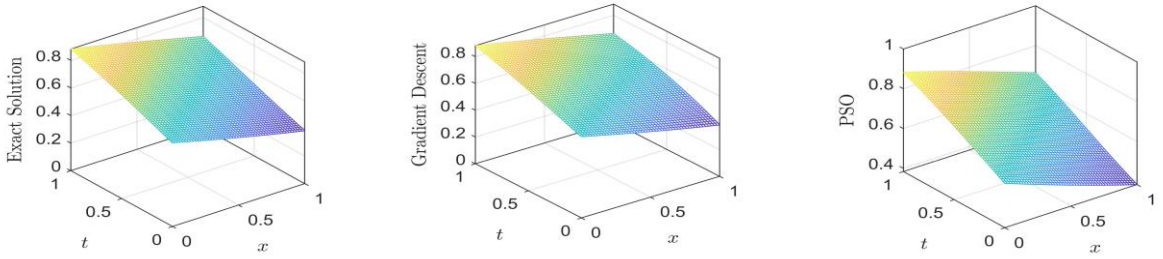


Figure 4. The exact solution, the GD and PSO solutions for Example (4.2) with 125 times training

Example (4.3): (Wazwaz, 2009) Let us consider the nonhomogeneous inviscid Burgers equation

$u_x + uu_t = x + xt^2$, $(x,t) \in [0,1] \times [0,1]$ (16)

and exact solution is $u(x,t) = xt$. (17)

To train the network, mesh grid points $N_x = 101$ and $N_t = 101$ run for 120 iterations.

Table 3. Comparison of the error norms MAE, MSE and MSRE for the GD and PSO approaches

	Types of Errors	ANN-GD	ANN-PSO
MAE	Min	5.7196e-04	5.0637e-03
	Worst	1.4041e-01	1.0892e-01
	Mean	4.3632e-02± 3.1666e-02	4.1464e-02± 2.8448e-02
MSE	Min	7.6250e-04	7.2376e-03
	Worst	1.8132e-01	1.4087e-01
	Mean	5.6303e-02± 4.0882e-02	5.3330e-02± 3.6610e-02
MSRE	Min	8.7076e-07	8.2653e-06
	Worst	2.0707e-04	1.6087e-04
	Mean	6.4297e-05± 4.6686e-05	6.0902e-05± 4.1808e-05

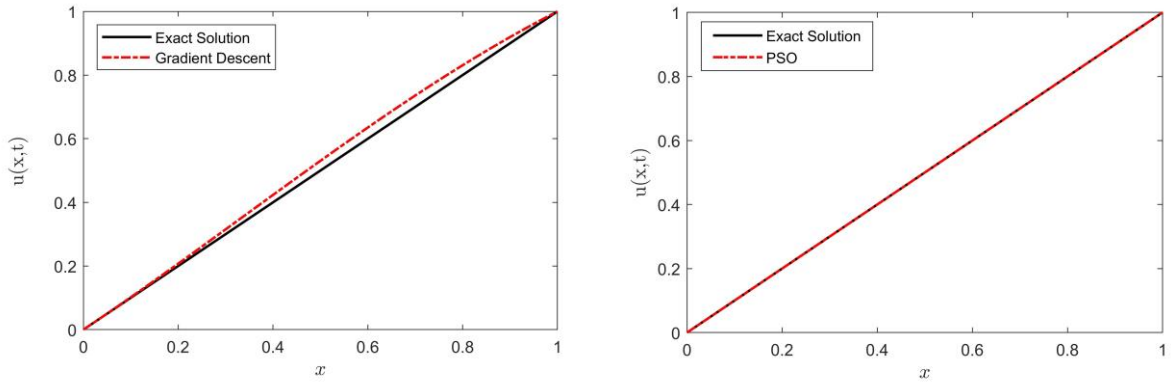


Figure 5. Solutions of Example (4.3) obtained by (a) the GD method and (b) PSO

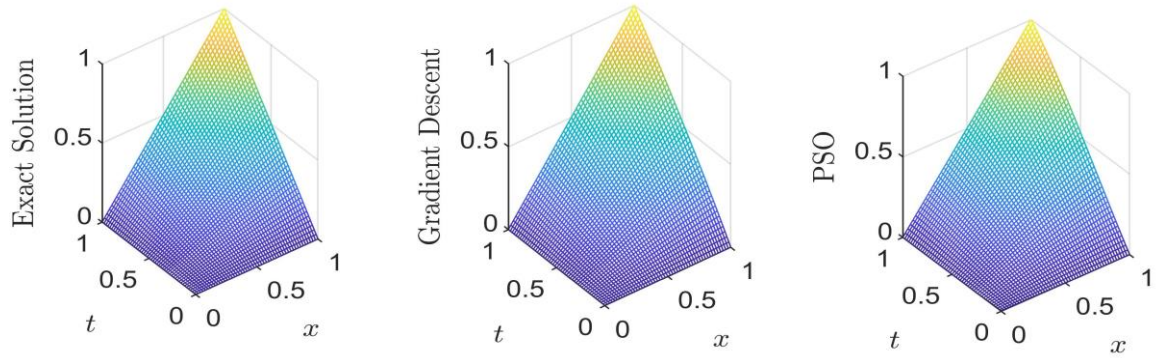


Figure 6. The exact solution, the GD and PSO solutions for Example (4.3) with 120 times training

The computed error norms of different solutions for Example (4.3) are listed in Table 3 and the behaviours of the solutions obtained by the GD and PSO are sketched in Figure 3. The computed results revealed that the current method seems to be accurate and reliable with minimal computational effort.

5. Conclusions

Solving PDEs requires computationally intensive numerical techniques. Therefore, in this study, an artificial neural network method has been presented to capture the behaviour of some linear and nonlinear advection-diffusion equations in a more efficient way. The neural network method has been seen to allow converting any partial differential equation into a single objective unconstrained optimization problem. To solve the optimization problem in obtaining the unknown parameters, the gradient descent and particle swarm optimization methods have been found out to be successfully applicable. In order to validate the performance of the proposed approach, the errors between the computed solutions, the exact solution and the literature have been compared. Ultimately, the computed results have revealed that the proposed method produces accurate results and needs less computational effort and storage space than the conventional techniques; hence, it is concluded that solving the PDEs through the presented method is a reliable and flexible alternative.

References

Baymani, M., Kerayechian, A., & Effati, S. (2010). Artificial Neural Networks Approach for Solving Stokes Problem. Applied Mathematics, 1(04), 288. https://doi.org/10.4236/am.2010.14037

Dehghan, M. (2005). On the Numerical Solution of the One-Dimensional Convection-Diffusion Equation. Mathematical Problems in Engineering, 1, 61–74. https://doi.org/10.1155/MPE.2005.61

Eskiizmirliler, S., Gunel, K., & Polat, R. (2020). On the Solution of the Black–Scholes Equation Using Feed-Forward Neural Networks. Computational Economics, 41, 697-704.

https://doi.org/10.1007/s.10614-020-10070-w

Hayati, M., and Karami, B. (2007). Feedforward neural network for solving partial differential equations. Journal of Applied Sciences, 7(19), 2812–2817.

https://doi.org/10.3923/jas.2007.2812.2817

Lagaris, I.E., Likas, A., & Fotiadis, D.I. (1998). Artificial neural networks for solving ordinary and partial differential equations. IEEE Transactions on Neural Networks, 9(5), 987-1000.

https://doi.org/10.1109/72.712178

Mohebbi, A., and Dehghan, M. (2010). High-order compact solution of the one-dimensional heat and advection-diffusion equations. Applied Mathematical Modelling, 34(10), 3071-3084.

https://doi.org/10.1016/j.apm.2010.01.013

Raissi, M. (2018). Deep hidden physics models: Deep learning of nonlinear partial differential equations. Journal of Machine Learning, 19(1), 932-955. https://jmlr.org/papers/v19/18-046.html

Wazwaz, A. M. (2009). Partial Differential Equations and Solitary Waves Theory. Higher Education Press, Beijing and Springer-Verlag Berlin Heidelberg, 323-324. https://doi.org/10.1007/978-3-642-00251-9

Hermite-Bell Based Bernoulli Polynomials

Uğur Duran[1*], Mehmet Acikgoz[2]

[1] *İskenderun Technical University, Department of Basic Sciences and Engineering, Hatay, Turkey*
[2] *Gaziantep University, Department of Mathematics, Gaziantep, Turkey*
ugur.duran@iste.edu.tr, acikgoz@gantep.edu.tr

Abstract

In this study, we consider Hermite-Bell based Stirling polynomials of the second kind and derive some useful relations and properties including some summation formulas related to the Bell polynomials and Stirling number of the second kind. Then, we introduce Hermite-Bell based Bernoulli polynomials of order α and investigate multifarious correlations and formulas including some summation formulas and derivative properties. Also, we acquire diverse implicit summation formulas and symmetric identities for Hermite-Bell based Bernoulli polynomials of order α. Moreover, we analyze some special cases of the results.

Keywords: Hermite polynomials, Bernoulli polynomials, Bell polynomials, mixed type polynomials

1. Introduction

Special polynomials and numbers possess many importance in multifarious areas of sciences, such as physics, mathematics, applied sciences, engineering and other related research fields covering, differential equations, number theory, functional analysis, quantum mechanics, mathematical analysis, mathematical physics, and so on, see all references and also each of the references cited therein. For example; Bernoulli polynomials and numbers are closely related with the Riemann zeta function which possesses a connection with the distribution of prime numbers, (Srivastava et al., 2012 and Srivastava et al., 1984). Some of the most significant polynomials in the theory of special polynomials are the Bell, Euler, Bernoulli, Hermite, and Genocchi polynomials. Recently, the aforesaid polynomials and their diverse generalizations have been densely considered and investigated by many physicists and mathematicians, see all references and see also the references cited therein.

2. Preliminaries

The Stirling polynomials $S_2\left(n,k:x\right)$ and numbers $S_2\left(n,k\right)$ of the second kind are given by the following exponential generating functions (Bell, 1934, Carlitz, 1980, Khan et al., 2008, Kim et al., 2017):

$$\sum_{n=0}^{\infty}S_2\left(n,k:x\right)\frac{t^n}{n!}=\frac{\left(e^t-1\right)^k}{k!}e^{tx} \quad \text{and} \sum_{n=0}^{\infty}S_2\left(n,k\right)\frac{t^n}{n!}=\frac{\left(e^t-1\right)^k}{k!}. \tag{1}$$

In combinatorics, Stirling number of the second kind $S_2\left(n,k\right)$ counts the number of ways in which n distinguishible objects can be partitioned into k indistinguishable subsets when each subset has to contain at least one object. The Stirling numbers of the second kind can also be derived by the following recurrence relation for $\zeta\in\mathbb{N}_0$:

* *Corresponding Author*

$$x^n = \sum_{k=0}^{n} S_2(n,k)(x)_k, \tag{2}$$

where $(x)_n = x(x-1)(x-2)\cdots(x-(n-1))$ for $n \in \mathbb{N}$ with $(x)_0 = 1$.

For each integer $k \in \mathbb{N}_0$, $S_k(n) = \sum_{l=0}^{n} l^k$ is named the sum of integer powers. The exponential generating function of $S_k(n)$ is as follows (Pathan et al., 2015):

$$\sum_{k=0}^{\infty} S_k(n) \frac{t^k}{k!} = \frac{e^{(n+1)t} - 1}{e^t - 1}. \tag{3}$$

The bivariate Bell polynomials are defined as follows:

$$\sum_{n=0}^{\infty} Bel_n(x;y) \frac{t^n}{n!} = e^{y\left(e^t - 1\right)} e^{xt} \tag{4}$$

When $x = 0$, $Bel_n(0;y) := Bel_n(y)$ called the classical Bell polynomials (also called exponential polynomials) given by means of the following generating function (Boas et al., 1958, Kim et al., 2015):

$$\sum_{n=0}^{\infty} Bel_n(y) \frac{t^n}{n!} = e^{y\left(e^t - 1\right)}. \tag{5}$$

The Bell numbers Bel_n are attained by taking $y = 1$ in (2.5), that is $Bel_n(0;1) = Bel_n(1) := Bel_n$ and are given by the following exponential generating function (Boas et al., 1958, Kim et al., 2015):

$$\sum_{n=0}^{\infty} Bel_n \frac{t^n}{n!} = e^{\left(e^t - 1\right)}. \tag{6}$$

The Bell polynomials considered by Bell (Bell., 1934) appear as a standard mathematical tool and arise in combinatorial analysis. Since the first consideration of the Bell polynomials, these polynomials have been intensely investigated and studied by several mathematicians, (Boas et al., 1958, Kim et al., 2015) and see also the references cited therein.

The usual Bell polynomials and Stirling numbers of the second kind satisfy the following relation (Kim et al., 2015)

$$Bel_n(y) = \sum_{m=0}^{n} S_2(n,m) y^m. \tag{7}$$

The Bernoulli polynomials $B_n^{(\alpha)}(x)$ of order α are defined as follows (Acikgoz et al., 2018, Dere et al., 2013, Kim et al., 2019, Kim et al., 2021):

$$\sum_{n=0}^{\infty} B_n^{(\alpha)}(x) \frac{t^n}{n!} = \left(\frac{t}{e^t - 1}\right)^{\alpha} e^{xt} \quad (|t| < 2\pi). \tag{8}$$

Setting $x = 0$ in (8), we get $B_n^{(\alpha)}(0) := B_n^{(\alpha)}$ known as the Bernoulli numbers of order α. We also note that when $\alpha = 1$ in (8), the polynomials $B_n^{(\alpha)}(x)$ and numbers $B_n^{(\alpha)}$ reduce to the classical Bernoulli polynomials $B_n(x)$ and numbers B_n.

The Hermite polynomials $H_n(x,y)$ (Dere et al., 2012, Pathan et al., 2015) are defined by

$$H_n(x,y) = n! \sum_{r=0}^{[\frac{n}{2}]} \frac{y^r x^{n-2r}}{r!(n-2r)!}. \tag{9}$$

It is easily seen from (9) that

$$H_n(x,0) = x^n.$$

The generating function for Hermite polynomials $H_n(x,y)$ are given by

$$e^{xt+yt^2}=\sum_{n=0}^{\infty}H_n(x,y)\frac{t^n}{n!}. \tag{10}$$

3. Hermite-Bell Based Stirling Polynomials of the Second Kind

We consider Hermite-Bell mixed polynomials by

$$\sum_{n=0}^{\infty}HB_n\left(x,y,z\right)\frac{t^n}{n!}=e^{xt+yx^2+z\left(e^t-1\right)}, \tag{11}$$

which reduce to usual Hermite polynomials (10) when $z=0$ and familiar Bell polynomials (5) when $y=0$.

Definition 1 *The Hermite-Bell based Stirling polynomials of the second kind are introduced by the following generating function:*

$$\sum_{n=0}^{\infty}{}_{HB}S_2\left(n,k:x,y,z\right)\frac{t^n}{n!}=\frac{\left(e^t-1\right)^k}{k!}e^{xt+yx^2+z\left(e^t-1\right)}. \tag{12}$$

Remark 1 *Replacing* $x=0$ *in (12), we attain extended Bell-Stirling polynomials* ${}_{HB}S_2\left(n,k:y,z\right)$ *of the second kind, which are also a new generalization of the usual Stirling numbers of the second kind in (1) as follows:*

$$\sum_{n=0}^{\infty}{}_{HB}S_2\left(n,k:y,z\right)\frac{t^n}{n!}=\frac{\left(e^t-1\right)^k}{k!}e^{yt^2+z\left(e^t-1\right)}. \tag{13}$$

Remark 2 *Replacing* $x=y=0$ *in (12), we obtain Bell-Stirling polynomials* ${}_{Bel}S_2\left(n,k:z\right)$ *of the second kind, which are also a new generalization of the usual Stirling numbers of the second kind in (1), as follows:*

$$\sum_{n=0}^{\infty}{}_{Bel}S_2\left(n,k:z\right)\frac{t^n}{n!}=\frac{\left(e^t-1\right)^k}{k!}e^{z\left(e^t-1\right)}.$$

Proposition 1 *The following correlation holds for non-negative integer* n*:*

$${}_{HB}S_2\left(n,k:x,y,z\right)=\sum_{u=0}^{n}\binom{n}{u}S_2\left(u,k\right)HB_{n-u}\left(x;y,z\right). \tag{14}$$

Proposition 2 *The following relations hold for non-negative integers* n *and* k *with* $n\geq k$*:*

$${}_{HB}S_2\left(n,k:x,y,z\right)=\sum_{l=0}^{n}\binom{n}{l}\,{}_{HB}S_2\left(l,k:y,z\right)x^{n-l} \tag{15}$$

and

$${}_{HB}S_2\left(n,k:x,y\right)=\sum_{l=0}^{n}\binom{n}{l}\,{}_{Bel}S_2\left(l,k:z\right)H_{n-l}\left(x,y\right) \tag{16}$$

Proposition 3 *The following summation formulae for Hermite-Bell based Stirling polynomials of the second kind hold for non-negative integers* n *and* k *with* $n\geq k$*:*

$${}_{HB}S_2\left(n,k:x_1+x_2,y,z\right)=\sum_{u=0}^{n}\binom{n}{u}\,{}_{HB}S_2\left(u,k:x_1,y,z\right)x_2^{n-u} \tag{17}$$

and

$${}_{HB}S_2\left(n,k:x,y,z_1+z_2\right)=\sum_{u=0}^{n}\binom{n}{u}\,{}_{HB}S_2\left(u,k:x,y,z_1\right)HB_{n-u}\left(z_2\right). \tag{18}$$

Proposition 4 *The following relation* is *valid for non-negative integer* n*:*

$$ {}_{HB}S_2\left(n,k_1+k_2:x,y,z\right)=\frac{k_1!k_2!}{\left(k_1+k_2\right)!}\sum_{u=0}^{n}\binom{n}{u}_{HB} S_2\left(u,k_1:x,y,z\right)S_2\left(n-u,k_2\right) \tag{19}$$

Proposition 5 *The following relation holds for non-negative integer* n:

$$ S_2\left(n,k\right)=\sum_{u=0}^{n}\binom{n}{u}_{HB} S_2\left(u,k:x,y,z\right)HB_{n-u}\left(-x,-y,-z\right) \tag{20}$$

4 Hermite-Bell based Bernoulli Polynomials of order α

In this section, we introduce Hermite-Bell based Bernoulli polynomials of order α and investigate multifarious correlations and formulas.

Definition 2 *The Hermite-Bell based Bernoulli polynomials of order* α *are defined by the following exponential generating function:*

$$ \sum_{n=0}^{\infty}{}_{HB}B_n^{(\alpha)}\left(x;y\right)\frac{t^n}{n!}=\left(\frac{t}{e^t-1}\right)^{\alpha}e^{xt+yt^2+z\left(e^t-1\right)} \tag{21}$$

Remark 3 *In the special case* $x=0$ *in (21), we acquire the extended Bell-Bernoulli polynomials of order* α*, which are also new extensions of the Bernoulli numbers of order* α *in (8), as follows:*

$$ \sum_{n=0}^{\infty}{}_{HB}B_n^{(\alpha)}\left(y,z\right)\frac{t^n}{n!}=\left(\frac{t}{e^t-1}\right)^{\alpha}e^{yt^2+z\left(e^t-1\right)}. \tag{22}$$

We also note that

$${}_{HB}B_n^{(1)}\left(x;y,z\right):={}_{HB}B_n\left(x;y,z\right)$$

which we call the Hermite-Bell based Bernoulli polynomials.

We now provide the following theorems without their proofs that can be done by using (21) and series manipulation methods.

Theorem 1 *Each of the following summation formulae hold for* $n\in\mathbb{N}_0$*:*

$$ {}_{HB}B_n^{(\alpha)}\left(x;y,z\right)=\sum_{k=0}^{n}\binom{n}{k}B_k^{(\alpha)}HB_{n-k}\left(x,y,z\right) \tag{23}$$

$$ {}_{HB}B_n^{(\alpha)}\left(x;y,z\right)=\sum_{k=0}^{n}\binom{n}{k}B_k^{(\alpha)}\left(x;y\right)Bel_{n-k}\left(z\right) \tag{24}$$

$$ {}_{HB}B_n^{(\alpha)}\left(x;y,z\right)=\sum_{k=0}^{n}\binom{n}{k}_{HB}B_k^{(\alpha)}\left(y,z\right)x^{n-k} \tag{25}$$

Theorem 2 *The following relationship is valid for* $n\in\mathbb{N}_0$:

$$ {}_{HB}B_n^{(\alpha_1+\alpha_2)}\left(x_1+x_2;y_1+y_2,z_1+z_2\right)=\sum_{k=0}^{n}\binom{n}{k}_{HB}B_k^{(\alpha_1)}\left(x_1;y_1,z_1\right){}_{HB}B_{n-k}^{(\alpha_2)}\left(x_2;y_2,z_1\right) \tag{26}$$

Theorem 3 *The difference operator formulas for the Hermite-Bell based Bernoulli polynomials* hold for $n\in\mathbb{N}$:

$$ \frac{\partial}{\partial x}\,{}_{HB}B_n^{(\alpha)}\left(x;y,z\right)=n\,{}_{HB}B_{n-1}^{(\alpha)}\left(x;y\right) \tag{27}$$

and

$$ \frac{\partial}{\partial z}\,{}_{HB}B_n^{(\alpha)}\left(x;y,z\right)=\,{}_{HB}B_n^{(\alpha)}\left(x;y,z\right)-\,{}_{HB}B_n^{(\alpha)}\left(x+1;y,z\right). \tag{28}$$

Theorem 4 *The following summation formula holds for* $n\in\mathbb{N}_0$*:*

$$HB_n(x;y,z)=\frac{{}_{HB}B_{n+1}(x+1;y,z)-{}_{HB}B_{n+1}(x;y,z)}{n+1}=\frac{1}{n+1}\sum_{k=0}^{n}\binom{n+1}{k}{}_{HB}B_k^{(\alpha)}(x;y,z) \tag{29}$$

Theorem 5 *The following explicit formula holds for* $n\in\mathbb{N}_0$*:*

$${}_{HB}B_n(x;y,z)=\sum_{k=0}^{\infty}\sum_{l=0}^{k-1}y^k\binom{k-1}{l}(-1)^{k-l-1}\frac{H_{n+1}(l+x,y)}{n+1}$$

Theorem 6 *The following formula* is valid for $n\in\mathbb{N}_0$ and $k\in\mathbb{N}$:

$$HB_n(x;y,z)=\frac{n!k!}{(n+k)!}\sum_{l=0}^{n+k}\binom{n+k}{l}{}_{HB}B_l^{(-k)}(x;y,z)S_2(n+k-l,m) \tag{30}$$

Theorem 7 *The following correlation holds for non-negative integers* n*:*

$${}_{HB}B_n^{(\alpha)}(x;y,z)=\sum_{l=0}^{n}\sum_{k=0}^{\infty}\binom{n}{l}(x)_k\,S_2(l,k)\,{}_{HB}B_{n-l}^{(\alpha)}(y) \tag{31}$$

Theorem 8 *The following summation formula holds for non-negative integers* k *and* n *with* $n\geq k$*:*

$$HB_n(x_1+x_2,y_1+y_2,z_1+z_2)=\frac{n!k!}{(n+k)!}\sum_{l=0}^{n+k}\binom{n+k}{l}{}_{HB}B_l^{(k)}(x_2;y_2,z_2)\;{}_{HB}S_2(n+k-l,k:x_1,y_1,z_2) \tag{32}$$

Recently, implicit summation formulas and symmetric identities for special polynomials have been studied by some mathematicians, *cf.* (Khan et al., 2008, Pathan et al., 2015) and see the references cited therein.

We note that the following series manipulation formulas hold (Pathan et al., 2015):

$$\sum_{N=0}^{\infty}f(N)\frac{(x+y)^N}{N!}=\sum_{n,m=0}^{\infty}f(n+m)\frac{x^n}{n!}\frac{y^m}{m!} \tag{33}$$

Theorem 9 *The following implicit summation formula holds:*

$${}_{HB}B_{k+l}^{(\alpha)}(x;y,z)=\sum_{n,m=0}^{k,l}\binom{k}{n}\binom{l}{m}(x-\omega)^{n+m}\,{}_{HB}B_{k+l-n-m}^{(\alpha)}(\omega;y,z) \tag{34}$$

Proof. Upon setting t by $t+u$ in (21), we derive

$$\left(\frac{t+u}{e^{t+u}-1}\right)^{\alpha}e^{y(t+u)^2+z\left(e^{t+u}-1\right)}=e^{-\omega(t+u)}\sum_{k,l=0}^{\infty}{}_{HB}B_{k+l}^{(\alpha)}(\omega;y,z)\frac{t^k}{k!}\frac{u^l}{l!}.$$

Again replacing ω by x in the last equation, and using (33), we get

$$e^{-x(t+u)}\sum_{k,l=0}^{\infty}{}_{HB}B_{k+l}^{(\alpha)}(x;y,z)\frac{t^k}{k!}\frac{u^l}{l!}=\left(\frac{t+u}{e^{t+u}-1}\right)^{\alpha}e^{y(t+u)^2+z\left(e^{t+u}-1\right)}$$

By the last two equations, we obtain

$$\sum_{k,l=0}^{\infty}{}_{HB}B_{k+l}^{(\alpha)}(x;y,z)\frac{t^k}{k!}\frac{u^l}{l!}=e^{(x-\omega)(t+u)}\sum_{k,l=0}^{\infty}{}_{HB}B_{k+l}^{(\alpha)}(\omega;y,z)\frac{t^k}{k!}\frac{u^l}{l!},$$

which implies the asserted result (34).

Theorem 10 *The following symmetric identity holds for* $a,b\in\mathbb{R}$ *and* $n\geq 0$*:*

$$\sum_{k=0}^{n}\binom{n}{k}{}_{HB}B_{n-k}^{(\alpha)}(bx;y,z)\;{}_{HB}B_k^{(\alpha)}(ax;y,z)a^{n-k}b^k=\sum_{k=0}^{n}\binom{n}{k}{}_{HB}B_k^{(\alpha)}(bx;y,z)\;{}_{HB}B_{n-k}^{(\alpha)}(ax;y,z)a^kb^{n-k}$$

Theorem 11 *Let* $a,b\in\mathbb{R}$ *and* $n\geq 0$*. Then the following identity holds:*

$$\sum_{k=0}^{n}\sum_{i=0}^{a-1}\sum_{j=0}^{b-1}\binom{n}{k}_{HB} B_k^{(\alpha)}\left(i+\frac{b}{a}j+bx_1;y,z\right)_{HB} B_{n-k}^{(\alpha)}\left(ax_2;y,z\right)a^k b^{n-k}$$

$$=\sum_{k=0}^{n}\sum_{i=0}^{a-1}\sum_{j=0}^{b-1}\binom{n}{k}_{HB} B_k^{(\alpha)}\left(\frac{a}{b}i+j+ax_2;y,z\right)_{HB} B_{n-k}^{(\alpha)}\left(bx_1;y,z\right)b^k a^{n-k}.$$

Theorem 12 *The following symmetric identity*

$$\sum_{l=0}^{n}\sum_{k=0}^{l}\binom{n}{l}\binom{l}{k}S_{n-l}\left(b-1\right)_{HB} B_k^{(\alpha)}\left(bx_1;y\right)_{HB} B_{l-k}^{(\alpha+1)}\left(ax_2;y,z\right)a^{n+k+1-l}b^{l-k}$$

$$=\sum_{l=0}^{n}\sum_{k=0}^{l}\binom{n}{l}\binom{l}{k}S_{n-l}\left(a-1\right)_{HB} B_k^{(\alpha)}\left(ax_2;y\right)_{HB} B_{l-k}^{(\alpha+1)}\left(bx_1;y,z\right)b^{n+k+1-l}a^{l-k}$$

holds for $a,b\in\mathbb{Z}$ and $n\geq 0$.

References

Acikgoz, M, Ates, R., Duran, U., Araci, S. (2018). Applications of q-Umbral Calculus to Modified Apostol Type q-Bernoulli Polynomials. Journal of Mathematics and Statistics, 14(1), 7-15.

Bell, E. T. (1934). Exponential polynomials. Ann. Math. 35, 258-277.

Boas, R. P., Buck, R. C. (1958). Polynomial expansions of analytic functions, Springer-Verlag, Berlin, Gottingen, Heidelberg.

Carlitz, L. (1980). Some remarks on the Bell numbers. Fibonacci Quart. 18, 66-73.

Dere, R., Simsek, Y., Srivastava, H. M. (2013). A unified presentation of three families of generalized Apostol type polynomials based upon the theory of the umbral calculus and the umbral algebra, J. Number Theory, 133, 3245-3263.

Dere, R., Simsek, Y. (2012). Applications of umbral algebra to some special polynomials, Adv. Stud. Contemp. Math., 22 (3), pp. 433-438.

Dere, R., Simsek, Y., (2015). Hermite base Bernoulli type polynomials on the umbral algebra. Russ. J. Math. Phys. 22, 1-5.

Duran, U., Araci, S., Acikgoz, M. (2021). Bell-Based Bernoulli Polynomials with Applications. Axioms. 10, 29.

Khan, S., Pathan, M. A., Hassan, N. A. M., Yasmin, G. (2008). Implicit summation formulae for Hermite and related polynomials, J. Math. Anal. Appl. 344, 408-416.

Khan, N., Husain, N. (2021). Analysis of Bell based Euler polynomials and thier application, arXiv:2104.09129v1 [math.NT].

Kim, D. S., Kim, T. (2015). Some identities of Bell polynomials. Sci. China Math. 58, 2095-2104.

Kim, T., Kim, D. S.,, Jang, G. W., Jang, L. C. (2019). A generalizations of some results for Appell polynomials to Sheffer polynomials, J. Comput. Anal. Appl. 26, 889-898.

Kim, D. S., Kim, T. (2015). Umbral calculus associated with Bernoulli polynomials, J. Number Theory, 147, 871-882.

Kim, D. S., Kim, T., Lee, S. H., Rim, S. H. (2013). Some identities of Bernoulli, Euler and Abel polynomials arising from umbral calculus. Adv Differ Equ 2013, 15.

Kim, T., Kim, D. S., Jang, L. C., Lee, H., Kim, H. Y. (2021). Complete and incomplete Bell polynomials associated with Lah-Bell numbers and polynomials. Adv Differ Equ 2021, 101.

Kim, T., Kim, D. S., Kwon, H. I., Rim, S. H. (2017). Some identities for umbral calculus associated with partially degenerate Bell numbers and polynomials, Journal of Nonlinear Sciences and Applications, 10(6), 2966-2975.

Kim, T., Kim, D. S., Jang, G. W., Kwon, J. (2017). Fourier series of higher-order ordered Bell functions, Journal of Nonlinear Sciences and Applications, 10(7), 3851–3855.

Kim, T., Kim, D. S., Jang, G. W., Jang, L. C. (2017). Degenerate ordered Bell numbers and polynomials associated with umbral calculus, Journal of Nonlinear Sciences and Applications, 10(10), 5142–5155.

Kim, T., Kim, D. S., Dolgy, D. V., Kwon, J. (2018). Fourier series of finite product of Bernoulli and ordered Bell functions, Journal of Nonlinear Sciences and Applications, 11 (2018), no. 4, 500-515.

Pathan, M. A., Khan, W. A. (2015). Some implicit summation formulas and symmetric identities for the generalized Hermite-Bernoulli polynomials, Mediterr. J. Math. 12, 679-695.

Roman, S. (1984). The Umbral Calculus, Academic Press, Inc., New York.

Srivastava, H. M., Choi, J. (2012). Zeta and q-Zeta functions and associated series and integrals; Elsevier Science Publishers: Amsterdam, The Netherlands, 674 p.

Srivastava, H. M., Pinter, A. (2004). Remarks on some relationships between the Bernoulli and Euler polynomials, Appl. Math. Lett. 17, 375-380.

Srivastava, H. M., Manocha, H. L. (1984). A treatise on generating functions, Ellis Horwood Limited. Co. New York.

Wang, W., Wang, T. (2009). Identities on Bell polynomials and Sheffer sequences, Discrete Mathematics, 309(6), 1637-1648.

Genetic Algorithm Responses of Advection-Diffusion Processes

Esra Yaganoglu, Murat Sarı*, Hande Uslu

Yıldız Technical University, Mathematics Department, Istanbul, Turkey

Abstract

This work provides a new evolutionary approach based on a modified Genetic Algorithm to produce the advection diffusion responses of natural processes encountered in various branches of science. Since evolutionary processes; crossover, mutation and selection, yield the fittest survival in nature, the algorithm mimics the same procedures to find the optimum candidate solution of an advection-diffusion model. The proposed method is seen to be a very effective for the model by giving the consistent error intervals. Besides, three cases of the advection diffusion equation with different Peclet number values are examined and the results are compared with the exact solution of the model equation to show how the algorithm is reliable in terms of challenging cases of the processes in nature.

Keywords: Genetic Algorithm, Advection Diffusion Equation

1. Introduction

During the last decades, there is a growing interest on evolutionary algorithms (EA), which is inspired from evolutionary processes in nature. Genetic Algorithm (GA), one of evolutionary computing methods, have recently been used to solve optimization problems, since they have been found practical to be applied for numerous type of problems and can be get close to optimal solutions in considerably short time.

The fundamental idea behind EA's is first introduced by Holland (1962) who tried to understand the principles underlying natural adaptive systems. With the guidance of his writings on Adaptive Systems Theory, the progress of evolutionary models and the use of geneticlike operators for optimization problems have been proceeded since 1960s. First usage of the term "Genetic Algorithm" and first publishment on the application of the method belongs to Bagley (1967). While his colleagues Cavicchio (1970) and Hollstien (1971) had been performed experimental studies, Holland developed his Schema Theorem which provides a macroscopic model testing the effectiveness of the algorithm and explains why GA effectively works. Holland (1975) then put all his works together on his novel book "Adaptation in Natural and Artificial Systems" which gives him the name "Father of Genetic Algorithm" and this book had been accepted the first published guidance in the heuristic algorithm world. By the improvement of computational resources and advanced programming languages in 1980s, a great number of iterations could have been used in calculations for more larger population. This had been accepted the key feature leading the advancement of the method and the utility of evolutionary methods for a solution to differential equations.

First use of EAs for ordinary differential equations has been carried out by the study of Koza (1992) and Diver (1993) by cobsiderable proper results. Also in later years an improved RNA GA (Xin et al, 2017) and Taylor Series assisted GA (Gutierrez et al, 2018) have been presented for solutions of ODEs. Heat transfer problems has became the next field on which GA has began to be used regularly in 1990s.

* *sarim@yildiz.edu.tr*

Davalos et al (1996,1998) have been used the methodology for through the solution of one and two dimensional transient conduction problems. Later in 2000s higher performance of computational environment and accumulated theoretical knowledge have led to a dramatical growth in GA studies for solving partial differential equations. Recently, the algorithm has been advanced in terms of both diversity and complexity of the method by the variety of genetic operators and fine-tuned meshes.

The usage of genetic algorithm as an alternative approach for solving partial differential equations has been got attention with various studies. Poisson Equation (Jebari, 2013); Schrödinger Equation (Saha et al, 2001), (Sugawara, 2001); inverse-heat problems (Karr et al, 2000), (Liu et al, 2008), (Raudensky et al, 1995); and heat-conduction problems (Kadri et al, 2014), (Tsourkas et al, 2003, 2005) have been accepted the main studies which uses GA on heat transfer problems. Various structures of GA have been used for heat transfer equations involving only conduction processes. Study of Xiaohua et al (2008) is the only publication that presents GA approach for both convection and diffusion processes. They have been proposed chaos-gray coded GA for pollution source identification in Convection Diffusion Equation which has the same mathematical structure with Advection Diffusion Equation (ADE).

Within this study, an improved GA has been constructed to obtain the advection diffusion responses of natural processes. This version of GA differs from the classical one in terms of the representation of the candidate solutions, breeding and selection techniques. The model equation is given and the applied algorithm has been discussed in the following sections.

2. Problem Definition

Consider one dimensional homogenous advection-diffusion initial boundary value problem with the following initial and boundary conditions;

$$\frac{\partial u(x,t)}{\partial t} + v\,\frac{\partial u(x,t)}{\partial x} = D\,\frac{\partial^2 u(x,t)}{\partial x^2} \quad x \in \Omega_x = [0,l]\,, t \in \Omega_t = [0,\infty] \tag{1}$$

with initial and boundary conditions

$u = f(x)$

$u(0,t) = g_1(t)$

$u(l,t) = g_2(t)$

where u is the concentration, v is the flow velocity and D the diffusivity constant. This equation describes one dimensional transport of a point mass through both advection and diffusion processes, without any source and sink terms. Mass is both transported and spread and the concentration depends on the quantity of advection and diffusion terms. If advection term is significantly higher than the diffusion term, then advection dominates and diffusion is negligible. The center of mass is transported in one way with negligible dispersion. If diffusion dominates and advection is negligible, it means that mass spreads almost symmetrically till homogeneity. Finally, if advection and diffusion terms are not significantly different, the transportation of the mass occurs for both in other words the concentration of the mass moves and spreads by the time. The relative behavior of advection versus diffusion is represented by Peclet number which has the following formulation for the given variables;

$$Pe = \frac{v\,l}{D}$$

For observing the real-life situations to presented model equation, three cases with different Peclet numbers has been examined by the current method. Table 1 shows the selected parameters for Peclet numbers as follows;

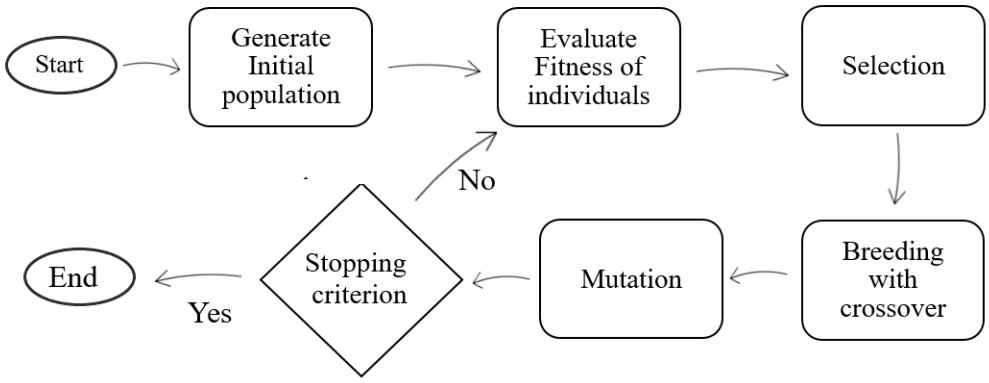
Table 1 Test Cases with characteristic length of l=8

	v (velocity)	D (diffusivity)	Peclet Number
Case 1	1	0,005	1600
Case 2	1	1	8
Case 3	0.005	1	0.04

3. Genetic Algorithm

Genetic Algorithm is a global search algorithm based on random exploration of the solution space. Compared to conventional methods, GA enables to explore wider range of possible solutions. Among some number of initial candidate solutions, the algorithm attempts to find the optimal solution by a process similar to natural selection. Genetic Algorithm has been considered as Heuristic Search method which is an approach for problem solving that proposes a practical method which does not guarantee the perfect solution, but sufficient for reaching a short-term goal or approximation. So, GA searches for the best among the limited number of solutions and does not guarantee the best solution, but the optimal solution

The structure of Genetic Algorithms is borrowed from natural processes, so inspired from the idea of evolution the algorithm contains five stages; representation of initial population, fitness assessment, selection, crossover and mutation as can be seen the following flowchart of the algorithm.



3.1. Initialization of Population

The algorithm starts with a set of initial guess for solution which is generated randomly from the beginning. Each candidate solution is called 'individual' or 'chromosome'. The choice of the appropriate representation for individuals is important and must be problem-specific. In literature, binary or real-valued representation of the candidate solution have been used for differential equations depending on design of the fitness function. Classical GA binary-coding are preferred in general. But since discrete numeric values are directly used to assess fitness, candidate solutions have been preferred to be real-valued arrays instead of commonly used binary strings. Solution to one dimensional ADE is a surface with two independent variables, so the candidate solutions have been represented by two-dimensional arrays. The matrix $A^{(k)} \in \mathbb{R}^{mxn}$ refers to k th candidate solution, where m and n are the chromosome sizes in space and time, respectively.

Size of the initial population is another parameter of the algorithm. Larger population size has expected to be useful for sharing the desired genetic information, since GA assumes that individuals carries some useful information, and with breeding process it will be transferred to next generation. So it was preferred to examine the effect of the initial population size with four different values.

3.2. Fitness Assessment

GA is a self-adapting method, so it proceeds by learning from previous knowledge, the performance of the former generation. To assess the performance, fitness function measures how well a candidate solution fits the desired solution and the design of it must be appropriate to structure of the problem. For ADE, fitness function was calculated directly by the equation itself. Time and space derivatives were discretized by central finite difference approximations and quadratic extrapolant formulas were applied at boundaries;

$$u_t(x_i,t_j) \approx \frac{-3u_i^j + 4u_i^{j+1} - u_i^{j+2}}{2\Delta t} \qquad u_x(x_i,t_j) \approx \frac{-3u_i^j + 4u_{i+1}^j - u_{i+2}^j}{2\Delta x}$$

$$u_{xx}(x_i,t_j) \approx \frac{u_i^j - 2u_{i+1}^j + u_{i+2}^j}{\Delta x^2}$$

Fitness of each chromosome were calculated by substituting derivatives by finite difference approximations through ADE and the result were expected to be close to zero as candidate solution gets close to exact solution. The smaller fitness result means the better fit to the solution and smaller error. Hence, the fitness values were considered as errors and overall errors were calculated as follows;

$$F(i,j)^{(k)} = \frac{u_i^{j+1} - u_i^{j-1}}{2\Delta t} + v\frac{u_{i+1}^j - u_{i-1}^j}{2\Delta x} - D\frac{u_i^{j+1} - 2u_i^j + u_i^{j-1}}{\Delta t^2} \rightarrow 0$$

$$F_{total}{}^{(k)} = \frac{1}{N_x N_t}\sum_{i=1}^{N_x}\sum_{j=1}^{N_t}(F(i,j) - F_{avg}{}^{(k)})^2$$

where $F(i,j)^{(k)}$ is the error of one component of k th chromosome and $F_{avg}{}^{(k)}$ is the average error of the k th chromosome. The choice of variance as total error measurement is because of determination and elimination of individuals with excessive errors of genes which increases the variance. Chromosome with outlier genes are intended to be eliminated to speed up the convergence with this choice.

3.3. Selection

Survival of the fittest individuals are maintained through selection step of the algorithm. Once the population has been evaluated, p number of individuals which have the smallest errors, have been selected to be parents of the next generation with probability of $p/2$ are moms and $p/2$ are dads. At the next stage, they were chosen randomly from mating pool and mated to produce new offsprings.

3.4. Crossover

Genetic information of one generation to the next one is transferred through crossover operator. As Holland (1975) states, the effectiveness of GA is highly determined by crossover. Two parents have been selected from the mating pool randomly. Recombination of the sections of parents have assembled to each other and two new members of population have been obtained.
In our study cross-over operator differs from simple GA in two ways;

- Horizontal and vertical crossovers; For one-dimensional arrays, crossover is performed by swapping parts of two parental chromosomes and two new individuals are created. Two-dimensional chromosome operator is improved in such a way that both horizontal and vertical crossovers are applied. So number of offsprings from one couple is increased with two types of crossovers which is expected to promote diversity of the population. Consider A_{mxn} and B_{mxn} as parent individuals, then the surface chromosomes of parents are cut vertically at the i th point. Thus, vertical offsprings v_1 and v_2 are generated as follows. Horizontal crossover results in two new individuals in a similar way.

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,i} & A_{1,i+1} & \cdots & A_{1,n} \\ A_{2,1} & & A_{2,i} & A_{2,i+1} & & A_{2,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ A_{m,1} & & A_{m,i} & A_{m,i+1} & & A_{m,n} \end{bmatrix}, \quad B = \begin{bmatrix} B_{1,1} & \cdots & B_{1,i} & B_{1,i+1} & \cdots & B_{1,n} \\ B_{2,1} & & B_{2,i} & B_{2,i+1} & & B_{2,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ B_{m,1} & & B_{m,i} & B_{m,i+1} & & B_{m,n} \end{bmatrix}$$

$$v_1 = \begin{bmatrix} A_{1,1} & \cdots & A_{1,i} & B_{1,i+1} & \cdots & B_{1,n} \\ A_{2,1} & & A_{2,i} & B_{2,i+1} & & B_{2,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ A_{m,1} & & A_{m,i} & B_{m,i+1} & & B_{m,n} \end{bmatrix}, \quad v_2 = \begin{bmatrix} B_{1,1} & \cdots & B_{1,i} & A_{1,i+1} & \cdots & A_{1,n} \\ B_{2,1} & & B_{2,i} & A_{2,i+1} & & A_{2,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ B_{m,1} & & B_{m,i} & A_{m,i+1} & & A_{m,n} \end{bmatrix}$$

- n-point crossover; Conventional genetic algorithms select the crossover point randomly among $(m-1)$ and $(n-1)$ alternatives. Probably, the selection of one crossover point may eliminate better combinations of the sections. So, the idea of including all possible offsprings with all possible crossovers to the population, is expected to contribute the diversity. With n-point crossover, population becomes huge-sized, and the bests are selected among them at each generation. By means of these modifications of the simple genetic algorithm, solution space is expected to be represented better.

3.5. Mutation

During the iteration process, the difference among individuals has been expected to decrease with the elimination of the outliers. Avoiding the convergence to non-optimal solutions, specified percentage of the genes, which was defined as mutation rate, are modified randomly to keep searching the neighborhood of the solution. This may help the algorithm to catch the optimal solution. Over-mutation may cause the candidate solution to get away from optimal. So, mutation rate was kept in the interval $[0.01-0.05]$ in literature. It was determined as 5% in this study.

4. Results and Discussion

In this section, the proposed method has been applied for three different cases of one dimensional homogenous advection-diffusion initial boundary value problem with Dirichlet boundary conditions. The results from cases has been evaluated in terms of optimal chromosome sizes with low error terms. Besides, as a contribution to improved crossover operator, n-point crossover has been compared with one-point crossover for each cases.

Furthermore, since Genetic Algorithm has been claimed to be an alternative to numerical methods in terms of ease of implementation, application and low computational cost, running time will be given for comparison purpuses. All experiments are performed with MATLAB on a personal computer with Intel(R) Core (TM) i5-8250U 1.60GHz - 1.80 GHz CPU and 8.00 GB RAM. Initial population size has been determined as p=1000 and number of generations as N=100. The error of cases of the equation are shown with following Figures and Tables:

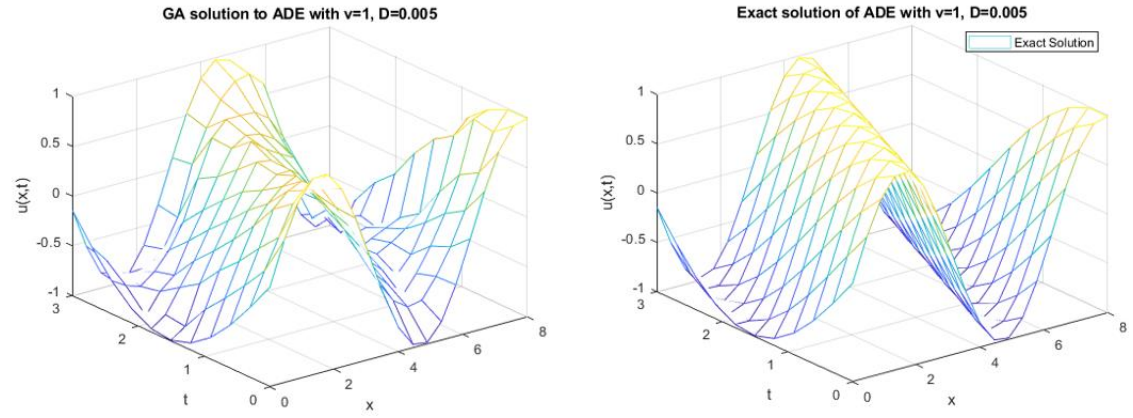


Figure 1 Comparison of GA results with the exact solution for Case 1 when number of iteration (N) is 100, initial population size (p) is 1000, with chromosome sizes of [19,11]

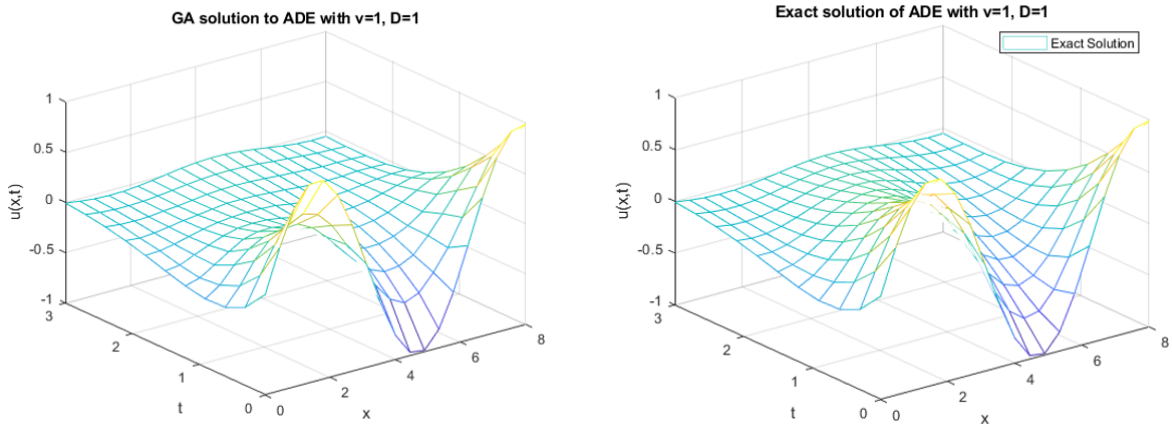


Figure 2 Comparison of GA results with the exact solution for Case 2 when number of iteration (N) is 100, initial population size (p) is 1000, with chromosome sizes of [19,11]

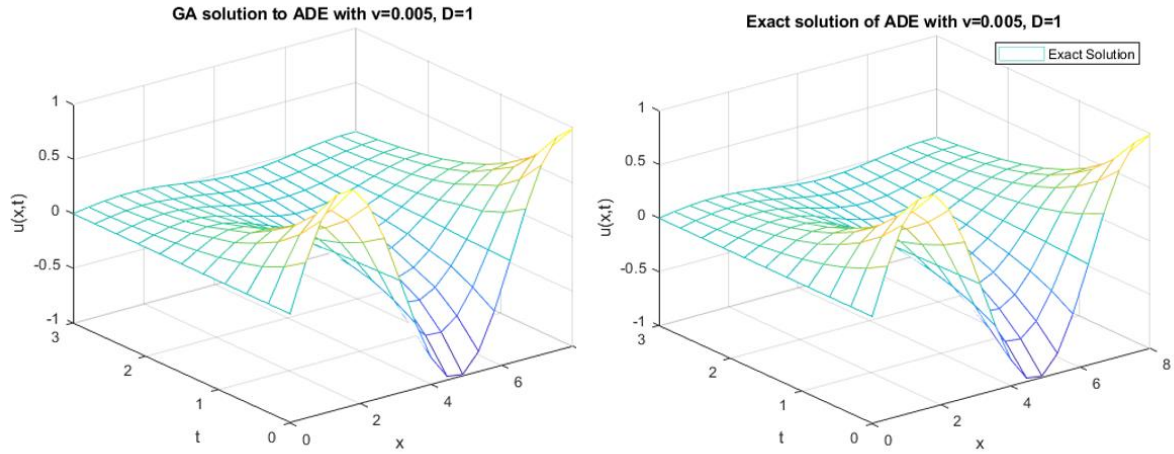


Figure 3 Comparison of GA results with the exact solution for Case 3 when number of iteration (N) is 100, initial population size (p) is 1000, with chromosome sizes of [19,11]

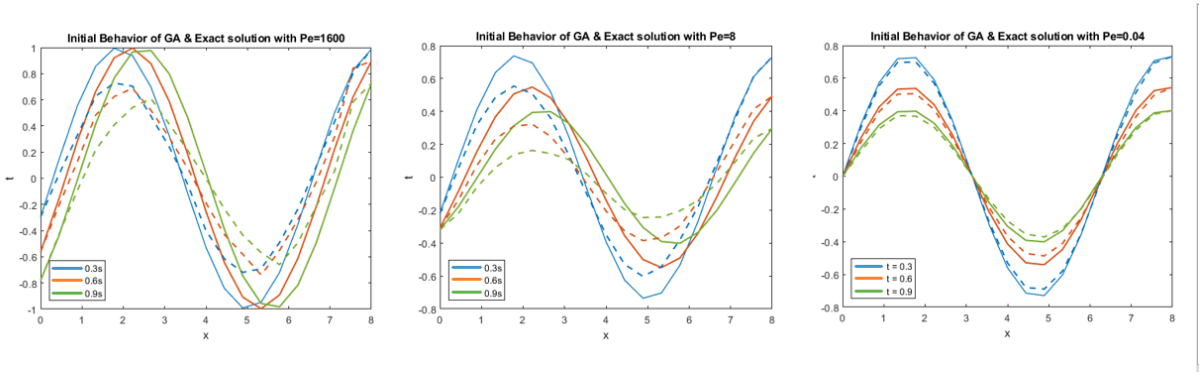


Figure 4 Comparison of GA results with the exact solution at first three time steps [0.3s, 0.6s, 0.9s] where continuous lines represents exact solution and the dotted lines represents GA solution.

As can be seen in pair of graphs, comperatively, the behaviour of the solution has seen to be very similar with the exact solution in all three test cases. In advection dominated case 1 with the Peclet number 1600, the algorithm has seen to perform error less than 10^{-2} compared to the others. While case 2 has a considerable error of 10^{-3}, the diffusion dominated case 3, with Peclet number 0.04, has a significant convergence to capture the exact solution with error of 10^{-4}. So GA has accomplished to converge to the solution even with small set of generations as can be observed numerically from the following table.

Table 2 Total errors of optimal solutions with comparison of n-point and 1-point crossover operator where N represents number of generation and p is the initial population size.

		Case 1, Pe=1600			Case 2, with Pe=8			Case 3, Pe=0,04		
		p=100	p=500	p=1000	p=100	p=500	p=1000	p=100	p=500	p=1000
N=50	**n-point**	**0.0875**	**0.0532**	**0.0510**	**0.0173**	**0.0101**	**0.0099**	**0.0128**	**0.0019**	**4.52e-04**
	1-point	0.1104	0.1049	0.0935	0.0263	0.0219	0.0168	0.0174	0.0066	1.19e-03
N=100	**n-point**	**0.0870**	**0.0503**	**0.0497**	**0.0161**	**0.0094**	**0.0062**	**0.0066**	**0.0017**	**3.31e-04**
	1-point	0.1090	0.1042	0.0921	0.0242	0.0212	0.0123	0.0141	0.0043	6.19e-04
N=150	**n-point**	**0.0852**	**0.0499**	**0.0432**	**0.0158**	**0.0092**	**0.0061**	**0.0048**	**0.0016**	**3.02e-04**
	1-point	0.1035	0.1021	0.0893	0.0240	0.0205	0.0098	0.0131	0.0038	4.85e-04

Experiments have been performed with various number of generations, initial population size and Peclet numbers. To show the decreasing pattern which has been observed through the results, parameters have randomly selected. It can be argued that the performance of the algorithm is likely to increase for smaller Peclet numbers. Furthermore, error decreases with increased number of generations and larger initial population sizes which is expected to contribute to diversity. Furthermore, crossover operator has been developed with n-point crossover. Comparing the results with the one-point crossover errors, n-point operator almost halves the error values.

Sizes of the parameters affects the running time of algorithm which is approximately 30 minutes with initial population of size 1000 and 150 generations, for all three cases. Considerable convergence could be obtained even with small sizes of generations and candidate solutions and low computational cost whose the performance criteria for numerical methods. This is the main advantage for Genetic Algorithm compared to numerical methods for practical usage for engineering problems.

However, chromosome size in both dimensions plays a critical role in terms of both accuracy and the performance of the method. During experimental studies, it has seen that average error increases chomosome sizes greater than twenty and surface of the solution damages with outliers. Even though the step sizes are needed to be minimized for accuracy in numerical methods, the current method have

a tendency of giving accurate results for different step sizes. Therefore, the presented work has been seen the reliable approach without decreasing step sizes.

5. Conclusion and Recommendations

In this study, a modified genetic algorithm has been explored for solving one dimensional ADE. To observe the consistency of the algortihm with nature, three different cases have been created by different Peclet numbers representing relative behavior of advection and diffusion in the real model. These three cases have been presented in a comparative way by discussing the parameters of the algorithm. The fast convergence to near-optimal solutions with very low computational costs had been observed for the real model. Moreover, the flexible implementation of the algorithm which are not restricted with structural limitations has seen to be practical to be implemented for more complex problems. Even though the considered example of ADE is seem to be very basic, it can be definetely concluded that the algorithm can be easily applied for models including nonlinearities, source terms or additional dimensions in nature.

References

Bagley, J. D. (1967). The behaviour of adaptive systems which employ genetic and correlational algorithms (Doctoral Dissertation). University of Michigan, Ann Arbor; Dissertation Abstracts" Int. 28 (12) (1967) 5106B; University Microfilms No. 68-7556.

Cavicchio, D. J. (1970). Adaptive Search using Simulated Evolution (Doctoral dissertation). Retrieved from https://hdl.handle.net/2027.42/4042

Davalos, R.V., Rubinsky, B. (1996). An Evolutionary-Genetic Approach to Heat Transfer Analysis, ASME J. Heat Transfer, vol. 118, no. 3, pp. 528_531, 1996. https://doi.org/10.1115/1.2822663

Davalos, R.V., Rubinsky, B. (1998). The Use of Evolutionary-Genetic Analogy in Numerical Analysis, Communications in Numerical Methods in Engineering, Vol. 14, 151-160
https://doi.org/10.1002/(SICI)1099-0887(199802)14:2<151::AID-CNM136>3.0.CO;2-N

Diver, D. A. (1993). Applications of Genetic Algorithms to the Solution of Ordinary Differential Equations, Journal of Physics. A General Physics: Math. Gen. 26 3503 https:// doi.org/10.1088/0305-4470/26/14/017

Gutierrez, D., Lopez-Aguayo, S., (2018). Solving ordinary differential equations using genetic algorithms and the Taylor series matrix method, Journal of Physics Communications, 2 -115010, https://doi.org/10.1088/2399-6528/aaedd2

Holland, J. H. (1962). Outline for a Logical Theory of Adaptive Systems. Journal of the ACM, 9, 297-314. http://dx.doi.org/10.1145/321127.321128

Holland, J. H. (1975). Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor. (2nd Edition, MIT Press, 1992.)

Hollstein, R. B. (1971). Artificial Genetic Adaptation in Computer Control Systems (Doctoral Dissertation). University of Michigan.

Jebari, K., Madiafi, M., Moujahid, A. (2013). Solving Poisson Equation by Genetic Algorithms, International Journal of Computer Applications (0975 – 8887) 83-5, https://doi.org/10.5120/14441-2597

Kadri, M. B., Waqar A. Khan, W. A. (2014). Application of Genetic Algorithms in Nonlinear Heat Conduction Problems, The Scientific World Journal 2014(6):451274. https://doi.org/10.1155/2014/451274

Karr, C. L., Yakushinb, I., Nicolosia, K. (2000). Solving inverse initial-value, boundary-value problems via genetic algorithm, Engineering Applications of Artificial Intelligence 13 (2000) 625–633. https://doi.org/10.1016/S0952-1976(00)00025-7

Koza, J. R. (1992). Genetic Programming on the programming of computers by means of natural selection. Cambridge, MA: MIT Press, 1992.

Liu, F. B. (2008) A modified genetic algorithm for solving the inverse heat transfer problem of estimating plan heat source, International Journal of Heat and Mass Transfer 51 (2008) 3745–3752 https://doi.org/10.1016/j.ijheatmasstransfer.2008.01.002

Raudenský, M., Woodbury, K. A., Kral, J., Brezina, T. (1995). Genetic Algorithm in solution of inverse Heat Conduction problems, Numerical Heat Transfer, Part B: Fundamentals: An International Journal of Computation and Methodology, 28:3, 293-306 https://doi.org/10.1080/10407799508928835

Saha, R., Chaudhury, P., Bhattacharyya S. P. (2001). Direct solution of Schrödinger equation by genetic algorithm: test cases, Physics Letters A 291 (2001) 397–406 https://doi.org/10.1016/S0375-9601(01)00704-6

Sugawara, M. (2001). Numerical solution of the Schrödinger equation by neural network and genetic algorithm, Computer Physics Communications 140 (2001) 366–380. https://doi.org/10.1016/S0010-4655(01)00286-7

Tsourkas, P., Rubinsky, B. (2003). Evolutionary-genetic algorithm for solving 2-D steady-state conduction problems, Numerical Heat Transfer B, vol. 43, no. 2, pp. 99–115, 2003. https://doi.org/10.1080/713836171

Tsourkas, P., Rubinsky, B. (2005). A Parallel Genetic Algorithm for Heat Conduction problems, Numerical Heat Transfer, Part B: Fundamentals: An International Journal of Computation and Methodology, 47:2, 97-110, https://doi.org/10.1080/10407790490515774

Xiaohua, Y., Zhifeng, Y., Xinan, Y., Jianqiang, L. (2008). Chaos gray-coded genetic algorithm and its application for pollution source identifications in convection–diffusion equation, Communications in Nonlinear Science and Numerical Simulation 13 1676–1688 https://doi.org/10.1016/j.cnsns.2007.03.003

Xin, Y. S., Jun, W. L., Xiao, Z., Jing, X. Z. (2017). An improved RNA genetic algorithm for the parameter estimation multiple solutions of ordinary differential equations, Procedia Engineering 174 (2017) 477 – 481, https://doi.org/10.1016/j.proeng.2017.01.169

UHF band Yagi-Uda TV Antenna

Sunil Narayan[1*], Ashwin Chand[1]

[1] School of Information Technology, Engineering, Mathematics, and Physics, The University of the South Pacific, Laucala Campus, Suva, Fiji
spnz2014@gmail.com, ashwinchand4@gmail.com

Abstract

Antenna is a metallic device which is used for transmitting and receiving an electromagnetic signal. Yagi-Uda antennas are known to be difficult to design and optimize due to their sensitivity at high gain, and the inclusion of numerous parasitic elements. Still, this antenna is familiar as the commonest kind of terrestrial TV antenna to be found on the rooftops of houses and have a wide application in other communication industries. It can be used at frequencies between about 30 MHz and 3 GHz. This paper presents simulated results of a UHF band Yagi-Uda antenna designed to operate in the UHF TV band ranging from 470 MHz to 890 MHz, using YO 6.5 Yagi Optimizer software. The designed antenna radiates an end-fire fan beam pattern with bandwidth of about 57% for voltage standing wave ratio (VSWR) less than 2. The simulated result shows that the antenna exhibits good bandwidth and moderate gain properties with good impedance characteristics.

Keywords: Yagi Uda antenna, simulation, VSWR, Gain,

1. Introduction

A Yagi-Uda Antenna, commonly known simply as a Yagi antenna or Yagi, is a directional antenna system consisting of an array of dipole and additional closely coupled parasitic elements (Balanis, 2016). Yagi-Uda Antenna is a widely used antenna design due to its high forward gain capability, low cost and ease to construction (Luo et al., 2015).

The main reason antenna engineering is such an extensive field is due to television. It is seemed that TV receiving antennas go from the once standard 'rabbit ears', through to the much larger roof mounted VHF antennas to the smaller UHF antennas, and now with the advent of pay TV channels in most parts of the world, companies utilize microwaves for television broadcasting (Fezai et al., 2013), (Gutierrez, 2017), (Vadivel et al., 2018), (singh et al, 2015), (Kanzaria et al., 2015), (Prasad et al., 2014)..

The UHF television band range from 470MHz to 890MHz and the most stringent requirement for TV broadcast antennas is that the voltage standing-wave ratio should be less than 1:1:1 over the band. There are also other several requirements a TV receiving antenna must meet; it must have sufficient gain and good impedance match in order to produce a signal of enough strength for a clear picture. It must also be able to reject reflected and unwanted signals arriving from different directions (Balanis, 2016). Losses due to reflected and therefore out of phase signals can be greatly reduced by using highly directive antennas.

The particular antenna studied in this paper is the Yagi-Uda, first proposed in 1926 (Uda et al, 1954). This was antenna was chosen because it presents difficult design and optimization challenges, and is

* *Corresponding Author*

widely used antenna design due to its high capability, low cost and ease of construction. The Yagi-Uda is an end-fire array consisting typically of a single driven dipole with the reflector dipole behind the driven element (driven from its centre), and one or more parasitic director elements as shown in Figure 1. The highest gain can be achieved along the axis and on the side with the directors. The reflector element reflects power forwards and thus acts like a small ground plane. The design parameters consist of element lengths, inter-element spacing, and element diameter.

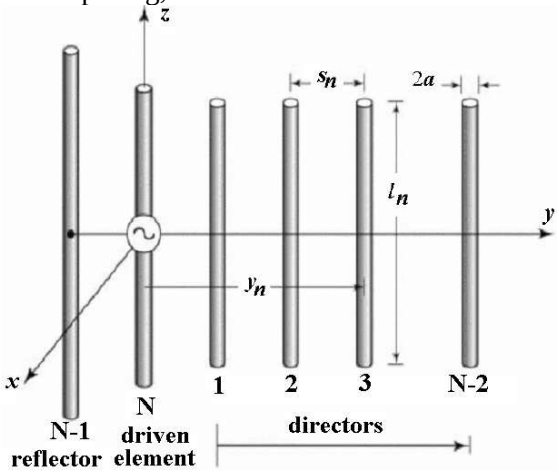


Figure 1. Antenna Components

The Yagi-Uda array antenna has the important practical advantage that a complex feed network is not required, only a single element is directly fed. In addition, the elements may be as simple as wires, rods or thin metallic tubes. So the Yagi-Uda array is popularly used in the application for radio beacons, radio links, and early radar systems in United States and Europe. It is ironical history in 1942, when the Japanese army invaded Singapore, they discovered that the Yagi-Uda array was being used as a radair antenna by the British army.

VSWR represents the degree with which an antenna is matched to the system impedance. It is the ratio between the highest voltage and the lowest voltage in the single envelope along a transmission line (Wu et al., 2013). The VSWR was desired to be less than 2 bandwidth and gain and impedance characteristic were to be maximized.

2. Yagi antenna design

There are formulas that we had used to decide both the length of the pieces and the spacing between them. The dimension of the elements is frequency-dependant. Usually a reflector is cut longer than the driver (< 0.5λ) and a director(0.5λ) is cut shorter than the driver, so that the main beam is enhanced in the same direction by both the reflector and director (Murali et al., 2014), (Perry et al., 2001). The dipole will be directly driven from a feed network, whereas the parasitic elements achieve excitation by the near field coupling from the driven element.

Figure 2 depicts a clear form of the Yagi-Uda antenna. The centre rod like structure on which the elements are mounted is called as boom. The element to which a thick black head is connected is the driven element to which the transmission line is connected internally, through that black stud. The single element present at the back of the driven element is the reflector, which reflects all the energy towards the direction of the radiation pattern. The other elements, before the driven element, are the directors, which direct the beam towards the desired angle.

Figure 2. Yagi Uda Antenna

The antenna can be designed from Figure 3 using the following design specifications as shown in Table 1.

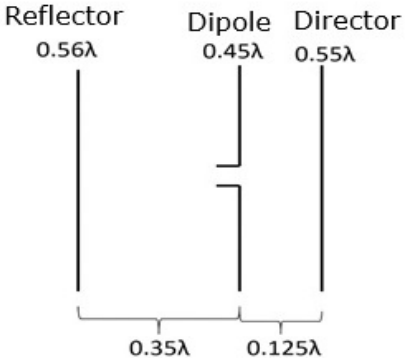


Figure 3. Antenna Specification

Table 1. Designing of Yagi-Uda antenna.

ELEMENT	SPECIFICATION
Length of the Driven Element	0.458λ to 0.5λ
Length of the Reflector	0.55λ to 0.58λ
Length of the Director 1	0.45λ
Length of the Director 2	0.40λ
Length of the Director 3	0.35λ
Spacing between Directors	0.2λ
Reflector to dipole spacing	0.35λ
Dipole to Director spacing	0.125λ

In a 6-element Yagi for UHF band TV antenna is designed using 10mm diameter untapered aluminium tubes insulated from metallic boom. The boom length and diameter are 0.86λ and 10mm respectively. Polyester is used as an insulating material. The designed antenna has vertical and horizontal polarization. Connector used would be N-female (it can also be TNV or BNC) since these types of connectors maintain good VSWR at high frequency as referred in Table 1: Designing UHF band Yagi-Uda TV antenna.

The maximum gain can be attained along the axis and on the side with the directors. The function of the reflector is to reflect power forwards and therefore it acts like a small ground plane. The spacing between each element is not identical and it can be considered as a non-linear array. The number of directors in the antenna depends on the gain requirements (Dubey et al., 2014). The flow of current on the active element of the yagi-uda antenna is determined by its length, frequency and coupling with nearby elements, while the current distribution in passive elements is governed by the boundary condition.

3. Simulation Results and Discussion

The analysis and design procedure is based on the full-wave electromagnetic solver based on YO 6.5 Yagi optimizer. Though the designed antennas performance on the actual operation was not checked but we suppose an almost similar result, as the simulation will come out when it is actually made. The reason being, several simulations programmes have been run and nearly same result was obtained from them. The simulated results for gain, front-back ratio, VSWR and input impedance against frequency are shown in Figure 4.

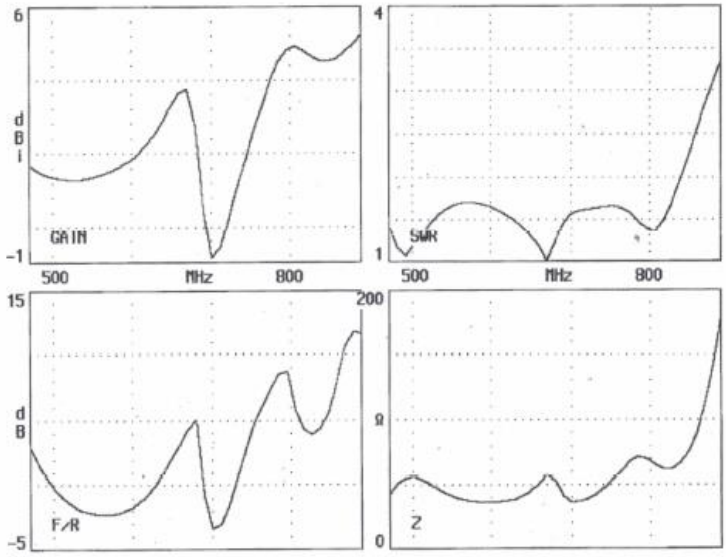


Figure 4. VSWR and Input Impedance against Frequency

From Figure 4 we notice the antenna provides a bandwidth of about 57% (450-840MHz) for VSWR < 2, gain of 3.6 dBi at centre frequency, and good impedance characteristic, as it is maintained between 40-70Ω over wide bandwidth. The antenna's matching frequency (669.68MHz) often changes depending on the environment where the antenna is mounted. It can either shift above or below the initial frequency but since the VSWR <2 is maintained for wide-bandwidth, this problem will be eliminated.

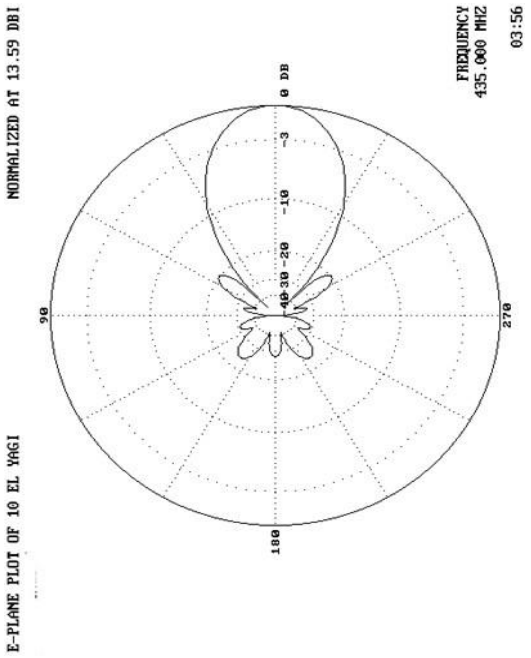


Figure 5a: Far Field radiation pattern in the E- plane

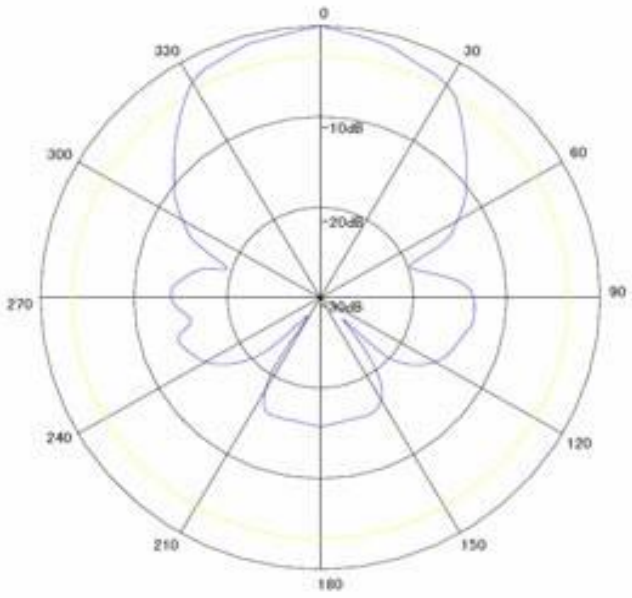


Figure 5b: Far Field radiation pattern in the H- plane

Figure 5a and 5b shows the calculated far field radiation pattern in the E and H-plane. The radiation pattern indicates a well-defined end fire with front to back ratio of more than 7dB. The simulated results of the antenna radiation pattern showed that the radiation pattern is fairly stable over the operating band.

4. Conclusion

A UHF band Yagi – Uda TV antenna was successfully demonstrated using Yo 6.5 Yagi Optimizer simulation. The antenna provides a bandwidth of about 57% for VSWR <2. The input impendence of 50Ω is maintained. The antenna designed had a gain of 3dBi at the center frequency. The proposed

antenna has a simple structure and is easy to construct. It provides an operating at the frequency range of 450 MHz - 890 MHz, centered at 680 MHz with a peak antenna gain of 3.61 dBi. Investigation on the location of the E and H-plane phase center over the entire antenna surface as opposed to only along its axis may also be conducted. Suitable techniques to reduce the cross polarization levels of these antennas may also be suggested

References

Balanis, C. A. (2015). Antenna theory: analysis and design. John Wiley & Sons.

Luo, Y., & Chu, Q. X. (2015). A Yagi–Uda antenna with a stepped-width reflector shorter than the driven element. IEEE Antennas and Wireless Propagation Letters, 15, 564-567.

Fezai, F., Menudier, C., Thévenot, M., & Monédière, T. (2013). Systematic design of parasitic element antennas—Application to a WLAN Yagi design. IEEE Antennas and Wireless Propagation Letters, 12, 413-416.

Gutierrez, F. (2017). Design of a wideband antenna for wireless network-on-chip in multimedia applications. Journal of Low Power Electronics and Applications, 7(2), 6.

Vadivel, M., Poongodhai, M., Madhumitha, R., Nivetha, V., & Banu, J. K. (2018). IOT based home visitor monitoring system using raspberry pi. International Research Journal of Engineering and Technology (IRJET), 5(3), 3132-3135.

Singh, S., Anap, P., Bhaigade, Y., Chavan, J.P. (2015). IP camera video surveillance using raspberry pi. International Journal of Advance Research in Computer and Communication Engineering, 4(2), 326-328.

Kanzariya, S., & Vora, P. V. (2015, March). Real time video monitoring system using raspberry pi. In National Conference on Emerging Trends in Computer, Electrical and electronics (ETCEE-2015).

Prasad, S., Mahalakshmi, P., Sunder, A.J.C., Swathi, R. (2014). Smart surveillance monitoring system using raspberry pi and pir sensor. International Journal of Computer Science and Information Technology, 5(6), 7107-7109.

Wu, J., Zhao, Z., Nie, Z., & Liu, Q. H. (2013). A broadband unidirectional antenna based on closely spaced loading method. IEEE transactions on antennas and propagation, 61(1), 109-116.

Dubey, S., & Zafar, S. (2014, March). Broadband microstrip Yagi array antenna for SC Band application. In 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science (pp. 1-4). IEEE.

Murali, S. D., Kavya, K. C. S., Kotamraju, S. K., Divya, R. S., Satish, Y., Chandra, K. P., & Sree, G. J. (2014). Size reduction of Yagi-Uda Antenna by altering the diameter and spacing between the elements. International Journal of Applied Engineering Research, 9(22), 12011-12022.

Perry, B. T., Coleman, C. M., Basch, B. F., Rothwell, E. J., & Ross, J. E. (2001, July). Self-structuring antenna for television reception. In IEEE Antennas and Propagation Society International Symposium. 2001 Digest. Held in conjunction with: USNC/URSI National Radio Science Meeting (Cat. No. 01CH37229) (Vol. 1, pp. 162-165). IEEE.

Uda, S. & Mushiak, Y. (1954). Yagi Uda Anttena, Maruzden, Tokyo

Degenerate Poisson-Charlier Polynomials

Uğur Duran[1*], Mehmet Acikgoz[2]

[1] *İskenderun Technical University, Department of Basic Sciences and Engineering, Hatay, Turkey*
[2] *Gaziantep University, Department of Mathematics, Gaziantep, Turkey*
ugur.duran@iste.edu.tr, acikgoz@gantep.edu.tr

Abstract

In this paper, we introduce both the generalized degenerate Gould-Hopper based degenerate Stirling polynomials of the second kind and the generalized degenerate Gould-Hopper based fully degenerate Bell polynomials. We study and investigate multifarious properties and relations of these polynomials such as explicit formulae, differentiation rules and summation formulas. Moreover, we derive several correlations with the degenerate Bernstein polynomials for these polynomials. Furthermore, we acquire several representations of the generalized degenerate Gould-Hopper based fully degenerate Bell polynomials via not only the fully degenerate Bell polynomials but also the generalized degenerate Gould-Hopper based degenerate Bernoulli, Euler and Genocchi polynomials.

Keywords: Poisson-Charlier polynomials, degenerate exponential function, special polynomials

1. Introduction

For $\lambda \in \mathbb{C}$, the λ-falling factorial $(x)_{n,\lambda}$ is defined by (Appell et al., 1926, Carlitz et al., 1975, Carlitz et al., 1956, Cheikh et al., 2003, Dattoli et al., 1999, Duran et al., 2021, Duran et al., 2018, Duran et al., 2019, Duran et al., 2020, Howard, 1996, Howard, 1979, Khan et al., 2016, Kim et al., 2018, Kim et al., 2018, Kim et al., 2019, Kim et al., 2017, Lim et al., 2016, Njionou, 2020, Khan et al., 2018):

$$(x)_{n,\lambda} = \begin{cases} x(x-\lambda)(x-2\lambda)\cdots(x-(n-1)\lambda), & n=1,2,\ldots \\ 1 & n=0. \end{cases} \tag{1.1}$$

In the case $\lambda = 1$, the λ-falling factorial reduces to the familiar falling factorial $(x)_n$ (Appell et al., 1926, Carlitz et al., 1975, Carlitz et al., 1956, Cheikh et al., 2003, Dattoli et al., 1999, Duran et al., 2021, Duran et al., 2018, Duran et al., 2019, Duran et al., 2020, Howard, 1996, Howard, 1979, Khan et al., 2016, Kim et al., 2018, Kim et al., 2018, Kim et al., 2019, Kim et al., 2017, Lim et al., 2016, Njionou et al., 2020, Khan et al., 2018, Rainville, 1960, Srivastava et al., 2012)

$$(x)_n = x(x-1)\cdots(x-n+1).$$

The Stirling numbers of the first kind $S_1(n,m)$ are defined by means of the falling factorial as follows

$$(x)_n = \sum_{m=0}^{n} S_1(n,m)\, x^m, \tag{1.2}$$

cf. (Carlitz et al., 1975, Duran et al., 2020, Rainville, 1960, Srivastava et al., 2012) and see also references cited therein.

The Δ_λ difference operator is defined by (see Dattoli et al., 1999, Duran et al., 2019, Duran et al., 2020, Njionou, 2020)

* *Corresponding Author*

$$\Delta_\lambda f(x) = \frac{1}{\lambda}(f(x+\lambda) - f(x)), \quad \lambda \neq 0. \tag{1.3}$$

The following Lemma will be useful in the derivation of several results.

Lemma 1 *The following elementary series manipulation hold:*

$$\sum_{n=0}^{\infty}\sum_{k=0}^{\infty} A(k,n) = \sum_{n=0}^{\infty}\sum_{k=0}^{\lfloor n/j \rfloor} A(k, n-jk), \tag{1.4}$$

where $\lfloor \cdot \rfloor$ is the Gauss notation, and represents the maximum integer which does not exceed the number in the square brackets.

The degenerate exponential function $e_\lambda^x(t)$ for a real number λ is given by (Duran et al., 2021, Kim et al., 2017, Lim, 2016)

$$e_\lambda^x(t) = (1+\lambda t)^{\frac{x}{\lambda}} \text{ and} e_\lambda^1(t) = e_\lambda(t). \tag{1.5}$$

It is readily seen that $\lim_{\lambda\to 0} e_\lambda^x(t) = e^{xt}$. From (1.1) and (1.5), we obtain the following relation

$$e_\lambda^x(t) = \sum_{n=0}^{\infty} (x)_{n,\lambda} \frac{t^n}{n!}, \tag{1.6}$$

which satisfies the following difference rule

$$\Delta_\lambda e_\lambda^x(t) = t e_\lambda^x(t). \tag{1.7}$$

The usual Bell polynomials and Stirling numbers of the second kind satisfy the following relation (Bell, 1934, Bouroubi et al., 2006, Carlitz, 1980, Kim et al., 2017)

$$B_n(x) = \sum_{m=0}^{n} S_2(n,m) x^m. \tag{1.8}$$

The degenerate Bell polynomials are given by the following Taylor series expansion at $t = 0$ as follows:

$$\sum_{n=0}^{\infty} B_{n,\lambda}(x) \frac{t^n}{n!} = e^{x\left(e_\lambda(t)-1\right)}. \tag{1.9}$$

The usual Bernoulli $B_n(x)$, Euler $E_n(x)$ and Genocchi $G_n(x)$ polynomials and the degenerate Bernoulli $B_{n,\lambda}(x)$, Euler $E_{n,\lambda}(x)$ and Genocchi $G_{n,\lambda}(x)$ polynomials are given as follows (Araci et al., 2016, Duran et al., 2018, Howard, 1979, Khan, 2016, Kurt et al., 2013, Mihoubi, 2008, Pathan, 2012, Ozarslan, 2013):

$$\sum_{n=0}^{\infty} B_n(x)\frac{t^n}{n!} = \frac{te^{xt}}{e^t - 1} \text{ and } \sum_{n=0}^{\infty} B_{n,\lambda}(x)\frac{t^n}{n!} = \frac{te_\lambda^x(t)}{e_\lambda(t) - 1} \quad (|t| < 2\pi)$$

$$\sum_{n=0}^{\infty} E_n(x)\frac{t^n}{n!} = \frac{2e^{xt}}{e^t + 1} \text{ and } \sum_{n=0}^{\infty} E_{n,\lambda}(x)\frac{t^n}{n!} = \frac{2e_\lambda^x(t)}{e_\lambda(t) + 1} \quad (|t| < \pi)$$

$$\sum_{n=0}^{\infty} G_n(x)\frac{t^n}{n!} = \frac{2te^{xt}}{e^t + 1} \text{ and } \sum_{n=0}^{\infty} G_{n,\lambda}(x)\frac{t^n}{n!} = \frac{2te_\lambda^x(t)}{e_\lambda(t) + 1} \quad (|t| < \pi).$$

The Gould-Hopper polynomials are defined via the following Mac Laurin series expansion (Appell, 1926):

$$\sum_{n=0}^{\infty} H_n^{(j)}(x,y)\frac{t^n}{n!} = e^{xt+yt^j}, \tag{1.10}$$

where $j \in \mathbb{N}$ with $j \geq 2$. Upon setting $j = 1$, the Gould-Hopper polynomials reduce to the representation of the Newton binomial formula. Also, choosing $j = 2$ in (1.10), we obtain the familiar Hermite polynomials denoted by $H_n(x, y)$.

Let $n, j \in \mathbb{Z}$ with $n \geqq 0$ and $j > 0$, and let $\lambda_1, \lambda_2 \in \mathbb{R} \setminus \{0\}$. The generalized degenerate Gould-Hopper polynomials $H_{n,\lambda_1,\lambda_2}^{(j)}(x, y)$ are defined by means of the following generating function (Duran et al., 2020):

$$\sum_{n=0}^{\infty} H_{n,\lambda_1,\lambda_2}^{(j)}(x, y)\frac{t^n}{n!} = e_{\lambda_1}^{x}(t)e_{\lambda_2}^{y}(t^j). \tag{1.11}$$

Diverse formulas and properties of the generalized degenerate Gould-Hopper polynomials $H_{n,\lambda_2,\lambda_3}^{(j)}(x, y)$ are investigated by Duran and Acikgoz in (Duran et al., 2020).

2 The Generalized Degenerate Gould-Hopper Based Degenerate Poisson-Charlier Polynomials

The Poisson-Charlier polynomials $c_n(z;\alpha)$, which are members of the family of Sheffer-type sequences, are defined explicitly by (Bouroubi et al., 2006, Cheikh et al., 2003, Dattoli et al., 1999)

$$c_n(z;\alpha) = \sum_{k=0}^{n} (-1)^k \binom{n}{k}\binom{\alpha}{k} k! z^{-k}.$$

The generating function for the Poisson-Charlier polynomials are given as follows (*cf*. Kim et al., 2015):

$$\sum_{n=0}^{\infty} c_n(z;\alpha)\frac{t^n}{n!} = \left(1+\frac{t}{\alpha}\right)^z e^{-t}. \tag{2.1}$$

Several properties and relations for the Poisson-Charlier polynomials have been studied and investigated by many mathematicians (Bouroubi et al., 2006, Cheikh et al., 2003, Dattoli et al., 1999) and also cited references therein.

Definition 2 *Let* $\lambda_1, \lambda_2 \in \mathbb{R} \setminus \{0\}$. *The generalized degenerate Gould-Hopper based degenerate Poisson-Charlier polynomials* $c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)$ *are defined by the following exponential generating function*

$$\sum_{n=0}^{\infty} c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)\frac{t^n}{n!} = \left(1+\frac{t}{\alpha}\right)^z e_{\lambda_1}^{x}(t)e_{\lambda_2}^{y}(t^j). \tag{2.2}$$

Remark 1

1. When $x = 0$, the polynomials $c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)$ in (2.2) reduce to the generalized Gould-Hopper based degenerate Poisson-Charlier polynomials in (2.3), which are also new generalizations of the Poisson-Charlier polynomials $c_n(z;\alpha)$ in (2.1), given by

$$\sum_{n=0}^{\infty} c_{n,\lambda_2}(z;\alpha,x)\frac{t^n}{n!} = \left(1+\frac{t}{\alpha}\right)^z e_{\lambda_2}^{y}(t^j). \tag{2.3}$$

2. When $y = 0$, the polynomials $c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)$ in (2.2) reduce to the generalized Gould-Hopper based degenerate Poisson-Charlier polynomials in (2.4), which are also new generalizations of the Poisson-Charlier polynomials $c_n(z;\alpha)$ in (2.1), given by

$$\sum_{n=0}^{\infty} c_{n,\lambda_1}^{[j]}(z;\alpha,y)\frac{t^n}{n!}=\left(1+\frac{t}{\alpha}\right)^z e_{\lambda_1}^{x}(t). \tag{2.4}$$

3. Upon setting $\lambda_1,\lambda_2\to 0$, we get the Gould-Hopper based generalized Poisson-Charlier polynomials $c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)$ (2.5), which are extensions of the Poisson-Charlier polynomials (2.1), shown by

$$\sum_{n=0}^{\infty} c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)c_n^{[j]}(z;\alpha,x,y)\frac{t^n}{n!}=\left(1+\frac{t}{\alpha}\right)^z e^{-t}e^{xt+yt^j}. \tag{2.5}$$

4. Choosing $x=-1$ and $y=0,$ we get the usual degenerate Poisson-Charlier polynomials, which are also new generalizations of the Poisson-Charlier polynomials $c_n(z;\alpha)$ in (2.1) given below:

$$\sum_{n=0}^{\infty} c_{n,\lambda_1}(z;\alpha)\frac{t^n}{n!}=\left(1+\frac{t}{\alpha}\right)^z e_{\lambda_1}^{-1}(t). \tag{2.6}$$

5. When $\lambda_1\to 0,$ $x=-1$ and $y=0$, we arrive at the usual Poisson-Charlier polynomials in (2.1).

We now provide the following theorems without their proofs.

Theorem 1 *For* $\lambda_1,\lambda_2,\lambda_3\in\mathbb{R}\setminus\{0\},$ *we have*

$$c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)=\sum_{l=0}^{n}\binom{n}{l}c_{l,\lambda_1}(z;\alpha)H_{n-l,\lambda_1,\lambda_2}^{(j)}(x+1,y). \tag{2.7}$$

Theorem 2 *The following summation formula is valid for* $\lambda_1,\lambda_2\in\mathbb{R}\setminus\{0\}$*:*

$$c_{n,\lambda_1,\lambda_2}^{[j]}(z_1+z_2;\alpha,x,y)=\sum_{m=0}^{n}\binom{n}{m}c_{n-m,\lambda_1,\lambda_2}^{[j]}(z_1;\alpha,x+1,y)c_{m,\lambda_1}(z_2;\alpha), \tag{2.8}$$

Theorem 3 *The following formula holds for* $\lambda_1,\lambda_2\in\mathbb{R}\setminus\{0\}$*:*

$$c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x_1+x_2,y_1+y_2)=\sum_{m=0}^{n}\binom{n}{m}c_{n-m,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x_1,y_1)H_{m,\lambda_1,\lambda_2}^{(j)}(x_2,y_2) \tag{2.9}$$

Theorem 4 *The following explicit formula holds for* $\lambda_1,\lambda_2\in\mathbb{R}\setminus\{0\}$*:*

$$c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)=\sum_{u=0}^{n}\sum_{k=0}^{\lfloor u/j\rfloor}\binom{n}{u}(z)_{n-u}(x)_{u-jk,\lambda_1}(y)_{k,\lambda_2}\frac{u!\alpha^{u-n}}{(u-jk)!k!} \tag{2.10}$$

Theorem 5 *The following relation holds for* $\lambda_1,\lambda_2\in\mathbb{R}\setminus\{0\}$*:*

$$\frac{d}{dz}c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)=n\sum_{u=0}^{n-1}\binom{n-1}{u}c_{u,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)\frac{(n-1-u)!(-1)^{n-1-u}}{(n-u)} \tag{2.11}$$

Theorem 6 *The following relation holds for* $\lambda_1,\lambda_2\in\mathbb{R}\setminus\{0\}$*:*

$$\frac{d}{dx}c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)=n\sum_{u=0}^{n-1}\binom{n-1}{u}c_{u,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)\frac{(n-1-u)!(-\lambda_1)^{n-1-u}}{(n-u)} \tag{2.12}$$

3 Multifarious Connected Formulas

Here, we perform to get several diverse relations for $c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)$ with some other degenerate polynomials including the generalized degenerate Gould-Hopper based degenerate Bernoulli, Genocchi

and Euler polynomials.

Definition 3 *(Duran et al., 2020) The generalized degenerate Gould-Hopper based degenerate Bernoulli* $B_{n,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x,y)$*, Euler* $E_{n,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x,y)$ *and Genocchi* $G_{n,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x,y)$ *polynomials are defined by the following exponential generating functions:*

$$\sum_{n=0}^{\infty} B_{n,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x,y)\frac{t^n}{n!} = \frac{t}{e_{\lambda_3}(t)-1} e_{\lambda_1}^{x}(t) e_{\lambda_2}^{y}(t^j) \tag{3.1}$$

$$\sum_{n=0}^{\infty} E_{n,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x,y)\frac{t^n}{n!} = \frac{2}{e_{\lambda_3}(t)+1} e_{\lambda_1}^{x}(t) e_{\lambda_2}^{y}(t^j) \tag{3.2}$$

$$\sum_{n=0}^{\infty} G_{n,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x,y)\frac{t^n}{n!} = \frac{2t}{e_{\lambda_3}(t)+1} e_{\lambda_1}^{x}(t) e_{\lambda_2}^{y}(t^j) \tag{3.3}$$

for $\lambda_1,\lambda_2,\lambda_3 \in \mathbb{R}\setminus\{0\}$.

We here provide a relation involving the polynomials $c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)$, $B_{n,\lambda,\lambda_1}(x:\omega)$ and $B_{n,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x,y)$ as follows.

Theorem 7 *The following correlation holds true:*

$$B_{n,\lambda,\lambda_1,\lambda_2,\lambda_3}^{[j]}(z:\omega,x,y) = \sum_{u=0}^{n}\sum_{s=0}^{u}\binom{n}{u}\binom{u}{s} c_{u-s,\lambda_1}(z;\alpha) B_{s,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x,y)\frac{(1)_{n-u+1,\lambda_3}}{n-u+1} \tag{3.4}$$

Proof. By (2.2) and (3.1), we get

$$\sum_{n=0}^{\infty} c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y)\frac{t^n}{n!} = \left(1+\frac{t}{\alpha}\right)^{z} e_{\lambda_1}^{x}(t) e_{\lambda_2}^{y}(t^j)\frac{t}{e_{\lambda_3}(t)-1}\frac{e_{\lambda_3}(t)-1}{t}$$

$$= \sum_{n=0}^{\infty}\left(\sum_{u=0}^{n}\binom{n}{u}\sum_{s=0}^{u}\binom{u}{s} c_{u-s,\lambda_1}(z;\alpha) B_{s,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x+1,y)\frac{(1)_{n-u+1,\lambda_3}}{n-u+1}\right)\frac{t^n}{n!},$$

which implies the desired result (3.4).

Theorem 8 *The following summation formulais valid.*

$$c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y) = \sum_{k=0}^{n}\sum_{m=0}^{k}\binom{n}{k}\binom{k}{m}\frac{(1)_{n-k,\lambda_3}}{2} c_{k-m,\lambda_1}(z;\alpha) E_{m,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x+1,y)$$

$$+\frac{1}{2}\sum_{k=0}^{n}\binom{n}{k} c_{n-k,\lambda_1}(z;\alpha) E_{k,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x+1,y). \tag{3.5}$$

Proof. From (2.2) and (3.2), the aimed result (3.5) can be directly obtained by utilizing similar method used in the proof of Theorem 7. Thus, we omit the proof.

Theorem 9 *The following relation holds true:*

$$c_{n,\lambda_1,\lambda_2}^{[j]}(z;\alpha,x,y) = \frac{1}{n+1}\sum_{k=0}^{n+1}\sum_{m=0}^{k}\binom{n+1}{k}\binom{k}{m}\frac{(1)_{n+1-k,\lambda_3}}{2}$$

$$\times c_{k-m,\lambda_1}(z;\alpha) G_{m,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x+1,y) + \sum_{k=0}^{n+1}\binom{n+1}{k}\frac{c_{n+1-k,\lambda_1}(z;\alpha) G_{k,\lambda_1;\lambda_2,\lambda_3}^{[j]}(x+1,y)}{2(n+1)}.$$

Proof. In view of (2.2) and (3.3), can be directly attained by utilizing similar method used in the proof

of Theorem 7. Thus, we omit the proof.

References

Appell, P. (1926). Kampé de Fériet, J., Fonctions hypergéométriques. Polynômes d'Hermite, Gauthier-Villars, Paris.

Araci, S., Khan, W. A., Acikgoz, M., Özel, C. (2016). Kumam, P. A new generalization of Apostol type Hermite-Genocchi polynomials and its applications, SpringerPlus. 2016, doi. 10.1186/s40064-016-2357-4.

Bell, E. T. (1934). Exponential polynomials. Ann. Math. 35, 258-277.

Bouroubi, S., Abbas, M. (2006). New identities for Bell's polynomials: New approaches. Rostock. Math. Kolloq. 61, 49-55.

Carlitz, L. (1980). Some remarks on the Bell numbers. Fibonacci Quart. 1980, 18, 66-73.

Carlitz, L. (1975). Degenerate Stirling, Bernoulli and Eulerian numbers. Utilitas Math. 1975, 15, 51-88.

Carlitz, L. (1956). A degenerate Staudt-Clausen theorem. Arch Math (Basel). 1956, 7, 28-33.

Cheikh, Y. B., Zaghouani A. (2003). Some discrete d -orthogonal polynomials sets. J. Comput. Appl. Math. 156, 253-263.

Dattoli, G.; Lorenzutta S.; Cesarano, C. (1999). Finite sums and generalized forms of Bernoulli polynomials, Rend. Math. Appl. 19, 385-391.

Duran, U.; Acikgoz, M. (2021). On some degenerate probability distributions related to the degenerate Stirling numbers of the second kind, in review.

Duran, U.; Acikgoz, M.; Araci, S. (2018). Hermite based poly-Bernoulli polynomials with a q -parameter. Adv. Stud. Contemp. Math. 28, 285-296.

Duran, U.; Sadjang, P.N. (2019). On Gould-Hopper-based fully degenerate poly-Bernoulli polynomials with a q -parameter. Mathematics, 7, 121.

Duran, U.; Acikgoz, M. (2020). Generalized Gould-Hopper based fully degenerate central Bell polynomials, Turkish Journal of Analysis and Number Theory, 7(5), 124-134.

Howard, F.T. (1996). Explicit formulas for degenerate Bernoulli numbers. Discrete Math., 162, 175-185.

Howard, F.T. (1979). Bell polynomials and degenerate Stirling numbers. Rend. Sem. Mat. Univ. Padova., 61, 203–219.

Khan, W. A. (2016). A note on degenerate Hermite poly-Bernoulli numbers and polynomials. J. Class. Anal., 8, 65-76.

Kim, T.; Kim, D.S.; Jang, L.-C.; Kwon, H.I. (2018). Extended degenerate Stirling numbers of the second kind and extended degenerate Bell polynomials. Utilitas Math., 106, 11-21.

Kim, D.S., Kim, T. (2015). Some identities of Bell polynomials. Sci. China Math., 58, 2095-2104.

Kim, T.; Kim, D.S. (2018). Degenerate Bernstein polynomials. RASCAM., doi: 10.1007/s13398-018-0594-9.

Kim, T., Kim, D.S.; Kim, H.Y.; Kwon, J. (2019). Degenerate Stirling polynomials of the second kind and some applications. Symmetry, 11, 1046.

Kim, T., Kim, D.S. (2017). On degenerate Bell numbers and polynomials. RASCAM., 111, 435-446.

Kim, T., Yao, Y.; Kim, D.S.; Jang, G.-W. (2018). Degenerate x -Stirling numbers and x -Bell polynomials. Russ. J. Math. Phys., 25, 44-58.

Kim, T., Kim, D.S.; Jang, G.-W. (2017). Extended Stirling polynomials of the second kind and extended Bell polynomials. Proc. Jangjeon Math. Soc., 20, 365-376.

Kim, T., Kim, D.S.; Dolgy, D.V. (2017). On partially degenerate Bell numbers and polynomials, Proc. Jangjeon Math. Soc., 20, 337-345.

Kim, T. (2017). A note on degenerate Stirling polynomials of the second kind. Proc. Jangjeon Math. Soc., 20, 319-331.

Kurt, B., Simsek, Y. (2013). On the Hermite based Genocchi polynomials. Adv. Stud. Contemp. Math., 23, 13-17.

Lim, D. (2016). Some identities of degenerate Genocchi polynomials. Bull. Korean Math. Soc. 53, 569-579.

Mihoubi, M. (2008). Bell polynomials and binomial type sequences. Discret. Math, 308, 2450-2459.

Njionou, S.P. (2020). Addition theorems for the Δ_{ω}-Appell polynomials and the associated classes of polynomial expansions. in review.

Ozarslan, M.A. (2013). Hermite-based unified Apostol-Bernoulli, Euler and Genocchi polynomials, Adv. Difference Equ, doi: 10.1186/1687-1847-2013-116.

Pathan, M.A. (2012). A new class of generalized Hermite-Bernoulli polynomials. Georgian Math. J. 19, 559-573.

Rainville, E.D. (1960). Special Functions, The Macmillan Company, New York.

Srivastava, H.M.; Choi, J. (2012). Zeta and q-Zeta functions and associated series and integrals; Elsevier Science Publishers: Amsterdam, The Netherlands, 674 p.

Khan, W.A., Nisar K.S., Araci, S., Acikgoz, M. (2018). Fully degenerate Hermite poly Bernoulli numbers and polynomials, Advances and Application in Mathematical Sciences,17(6), 461-478.

Vibration Control of a Pier Pile Using Deep Learning LSTM Network

Barış Namlı[1*], Cihan Bayındır[1,2]

[1] *İstanbul Technical University, Civil Engineering Department, İstanbul, Turkey*
[2]*Boğaziçi University, Civil Engineering Department, İstanbul, Turkey.*
barisnamli17@gmail.com, cbayindir@itu.edu.tr

Abstract

Balancing the offshore structures in the ocean or sea against the forces created by seismic movements and waves is extremely important for the serviceability and the safety of the structure. In order to achieve this, various approaches are currently being considered (Kandasamy et al., 2016; Madkour et. al, 2007). Region-specific parameters should be recorded and used when needed to assist these approaches. In addition, the use of artificial intelligence methods, which have been in demand recently, to reduce the oscillation in buildings, allows more effective results in the application that can be done in this field (Kim, 2009). Using the long-short-time memory (LSTM) algorithm (Hochreiter et. al, 1997), one of the deep learning methods, time series prediction can be performed. As a result of the prediction, better approaches can be developed for the future. This study shows that the vibration control of offshore platforms can be achieved against various types of loadings by the deep learning techniques, which is a branch of artificial intelligence. For this purpose, a long pile is analyzed by solving the equation of motion under forcing described by the Morison equation (Morison et al., 1950; Dean et al., 2000). Thus, a realistic wave load is applied to analyze system behavior in a more realistic setting. The applied wave loads are predicted using the LSTM deep learning network and applied to the system as negative feedback. It is shown that a significant reduction in the vibration amplitudes can be achieved by this approach. Our findings and their possible applications are also discussed.

Keywords: Vibration control, offshore structures, deep learning

1. Introduction

Controlling the vibrations that may occur in structures is extremely important both for the health of the structures and for the safety of the people. Reducing the vibration parameters such as displacements that may occur as a result of both the earthquake or sea waves, especially in offshore structures, is a situation that needs extra attention. Due to this fact, many studies have been carried out to reduce the vibration that may occur in offshore structures (Kandasamy et al., 2016; Madkour et al., 2007). However, approaches developed for vibration control seem to have some advantages and disadvantages (Kandasamy et al., 2016; Madkour et al., 2007). Therefore, it is crucial to choose the optimal approach that will suit the system, considering the characteristics and environmental conditions. Another point to be considered is that existing approaches can be improved due to the development of today's technology over time. Significantly, studies in artificial intelligence affect many points in human life as well as helping vibration control approaches (Hochreiter et al., 1997; Olah, 2015). Moreover, using artificial intelligence methods to reduce vibration has recently been one of the considered approaches (Kim, 2009; Madkour et al., 2007). In our study, we consider the vibration control of a pier pile using the feedback force forecasted by the LSTM deep learning network. The wave force on a pier pile can be calculated by the Morison equation (Morison et al., 1950; Dean et al., 2000). Using the 90% of the Morison force

[*] *Corresponding Author*

time series that will be learned by the LSTM algorithm, the remaining 10% is forecasted. Then this forecasted force is feedback into the system whose dynamics are modeled using the equation of motion solved by an ode45 solver. It is shown that LSTM based deep learning approach can be used to control and reduce vibration parameters of a pier pile or in general any structure.

2. Methodology

2.1. Vibration Model of a Pier Pile with a Feedback Force Control

In order to simulate the vibration dynamics of a pier pile, a simple vibration model is considered. The equation of motion can be given as

$$m\ddot{x} + c\dot{x} + kx = F(t) \tag{1}$$

where *x(t)* denotes the displacement from the equilibrium position, *m* is the mass, *c* is the damping coefficient, *k* is the stiffness, and *F(t)* is the time-dependent force. Throughout this study, the computational parameters are selected as *m=150kg, c=50Ns/m, k=2500 N/m.* Eq. (1) is solved using ode45 solver of the MATLAB software starting from the initial conditions of $x(0) = 0.5\ m$ and $\dot{x}(0) = -0.2\ m/s$. The forcing term on the right hand side of Eq. (1) is calculated using the Morison formula. For this purpose, a circular cylinder which can be a model for a pile foundation is considered. For a pile of length *dz,* the Morison formula can be summarized as

$$dF = dF_D + dF_I = \frac{1}{2} C_D \rho A |u| u + C_M \rho V \frac{Du}{Dt}. \tag{2}$$

Here the subscripts *D* and *I* denote the drag and inertia forces, respectively. C_D denotes the drag coefficient and C_M shows the inertia coefficient. The parameter ρ is the fluid density and *A* shows the area perpendicular to the flow. The parameter *V* represents the volume of the element of height *dz* and *u* shows the horizontal flow excursion velocity of the wavefield (Dean et al., 2000). Total equivalent force can be computed after integration as

$$F = \int_{-h}^{\eta} dF. \tag{3}$$

Here *h* is the depth and *η* is the water surface fluctuations. After the linearization of Eq. (3) for the case of a pile with diameter *D*, the force exerted by a monochromatic wave can be found as

$$F = C_D DnE \cos(kx_1 - \omega t)|\cos(kx_1 - \omega t)| + C_M \pi DE \frac{D}{H} \tanh(kh) \sin(kx_1 - \omega t) \tag{4}$$

where *n* is the ratio of group velocity to wave celerity. This parameter is taken as 1 for simulation of the shallow water conditions. In the same formula, the wave energy is $E=1/8\rho g H^2$ where *H* denotes the wave height. Other parameters are; *k* denotes the wavenumber, x_1 is the coordinate of pile location and the angular frequency is $\omega=2\pi/T$ where *T* denotes the period of the wave (Dean et al., 2000). Throughout this study, the computational parameters are selected to be $C_D=1$, $C_M=2$, *D=1m, H=1m,* $\rho=1025\ kg/m^3$, $g=9.81 m/s^2$, *h=10m, T=15s,* $x_1=0$ for representation of a realistic monochromatic wave loading. The wavenumber is calculated using $k = \omega / \sqrt{gh}$ for simulation of the shallow water conditions.

2.2. Long Short Term Memory (LSTM) Networks

Recurrent neural networks (RNN) deep learning methods were developed in the 1980s (Hochreiter et al.,1997; Olah, 2015). Thanks to this neural network architecture, predictions can be made using serial data sets. Especially in time-series analysis, the RNN structure is widely used (Hochreiter et. al, 1997; Olah, 2015). However, it was seen that gradient disappearance or gradient burst events frequently occur during the use of long series (Hochreiter et. al, 1997). The Long Short Term Memory (LSTM) architecture was developed in 1997 to fix these problems (Hochreiter et al., 1997). Thanks to the gates

of this neural network architecture, the information flow can be done smoothly, even between long series. Using the gate, it is decided which information will be forgotten and which data will be retained and transferred. These gates are a forgotten gate, input gate, and output gate, respectively. It also has its memory cell outside of the passages. It is seen that it is different from the classical RNN structure due to its features. The mathematical expressions of forget gate, update gate, output gate, and memory cell are below:

Forget Gate:

$$f_i^{(t)} = \sigma(b_i^f + \sum_y U_{i,j}^f x_j^{(t)} + \sum_y W_{i,j}^f h_j^{(t-1)}) \tag{5}$$

State Unit:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^t \sigma(b_i + \sum_y U_i x_j^{(t)} + \sum_y W_i h_j^{(t-1)}) \tag{6}$$

Input Gate:

$$g_i^{(t)} = \sigma(b_i^g + \sum_y U_i^g x_j^{(t)} + \sum_y W_i^g h_j^{(t-1)}) \tag{7}$$

Output Gate:

$$q_i^{(t)} = \sigma(b_i^o + \sum_y U_{i,j}^o x_j^{(t)} + \sum_y W_{i,j}^o h_j^{(t-1)}) \tag{8}$$

The Output:

$$h_i^{(t)} = q_i^{(t)} \tanh(s_i^{(t)}) \tag{9}$$

Here the b_i^f, b_i^g, b_i^o paramaters show the bias values in the input, output, and forget gates. The $U_{i,j}^f, U_{i,j}^g, U_{i,j}^o, U_{i,j}$ parameters denote the input weights belonging to the gates. The parameters $W_{i,j}^f, W_{i,j}^g, W_{i,j}^o, W_{i,j}$ indicate the recurrent weight of the gate to which they belong. The parameter *X* is the input time series (Hochreiter et al., 1997; Olah, 2015). The reader is referred to Hochreiter et al. (1997) and Olah (2015) for a more comprehensive discussion of the LSTM network and its applications.

We use the LSTM network for the prediction of the wave force induced by a monochromatic wave which is modeled using the Morison equation. Then, the predicted force time series, $F_D(t)$, is applied to the right-hand side of Eq. (1) by simple addition in the reverse direction of $F(t)$. In a real scenario, such a feedback force can be realized by energy dissipation devices, tuned-mass-dampers, or similar structures.

3. Results and Discussion

In Fig. 1 we depict the forecasted time-series of the Morison force obtained using the LSTM network without updates. For this forecast initial 90% of the time-series data, which corresponds to the first *45s* of *50s* data, is used. The forecast only relies on this first part of the data, no updates from the observations during the forecast phase are used. As indicated in Fig. 1, the LSTM network with no updates is successful in predicting the Morison force time series. The comparison between the forecast time-series and the actual time series and the RMS error are depicted in Fig. 2. As this figure confirms, the order of RMS error is much smaller compared to the order of the magnitude of the time series, thus a reliable prediction is performed using the LSTM network.

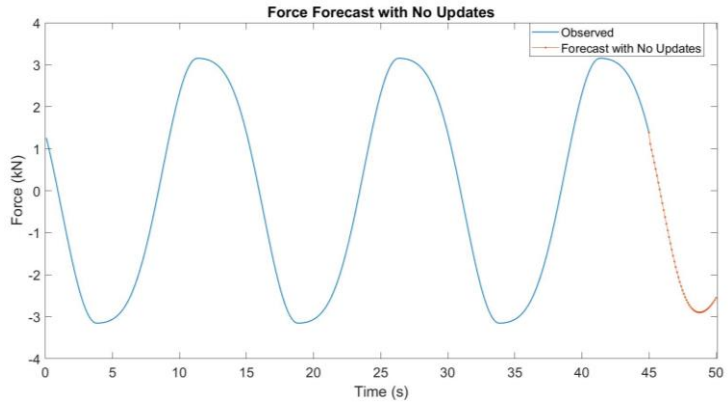


Figure 1. Forecast of the Morison force time-series by LSTM network without updates.

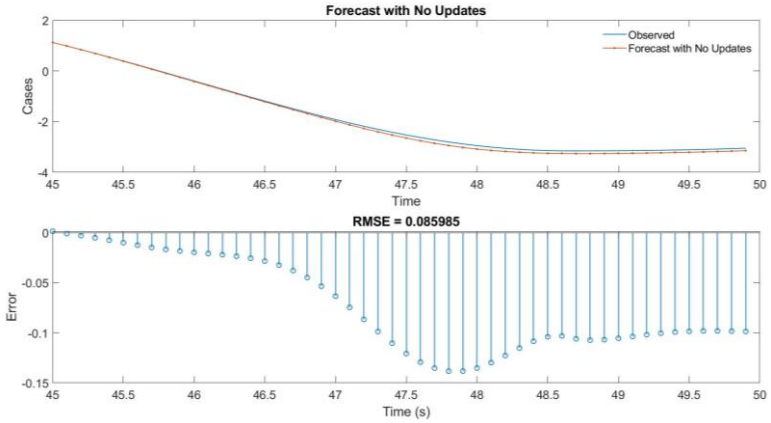


Figure 2. Forecast of the Morison force time-series by LSTM network without updates and RMS error.

Although the predictions depicted in Fig. 1 and Fig. 2 show reliable results for the prediction of the Morison force, it is known that LSTM based deep-learning predictions can give unreliable results in the case of irregular and spiky data. Thus, as a remedy, the LSTM network is used with updates, that is the forecasted part of the time series are updated by the new data observed during the forecast phase. With this motivation, we depict the Morison force time-series predictions obtained by LSTM network with updates in Fig. 3. As this figure confirm, the forecast results are similar to the forecasts depicted in Figs. 1 and 2 obtained using LSTM with no updates, however an improvement in the RMS error is still observable.

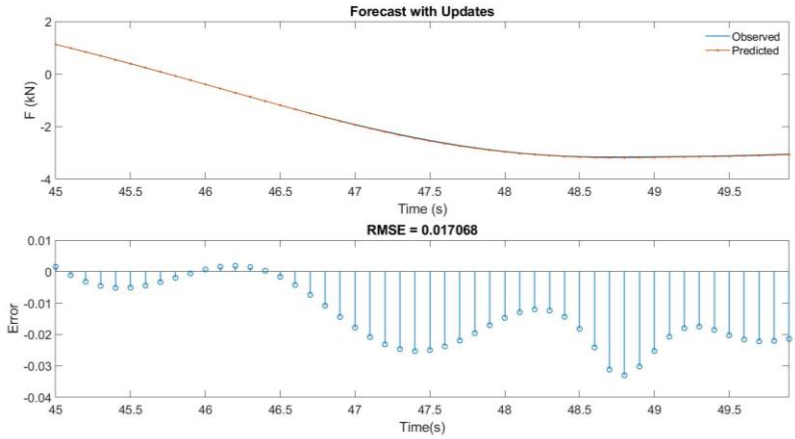


Figure 3. Forecast of the Morison force time-series by LSTM network with updates.

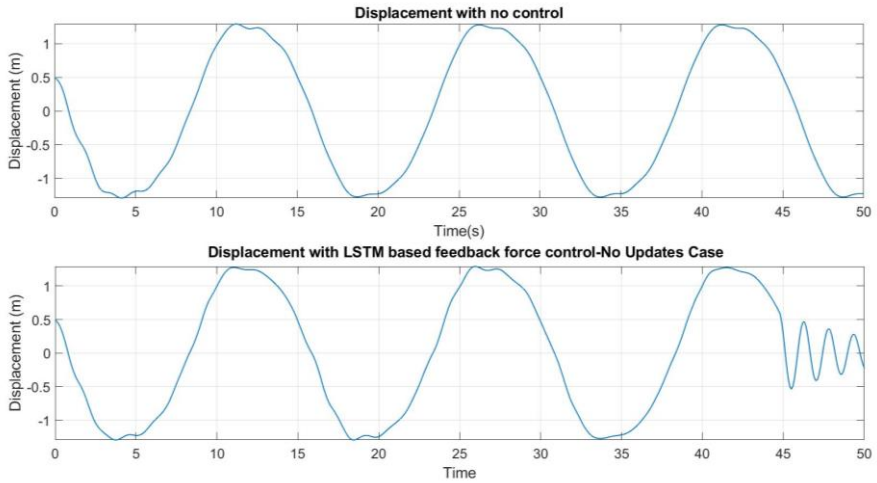


Figure 4. Displacement of a pier a) uncontrolled case b) controlled by feedback force predicted using LSTM without updates.

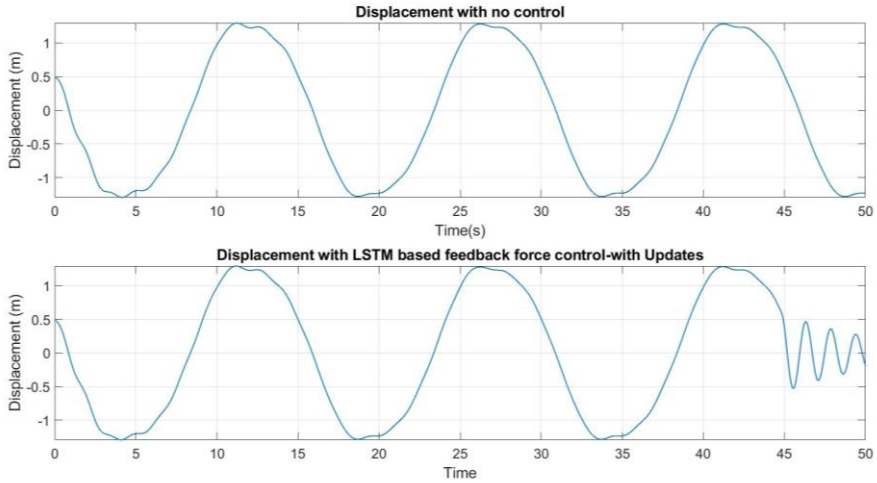


Figure 5. Displacement of a pier a) uncontrolled case b) controlled by feedback force predicted using LSTM with updates.

The importance of the LSTM network with updates would be very crucial for the case of irregular data with spikes. After successful prediction of the Morison force time series, we solve the equation of the motion of a pier pile given in Eq. (1) using the ode45 solver starting from the aforementioned initial conditions. We depict the displacements of the modelled pier pile obtained by using the ode solver with no control and with feedback force predicted using the LSTM with no updates and updates in Figs. 4 and 5, respectively. As Figs. 4 and 5 confirm, the feedback force forecasted by LSTM can be used to control the vibration dynamics and reduce the vibration amplitudes effectively. The feedback force applied for the vibration control forces is applied between the times of *t=45s-50s.* Due to the cancellation effect of the feedback force, the system behaves like an unforced damped harmonic oscillator after the time of *t=45s.* Due to the success of LSTM with no updates in predicting the time series of the Morison force, the results depicted in Figs. 4 and 5 are very similar. However, as mentioned before, for an irregular data set with many peaks, the LSTM with updates is expected to be more successful in vibration control applications. Although the results depicted here are for displacements, one can extend them to velocities, accelerations, and other displacement parameters. One research direction to follow is the use of different deep learning networks such as GRU networks with similar

purposes. Also, different vibration control strategies such as state variable feedback and output feedback approach can also be proposed together with the deep learning algorithms. Our findings can be successfully used to control the vibrations of coastal and ocean structures.

4. Conclusion

In this study, we have considered the vibration control problem of coastal and ocean structures by using the deep learning approach. More specifically, we considered the vibration dynamics of a pier under the excitation of a monochromatic wave loading. It is shown that using the LSTM network, the prediction of wave force can be performed and this prediction can be used as a feedback force to reduce/control the vibration amplitudes and frequencies. We showed that LSTM with no updates and with updates can be used for this purpose, however, it is expected that the LSTM with updates taken from measurements in the prediction phase will be more successful in the case of irregular data. Although such an LSTM with updates approach limits the successful application with the one-time step of the input data, it would be beneficial for many engineering purposes such as avoiding structural damage and saving lives. In near future, we aim to extend our findings to irregular waves, as well as other vibration control strategies such as state variable feedback or output feedback.

Acknowledgements

The author gratefully acknowledges the support of the İstanbul Technical University. This work was supported by the Research Fund of the İstanbul Technical University. Project Code: MGA-2020-42544. Project Number: 42544.

References

Dean, R. G., & Dalrymple, R. A. (2000). Water wave mechanics for engineers and scientists (Vol. 2). World Scientific Publishing Company. https://doi.org/10.1142/1232

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. https://doi.org/10.1007/s10710-017-9314-z

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9 (8), 1735-1780. DOI: 10.1162/neco.1997.9.8.1735 https://doi.org/10.1162/neco.1997.9.8.1735

Kandasamy, R., Cui, F., Townsend, N., Foo, C. C., Guo, J., Shenoi, A., & Xiong, Y. (2016). A review of vibration control methods for marine offshore structures. Ocean Engineering, 127, 279–297. https://doi.org/10.1016/j.oceaneng.2016.10.001

Kim, D. H. (2009). Neuro-control of fixed offshore structures under earthquake. Engineering Structures, 31 (2), 517-522. https://doi.org/10.1016/j.engstruct.2008.10.002

Madkour, A., Hossain, M. A., Dahal, K. P., & Yu, H. (2007). Intelligent learning algorithms for active vibration control. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37 (5), 1022-1033. https://doi.org/10.1109/TSMCC.2007.900640

Morison, J. R., Johnson, J. W., & Schaaf, S. A. (1950). The Force Exerted by Surface Waves on Piles. Journal of Petroleum Technology, 2 (5), 149–154. https://doi.org/10.2118/950149-G

Olah, C. (2015). Understanding LSTM networks. Accession Date: 22 August 2021. Available from: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Zhang, B. L., Han, Q. L., & Zhang, X. M. (2017). Recent advances in vibration control of offshore platforms. Nonlinear Dynamics, 89 (2), 755-771. https://doi.org/10.1007/s11071-017-3503-4

Analysis Methods and FPAA Implementation of Hyperchaotic Systems

Gülnur Yılmaz*, Enis Günay

Erciyes University, Electrical & Electronics Engineering Department, Kayseri, Turkey
gulnur.yilmaz@erciyes.edu.tr, egunay@erciyes.edu.tr

Abstract

In this study, hyperchaotic systems in the literature are investigated and analyzed. At first, systems with different structures are specified. Then, the methods used in the analysis of hyperchaotic systems are studied. Phase portraits, time series, Lyapunov exponents, bifurcation diagrams, and Poincare maps are analyzed. Furthermore, chaotic structure of the systems is demonstrated by power spectrum and 0-1 tests. After all analysis methods are completed, FPAA implementation is performed, and the implementation results are given at the end.

Keywords: Hyperchaos, Lyapunov exponents, bifurcation diagrams, Poincare maps, power spectrum, 0-1 tests, FPAA

1. Introduction

Chaotic systems have been in our lives since the last century and have a great contribution to the literature as the source of many studies in different fields. Sensitivity to initial conditions, deterministic but unpredictable structure are the most distinctive characteristics of these types of nonlinear systems.

Systems that have the same characteristics as chaotic systems but exhibit more complex dynamic behaviors and were first introduced to the literature at the end of the 1970s are called hyperchaotic systems. They differ from chaotic systems in terms of having at least 2 positive Lyapunov exponents and at least 4 system dimensions. Demonstrating more complex dynamic behaviors makes the hyperchaotic systems valuable for some implementation areas, especially in communication and cryptology, where security is of great importance.

When the historical development of hyperchaotic systems is examined, it is possible to get the same type of systems used in different research. It is known that hyperchaos was first introduced theoretically by Rössler (Rössler, 1979) in 1979. Experimentally, for the first time, Chua et al. (Chua & Kobayashi, 1986) observed hyperchaotic behavior on the electronic circuit designed in 1986. In this study, they obtained a 4th order hyperchaotic system from a simple circuit design by making modifications on the Chua circuit, which shows chaotic behavior. Furthermore, the Chua circuit has been the subject of different studies for generating hyperchaotic systems (Cannas & Cincotti, 2002; Fitch et al., 2012; Kapitaniak et al., 1994). Besides the Rössler and Chua systems, another system used for hyperchaos generation is the Lorenz system. By adding different control parameters and variables to the 3-dimensional Lorenz system, 4, 5, 6, and 7 dimensional systems were obtained and hyperchaotic behavior was observed (Li et al., 2005; Yang et al., 2015, 2018; Yang & Chen, 2013). On the other hand, 3-dimensional Lü chaotic system was transformed into a 4-dimensional system by adding fourth state variable in and a new hyperchaotic model was obtained (Chen et al., 2006). This system was modified by Bao et al. (Bao & Liu, 2008) by using different parameters and the hyperchaotic structure of the system was observed.

* *Corresponding Author*

Another hyperchaotic model, which has a similar structure to the Lü system was proposed by Gao et al. (Gao et al., 2006). In this study, 3-dimensional chaotic structure proposed by Chen is transformed into a 4-dimensional hyperchaotic system. Chen system has also been modified and different hyperchaotic models were created (Jia et al., 2010; H. Wang & Cai, 2009). In addition to these studies, there are other types of hyperchaotic models such as Qi (Yujun et al., 2010), Liu (F. Q. Wang & Liu, 2006), Wang (Z. Wang et al., 2012), Rabinovich (Liu et al., 2010), and Sprott (Ojoniyi & Njah, 2016) system.
In this study, dynamic structure of hyperchaotic systems is analyzed by phase portraits, time series, Lyapunov exponents, bifurcation diagrams, Poincare maps, power spectrum, and 0−1 test. Also, FPAA implementation, illustrating on Lorenz hyperchaotic system, is given and results are considered.

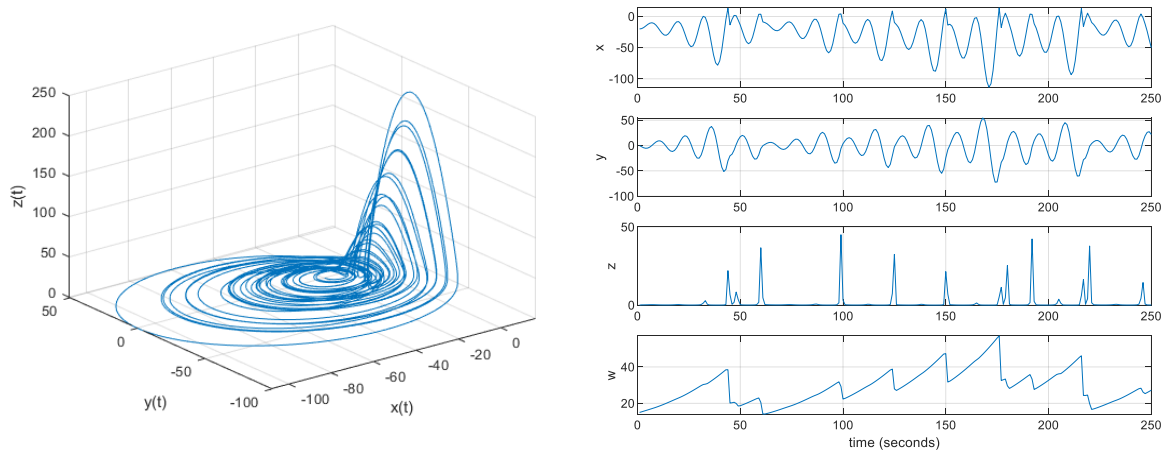
2. Methods Used in the Analysis of Hyperchaotic Systems

2.1. Phase-space Diagrams and Time Series

Phase-space diagrams are obtained by analyzing the behavior of state variables of dynamic systems relative to each other. These phase diagrams have a closed-loop shape in a periodic system while they have more complex structured shape in a chaotic system. As an example, Rössler (Rössler, 1979) hyperchaotic system is considered and 4-dimensional system model is given in Equation (1).

$$\begin{aligned} \dot{x} &= -(y+z) \\ \dot{y} &= x + ay + w \\ \dot{z} &= xz + b \\ \dot{w} &= -cz + dw \end{aligned} \tag{1}$$

where the fixed parameters are $(a, b, c, d) = (0.25, 3, 0.5, 0.05)$. Also, initial conditions of the system (1) are defined as $(x_0, y_0, z_0, w_0) = (-20, 0, 0, 15)$. According to these parameter values, 3-D phase-space graph and time series obtained for the *x-y-z-w* state variables are shown in Figure 1.a and b, respectively.



a) b)

Figure 1. a) 3-D phase-space representation for Rössler system b) Time series of Rössler system for *x*, *y*, *z*, *w* space variables

2.2. Lyapunov Exponents

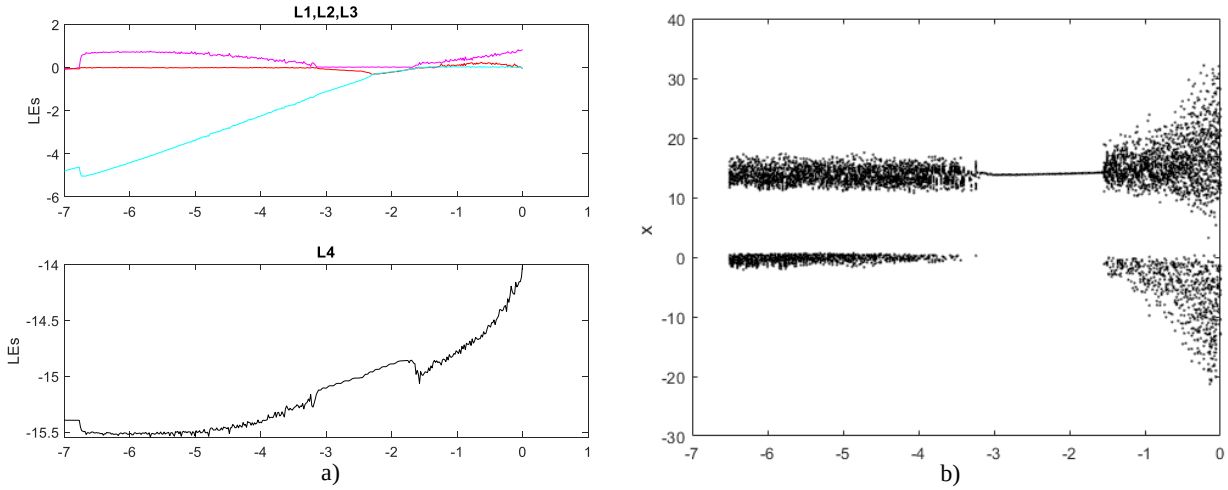
Lyapunov exponents are one of the analysis methods in which the dynamic structure of a system is measured numerically. The exponential divergence of the orbits for different initial conditions of a dynamic system is an indication that the system has a positive Lyapunov exponent, and these systems are in a chaotic structure (Wolf et al., 1985). A hyperchaotic system must have at least 2 positive Lyapunov exponents. As an example of this analysis method, Lorenz hyperchaotic system (X. Wang & Wang, 2008) given in Equation (2) has been considered.

$$\begin{aligned}\dot{x} &= a(y-x)+w\\ \dot{y} &= cx-y-xz\\ \dot{z} &= xy-bz\\ \dot{w} &= -yz+rw\end{aligned} \tag{2}$$

where *(a, b, c)* parameters are specified as *(10, 8/3, 28)*, respectively. Lyapunov exponents are calculated for the *r* parameter, which varies in the range of *[−7 0]*, and the resulting graph is shown in Figure 2.a. The system exhibits hyperchaotic behavior for the values of *r* greater than approximately *-1.5*.

2.3 Bifurcation Diagrams

Variation of the system parameters can cause changes in the dynamic behavior of the system. Bifurcation diagrams can be obtained by plotting the state variables of the system relative to the control parameter changes in a certain range. These diagrams give information about the dynamic behaviors of a system at what values the system enters chaos or has a periodic structure. For the hyperchaotic Lorenz system given in Equation (2), the bifurcation diagram is studied and given in Figure 2.b. Vertical and horizontal axes represent the *x* state variable and the *r* parameter, respectively. As seen from the graph, the system is periodic when *r* is approximately between *[-3.1, -1.5]* and it demonstrates hyperchaotic behavior for greater values.



a) b)

Figure 2. a) Lyapunov Exponents and b) Bifurcation diagrams of Lorenz hyperchaotic system

2.4 Poincare Maps

Poincare maps are obtained by transforming an *n*-dimensional continuous-time system into an *(n-1)*-dimensional discrete-time system and graphing in cross-sections. With this method, the dynamics of a system can be decided by determining the points where the orbit passes. If there is a single fixed point on the map, this indicates that the system follows the same trajectory continuously and it is periodic. However, if the system is chaotic, a shape consisting of random points is obtained. Poincare transform is applied to the Wang-Cang (Z. Wang et al., 2012) system given in Equation (3).

$$\begin{aligned}\dot{x} &= y\\ \dot{y} &= -x+yz+axzw\\ \dot{z} &= 1-y^2\\ \dot{w} &= z+bxz+cxyz\end{aligned} \tag{3}$$

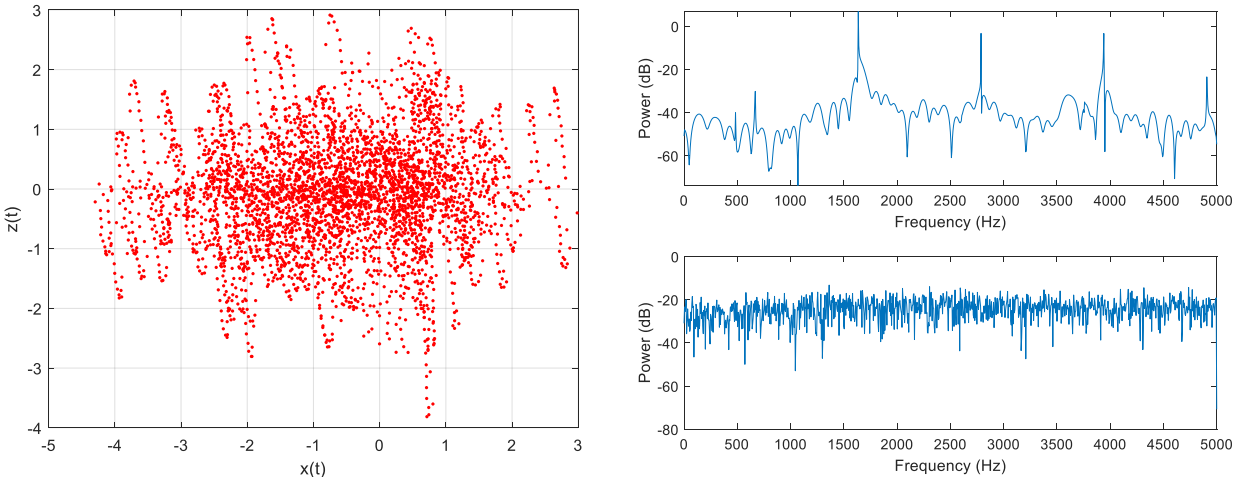
When the system parameters *(a, b, c)* are considered as respectively *(8, -2.5, -30)* and the initial values are *0.1* for all state variables, the system exhibits hyperchaotic behavior. Poincare Map where the chaotic structure can be observed is given in Figure 3.a.

2.5 Power Spectrum

Another method used to analyze the chaotic structure of a system is power spectrum analysis. When a power spectrum graph is examined in a periodic system, it is seen that it increases only in some frequency values, while in a chaotic system, a noise-like graph spreading over a wider frequency band is obtained. Power spectrum analysis has been realized for the Sprott B (Ojoniyi & Njah, 2016) system given in Equation (4). Results for periodic and hyperchaotic cases are depicted in Figure 3.b.

$$\begin{aligned} \dot{x} &= yz - v \\ \dot{y} &= x - y - w \\ \dot{z} &= 1 - xy \\ \dot{w} &= ax + y \\ \dot{v} &= x \end{aligned} \tag{4}$$

In the Sprott B system, when the control parameter a equals *-0.81*, the system is periodic. However, if a is considered as *0.16*, system dynamics change to hyperchaotic. Accordingly, the power spectrum graph obtained and shown in Figure 3.b. Top figure demonstrates the periodic case while the hyperchaotic case is given bottom.

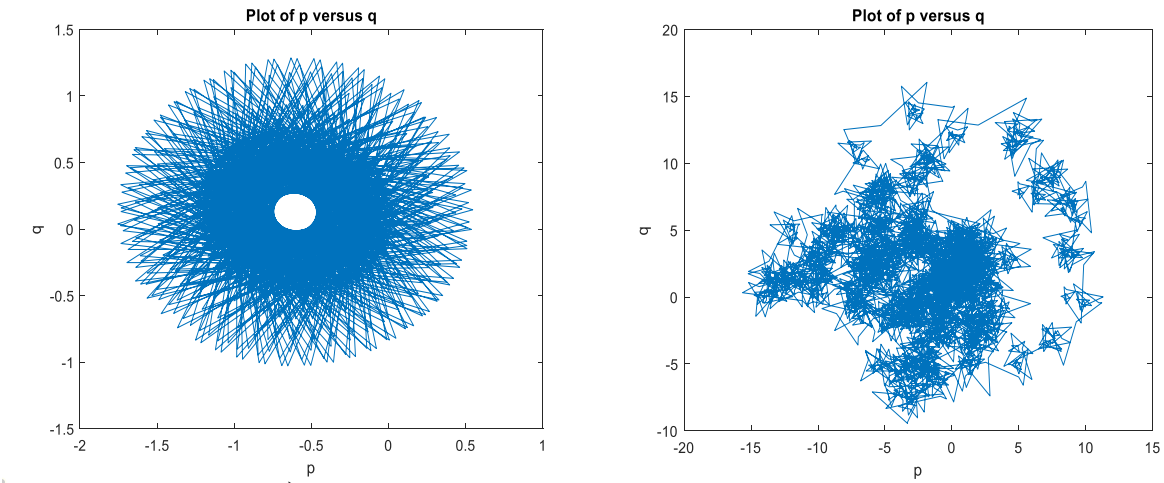


a) b)

Figure 3. a) Poincare maps of Wang-Cang System b) Power spectrum of Sprott B system

2.6. 0-1 Test

The 0−1 test is a method used for the analysis of chaotic systems applied on binary basis. Complex dynamics of a system can be specified by using only the time series without the need for analysis phase space representation or calculation of the Lyapunov exponent (Gottwald & Melbourne, 2004). p and q variables using time series data are calculated first and then the graph of p *vs* q is obtained. If the graph has a regular shape, the system is periodic. On the other hand, system behavior is hyperchaotic for irregular shapes. Sprott B hyperchaotic system, given in equation (4), is used for the 0−1 test. Parameter a is considered as *-0.81* and *0.16* for periodic and hyperchaotic structures, respectively. 0-1 test results are given in Figure 4.a and 4.b demonstrating regular and irregular shapes of Sprott B.



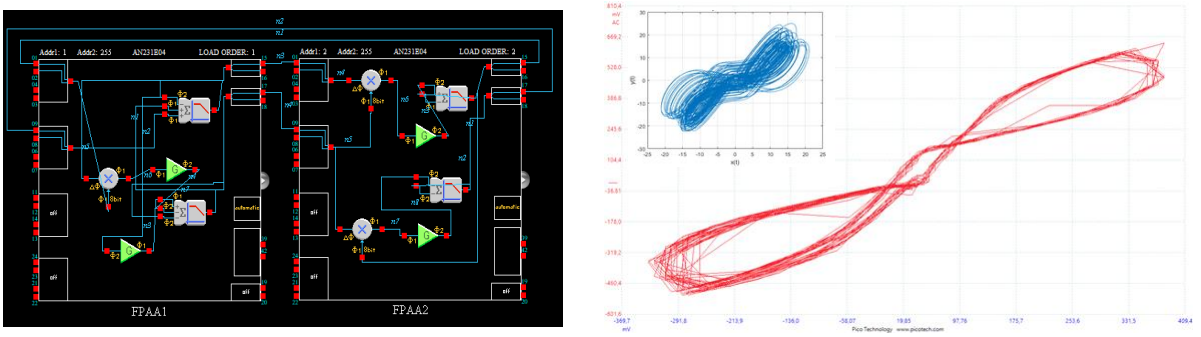
a) b)

Figure 4. 0-1 test for a) periodic-case b) hyperchaotic case of Sprott B system

3. FPAA Implementation

Field Programmable Analog Arrays (FPAAs) are a way of modeling and implementing analog dynamical systems on a programmable platform. FPAAs are reconfigurable devices which is an advantage to change dynamics of the model even during implementation (Kılıç, 2010). They also provide a simple experimental setup that eliminates analog circuit complexity (Hasler, 2020).

AN231E04 Quad type FPAA is used for implementing Lorenz hyperchaotic system in this study. System (2) is designed on Anadigm Designer 2 and given in Figure 5.a. Implementation result for *x-y* dimensions is shown in Figure 5.b. Also, simulation result for the *x-y* dimensions is given at top of the Figure 5.b and it is seen that simulation and implementation results are consistent.



a) b)

Figure 5. a) FPAA design on Anadigm Designer b) Implementation result of Lorenz Hyperchaotic system for *x-y* dimensions (graph given upper left corner is simulation result for *x-y* dimensions)

4. Conclusion

In this study, analysis methods of hyperchaotic systems were investigated and performed on Rössler, Lorenz, Wang-Cang, and Sprott B systems. Then FPAA implementation of the Lorenz system is carried out and results are provided. Moreover, consistency of simulation and implementation results are considered.

References

Bao, B. C., & Liu, Z. (2008). A hyperchaotic attractor coined from chaotic lü system. *Chinese Physics Letters*, *25*(7), 2396–2399. https://doi.org/10.1088/0256-307X/25/7/018

Cannas, B., & Cincotti, S. (2002). Hyperchaotic behaviour of two bi-directionally coupled Chua's circuits. *International Journal of Circuit Theory and Applications*, *30*(6), 625–637. https://doi.org/10.1002/cta.213

Chen, A., Lu, J., Lü, J., & Yu, S. (2006). Generating hyperchaotic Lü attractor via state feedback control. *Physica A: Statistical Mechanics and Its Applications*, *364*, 103–110. https://doi.org/10.1016/j.physa.2005.09.039

Chua, L. O., & Kobayashi, K. (1986). Hyperchaos: Laboratory Experiment and Numerical Confirmation. *IEEE Transactions on Circuits and Systems*, *33*(11), 1143–1147. https://doi.org/10.1109/TCS.1986.1085862

Fitch, A. L., Yu, D., Iu, H. H. C., & Sreeram, V. (2012). Hyperchaos in a memristor-based modified canonical Chua's circuit. *International Journal of Bifurcation and Chaos*, *22*(6). https://doi.org/10.1142/S0218127412501337

Gao, T., Chen, Z., Yuan, Z., & Chen, G. (2006). A hyperchaos generated from Chen's system. *International Journal of Modern Physics C*, *17*(4), 471–478. https://doi.org/10.1142/S0129183106008625

Gottwald, G. A., & Melbourne, I. (2004). A new test for chaos in deterministic systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *460*(2042), 603–611. https://doi.org/10.1098/rspa.2003.1183

Hasler, J. (2020). Large-Scale Field-Programmable Analog Arrays. *Proceedings of the IEEE*, *108*(8), 1283–1302. https://doi.org/10.1109/JPROC.2019.2950173

Jia, L. X., Dai, H., & Hui, M. (2010). A new four-dimensional hyperchaotic Chen system and its generalized synchronization. *Chinese Physics B*, *19*(10), 100501. https://doi.org/10.1088/1674-1056/19/10/100501

Kapitaniak, T., Chua, L. O., & Zhong, G. Q. (1994). Experimental Hyperchaos in Coupled Chua's Circuits. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *41*(7), 499–503. https://doi.org/10.1109/81.298367

Kılıç, R. (2010). *A practical guide for studying Chua's circuits*. http://www.worldscibooks.com/series/wssnsa_series.shtml

Li, Y., Tang, W. K. S., & Chen, G. (2005). Hyperchaos evolved from the generalized Lorenz equation. *International Journal of Circuit Theory and Applications*, *33*(4), 235–251. https://doi.org/10.1002/cta.318

Liu, Y., Yang, Q., & Pang, G. (2010). A hyperchaotic system from the Rabinovich system. *Journal of Computational and Applied Mathematics*, *234*(1), 101–113. https://doi.org/10.1016/j.cam.2009.12.008

Ojoniyi, O. S., & Njah, A. N. (2016). A 5D hyperchaotic Sprott B system with coexisting hidden attractors. *Chaos, Solitons and Fractals*, *87*, 172–181. https://doi.org/10.1016/j.chaos.2016.04.004

Rössler, O. E. (1979). An equation for hyperchaos. *Physics Letters A*. https://doi.org/10.1016/0375-9601(79)90150-6

Wang, F. Q., & Liu, C. X. (2006). Hyperchaos evolved from the Liu chaotic system. *Chinese Physics*, *15*(5), 963–968. https://doi.org/10.1088/1009-1963/15/5/016

Wang, H., & Cai, G. (2009). Controlling hyperchaos in a novel hyperchaotic system. In *Journal of Information and Computing Science* (Vol. 4, Issue 4).

Wang, X., & Wang, M. (2008). A hyperchaos generated from Lorenz system. *Physica A: Statistical Mechanics and Its Applications*, *387*(14), 3751–3758. https://doi.org/10.1016/j.physa.2008.02.020

Wang, Z., Cang, S., Ochola, E. O., & Sun, Y. (2012). A hyperchaotic system without equilibrium. *Nonlinear Dynamics*, *69*(1–2), 531–537. https://doi.org/10.1007/s11071-011-0284-z

Wolf, A., Swift, J. B., Swinney, H. L., & Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena*, *16*(3), 285–317. https://doi.org/10.1016/0167-2789(85)90011-9

Yang, Q., & Chen, C. (2013). A 5D hyperchaotic system with three positive lyapunov exponents coined. *International Journal of Bifurcation and Chaos*, *23*(6). https://doi.org/10.1142/S0218127413501095

Yang, Q., Osman, W. M., & Chen, C. (2015). A new 6D hyperchaotic system with four positive Lyapunov exponents coined. *International Journal of Bifurcation and Chaos*, *25*(4), 1550060. https://doi.org/10.1142/S0218127415500601

Yang, Q., Zhu, D., & Yang, L. (2018). A New 7D Hyperchaotic System with Five Positive Lyapunov Exponents Coined. *International Journal of Bifurcation and Chaos*, *28*(5). https://doi.org/10.1142/S0218127418500578

Yujun, N., Xingyuan, W., Mingjun, W., & Huaguang, Z. (2010). A new hyperchaotic system and its circuit implementation. *Communications in Nonlinear Science and Numerical Simulation*, *15*(11), 3518–3524. https://doi.org/10.1016/j.cnsns.2009.12.005

Analysis of the Service Quality Performance of Hospitals in the COVID-19 Vaccine Process

Melike Erdoğan[1], Ertuğrul Ayyıldız[2,3*]

[1]*Department of Industrial Engineering, Duzce University, Duzce, Turkey*
[2]*Department of Industrial Engineering, Karadeniz Technical University, Trabzon, Turkey*
[3]*Department of Industrial Engineering, Yildiz Technical University, İstanbul, Turkey*

Abstract

The vaccination process has finally started for the pandemic that the whole world has been trying to cope with in recent years. In this context, the vaccination process has started in many countries. Most of the people in Turkey have been vaccinated since January in order of priority by age or occupational group (healthcare workers, teachers... etc.). People can get an appointment from the online system of The Ministry of Health and be vaccinated at specified times they select. At this point, different service quality is revealed in the vaccination process for different hospitals. In this study, we try to determine the priorities of the criteria that hospitals should consider for the vaccine service quality measurement process. For this purpose, we apply the SERVPERF (Service Performance) and multi-criteria decision-making (MCDM) approach to weighting the service performance attributes of hospitals in the COVID-19 vaccination process. Using the dimensions of the SERVPERF model, MCDM analysis is developed to deal with all the qualitative and quantitative criteria in the decision process is obtained. In addition, fuzzy sets are adopted to reflect the uncertainty to the decision-making process in the best way. As a result of this paper, the most important performance criteria on the vaccination process will be determined for the people who will be vaccinated, and it will be determined which criteria should be given more importance by hospitals in providing vaccination services. With this study, a quantitative analysis of the service quality of hospitals in the vaccination process will be presented for the first time.

Keywords: COVID-19, Fuzzy Logic, Hospital, MCDM, SERVPERF, Vaccination

1. Introduction

As a result of the acceleration of vaccination activities all over the world, different service quality levels are revealed in the vaccination process for different hospitals. Thus, it has emerged that the evaluation of the perceived service performance quality in hospitals should be implemented within the vaccination process. In this way, it is expected that the new horizon will be opened for the measures that can be taken to eliminate drawbacks for people who have hospital-based hesitations during the vaccination process. In this paper, we aim to determine the priorities of the factors to be used in measuring the perceived performance quality in hospitals where vaccines are applied during the COVID-19 period. For this aim, the SERVPERF (Service Performance) and multi-criteria decision-making (MCDM) under fuzzy environment approaches are adopted in weighting the service performance attributes of hospitals in the COVID-19 vaccination process. Using the dimensions of the SERVPERF model and new criteria identified along with this paper, fuzzy MCDM analysis is developed to deal with all the qualitative and quantitative criteria in the decision process. Interval-valued fuzzy sets are also adopted to reflect the uncertainty to the decision-making process in the best way. Besides, by applying the AHP method in

* *Corresponding Author*

the Pythagorean fuzzy environment, it is aimed to get the best results in decision problems with conflicting criteria in an uncertain environment and vagueness.

Examples of newsworthy studies conducted with the SERVPERF method in order to measure the quality perception in hospitals can be given as follows. Akdere et al. (Akdere, Top, & Tekingündüz, 2020) applied the SERVPERF method to measure the service quality perceptions of patients in a public hospital in Turkey. Giao et al. (Giao, Thy, Vuong, Van Kiet, & Lien, 2020) used the SERVPERF model to identify factors affecting outpatient satisfaction for five private hospitals in China. Arab et al. (Arab, Tabatabaei, Rashidian, Forushani, & Zarei, 2012) used the SERVPERF questionnaire to determine hospital service quality from the patients' perspective and the relative importance of quality dimensions in predicting patient adherence. Subiyakto et al. (Kot & Syaharuddin, 2020) adopted the dimensions of SERVPERF to measure outpatient satisfaction with radiology facilities in public hospitals in a specific region of Indonesia. Dako et al. (Lim et al., 2018) adopted the SERVPERF approach to measure the perceptions of service quality for patients in a PET/CT Centre.

The SERVPERF approach has been used in some studies by hybridizing it with the MCDM method. For example, Zehmet and Jawad (Zehmed & Jawab, 2020b) applied the SERVPERF method in fuzzy environment and developed a three-step approach using Data Envelopment analysis to determine the relative quality of service at the level of bus routes. Lee and Kang (S. Lee & Kang, 2019) proposed an approach based on objective and subjective weights for the assessment of airway service quality, using SERVPERF with interval-valued fuzzy GRA. Abdolvand and Rahpeima (Abdolvand & Rahpeima, 2013) applied a hybrid approach consisting of SERVPERF, entropy method and TOPSIS method in evaluating an insurance company's branches based on customers' perceptions of their shopping experience. Jeronimo and Medeiros (Jerônimo & Medeiros, 2014) presented a model to guide organizations according to the relative importance of customer satisfaction, using the dimensions of the SERVPEF method and the fee dimension as inputs for the ELECTRE-III method. Zhi-gang (Zhi-gang, 2015) revealed a revised SERVPERF model and seventeen taxi service quality indexes for taxi service quality assessment and ranked participants' ranked perceptions of taxi service quality using the VIKOR method. Besides, Pythagorean fuzzy AHP has been a frequently used approach in decision making problems in the literature recently (Alkan, 2021; Bakioglu & Atahan, 2020; Boyacı & Şişman, 2021; Otay & Jaller, 2020; Seker & Kahraman, 2021; Tepe & Kaya, 2020). Based on all these literature studies, we tried to measure the service performance of hospitals by combining the SERVPERF tool with the interval-valued Pythagorean fuzzy AHP in this study. Our study points to an innovation in the literature in terms of the adopted methodology.

2. Proposed Methodology

2.1. SERVPERF

The SERVPERF model was first used by Cronin and Taylor in 1992 to measure the service quality score on 22 items based only on customer perception (Cronin & Taylor, 1992). The SERVPERF method apply perceived performance of customers as the measure of service quality directly (H. Lee & Kim, 2014). The SERVPERF method was developed in order to overcome the deficiencies of the SERVQUAL method developed in 1985 in the measurement of service quality, such as operationalization, validity, and combining expectations and perceptions (Zehmed & Jawab, 2020a).

2.2. Interval-Valued Pythagorean Fuzzy Analytic Hierarchy Process (PF-AHP)

The steps of the interval valued PF-AHP method are shown in the following (Ilbahar, Karaşan, Cebi, & Kahraman, 2018; Özkan, Kaya, Erdoğan, & Karaşan, 2020):

Step 1. Compromised pairwise comparison matrix $R= (r_{jt})_{mxn}$ is constructed according to experts' evaluations. The linguistic scale that is used for decision matrices is presented in Table 1.

Step 2. The differences matrix $D = (d_{ij})_{mxm}$ is found between lower and upper points of the membership and non-membership functions using Eqs. (1) and (2):

$$d_{ij_L} = {\mu_{ij}}_L^{\,2} - {v_{ij}}_L^{\,2} \tag{1}$$

Table 1. Scale for the interval-valued PF-AHP evaluations

	Interval Valued Fuzzy Number				**The corresponding crisp value**
Linguistic Terms	μ_L	μ_u	U_L	U_u	
Certainly Low Importance -CLI	0.00	0.00	0.90	1.00	1/9
Very Low Importance – VLI	0.10	0.2	0.8	0.9	1/7
Low Importance – LI	0.20	0.35	0.65	0.8	1/5
Below Average Importance -BAI	0.35	0.45	0.55	0.65	1/3
Equal Importance – EI	0.45	0.55	0.45	0.55	1
Above Average Importance – AAI	0.55	0.65	0.35	0.45	3
High Importance – HI	0.65	0.80	0.20	0.35	5
Very High Importance – VHI	0.80	0.90	0.10	0.20	7
Certainly High Importance – CHI	0.90	1.00	0.00	0.00	9

$$d_{ij_U} = {\mu_{ij}}_U^{\,2} - {v_{ij}}_U^{\,2} \tag{2}$$

Step 3. The interval multiplicative matrix $S = (s_{ij})_{mxm}$ is calculated via Eqs. (3) and (4):

$$s_{ij_L} = \sqrt{1000^{d_{ij_L}}} \tag{3}$$

$$s_{ij_U} = \sqrt{1000^{d_{ij_U}}} \tag{4}$$

Step 4. The indeterminacy value $H = (h_{ij})_{mxm}$ of the r_{jt} is calculated with using Eq. (5):

$$h_{ij} = 1 - \left({\mu_{ij}}_U^{\,2} - {\mu_{ij}}_L^{\,2}\right) - \left({v_{ij}}_U^{\,2} - {v_{ij}}_L^{\,2}\right) \tag{5}$$

Step 5. The indeterminacy degrees are multiplied with $S = (s_{ij})_{mxm}$ matrix to calculate the matrix of unnormalized weights $T = (\tau_{ij})_{mxm}$ via Eq. (6):

$$t_{ij} = \left(\frac{s_{ij_L} + s_{ij_U}}{2}\right) h_{ij} \tag{6}$$

Step 6. The priority weights w_i are obtained via Eq. (7).

$$w_i = \frac{\sum_{i=1}^{m} t_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{m} t_{ij}} \tag{7}$$

3. Real Case Study

In this study, a prioritization analysis is carried out for the criteria to be used in determining the vaccination performance of hospitals in the COVID-19 period. At this point, first of all, SERVPERF dimensions are discussed, and then the factors to be considered in the pandemic period for hospital evaluation are investigated. Considering the relevant literature and expert opinions, it has been determined that "Pandemic" and "Smartness" factors should also be taken into account in the evaluation of the criteria along with the SERVPERF dimensions. While deciding which hospitals to be evaluated,

the hospitals serving in Istanbul are taken into consideration and the most intensive hospitals are selected in terms of both COVID-19 treatment applications and vaccine application. Quality managers from five different hospitals are interviewed to determine the order of importance of the factors to be used in measuring the perceived performance quality in hospitals where vaccines are applied during the COVID-19 period. These quality experts are asked to evaluate the SERVPERF dimensions and the factors identified in this study. Experts are requested to make an evaluation by using to the scale given in Table 1 while determining the relative importance of the criteria to each other. As an example, the table of the first expert, which includes the evaluation of the criteria against each other, is added as follows.

Table 2. Evaluations of Expert-1

Expert-1	*Tangibles*	*Reliability*	*Assurance*	*Responsiveness*	*Empathy*	*Smartness*	*Pandemic*
Tangibles	EI	BAI	AAI	EI	HI	AAI	BAI
Reliability	AAI	EI	HI	AAI	HI	EI	AAI
Assurance	BAI	LI	EI	LI	AAI	BAI	LI
Responsiveness	EI	BAI	HI	EI	HI	AI	BAI
Empathy	LI	LI	BAI	LI	EI	BAI	LI
Smartness	BAI	EI	AAI	EI	AAI	EI	BAI
Pandemic	AAI	BAI	HI	AAI	HI	AAI	EI

After receiving evaluations from five different experts, it is first examined whether these evaluations are consistent. For this purpose, the consistency analysis suggested by Saaty (Saaty, 1980, 1990, 2003) is carried out for each expert's evaluation matrix. Consistency analysis is performed by referring to the crisp values corresponding to linguistic assessments in Table 1, and all pairwise comparison matrices are obtained consistent (Consistency Ratio (C.R.) <0.1) as shown in Table 3 for each expert (E).

Table 3. Consistency Ratio for Each Matrices

Matrix	**C.R.**
Pairwise Comparison Matrix of E-1	0,093255
Pairwise Comparison Matrix of E-2	0,081157
Pairwise Comparison Matrix of E-3	0,099490
Pairwise Comparison Matrix of E-4	0,047493
Pairwise Comparison Matrix of E-5	0,085952

After calculating the consistency ratios, the steps of the interval-valued Pythagorean fuzzy AHP are implemented and dimension weights are obtained for each expert evaluation. In order to find the final weights, the weights for the five experts are aggregated via geometric mean. Table 4 shows the weights found as a result of each expert assessment, and the final results.

Table 4. Weights of the Dimensions

Dimension	**E-1**	**EM-2**	**EM-3**	**E-4**	**E-5**	**Aggregated Weights**
Tangibles	0,157	0,240	0,136	0,289	0,159	**0,188**
Reliability	0,231	0,190	0,150	0,139	0,221	**0,182**
Assurance	0,063	0,046	0,116	0,070	0,091	**0,073**
Responsiveness	0,177	0,159	0,276	0,123	0,164	**0,173**
Empathy	0,038	0,041	0,038	0,071	0,077	**0,050**
Smartness	0,110	0,121	0,122	0,123	0,148	**0,124**
Pandemic	0,224	0,203	0,163	0,185	0,140	**0,181**

When the results obtained are evaluated, it is determined that the criterion that should be given the most importance in the evaluation of hospital performances for vaccination during the pandemic process is the "tangibles" criterion, followed by the "reliability" and "pandemic" criteria with very close importance levels. The least important criterion is determined as "empathy." As a result, the factors that affect the hospital performance the most and the least in the COVID-19 vaccination process have been successfully identified.

4. Conclusion and Future Suggestions

In this study, it is determined the priorities of the criteria that hospitals should consider for the service performance measurement for vaccine process. The SERVPERF and fuzzy MCDM approaches are applied for weighting the service performance attributes of hospitals in the COVID-19 vaccination process. For this aim, SERVPERF and interval-valued Pythagorean fuzzy AHP methods are adopted. Interval-valued fuzzy sets are adopted to reflect the uncertainty to the decision-making process in the best way. As a result of the analysis values obtained, it is determined that the factor that most affected the performance of the hospitals for the COVID-19 vaccination process is "tangibles" and the factor that affected the least is "empathy".

For future studies, the criteria can be further detailed by breaking down on the basis of sub-criteria and the results can be compared using different MCDM methods. In addition, service performance comparisons can be made on the basis of criteria and weights determined for different hospitals for vaccination process of COVID-19.

References

Abdolvand, M. A., & Rahpeima, A. (2013). Evaluating and ranking businesses' branches, based on clients' perception, a study in insurance industry. *Research Journal of Applied Sciences, Engineering and Technology*, *5*(4), 1323–1329. https://doi.org/10.19026/RJASET.5.4868

Akdere, M., Top, M., & Tekingündüz, S. (2020). Examining patient perceptions of service quality in Turkish hospitals: The SERVPERF model. *Total Quality Management and Business Excellence*, *31*(3–4), 342–352. https://doi.org/10.1080/14783363.2018.1427501

Alkan, N. (2021). Risk Analysis for Digitalization Oriented Sustainable Supply Chain Using Interval-Valued Pythagorean Fuzzy AHP. *Advances in Intelligent Systems and Computing*, *1197 AISC*, 1373–1381. https://doi.org/10.1007/978-3-030-51156-2_160

Arab, M., Tabatabaei, S. G., Rashidian, A., Forushani, A. R., & Zarei, E. (2012). The Effect of Service Quality on Patient loyalty: a Study of Private Hospitals in Tehran, Iran. *Iranian Journal of Public Health*, *41*(9), 71. Retrieved from /pmc/articles/PMC3494218/

Bakioglu, G., & Atahan, A. O. (2020). Evaluating the Influencing Factors on Adoption of Self-driving Vehicles by Using Interval-Valued Pythagorean Fuzzy AHP. *Advances in Intelligent Systems and Computing*, *1197 AISC*, 503–511. https://doi.org/10.1007/978-3-030-51156-2_58

Boyacı, A. Ç., & Şişman, A. (2021). Pandemic hospital site selection: a GIS-based MCDM approach employing Pythagorean fuzzy sets. *Environmental Science and Pollution Research*. https://doi.org/10.1007/S11356-021-15703-7

Cronin, J. J., & Taylor, S. A. (1992). Measuring Service Quality: A Reexamination and Extension. *Journal of Marketing*, *56*(3), 55. https://doi.org/10.2307/1252296

Giao, H. N. K., Thy, N. T. A., Vuong, B. N., Van Kiet, T., & Lien, L. T. P. (2020). Outpatient satisfaction at private general hospitals in Ho Chi Minh city, Vietnam. *Journal of Asian Finance, Economics and Business*, *7*(7), 323–334. https://doi.org/10.13106/JAFEB.2020.VOL7.NO7.323

Gul, M., & Ak, M. F. (2018). A comparative outline for quantifying risk ratings in occupational health and safety risk assessment. *Journal of Cleaner Production*, *196*, 653–664. https://doi.org/10.1016/j.jclepro.2018.06.106

Ilbahar, E., Karaşan, A., Cebi, S., & Kahraman, C. (2018). A novel approach to risk assessment for occupational health and safety using Pythagorean fuzzy AHP & fuzzy inference system. *Safety Science*, *103*, 124–136. https://doi.org/10.1016/j.ssci.2017.10.025

Jerônimo, T. de B., & Medeiros, D. (2014). Measuring quality service : The use of a SERVPERF scale as an input for ELECTRE TRI multicriteria model. *International Journal of Quality & Reliability Management*, *31*(6), 652–664. https://doi.org/10.1108/IJQRM-06-2012-0095

Karasan, A., Kaya, İ., Erdoğan, M., & Budak, A. (2020). Risk analysis of the autonomous vehicle driving systems by using pythagorean fuzzy AHP. *Advances in Intelligent Systems and Computing*, *1029*, 926–934. https://doi.org/10.1007/978-3-030-23756-1_110

Kot, S., & Syaharuddin, S. (2020). The government reform on healthcare facilities from the standpoint of service quality performance. *International Journal of Economics and Finance Studies*, *12*(1), 16–31. https://doi.org/10.34109/IJEFS.202012102

Lee, H., & Kim, C. (2014). Benchmarking of service quality with data envelopment analysis. *Expert Systems with Applications*, *41*(8), 3761–3768. https://doi.org/10.1016/j.eswa.2013.12.008

Lee, S., & Kang, D. (2019). Development of interval-valued fuzzy GRA with SERVPERF based on subjective and objective weights for evaluation of airline service quality: A case study of Korea low-cost carriers. *PLOS ONE*, *14*(8), e0219739. https://doi.org/10.1371/JOURNAL.PONE.0219739

Lim, J. S., Lim, K. S., Heinrichs, J. H., Al-Aali, K., Aamir, A., & Qureshi, M. I. (2018). The role of hospital service quality in developing the satisfaction of the patients and hospital performance. *Management Science Letters*, *8*(12), 1353–1362. https://doi.org/10.5267/J.MSL.2018.9.004

Otay, I., & Jaller, M. (2020). Multi-criteria and multi-expert wind power farm location selection using a pythagorean fuzzy analytic hierarchy process. *Advances in Intelligent Systems and Computing*, *1029*, 905–914. https://doi.org/10.1007/978-3-030-23756-1_108

Özkan, B., Kaya, İ., Erdoğan, M., & Karaşan, A. (2020). Evaluating blockchain risks by using a MCDM methodology based on pythagorean fuzzy sets. In *Advances in Intelligent Systems and Computing* (Vol. 1029). https://doi.org/10.1007/978-3-030-23756-1_111

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, *48*(1), 9–26. https://doi.org/10.1016/0377-2217(90)90057-I

Saaty, T. L. (2003). Decision-making with the AHP: Why is the principal eigenvector necessary. *European Journal of Operational Research*, *145*(1), 85–91. https://doi.org/10.1016/S0377-2217(02)00227-8

Seker, S., & Kahraman, C. (2021). Socio-economic evaluation model for sustainable solar PV panels using a novel integrated MCDM methodology: A case in Turkey. *Socio-Economic Planning Sciences*, *77*. https://doi.org/10.1016/J.SEPS.2020.100998

Tepe, S., & Kaya, İ. (2020). A fuzzy-based risk assessment model for evaluations of hazards with a real-case study. *Human and Ecological Risk Assessment: An International Journal*, *26*(2), 512–537. https://doi.org/10.1080/10807039.2018.1521262

Zehmed, K., & Jawab, F. (2020a). A combined approach based on fuzzy SERVPERF and DEA for measuring and benchmarking the quality of urban bus transport service at the route level. *Industrial Engineering and Management Systems*, *19*(2), 442–459. https://doi.org/10.7232/IEMS.2020.19.2.442

Zehmed, K., & Jawab, F. (2020b). A Combined Approach Based on Fuzzy SERVPERF and DEA for Measuring and Benchmarking the Quality of Urban Bus Transport Service at the Route Level. *Industrial Engineering & Management Systems*, *19*(2), 442–459. https://doi.org/10.7232/IEMS.2020.19.2.442

Zhi-gang, Y. (2015). Ranking Perceptions of Taxi Service Quality among Participants with VIKOR. *Journal of Transportation Systems Engineering and Information Technology*, *15*(4), 210. Retrieved from http://manu01.magtech.com.cn/Jwk_jtys/EN/

Organizational Configurations Boosting Enterprise Performance and Job Satisfaction*

Alperen Ekrem Çelikdin*

Planning & Logistics Manager, Sutas Co., Aksaray, Turkey

Abstract

It is known that organizations consist of a combination of several causal factors. Organizational Configurations theory had started with the contingency theory. Contingency theory is known as, an organization cannot be configured with the best single structure and Organizational Configurations theory has developed by the equifinality approach. Equifinality is defined as, "a system can reach the same final situation from different and original starting points". It has been stated that businesses with similar practices and strategies can be seen as clustered groups and these decisive organizational characteristics and strategies shape the performance of the enterprise. Therefore, the importance of Organizational Configurations approach is increasing. However, since the sustainable performance of the enterprises can be supported with the efforts of the satisfied employees, the Organizational Configurations that are desired to be formed should be shaped accordingly. With this study, the automotive spare parts sector of Central Anatolia Region was selected as the main sample and high performing Organizational Configurations was determined based on the Miles and Snow typologies theory. The relationship between the targeted output and the set of independent variables that make up this output was determined by Fuzzy Set Qualitative Comparative Analysis, which aims explaining with the set theoretical approach. While providing high enterprise performance, the arrangements to be made in order to have high job satisfaction at the same time according to Decision Makers choices, are modelled by the help of Compromise Programming which is one of the Operations Research methods. The results of the study support the Miles and Snow typologies. Finally, the results show that, enterprises that manage their entrepreneurial, engineering and administrational problems with solutions suitable for environment and competition conditions, authorizing the employees and encouraging them to participate in the management processes, reach high performance and job satisfaction results.

Keywords: Organizational Configurations, Performance, Job Satisfaction, fsQCA, Compromise Programming, Multi Criteria Decision Making

** Produced from the doctoral thesis of the same title.*

1. Introduction

Within the scope of structure, process and strategy dimensions, the configuration groups formed by all these alternatives are called Organizational Configurations (OC). The choices of these configurations can affect business performance and employee job satisfaction. While a decentralized and innovative organization can both increase business performance and increase the level of employee job satisfaction, an organization that maintains its central and current status will increase business performance and decrease employee job satisfaction in certain situations. There is not yet any empirical study on the organizational configurations approach in Turkey. For this reason, with this study, the first field research was carried out with the aim of both closing this gap and providing a happy working environment without compromising the goals. It is aimed to offer a solution to the ongoing problem of businesses to be completely performance-oriented in order to compete and to neglect the job satisfaction of

* *Corresponding Author*

employees, which are their most important resources. Since multi-disciplinary studies provide a multi-dimensional perspective, an interdisciplinary contribution has been made to the business and management literature with the operations research method.

In this study, how OC affects employee job satisfaction and business performance, in which cases business performance increases, in which situations the level of employee job satisfaction increases, what is/are OC that increases both criteria will be examined. Accordingly, the study consists of three parts. In the first part, it is aimed to contribute to the field of management science with an interdisciplinary method in order to express the importance of the study. Updates and eclectic contributions, and how a tool can be provided to implementing managers in case of success are mentioned. The concept of OC and the antecedent theories that revealed this concept are mentioned. The concept of business performance, its measurement, the concept of employee job satisfaction, its measurement and dimensions, and the relations of all these concepts with each other are mentioned. Finally, the operations research method to be used in the selection of configurations that improve both business performance and employee job satisfaction is mentioned.

2. Problem Definition

Organizational configurations, organizational clusters that share features such as business strategies, organizational structures and processes, occur together, and in which the harmony between all necessary variables in terms of structure and content are gathered in a single orientation (Miller, 1987; Mintzberg, 1979; Miller, 1996; Siggelkow, 2002). In the study, although there are many OC's found in the literature, the reason why Miles and Snow's typologies were chosen is that they have high generalizability, very strong theoretical bases and empirical responses, and receive very high support from confirmatory studies (Hambrick, 1983; Mitchell and Zmud, 2006; Johansen, 2007; Pleshko and Nickerson, 2008; Smith et al., 1989).

Miles and Snow's typology includes four dimensions. Defenders tend to balance by keeping control of the market segment they have identified. In order to achieve these goals, they appeal to a narrow segment of the total market by producing only a limited variety of products. In this limited framework, they fiercely advocate economic measures such as price competition and high quality to prevent competitors from entering this market. The Prospectors move in many directions against their chosen environment, diametrically opposite to the defenders. In general, they are more dynamic than other companies in their industry. Instead of a narrow product range and market understanding, as advocates do, they show an approach to developing new products and seizing new market opportunities. The Analyzers operate with an ideal mix of analysts, defenders, and prospectors. The Reactors are businesses that try to adapt to their environment with incompatible and inconsistent responses. The reason for the inconsistent responses is that they cannot provide a solid structure and strategy relationship.

In order to determine as many different configurations as possible, which is one of the requirements of the method, data has been collected from as many enterprises as time and resource constraints allow.

In the selection of the sample, enterprises operating in the branch of "Parts and accessories for motor vehicles not elsewhere classified", classified with the Nace code of 29.32.30, competing to supply spare parts to the same automotive main industry companies were selected. In order to increase the consistency of the study, data were collected from businesses operating in the Central Anatolian region. Some statistics on the distribution of the sample by provinces are shown in Table 1, participants positions are given by Table 2.

In order to determine the multi-dimensional structure and complex relationships of organizational configurations, companies that make fabricated production and have at least 50 employees been chosen rather than workshop-type manufacturing enterprises. The employee distribution of the 25 enterprises where the study was conducted is given in Table 3.

Table 1. Sample space

City	Number of Enterprises	Number of Participants
Ankara	7	23
Aksaray	2	14
Konya	14	69
Nevşehir	1	4
Kayseri	1	2
Total	25	112

Table 2. Participant's position

Position	Number of Participants
General Manager/CEO	8
Manager	16
Executive	15
Specialist	71
Blue Collar Foreman	2
Total	112

Table 3. Distribution of number of employees of the businesses where the study was done

No of Employees	No of Enterprises	No of White Collars
0-50	1	2
50-100	7	30
100-150	4	27
150-200	6	35
200-250	2	23
>250	5	100
Total	25	*217*

As can be seen in Figure 1, employee job satisfaction and change in business performance according to organizational configurations are determined as dependent variables. The dimensions that will form the organizational configurations with the deductive approach according to the Miles and Snow typologies are given as independent variables.

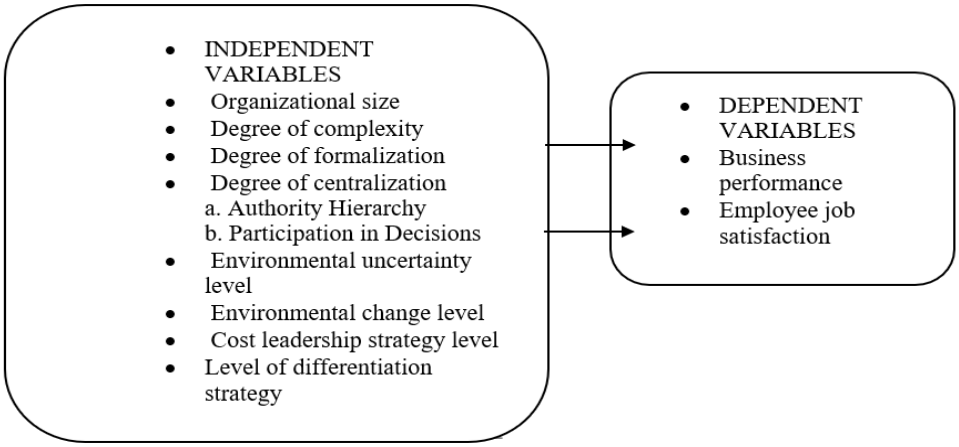


Figure 1. Research model

There are eleven hypotheses in the original doctoral thesis. Due to space constraints, the eleventh hypothesis of the study, which includes a multidisciplinary approach, and the proposed method to test this hypothesis will be examined.

3. Proposed Method

In some studies, in the literature, it is stated that there is a positive relationship between business performance and employee job satisfaction (Judge, et al., 2001), some other studies states job satisfaction is a cause of performance (Locke, 1976). While in other studies, performance causes job satisfaction (Lawler & Porter, 1967). In order to contribute to the literature in terms of achieving two different goals in a balanced way, the eleventh hypothesis is to try to model the adjustments that will

simultaneously bring the organizational configurations, business performance and employee job satisfaction multi-purpose goals according to internal and external environmental conditions, to the highest point within the specified organizational configuration.

H_{11}: Within the determined organizational configurations, there are compromise solutions that simultaneously meet the objectives of business performance and employee job satisfaction. Correlation analysis was conducted to determine the relationships between organizational structure, environment and strategy dimensions (independent variables), business performance and employee job satisfaction. As a result of the correlation analysis, a significant relationship was found between the dimensions. In the multi-objective decision-making model, correlation coefficients were used while constructing the objective functions.

Qualitative comparative analysis (QCA) method was used to create constraints in the multi-objective decision-making model, to identify configurations, to explain social phenomena and causal complexity. It was developed by Charles Ragin (1987,2008) both as a method and an approach. It uses 0-1 algebra and set theory. Fuzzy Set QCA software simplifies causal statements. The set theoretical approach examines the extent to which the variables in causal expressions have membership in the related feature set with the fuzzy logic approach. The measured variables are calibrated using anchor points to fuzzy number values between 0-1 using in-depth information. This calibration requires the determination of three qualitative benchmarks (full membership, full non-cluster and full uncertainty). Consistency shows how close one is getting to a perfect subset relationship. Cluster-theoretic inclusion measures the degree to which a type of cause or causal combination "expresses" an output.

Table 4. Anchor Points in Calibrating Dependent Variables

	Full non-membership (0)	Cross-over point (0.5)	Full membership (1)
Perceived Performance	3,14	3,81	5,00
Job Satisfaction	3,10	3,70	5,00
Organizational Size	50	250	1000

The comparison table (which can also be referred to as the comparison table) analysis specified in the study was made separately for the same independent variables for the results of business performance and employee job satisfaction, in accordance with the procedure specified with the help of the fsQCA program. Job satisfaction results were obtained separately. Consistency cut-off point was determined as 0.94 (must be >0.75) in the analysis for business performance. Causal prescriptions (rows) above this value result in the dependent variable taking a value of 1 and having a membership of 0.5 and above in the high-performance enterprises cluster according to the consistency values, while the rows below this value take a value of 0 and remain out of this cluster. The comparison chart containing these results is given in Table 5. As can be seen from Table 5, 11 different causal prescriptions have been reached, of which 6 summarize organizational configurations that provide high performance, and 5 summarize organizational configurations that cause low performance. Same table for the employee job satisfaction can be seen from Table 6.

Compromise programming (Zeleny, 1974) was used to determine the ideal configuration that provides both business performance and employee job satisfaction.

Min D, subject to Constraints;

$Z_1\ (Maximize\ enterprise\ performance) =$ (1)

$a_{\ performance}$*(0.265*Formalization+0.231*Participation to decisions-0.055*Hierarchy of authorization-0.118*Organizational Complexity- 0.176*Environmental Uncertainty+0.043*Level of environmental change+0.085*Diffirentiation+0.16*Cost leadership) + 2.728*D ≥ $a_{\ performance}$*3.315

Table 5. Enterprises with high business performance

Formalization	Participation In Decisions	Hierarchy of Authority	Organizational Complexity	Environmental Uncertainty	Environmental Change Level	Differentiation Strategy	Cost Leadership	Organizational Size	Frequency	Perceived Performance	Consistency
1	0	0	0	0	0	0	1	1	1	1	0,96
1	1	1	0	0	0	1	1	1	1	1	0,96
1	1	0	0	0	0	0	1	1	1	1	0,96
0	1	1	0	1	1	1	1	0	1	1	0,96
1	1	1	1	1	1	1	0	0	1	1	0,96
1	0	1	0	0	1	1	1	0	1	1	0,94
1	1	1	0	0	0	1	1	0	1	0	0,94
1	1	0	0	0	0	0	1	0	1	0	0,93
1	0	1	1	0	1	1	1	0	1	0	0,92
1	1	0	0	0	1	0	1	0	1	0	0,90
1	0	1	1	1	0	1	0	0	1	0	0,89

Table 6. Enterprises with high employee job satisfaction

Formalization	Participation In Decisions	Hierarchy of Authority	Organizational Complexity	Environmental Uncertainty	Environmental Change Level	Differentiation Strategy	Cost Leadership	Organizational Size	Frequency	Job Satisfaction	Consistency
1	1	0	0	0	0	0	1	0	1	1	0,93
1	1	1	0	0	0	1	1	0	1	1	0,92
1	1	0	0	0	1	0	1	0	1	1	0,92
1	1	0	0	0	0	0	1	1	1	1	0,88
1	0	0	0	0	0	0	1	1	1	1	0,86
1	1	1	1	1	1	1	0	0	1	1	0,85
0	1	1	0	1	1	1	1	0	1	1	0,83
1	1	1	0	0	0	1	1	1	1	0	0,81
1	0	1	1	1	0	1	0	0	1	0	0,81
1	0	1	0	0	1	1	1	0	1	0	0,76
1	0	1	1	0	1	1	1	0	1	0	0,75

$Z_2(Maximize\ job\ satisfaction) =$ (2)

$a_{job\ satisfaction}$*(0.492* Formalization +0.482* Participation to decisions -0.306* Hierarchy of authorization -0.12* Organizational Complexity +0.084* Environmental Uncertainty -0.394* Level of environmental change -0.114* Diffirentiation +0.11* Cost leadership) + 5.1679*D ≥ $a_{job\ satisfaction}$*4.738 ; Goal weights : $a_{performance}$ and $a_{job\ satisfaction}$

Formalization ≥ 3; formalization ≤5; participation in decisions ≥ 1;Participation in decisions ≤ 5; hierarchy of authority≤ 3; hierarchy of authority ≥ 1; organizational complexity ≥ 1;organizational complexity ≤ 3; environmental uncertainty level ≥ 1;environmental uncertainty level ≤ 3; environmental change level ≥ 1; environmental change level ≤ 3; differentiation ≥ 1; differentiation ≤ 3; cost leadership ≥ 3; cost leadership ≤ 5; All variables ≥ 0, Equal goal weights.

≥ 1 Constraint: Due to both the structure of the data collection tool and the calibration logic, the choices of the people working in the data collection enterprise can be both equal to 1 and greater than 1. Since the questionnaire also starts from 1, it cannot take smaller values.

≥ 1 and ≤ 3 Constraint: Considering the selected calibration benchmarks, as stated in the literature, for a condition to be out of the set, the condition must be at or below the fully fuzzy point 0.5 (i.e., it corresponds to 3 in the study). (Ragin, 2008).

≥ 3 and ≤ 5 Constraint: Considering the selected calibration benchmarks, as stated in the literature, for a condition to be included in the set, the condition of being included in the set must be 0.5 (that is, equivalent to 3 in the study) value and above. (Ragin, 2008).

4. Results and Discussion

Compromise programming model solved with Lindo software. Hypothesis 11 was accepted as configurations that provide both business performance and employee job satisfaction were identified. Compromise solution can be seen from Table 7. With Configuration 1, if it is aimed to ensure both business performance and employee job satisfaction, it is necessary to increase the level of formalization and to perform all operations with written and well-defined workflows. Instead of centralizing the administration, it is necessary to bring it to a decentralized structure by spreading the authority to the lower parts of the organization. However, employees need to be more effective and empowered in managerial decisions, which is more related to employee job satisfaction. Rather than multi-level and complex structuring of the organization vertically and horizontally, a simpler organization should be established and organizational complexity should be brought to the lowest level. Since Configuration 1 operates in an external environment that is predictable, static, uncertain and the level of change is not rapid, its strategies also differentiate its products and services, albeit at a low level, by making the highest level of products and services at the lowest cost, by reducing the costs at the highest level.

Table 7. Compromise solution for Configuration 1

Variables	Z_1 $\alpha_{perf.} = 0{,}5$	Z_2 $\alpha_{job\ satis.} = 0{,}5$	Z_1 $\alpha_{perf.} = 0{,}7$	Z_2 $\alpha_{job\ satis.} = 0{,}3$	Z_1 $\alpha_{perf.} = 0{,}3$	Z_2 $\alpha_{job\ satis.} = 0{,}7$
Formalization	5	5	5	5	5	5
Participation in decisions	5	5	5	5	5	5
Hierarchy of authority	1	1	1	1	1	1
Organizational complexity	1	1	1	1	1	1
Environmental uncertainty level	1	1	1	1	2	2
Environmental change level	1	1	1	1	1	1
Differentiation	2	2	3	3	1	1
Cost leadership	5	5	5	5	5	5
OBJECTIVES	3.144	4.456	3.229	4.342	2.883	4.654
	D = 0.0313416		**D = 0.022988**		**D = 0.012669**	

5. Conclusion

Due to resource and time constraints, the study has been limited to the Central Anatolian region, it can be expanded and done throughout Turkey. While examining the structure-process and strategy dimensions, adding technology to the model will allow finding and explanation in technology detail.

For employee job satisfaction, analyses involving blue-collar employees can be added. In addition to the Miles and Snow typologies, a synthesis study can be conducted that includes all the configurations of Mintzberg and other strategists. In this study, the deductive (theory-based) method was chosen. In a sector with a wide field of activity (for example, textile) in Turkey, an inductive configuration study according to Turkish business practices can be carried out with cluster analysis.

References

Hambrick, D. (1983). Some tests of the effectiveness and functional attributes of Miles and Snow's strategic types, The Academy of Management Journal , 26 (1), 5–26.

Johansen, M. S. (2007). The effect of female strategic managers on organizational performance. Public Organization Review, 7 (3), 269–279.

Judge, T. A., Thoresen, C. J., Bono, J. E., Patton, G. K. (2001). The job satisfaction-Job performance relationship: a qualitative and quantitative review. Psychological Bulletin, 127 (3), 376-407.

Lawler, E. E., Porter, L. W. (1967). The effect of performance on job satisfaction. Industrial Relations, 7(1), 20–28.

Locke, E. A. (1976). The nature and causes of job satisfaction' in Dunnette. M. D. (Ed.). Handbook of Industrial and Organizational Psychology,(ss 1297- 1349), Chicago: Rand McNally

Miller, D. (1987). The genesis of configuration. The Academy of Management Review, 12 (4),686-701.

Miller, D. (1996). Configurations revisited. Strategic Management Journal , 17, 505-512.

Mitchell, V., & Zmud, R. (2006). Endogenous adaptation: The effects of technology position and planning mode on it-enabled change. Decision Sciences , 37 (3), 325–355.

Pleshko, L., & Nickerson, I. (2008). Strategic orientation, organizational structure, and the associated effects on performance in industrial firms. Academy of Strategic Management Journal , 7, 95–110.

Ragin, C. C. (1987). The comparative method: Moving beyond qualitative and quantitative strategies. University of California Press, Berkeley

Ragin, C.C. (2008). Redesigning the social inquiry, fuzzy sets and beyond, The University of Chicago Press, Chicago

Siggelkow, N. (2002). Evolution toward fit. Administrative Science Quarterly, 47,125-159

Smith, G. K., Guthrie, J.P., Chen, M.J. (1989). Strategy, size and performance. Organization studies, 10 (1), 63-81.

Zeleny, M. (1974). A concept of compromise solutions and the method of the displaced ideal. C*omputers and Operations Research* , 1, 479-496.

Author Index

2nd International Conference on Applied Mathematics in Engineering

September 1-3, 2021 – Balikesir, Turkey

ISBN:
978-605-70379-8-5