# IRANIAN JOURNAL OF MATHEMATICAL CHEMISTRY

University of Kashan

I
J
M
C

**Editor-in-Chief:**
**A R Ashrafi** (University of Kashan, Iran)

**Managing Director:**
**B Bazigaran** (University of Kashan, Iran)

**Language Editor:**
**S Mosazadeh** (University of Kashan, Iran)

**Technical Manager:**
**M Pourbabaee (University of Kashan, Iran)**

**Editorial Board:**
A R Ashrafi (University of Kashan, Iran)
B Bazigaran (University of Kashan, Iran)
M R Darafsheh (University of Tehran, Iran)
M Diudea (Babes-Bolyai University, Romania)
T Došlić (University of Zagreb, Croatia)
G H Fath-tabar (University of Kashan, Iran)
A Gholami (University of Qom, Iran)
I Gutman (University of Kragujevac, Serbia)
A Iranmanesh (Tarbiat Modares University, Iran)
M A Iranmanesh (Yazd University, Iran)
P E John (Technical University of Ilmenau, Germany)
S Klavžar (University of Maribor, Slovenia)
X Li (Nankai University, China)
H R Maimani (Shahid Rajaee Teacher Training University. Iran)
S Masoum (University of Kashan, Iran)
O Ori (Actinum Chemical Research, Italy)
M V Puts (West University of Timisoara, Romania)
M Salavati-Niasari (University of Kashan, Iran)
B Shareghi-Boroujeni (University of Shahrkord, Iran)
B Taeri (Isfahan University of Technology, Iran)
H Yousefi-Azari (University of Tehran, Iran)
S Yousefi (Malek-Ashtar University of Technology, Iran)
B Zhou (South China Normal University, China)

**Past Editor**
A Graovac (Ruder Bosǩovic institute, Croatia)
M Deza (École Normale Supérieure, France)

**http://IJMC.Kashanu.ac.ir**

# Iranian Journal of Mathematical Chemistry

Iranian Journal of Mathematical Chemistry (IJMC) publishes quality original research papers and survey articles in Mathematical Chemistry and related areas that are of the highest possible quality.

Research papers and review articles are selected through a normal refereeing process by a member of editorial board.

It is intended that the journal may act as an interdisciplinary forum for publishing chemically important mathematical papers. Papers published in this journal must have a clear connection to chemistry with non-trivial mathematics.

**Aims and Scopes:** Iranian Journal of Mathematical Chemistry is a bi publication of the University of Kashan. It contains research and expository papers as well as short communications on chemically important mathematical problems.

**Preparation of Manuscripts:** An abstract of 150 words or less and keywords clarifying the subject of the manuscripts are required. Authors should submit the paper electronically in word file to ijmc@kashanu.ac.ir.

The language of the journal is English. Authors who are less familiar with English language are advised to seek assistance from proficient scholars to prepare manuscripts that are grammatically and linguistically free from errors. Manuscripts should be typed with double spacing and wide margins. Papers submitted for consideration for publication have to be prepared concisely and must not exceed 30 typewritten pages, (font 12, 1.15 spacing, including tables & graphs). Diagrams and figures should be submitted as original drawings. Figures and tables are to be numbered in the sequence in which they are cited in the manuscript. Every table must have a caption that explains its content. Tables and diagrams are to be prepared in separate pages and placed at the end of the manuscript. References to the literature should be numbered in square brackets in the order in which they appear in the text. A complete list of references should be presented in numerical order at the end of the manuscript. References to journals, books, proceedings and patents must be presented in accordance with the following examples:

**Journals:** I. Gutman, B. Zhou and B. Furtula, The Laplacian-energy like invariant is an energy like invarint MATCH Commun. Math. Comput. Chem., 64(1) (2010), 85-96.

**Patents:** Primack, H.S.; Method of Stabilizing Polyvalent Metal Solutions, U.S. patent No. 4, 373, 104(1983).

**Indexing/ Abstracting:** The Iranian Journal of Mathematoical Chemistry is indexed/abstracted in the following:

- Islamic World Science Citation Center (ISC)
- Zentralblatt für Mathematik
- Web of Science (ISI)

# EditorialBoard

| | |
|---|---|
| **A R Ashrafi** | University of Kashan, I R Iran<br>E-mail:ashrafi@kashanu.ac.ir |
| **B Bazigaran** | University of Kashan, I R Iran<br>E-mail:bazigaran@kashanu.ac.ir |
| **M R Darafsheh** | University of Tehran, I RIran<br>E-mail:darafsheh@ut.ac.ir |
| **M Diudea** | Babes–Bolyai University, Romania<br>E-mail:diudea@chem.ubbcluj.ro |
| **T Došlić** | University of Zagreb, Croatia<br>E-mail:doslic@faust.irb.hr |
| **G H Fath-Tabar** | University of Kashan, I R Iran<br>E-mail:fathtabar @kashanu.ac.ir |
| **A Gholami** | University of Qom, I R Iran<br>E-mail:a.gholami@qom.ac.ir |
| **I Gutman** | University of Kragujevac, Serbia<br>E-mail:gutman@kg.ac.yu |
| **A Iranmanesh** | Tarbiat Modares University, I R Iran<br>E-mail:iranmanesh@modares.ac.ir |
| **M A Iranmanesh** | Yazd University, I R Iran<br>E-mail:iranmanesh@yazduni.ac.ir |
| **P E John** | Technical Univesity of IImenau, Germany<br>E-mail:peter.john@tu-ilmenau.de |
| **S Klavžar** | University of Maribor, Slovenia<br>E-mail:sandi.klavzar@fmf.uni-lj.si |
| **X Li** | Nankai University,Tianjin300071,Tianjin, P RChina<br>E-mail:lxl@nankai.edu.cn |
| **H R Maimani** | Shahid Rajaee Teacher Training University, I R Iran<br>E-mail:maimani@ipm.ir |
| **S Masoum** | University of Kashan, I RIran<br>E-mail:masoum@kashanu.ac.ir |
| **O Ori** | Actinum Chemical Research, Rome, Italy<br>E-mail:ottorino.ori@ gmail.com |
| **M V Puts** | West University of Timisoara, Timisoara, Romania<br>E-mail:mv_putz@yahoo.com |
| **M Salavati–Niasar** | University of Kashan, I R Iran<br>E-mail:Salavati@kashanu.ac.ir |
| **B Shareghi–Boroujeni** | Shahrekord University, I R Iran<br>E-mail:share.beh@sci.sku.ac.ir |
| **B Taeri** | Isfahan University of Technology, I R Iran<br>E-mail: b.taeri@cc.iut.ac.ir |
| **H Yousefi-Azari** | University of Tehran, I R Iran<br>E-mail:hyousefi@ut.ac.ir |
| **S Yousefi** | Malek–Ashtar University of Technology, I R Iran<br>E-mail:yousefi100@yahoo.com |
| **B Zhou** | South China Normal University, P RChina<br>E-mail:zhoubo@scnu.edu.cn |

## Past Editors

| | |
|---|---|
| **A Graovac** | Ruđer Bošković Institute, Croatia<br>E-mail:graovac@irb.hr |
| **M Deza** | École Normale Supérieure, France<br>E-mail:Michel.Deza@ens.fr |

**Editor-in-Chief:**

Ali Reza Ashrafi
Department of Pure Mathematics, Faculty of Mathematical Sciences, University of Kashan, Kashan87317-53153, I R IRAN
mail:ashrafi@kashanu.ac.ir

**Managing Director:**

Behnam Bazigaran
Department of Pure Mathematics, Faculty of Mathematical Sciences, University of Kashan, Kashan87317-53153, I R IRAN
E-mail:bazigaran@kashanu.ac.ir

**Language Editor:**

Seyfollah Mosazadeh
Department of Pure Mathematics, Faculty of Mathematical Sciences, University of Kashan, Kashan 87317-53153, I R IRAN
E-mail: s.mosazadeh@kashanu.ac.ir

**Technical Manager:**

Marzieh Pourbabaee
Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Kashan, Kashan 87317-53153, I R IRAN
Email:m.pourbabaee@kashanu.ac.ir

# Iranian Journal of Mathematical Chemistry

## Vol. 8, No. 2 June 2017

**CONTENTS** **pages**

Iranian J. Math. Chem. 8(2) June (2017) 93−105

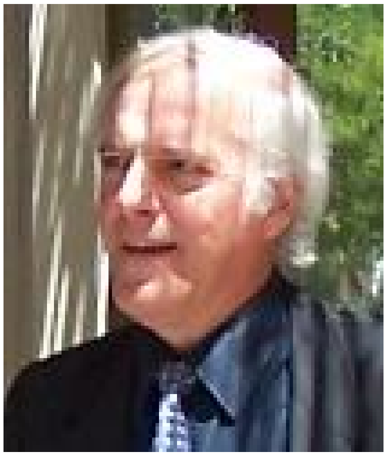
Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

# *Autobiography of Roberto Todeschini*

**ROBERTO TODESCHINI**•

Milano Chemometrics and QSAR Research Group, Department of Earth and Environmental Sciences, University of Milano Bicocca, Milan, Italy
Website: http://michem.disat.unimib.it/chm/



Born in Rome (Italy) in 1949, Roberto Todeschini graduated in chemistry at the University of Milan (Italy) in 1972 with a thesis in theoretical chemistry. He got married to Marilena in 1973 and soon had a daughter Alessia (1975) and some years later a son Edoardo (1982).

In 1975, he was appointed to the position of researcher at the same university in Milano and worked in the field of theoretical chemistry, with special focus on conformational analysis, until the first ‘80s. Then, since 1985 his research interests have shifted towards chemometrics, multivariate analysis and applied statistics, QSAR, molecular descriptors, multicriteria decision making, software development, starting to publish on these topics. The new research interests arose from the participation to an educational project carried out at Escuela Politecnica del Chimborazo (Riobamba, Ecuador) in the framework of an official cooperation action between the Foreign Ministers of Ecuador and Italy, aimed to propose over there up-to-date but low cost researches. In the beginning of ‘90s, Todeschini became associate professor at the University of Milano-Bicocca (Milan, Italy) and then, in 2001, full professor of chemometrics at the Department of Earth and Environmental Sciences in the same Institution, where he founded the *Milano Chemometrics and QSAR Research Group*. The actual staff of the research group includes Davide Ballabio, Viviana Consonni and Francesca Grisoni, while past-researchers of the

---

•Corresponding author (Email: roberto.todeschini@unimib.it)

DOI: 10.22052/ijmc.2017.43095

group are Paola Gramatica, Manuela Pavan, Andrea Mauri, Matteo Cassotti, Kamel Mansouri, and Faizan Sahigara.

He was one of the founders of the International Academy of Mathematical Chemistry in 2004 and served as President of the Academy from 2008 to 2013. He is also President of the Italian Chemometric Society and "ad honorem" professor of the University of Azuay (Cuenca, Ecuador) since 2006.

He is author of more than 200 publications in international peer-reviewed journals and co-author of the books: "The Data Analysis Handbook", by I.E. Frank and R. Todeschini (Elsevier, 1994); "Handbook of Molecular Descriptors", by R. Todeschini and V. Consonni (Wiley-VCH, 2000); "Molecular Descriptors for Chemoinformatics" by R. Todeschini and V. Consonni (Wiley-VCH, 2009); "Handbook of Bibliometric Indicators" by R. Todeschini and A. Baccini (Wiley-VCH, 2016). The first book about the molecular descriptors (2000) is among the most cited scientific books, with more 3300 in 2016.

Todeschini, together with his research group, organized several national and international meetings and schools about molecular descriptors, multi-criteria decision making, chemometrics, experimental design.

The complete list of publications of Todeschini can be accessed in Google Scholar at "https://scholar.google.com/citations?user=MNyEfBcAAAAJ" with more than 11000 citations and an h-index of 44 (January 2017).His main scientific results were organized in different topics and presented below.

## 1. CHEMOMETRICS

The interests of Todeschini about chemometrics include all its theoretical aspects. Critical reviews about $Q^2$ metrics proposed in literature for evaluating prediction ability of regression models were published and earned a big success. More recently, an extended comparison about the different proposals to estimate $Q^2$ was published and a reliable solution was proposed.

About classification methods, he proposed three new methods, called CAIMAN, N3 and BNN; in particular, the last two methods, proposed together with Ballabio, Cassotti and Consonni, were demonstrated to give very good quality classification results. Moreover, studies were also made on the KNN classification method, about its dependence on scaling and distance measures and for evaluating missing values.

Several studies were also performed on the concept of the applicability domain of classification and regression models and on the topic of variable selection, for which some novel methods were also proposed.

In 1993, Todeschini along with Ildiko Frank released the software SCAN (Software for Chemometric Analysis) andin 1994 Todeschini, still with Ildiko Frank, published the book "The Data Analysis Handbook" (Elsevier).

## 2. MOLECULAR DESCRIPTORS

Todeschini proposed in 1994 a set of 3D molecular descriptors, called WHIM, based on the Principal Component Analysis and, in 2000 with Consonni, the GETAWAY descriptors, based on the statistical concept of leverage. Starting from the publication of the first book on molecular descriptors, already mentioned afore, Consonni and Todeschini published several reviews and book chapters in this field. About topological descriptors, they published a paper where a generalization of graph energy was proposed analysing the spectral indices defined in literature as well as the different kinds of topological vertex degree. Still with Consonni, a review was also dedicated to analyse all the weighted and unweighted topological matrices defined in literature.

The software DRAGON was also developed by Todeschini and his research group since 2000, collecting from time to time always more molecular descriptors.

## 3. STATISTICS

Among the main results in statistics, the K multivariate correlation index was proposed in two papers, giving the exact definition of an index able to give a measure of global correlation in a multivariate dataset.

Several papers were also dedicated to similarity/diversity measures. Among these, a paper with Consonni, Peter Willett *et al*., on all the binary similarity measures proposed in the literature, some papers about canonical distances and similarities between datasets, an original variant of the Mahalanobis distance, a similarity measure for DNA sequences, a new similarity measure taking into account higher order similarities, a Hausdorff-like similarity measure to evaluate the similarity between sets (i.e. molecules) when they are partitioned in different constituents (i.e. ionic liquids, metabolites, sub-structures) and a review about similarity/diversity measures for the Encyclopaedia of Analytical Chemistry by Wiley & Sons.

## 4. QUANTITATIVE STRUCTURE-ACTIVITY (-PROPERTY) RELATIONSHIPS (QSAR/QSPR), DRUG DESIGN

Several chemometric applications were developed by Todeschini's research group in QSAR/QSPR and related fields.

In the environmental and human health fields, studies were performed on biodegradability, aquatic toxicity, tropospheric degradation, bioconcentration factor, bioaccumulation, water quality, non-ionic surfactants, mutagenicity and environmental priority settings. In chemistry, studies were performed on the phisico-chemical and toxicological behaviour of PAH, PCB, PCDD and PCDF compounds. In food chemistry, studies were performed on olive oils, sweetness, dairy cream, spirits, and food quality analysis, while in drug design, studies were performed on Cytochrome P450, Ca2+ channel antagonists, non-peptide angiotensin II receptor antagonists and virtual compound screening for sets of pharmacological targets.

## 5. MULTICRITERIA DECISION MAKING

He was the editor, together with Manuela Pavan, of a book for Elsevier about the ranking methods and decisions based on multi-criteria. In this field, papers were published about indices for the evaluation of Hasse diagrams, generalization of the Power-Weakness Ratio (PWR) measure and Hasse theory.

## 6. BIBLIOMETRIC INDICATORS

In 2016 he published the "Handbook of Bibliometric Indicators" together with Alberto Baccini (Wiley-VCH), following the same encyclopaedic approach used in the past for the books about molecular descriptors.

This topic is actually particularly warm, due to the large interest of researchers, institutions, countries, etc. for the research evaluation.

## BIBLIOGRAPHIC REFERENCES SINCE 2000

1. Ballabio, D., Biganzoli, F., Todeschini, R. and Consonni, V. (2016) Qualitative consensus of QSAR ready biodegradability predictions. *Toxicological and Environmental Chemistry*, **in press**.
2. Cassotti, M., Grisoni, F., Nembri, S. and Todeschini, R. (2016) Application of the weighted Power-Weakness Ratio (wPWR) as a fusion rule in ligand-based virtual screening. *MATCH Commun. Math. Comput. Chem.*, **76**, 359–376.
3. Grisoni, F., Consonni, V., Vighi, M., Villa, S. and Todeschini, R. (2016) Investigating the mechanisms of bioconcentration through QSAR classification trees. *Environmental International*, **88**, 198–205.
4. Grisoni, F., Consonni, V., Vighi, M., Villa, S. and Todeschini, R. (2016) Expert QSAR system for predicting the bioconcentration factor under the REACH regulation. *EnvironmentalResearch*, **148**, 507–512.

5. Grisoni, F., Reker, D., Schneider, P., Friedrich, L., Consonni, V., Todeschini, R., Koberle, A., Werz, O. and Schneider, G. (2016) Matrix-based molecular descriptors for prospective virtual compound screening. *MolecularInformatics*, **35**.
6. Mauri, A., Consonni, V. and Todeschini, R. (2016) Molecular Descriptors, in *Handbook of Computational Chemistry* (ed. T. Puzyn), Springer.
7. Mauri, A., Ballabio, D., Todeschini, R. and Consonni, V. (2016) Mixtures, metabolites, ionic liquids: anew measure to evaluate similarity between complex chemical systems. *Journal of Cheminformatics*, **8**, 1-3.
8. Nembri, S., Grisoni, F., Consonni, V. and Todeschini, R. (2016) *In silico* prediction of Cytochrome P450 - Drug interaction: QSARs for CYP3A4 and CYP2C9. *International Journal of MolecularSciences*, **17**, 1–19.
9. Rojas, Ch., Ballabio, D., Consonni, V., Tripaldi, P., Mauri, A. and Todeschini, R. (2016) Quantitative Structure-Activity Relationships to predict sweet and non-sweet tastes. *TheoreticalChemistry Accounts*, 135–166.
10. Todeschini, R., Ballabio, D., Grisoni, F. and Consonni, V. (2016) A new concept of second-order similarity and the role od distance/similarity measures in local classification methods. *Chemometrics &Intell. Lab. Syst.*, **157**, 50–57.
11. Todeschini, R. and Baccini, A. (2016) *Handbook of Bibliometric Indicators*, Wiley-VCH, Weinheim (Germany), 512 pp.
12. Todeschini, R., Ballabio, D. and Grisoni, F. (2016) Beware of unreliable $Q^2$! A comparative study of regression metrics for predictivity assessment of QSAR models. *Journal of Chemical Information and Modeling*, **56**, 1905-1913.
13. Cassotti, M., Ballabio, D., Todeschini, R. and Consonni, V. (2015) A similarity-based QSAR model for predicting acute toxicity towards the fathead minnow (*Pimephalespromelas*). *SAR & QSAR in EnvironmentalResearch*, **26**, 217–243.
14. Grisoni, F., Consonni, V., Nembri, S. and Todeschini, R. (2015) How to weight Hasse matrices and reduce incomparabilities. *Chemometrics &Intell. Lab. Syst.*, **147**, 95–104.
15. Grisoni, F., Consonni, V., Villa, S., Vighi, M. and Todeschini, R. (2015) QSAR models for bioconcentration: is the increase in the complexity justified by more accurate predictions? *Chemosphere*, **127**, 171–179.
16. Mansouri, K. and et al. (2015) CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. *EnvironmentalHealthPerspectives*, **124**, 1023–1033.
17. Todeschini, R., Ballabio, D. and Consonni, V. (2015) Distances and Other Dissimilarity Measures in Chemometrics, in*Encyclopedia of Analytical Chemistry* Wiley & Sons, pp. 1–60.
18. Todeschini, R., Nembri, S. and Grisoni, F. (2015) Weighted Power-Weakness Ratio for multi-criteria decision making. *Chemometrics &Intell. Lab. Syst.*, **146**, 329–336.

19. Todeschini, R., Ballabio, D., Cassotti, M. and Consonni, V. (2015) N3 and BNN: Two new similarity based classification methods in comparison with other classifiers. *Journal of Chemical Information and Modeling*, **55**, 2365–2375.
20. Ballabio, D., Consonni, V., Mauri, A., Claeys-Bruno, M., Sergent, M. and Todeschini, R. (2014) A novel variable reduction method adapted from space-filling designs. *Chemometrics &Intell. Lab. Syst.*, **136**, 147–154.
21. Buscema, M., Consonni, V., Ballabio, D., Mauri, A., Massini, G., Breda, M. and Todeschini, R. (2014) K-CM: a new artificialneural network. Application to supervised pattern recognition. *Chemometrics &Intell. Lab. Syst.*, **138**, 110–119.
22. Cassotti, M., Ballabio, D., Consonni, V., Mauri, A., Tetko, I.V. and Todeschini, R. (2014) Prediction of acute aquatic toxicity toward *daphnia magna* using GA-*k*NN methods. *ATLA*, **42**, 31–41.
23. Cassotti, M., Grisoni, F. and Todeschini, R. (2014) Reshaped Sequential Replacement algorithm: an efficient approach to variable selection. *Chemometrics &Intell. Lab. Syst.*, **133**, 136–148.
24. Cherkasov, A., Muratov, E., Fourches, D., Varnek, A., Baskin, I., Cronin, M.T.D., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V., Cramer, R., Benigni, R., Yang, C., Richrad, A., Terfloth, L., Gasteiger, J. and Tropsha, A. (2014) QSAR Modeling: Where have you been? Where are yougoing to? *Journal MedicinalChemistry*, **57**, 4997–5010.
25. Grisoni, F., Cassotti, M. and Todeschini, R. (2014) Reshaped Sequential Replacement algorithm for variable selection in QSPR modelling: comparison with other benchmark methods. *Journal of Chemometrics*, **28**, 249–259.
26. Sahigara, F., Ballabio, D., Todeschini, R. and Consonni, V. (2014) Assessing the validity of QSARs for ready biodegradability of chemicals: An Applicability Domain perspective. *Current Computer-AidedDrug Design*, **10**, 137–147.
27. Swapnil, C., Nicholls, I., Karlsson, B., Rosengren, A., Ballabio, D., Consonni, V. and Todeschini, R. (2014) Towards Global QSAR Model Building for Acute Toxicity: Munro Database Case Study. *International Journal of MolecularSciences*, **15**, 18162–18174.
28. Tetko, I.V., Schramm, K.-W., Knepper, T., Peijnenburg, W.J.G.M., Hendriks, A.J., Nicholls, I.A., Öberg, T., Todeschini, R., Schlosser, E. and Brandmaier, S. (2014) The Experimental and Theoretical Studies within the FP7 Environmental ChemOinformatics Marie Curie Initial Training Network 'ECO'. *ATLA*, **42**, 1–5.
29. Todeschini, R., Consonni, V., Ballabio, D., Mauri, A., Cassotti, M., Lee, S., West, A. and Cartlidge, D. (2014) QSPR study of rheological and mechanical properties of Chloroprene rubber accelerators. *RubberChemistry and Technology*, **87**, 219–238.

30. Mansouri, K., Ringsted, T., Ballabio, D., Todeschini, R. and Consonni, V. (2013) Quantitative Structure-Activity Relationship models for ready biodegradability of chemicals. *Journal of Chemical Information and Modeling*, **53**, 867–878.
31. Sahigara, F., Ballabio, D., Todeschini, R. and Consonni, V. (2013) Defining a novel k-Nearest Neighbours approach to assess the applicability of a QSAR model for reliable predictions. *Journal of Chemoinformatics*, **5**, 1–9.
32. Todeschini, R., Ballabio, D., Consonni, V., Sahigara, F. and Filzmoser, P. (2013) Locally-centred Mahalanobis distance: a new distance measure with salient features towards outlier detection. *Anal. Chim. Acta*, **787**, 1–9.
33. Consonni, V. and Todeschini, R. (2012) Multivariate Analysis of Molecular Descriptors, in *Statistical Modelling of Molecular Descriptors in QSAR/QSPR* (eds. M. Dehmer, K. Varmuza and D. Bonchev), Wiley-Blackwell, Weinheim (Germany), pp. 111–147.
34. Consonni, V. and Todeschini, R. (2012) New similarity coefficients for binary data. *MATCH Commun. Math. Comput. Chem.*, **68**, 581–592.
35. Ippolito, A., Todeschini, R. and Vighi, M. (2012) Sensitivity assessment of freshwater macroinvertebrates to pesticides using biological traits. *Ecotoxicology*, **21**, 336–352.
36. Mansouri, K., Consonni, V., Durjava, M.K., Kolar, B., Öberg, T. and Todeschini, R. (2012) Assessing bioaccumulation of polybrominated diphenyl ethers for aquatic species by QSAR modeling. *Chemosphere*, **89**, 433–444.
37. Nielsen, N.J., Ballabio, D., Tomasi, G., Todeschini, R. and Christensen, J.H. (2012) Chemometric analysis of GC-FID chromatograms (CHEMFID): A novel method for classification of petroleum products. *J. Chromat. A*, **1238**, 121–127.
38. Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V. and Todeschini, R. (2012) Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules*, **17**, 4791–4810.
39. Todeschini, R., Consonni, V., Xiang, H., Holliday, J., Buscema, M. and Willett, P. (2012) Similarity coefficients for binary chemoinformatics data: overview and extended comparison using simulated and real datasets. *Journal of Chemical Information and Modeling*, **52**, 2884–2901.
40. Consonni, V. and Todeschini, R. (2011) Structure - Activity Relationships by autocorrelation descriptors and genetic algorithms, in*Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques* (eds. H. Lohdi and Y. Yamanishi), IGI Global Publishers, Hershey, PA (USA), pp. 60–93.
41. Sushko, I., Novotarskyi, S., Körner, R., Pandey, A.K., Rupp, M., Teetz, W., Brandmaier, S., Abdelaziz, A., Prokopenko, V.V., Tanchuk, V.Y., Todeschini, R., Varnek, A., Marcou, G., Ertl, P., Potemkin, V., Grishina, M., Gasteiger, J., Schwab,

C., Baskin, I., Palyulin, V.A., Radchenko, E.V., Welsh, W.J., Kholodovych, V., Chekmarev, D., Cherkasov, A., Aires-de-Sousa, J., Zhang, Q.-Y., Bender, A., Nigsch, F., Patiny, L., Williams, A., Tkachenko, V. and Tetko, I.V. (2011) Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J. Computer-AidedMol. Des.*, **25**, 533–554.

42. Todeschini, R. (2011) The *j*-index: a new bibliometric index and multivariate comparisons between other common indices. *Scientometrics*, **87**, 621–639.
43. Ballabio, D., Consonni, V., Mauri, A. and Todeschini, R. (2010) Canonical Measure of Correlation (*CMC*) and Canonical Measure of Distance (*CMD*) between sets of data. Part 3. Variableselection in classification. *Anal. Chim. Acta*, **657**, 116–122.
44. Ballabio, D. and Todeschini, R. (2010) Geographical characterization of olive oil by means of multivariate classification: application of CAIMAN, in *Olives and olive oil in health and disease prevention* (eds. V. R. Preedy and R. R. Watson), Elsevier, Amsterdam, pp. 131–139.
45. Consonni, V. and Todeschini, R. (2010) Molecular Descriptors, in *Recent Advances in QSAR Studies: Methods and Applications*, Vol. 8 (eds. M. T. D. Cronin, J. Leszczynski and T. Puzyn), Springer, Amsterdam (The Netherlands), pp. 29–102.
46. Consonni, V., Ballabio, D. and Todeschini, R. (2010) Evaluation of model predictive ability by external validation techniques. *Journal of Chemometrics*, **24**, 194–201.
47. Consonni, V., Ballabio, D. and Todeschini, R. (2010) Enhancing Chemical Information in QSAR: Generalized Graph-Theoretical Matrices, in *Novel Molecular Structure Descriptors - Theory and Applications II* (eds. I. Gutman and B. Furtula), University of Kragujevac, Kragujevac (Serbia), pp. 21–55.
48. Fernandez-Varela, R., Gomez-Carracedo, M.P., Ballabio, D., Andrade, J.M., Consonni, V. and Todeschini, R. (2010) Self Organizing Maps For Analysis Of Polycyclic Aromatic Hydrocarbons 3-Way Data From Spilled Oils. *AnalyticalChemistry*, **82**, 4264–4271.
49. Sushko, I., Novotarskyi, S., Körner, R., Pandey, A.K., Cherkasov, A., Li, J., Gramatica, P., Hansen, K., Schroeter, T., Müller, K.-R., Xi, L., Liu, H., Yao, X., Öberg, T., Hormozdiari, F., Dao, P., Sahinalp, C., Todeschini, R., Polishchuk, P., Artemenko, A., Kuz'min, V., Martin, T.M., Young, D.M., Fourches, D., Muratov, E., Tropsha, A., Baskin, I., Horbath, D., Marcou, G., Varnek, A., Prokopenko, V.V. and Tetko, I.V. (2010) ApplicabilityDomains for ClassificationProblems: Benchmarking of Distance to Models for AmesMutagenicity Set. *Journal of Chemical Information and Modeling*, **50**, 2094–2111.

50. Todeschini, R. and Consonni, V. (2010) New local vertex invariants and molecular descriptors based on functions of the vertex degrees. *MATCH Commun. Math. Comput. Chem.*, **64**, 359–372.
51. Todeschini, R., Ballabio, D. and Consonni, V. (2010) Novel Molecular Descriptors Based on Functions of New Vertex Degrees, in *Novel Molecular Structure Descriptors - Theory and Applications I* (eds. I. Gutman and B. Furtula), University of Kragujevac, Kragujevac (Serbia), pp. 73–100.
52. Ballabio, D. and Todeschini, R. (2009) Multivariate Classification for Qualitative Analysis, in *Infrared Spectroscopy for Food Quality Analysis and Control* (ed. S. Da-Wen), Elsevier, Amsterdam, pp. 83–104.
53. Ballabio, D., Manganaro, A., Consonni, V., Mauri, A. and Todeschini, R. (2009) Introduction to MOLE DB – on-line MolecularDescriptors Database. *MATCH Commun. Math. Comput. Chem.*, **62**, 199–207.
54. Ballabio, D., Consonni, V. and Todeschini, R. (2009) The Kohonen and CP-ANN toolbox: a collection of MATLAB modules for Self Organising Maps and Counterpropagation Artificial Neural Networks. *Chemometrics &Intell. Lab. Syst.*, **98**, 115–122.
55. Consonni, V., Ballabio, D., Manganaro, A., Mauri, A. and Todeschini, R. (2009) Canonical Measure of Correlation (*CMC*) and Canonical Measure of Distance (*CMD*) between sets of data. Part 2. Variablereduction. *Anal. Chim. Acta*, **648**, 52–59.
56. Consonni, V., Ballabio, D. and Todeschini, R. (2009) Comments on the definition of the $Q^2$ parameter for QSAR validation. *Journal of Chemical Information and Modeling*, **49**, 1669–1678.
57. Pavan, M. and Todeschini, R. (2009) Multicriteria Decision Making Methods, in *Comprehensive Chemometrics*, Vol. 1 (eds. B. Walczak, R. Taulér and S. Brown), Elsevier, Amsterdam (The Netherlands), pp. 591–629.
58. Piazza, L., Gigli, J., Rojas, Ch., Ballabio, D., Todeschini, R. and Tripaldi, P. (2009) Dairy Cream Response In Instrumental Texture Evaluation Processed By Multivariate Analysis. *Chemometrics &Intell. Lab. Syst.*, **96**, 258–263.
59. Todeschini, R., Consonni, V. and Gramatica, P. (2009) Chemometrics in QSAR, in *Comprehensive Chemometrics, vol. 4*, Vol. 4 (eds. S. Brown, B. Walczak and R. Taulér), Elsevier, Oxford (UK), pp. 129–172.
60. Todeschini, R. and Consonni, V. (2009) *Molecular Descriptors for Chemoinformatics (2 volumes),* Vol. 41, WILEY-VCH, Weinheim (Germany), 1257 pp.
61. Todeschini, R., Consonni, V., Manganaro, A., Ballabio, D. and Mauri, A. (2009) Canonical Measure of Correlation (*CMC*) and Canonical Measure of Distance

(*CMD*) between sets of data. Part 1. Theory and simple chemometric applications. *Anal. Chim. Acta*, **648**, 45–51.

62. Consonni, V. and Todeschini, R. (2008) New Spectral Indices for Molecule Description. *MATCH Commun. Math. Comput. Chem.*, **60**, 3–14.
63. Gutman, I., Indulal, G. and Todeschini, R. (2008) Generalizing the McClelland Bounds for Total □-Electron Energy. *ZeitschriftfürNaturforschung A*, **63a**, 280–282.
64. Manganaro, A., Ballabio, D., Consonni, V., Mauri, A., Pavan, M. and Todeschini, R. (2008) The DART (Decision Analysis by Ranking Techniques) software, in *Scientific Data Ranking Methods: Theory and Applications* (eds. M. Pavan and R. Todeschini), Elsevier, Amsterdam (The Netherlands), pp. 193–207.
65. Mauri, A., Ballabio, D., Consonni, V., Manganaro, A. and Todeschini, R. (2008) Peptides multivariate characterisation using a molecular descriptor based approach. *MATCH Commun. Math. Comput. Chem.*, **60**, 671–690.
66. Pavan, M. and Todeschini, R. (2008) Total order ranking methods, in *Scientific Data Ranking Methods: Theory and Applications*(eds. M. Pavan and R. Todeschini), Elsevier, Amsterdam (The Netherlands), pp. 51–72.
67. Todeschini, R. and Pavan, M., Eds. (2008) *Scientific Data Ranking Methods: Theory and Applications.*Elsevier, Amsterdam (The Netherlands), 180 pp.
68. Tetko, I.V., Sushko, I., Pandey, A.K., Zhu, H., Tropsha, A., Papa, E., Õberg, T., Todeschini, R., Fourches, D. and Varnek, A. (2008) Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: Focusing on applicability domain and overfitting by variable selection. *Journal of Chemical Information and Modeling*, **48**, 1733–1746.
69. Todeschini, R., Ballabio, D., Consonni, V. and Mauri, A. (2008) A new similarity/diversity measure for the characterization of DNA sequences. *CroaticaChemica Acta*, **81**, 657–664.
70. Ballabio, D., Consonni, V. and Todeschini, R. (2007) Classification of multiway analytical data based on MOLMAP approach. *Anal. Chim. Acta*, **605**, 134–146.
71. Ballabio, D., Kokkinofta, R., Todeschini, R. and Theocharis, C.R. (2007) A classification model built by means of Artificial Neural Networks for the characterization of the traditional Cypriot spirit Zivania. *Chemometrics &Intell. Lab. Syst.*, **87**, 78–84.
72. Todeschini, R., Ballabio, D., Consonni, V., Mauri, A. and Pavan, M. (2007) CAIMAN (Classification And Influence Matrix Analysis): A new approach to the classification based on leverage-scaled functions. *Chemometrics &Intell. Lab. Syst.*, **87**, 3–17.

73. Todeschini, R., Ballabio, D., Consonni, V. and Mauri, A. (2007) A new similarity/diversity measure for sequential data. *MATCH Commun. Math. Comput. Chem.*, **57**, 51–67.
74. Ballabio, D., Mauri, A., Todeschini, R. and Buratti, S. (2006) Geographical classification of wine and olive oil by means of CAIMAN (Classification And Influence Matrix Analysis). *Anal. Chim. Acta*, **570**, 249–258.
75. Ballabio, D., Cosio, M.S., Mannino, S. and Todeschini, R. (2006) A chemometric approach based on a novel similarity/diversity measure for the characterization and selection of electronic nose sensors. *Anal. Chim. Acta*, **578**, 170–177.
76. Mauri, A., Consonni, V., Pavan, M. and Todeschini, R. (2006) DRAGON software: an easy approach to molecular descriptor calculations. *MATCH Commun. Math. Comput. Chem.*, **56**, 237–248.
77. Pavan, M., Consonni, V., Gramatica, P. and Todeschini, R. (2006) New QSAR modelling approach based on ranking models by Genetic Algorithms - Variable Subset Selection (GA-VSS), in *Partial Order in Environmental Sciences and Chemistry* (eds. R. Brüggeman and L. Carlsen), SpringerVerlag, pp. 185–224.
78. Todeschini, R. (2006) Molecular Descriptors and Chemometrics. *G. I. T. Laboratory Journal*, **5**, 40–42.
79. Todeschini, R., Consonni, V., Mauri, A. and Ballabio, D. (2006) Characterization of DNA primary sequences by a new similarity/diversity measure based on the partial ordering. *Journal of Chemical Information and Modeling*, **46**, 1905–1911.
80. Pavan, M., Consonni, V. and Todeschini, R. (2005) Partial Ranking Models by Genetic Algorithms Variable Subset Selection (GA-VSS) approach for environmental priority settings. *MATCH Commun. Math. Comput. Chem.*, **54**, 583–609.
81. Tetko, I.V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V.A., Radchenko, E.V., Zefirov, N.S., Makarenko, A.S., Tanchuk, V.Y. and Prokopenkov, V.V. (2005) Virtual ComputationalChemistryLaboratory -- Design and Description. *J. Computer-AidedMol. Des.*, **19**, 453–463.
82. Pavan, M. and Todeschini, R. (2004) New indices for analyzing partial ranking diagrams. *Anal. Chim. Acta*, **515**, 167–181.
83. Pavan, M., Mauri, A. and Todeschini, R. (2004) Total ranking models by the Genetic Algorithms Variable Subset Selection (GA-VSS) approach for environmental priority settings. *Analytical and BioanalyticalChemistry*, **380**, 430–444.
84. Todeschini, R., Consonni, V., Mauri, A. and Pavan, M. (2004) New fitness functions to avoid bad regression models in variable subset selection by Genetic Algorithms, (eds. M. Ford, D. Livingstone, J. Deardean and H. van de Waterbeemd), Blakwell, Oxford (UK), pp. 323–325.

85. Todeschini, R., Consonni, V., Mauri, A. and Pavan, M. (2004) Detecting "bad" regression models: multicriteria fitness functions in regression analysis. *Anal. Chim. Acta*, **515**, 199–208.
86. Todeschini, R., Consonni, V. and Pavan, M. (2004) A Distance Measure between Models: a Tool for Similarity/Diversity Analsysis of Model Populations. *Chemometrics &Intell. Lab. Syst.*, **70**, 55–61.
87. Backhaus, T., Altenburger, R., Arrhenius, A., Blanck, H., Faust, M., Finizio, A., Gramatica, P., Grothe, M., Junghans, M., Meyer, W., Pavan, M., Porspring, T., Scholze, M., Todeschini, R., Vighi, M., Walter, H. and Grimme, L.H. (2003) The BEAM-project: prediction and assessment of mixture toxicities in the aquatic environment. *Continental ShelfResearch*, **23**, 1757–1769.
88. Lleti, R., Sarabia, L., Ortiz, M.C., Todeschini, R. and Colombini, M.P. (2003) Application of the Kohonen Artificial Neural Network in the identification of Proteinaceous Binders in Samples of Panel Painting Using Gas Chromatography-Mass Spectrometry. *The Analyst*, **181**, 281–286.
89. Mezzanotte, V., Castiglioni, F., Todeschini, R. and Pavan, M. (2003) Study on anaerobic and aerobic degradation of different non-ionic surfactants. *Bioresource Technology*, **87**, 87–91.
90. Todeschini, R., Consonni, V. and Pavan, M. (2003) MobyDigs: Software for Regression and Classification Models by Genetic Algorithms, in *Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks* (ed. R. Leardi), Elsevier, Amsterdam (The Netherlands), pp. 141–167.
91. Todeschini, R. and Consonni, V. (2003) Descriptors from Molecular Geometry, in *Handbook of Chemoinformatics - Vol.3*, Vol. 3 (ed. J. Gasteiger), WILEY-VCH, Weinheim (GER), pp. 1004–1033.
92. Todeschini, R., Consonni, V. and Pavan, M. (2003) Distance measure between models: a tool for model similarity/diversity analysis, in *Designing Drugs and Crop Protectants: processes, problems and solutions.* (eds. M. Ford, D. Livingstone, J. Deardean and H. van de Waterbeemd), Blakwell, Oxford (UK), pp. 467–469.
93. Consonni, V., Todeschini, R. and Pavan, M. (2002) Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *Journal of Chemical Information and Computer Sciences*, **42**, 682–692.
94. Consonni, V., Todeschini, R., Pavan, M. and Gramatica, P. (2002) Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies. *Journal of Chemical Information and Computer Sciences*, **42**, 693–705.

95. Benicori, T., Consonni, V., Gramatica, P., Pilati, T., Rizzo, S., Sannicolò, F., Todeschini, R. and Zotti, G. (2001) Steric Control of Conductivity in Highly ConjugatedPolythiophenes. *Chemistry of Materials*, **13**, 1665–1673.
96. Di Marzio, W., Galassi, S., Todeschini, R. and Consolaro, F. (2001) Traditional versus WHIM molecular descriptors in QSAR approaches applied to fish toxicity studies. *Chemosphere*, **44**, 401–406.
97. Gramatica, P., Vighi, M., Consolaro, F., Todeschini, R., Finizio, A. and Faust, M. (2001) QSAR approach for the selection of congeneric compounds with a similar toxicological mode of action. *Chemosphere*, **42**, 873–883.
98. Vighi, M., Gramatica, P., Consolaro, F. and Todeschini, R. (2001) QSAR and Chemometric Approaches for Setting Water Quality Objectives for Dangerous Chemicals. *Ecotoxicology and EnvironmentalSafety*, **49**, 206–220.
99. Capitan-Vallvey, L.F., Navas, N., del Olmo, M., Consonni, V. and Todeschini, R. (2000) Resolution of mixtures of three nonsteroidal anti-inflammatory drugs by fluorescence using partial least squares multivariate calibration with previous wavelength selection by Kohonen artificial neural networks. *Talanta*, **52**, 1069–1079.
100. Todeschini, R. and Consonni, V. (2000) *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim (Germany), 668 pp.

Iranian J. Math. Chem. 8(2) June (2017) 107−135

Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

# A Novel Topological Descriptor Based on the Expanded Wiener Index: Applications to QSPR/QSAR Studies

**AFSHAN MOHAJERI**[1,•], **POUYA MANSHOUR**[2] AND **MAHBOUBEH MOUSAEE**[1]

[1]Department of Chemistry, College of Sciences, Shiraz University, Shiraz 71454, Iran
[2]Department of Physics, Persian Gulf University, Bushehr 75169, Iran

**ARTICLE INFO**

**Article History:**

Received 6 April 2015
Accepted 3 March 2016
Published Online: 10 March 2017
Academic Editor: Ali Reza Ashrafi

**Keywords:**

Topological index
Graph theory
Expanded Wiener index
QSPR
QSAR

**ABSTRACT**

In this paper, a novel topological index, named *M*−index, is introduced based on expanded form of the Wiener matrix. For constructing this index the atomic characteristics and the interaction of the vertices in a molecule are taken into account. The usefulness of the *M*−index is demonstrated by several QSPR/QSAR models for different physico−chemical properties and biological activities of a large number of diversified compounds. Moreover, the applicability of the proposed index has been checked among isomeric compounds. In each case the stability of the obtained model is confirmed by the cross-validation test. The results of present study indicate that the *M*−index provides a promising route for developing highly correlated QSPR/QSAR models. On the other hand, the *M*−index is easy to generate and the developed QSPR/QSAR models based on this index are linearly correlated. This is an interesting feature of the *M*−index when compared with quantum chemical descriptors which require vast computational cost and exhibit limitations for large sized molecules.

## 1. INTRODUCTION

Graph theory is one of the most useful tools for studying systems in various fields such as chemistry,physics, computer science, economy, and biology [1–3]. This powerful concept, which introduced originally by agreat mathematician, Leonhard Euler [4], has been a useful approach to predict some key features of such systems. Chemical graph theory is a branch of graph theory that is concerned with analyses of all consequences of connectivity in a chemical graph. Chemical graph serves as a convenient model for any real or abstracted

---

[•] Corresponding author (Email: amohajeri@shirazu.ac.ir)

DOI: 10.22052/ijmc.2017.27307.1101

chemical system. It can represent different chemical objects as molecules, reactions, crystals, polymers, and clusters [5–8].

Successful use of chemical graph to quantitative structure–property and structure–activity relationships (QSPR/QSAR) has led to the emergence of several molecular descriptors [9–13]. The molecular descriptors derived there from are commonly named topological indices [5]. Regardless of the descriptors used in the development of QSPR/QSAR models, all of them share in common a basic approach; molecules are represented by vectors constructed in turn by molecular parameters, which are supposed to contain relevant information about molecular structure. So far, hundreds of topological indices have been proposed in chemical literature [14]. Among these, topological indices such as the molecular connectivity indices of the Wiener [15], Balaban [16], Randić [17], and Hosoya [18] indices have received greater attention due to their application in chemistry.

The Wiener index, $W$, is one of the most frequently used graph descriptors in QSPR/QSAR models. Its applicability for predicting physico–chemical and pharmacological properties of organic compounds is well documented and was outlined in quite afew reviews [19–21]. In the past decades, a remarkably large number of modifications and extensions of the Wiener index was put forward and studied by mathematical chemists [22,23]. These indices are generally based on the adjacency matrix or on the distance matrix.

However, most of introduced indices lack certain information such as the features and interaction of vertices (atoms). Recent attentions of graph theoretical chemists have been focused on resolving this problem [24]. In this respect, Yang et al. [25−28] used the electronegativity, the energy, the length between vertices, and principal quantum number to reform the distance matrix of Wiener so that more information of the distance matrix were included in the molecular graph. They have applied these modified Wiener in prediction of the retention indices of gas chromatography, the standard formation enthalpy and gaseous solubility.

However, owing to the complexity of the molecular structure, it seems to be impossible to expect that a single set of descriptors would contain all the relevant structural information. Hence, introducing topological indices which can predict a wide range of physico–chemical properties requiring a minimum number of inputs is the goal of many studies. As far as we know the Wiener index and its modification have been widely used to account for many physico–chemical properties. In the present study, we intend to propose a new topological index based on the reciprocal form of the expanded distance matrix. The expanded form of the Wiener index was suggested by Tratch et al. [29] for characterization of molecular graphs and structure–property correlations. This index is more sensitive to the structural characteristic of alkane molecule as compared with normal Wiener index and also can differentiate several graphs having just the same value of the very powerful

Balaban index. On the other hand, the inclusion of the interactions among vertices (atoms) as well as the vertex properties is a crucial issue in describing the whole properties of a molecule.

Accordingly, based on the above–mentioned issues, in the present study we have introduced a new topological index and named it *M*–index. This newly proposed index includes topological properties of the vertices (atom parameter) and the interaction of such vertices in a molecular graph (bond parameter). The applicability of *M*–index for the estimation of physical, chemical, and pharmacological properties such as boiling point, enthalpy of formation, refractive index, retention index, toxicity, Gibbs free energy, heat capacity, and etc. has been investigated. We have checked the appropriateness of this index for simple molecular compounds such as un–branched alkanes or cycloalkanes as well as more complicated systems with various functional groups and isomers.

## 2. DEFINITION OF THE M–INDEX

### 2.1 EXPANDED WIENER NUMBER

In 1971, Hosoya proposed a modified Wiener number which can be applied to chain and cyclic molecules [18], as follows:

$$W = \frac{1}{2}\sum\nolimits_{i,j}^{N} d_{ij} \tag{1}$$

where $N$ is the total number of the atoms in a molecule and $d_{ij}$'s are the elements of a matrix, called *distance matrix*. $d_{ij}$ is defined as the shortest distance between any two given atoms $i$ and $j$ in a molecule which is clearly equal to zero for all diagonal elements of $i = j$. Toobtaina higher discriminating ability of the Wiener number, Tratch et al. [29] proposed a novel topological index, called expanded Wiener number which is defined as

$$\tilde{W} = \sum\nolimits_{i<j} \tilde{d}_{ij} = \sum\nolimits_{i<j} \mu_{ij} \nu_{ij} d_{ij}. \tag{2}$$

In general, the vertices $i$ and $j$ may be connected by several, $\mu_{ij}$, shortest paths and for each of these paths a set of $\nu_{ij}$ shortest super–paths of the length equal or greater than $d_{ij}$ must be taken into account. It may be easily shown that the number of shortest superpaths is just the same for each of the shortest paths connecting $i$ with $j$. However, because of computationally extensive nature of the Eq. 2, an alternative method was proposed to compute the expanded Wiener number,

$$\tilde{W} = \sum\nolimits_{i<j} \tilde{f}_{ij} = \sum\nolimits_{i<j} \mu_{ij} \frac{d_{ij}(d_{ij}+1)(d_{ij}+2)}{6} \tag{3}$$

in which the total length of all subpaths for every pair of vertices is taken into account [29]. Note that the resulting expanded Wiener numbers calculated from Eqs. 2 and 3 are the same.

## 2.2 MODIFIED ADJACENCY MATRIX

The Wiener index considers the length of shortest paths only, and the properties of vertices as well as their interactions are not included. In this article, we try to introduce a novel topological index by focusing on some information about the structural details in the molecular graph such as the features and interaction of vertices. For this purpose, we modify the conventional form of adjacency matrix $A$, by inclusion of the bond parameters. The elements of modified adjacency matrix $A^*$ are defined as: $A^*_{ij} = 0$, in the cases where $i = j$ and where two vertices are not connected (non–adjacent vertices), otherwise $A^*_{ij} = b_{ij} X_{ij}$. The parameter $b_{ij}$ represents the bond order between atoms $i$ and $j$, and is 1, 2, and 3, respectively for the single, double, and triple bonds. Moreover, we have used the definition of Yang et al. [25−28] for the bonding characteristics, $X_{ij} = (1 + \Delta I_{ij}) / R_{ij}$. In this formula, $\Delta I_{ij} = |I_i - I_j|$ stands for the electronegativity difference between atoms $i$ and $j$, and $R_{ij} = n_i + n_j$, where $n_i$ and $n_j$ are the maximum principal quantum numbers of the atoms $i$ and $j$. According to its definition, $X_{ij}$ is a measure for the bonding ability between vertices $i$ and $j$, i.e., the smaller the value of $X_{ij}$, the weaker the bonding ability between atoms $i$ and $j$.

## 2.3 ELEMENTS OF THE $\tilde{M}$ –MATRIX

The modified version of adjacency matrix has been utilized for constructing the *M*–index in which not only the characteristics of an individual atom but also the role of that atom in establishing the connection with other atoms in a molecular graph are taken into account. The diagonal elements of the $\tilde{M}$ –matrix contain the electronegativity of atom $i$ as the characteristic of that atom and the sum of the $i$–th row of the $A^*$ matrix divided by $k_i$ (the number of neighboring atoms of atom $i$) as the average role of atom $i$ in establishing connections with other atoms in the molecule.

$$\tilde{M}_i = I_i + \frac{\sum_j A^*_{ij}}{k_i}. \tag{4}$$

If the vertices $i$ and $j$ are adjacent, then $\tilde{M}_{ij}$ is obtained by employing the inverse of the off–diagonal elements of the expanded Wiener index. Therefore, the elements of $\tilde{M}$ –matrix are summarized as below

$$\tilde{M}_{ij} = \begin{cases} \tilde{M}_i & i = j \\ \dfrac{1 + I_i I_j A^*_{ij}}{\tilde{f}_{ij}} & i, j \text{ are adjacent} \\ \dfrac{1}{\tilde{f}_{ij}} & i, j \text{ are non-adjacent.} \end{cases} \tag{5}$$

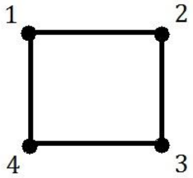
This matrix includes not only the characteristics of atoms but also the pattern of their connections in the molecular graph.

### 2.4 THE *M*–INDEX

Taking into account the symmetric property of the $\tilde{M}$ matrix, we introduce our proposed $M$–index as the sum of the upper triangular and diagonal elements of the $\tilde{M}$ matrix:

$$M = \sum_{i=1}^{N} \tilde{M}_i + \sum_{i<j}^{N} \tilde{M}_{ij}. \tag{6}$$

Now, we take an example to illustrate how to get *M*–index for a simple cyclic graph containing four vertices



The expanded distance matrix of such a graph is obtained by use of Eq. 3:

$$\tilde{f}_{ij} = \begin{pmatrix} 0 & 1*(1*2*3)/6 & 2*(2*3*4)/6 & 1*(1*2*3)/6 \\ 1*(1*2*3)/6 & 0 & 1*(1*2*3)/6 & 2*(2*3*4)/6 \\ 2*(2*3*4)/6 & 1*(1*2*3)/6 & 0 & 1*(1*2*3)/6 \\ 1*(1*2*3)/6 & 2*(2*3*4)/6 & 1*(1*2*3)/6 & 0 \end{pmatrix}.$$

Then, we can find $A^*$,

$$A^* = \begin{pmatrix} 0 & b_{12}X_{12} & 0 & b_{14}X_{14} \\ b_{21}X_{21} & 0 & b_{23}X_{23} & 0 \\ 0 & b_{32}X_{32} & 0 & b_{34}X_{34} \\ b_{41}X_{41} & 0 & b_{43}X_{43} & 0 \end{pmatrix}.$$

The diagonal elements, $\tilde{M}_i$ s, can be obtained through Eq. 4.

$$\begin{cases} \tilde{M}_1 = I_1 + (1/2)*(b_{12}X_{12} + b_{14}X_{14}) \\ \tilde{M}_2 = I_2 + (1/2)*(b_{21}X_{21} + b_{23}X_{23}) \\ \tilde{M}_3 = I_3 + (1/2)*(b_{32}X_{32} + b_{34}X_{34}) \\ \tilde{M}_4 = I_4 + (1/2)*(b_{41}X_{41} + b_{43}X_{43}) \end{cases}$$

Finally, the whole $\tilde{M}$ matrix is expressed as

$$\tilde{M} = \begin{pmatrix} \tilde{M}_1 & 1+I_1I_2b_{12}X_{12} & 1/8 & 1+I_1I_4b_{14}X_{14} \\ 1+I_2I_1b_{21}X_{21} & \tilde{M}_2 & 1+I_2I_3b_{23}X_{23} & 1/8 \\ 1/8 & 1+I_3I_2b_{32}X_{32} & \tilde{M}_3 & 1+I_3I_4b_{34}X_{34} \\ 1+I_4I_1b_{41}X_{41} & 1/8 & 1+I_4I_3b_{43}X_{43} & \tilde{M}_4 \end{pmatrix}$$

If we suppose that all vertices of the above graph are carbon the resulting graph represents a cyclobutane molecule in which $b_{12} = b_{23} = b_{34} = b_{14} = 1$, $I_1 = I_2 = I_3 = I_4 = 2.55$, and $n_1 = n_2 = n_3 = n_4 = 2$. These quantities give the value of $M \approx 21.96$ for the cyclobutane.

## 3. RESULTS AND DISCUSSION

We have carried out comprehensive studies on the physico–chemical properties of a large number of diversified compounds. In this respect, the applicability of the proposed index has been checked for a wide range of properties including: partition coefficient, molar refraction, molar volume, parachor, polarizability, standard enthalpy of formation, toxicity,boiling point, heat capacity, refractive index, and Gibbs free energy. On the other hand, our analysis was based on different category of compounds such as: alkanes, cycloalkanes, silicon/titanium halides, methyl halides, alcohols, aldehydes and ketones, carboxylic acids, as well as isomeric systems. The values of $M$–index were computed for each compound with a view to study their correlation potential in developing QSPR/QSAR models.

### 3.1. UN–BRANCHED ALKANES/CYCLOALKANES

In the first test, the proposed $M$–index is evaluated for the prediction of some physico–chemical properties of a series of simple un–branched alkanes and cycloalkanes. Many properties of alkanes vary in a regular manner with molecular mass and because the alkanes are nonpolar, complexities due to polarity, polarizability, and hydrogen bonding are avoided. Thus, the physico–chemical properties of alkanes are dominated by their inherent structural features, such as molecular dimension or shape. Here, we are mainly concerned with the size effect and consider hydrogen–depleted graphs, i.e. we do not take into account the hydrogen atoms as vertices of the graph.

Table 1 collects the calculated *M*–index for 19 un–branched alkanes together with the experimental data for the logarithm of partition coefficient in octanol/water ($\log p$), molar refraction (*MR*), molar volume (*MV*), parachor (*PR*), and polarizability ($\alpha$) [30]. For all studied properties, we have reported the correlation coefficient ($R$) and the standard error ($s$). Moreover, the resultant models were validated for generalization and productivity by leave–one-out cross–validation (LOO–CV) method, and calculated $R_{cv}$ and $s_{sv}$ are also given. The statistical significance of the obtained models was confirmed by a high $R$ with a close $R_{cv}$ in value and a small $s$ with a close $s_{sv}$ in value.

**Table 1.** The calculated $M$–index and the experimental values of five physico–chemical properties for un–branched alkanes.[a]

| Compd. | $M$ | $\log p$ | $MR$ | $MV$ | $PR$ | $\alpha$ |
|---|---|---|---|---|---|---|
| Methane | 0 | 1.53 | 11.31 | 61.50 | 111.70 | 4.48 |
| Ethane | 8.23 | 2.06 | 15.94 | 78.00 | 151.50 | 6.32 |
| Propane | 13.90 | 2.60 | 20.58 | 94.50 | 191.20 | 8.15 |
| Butane | 19.68 | 3.14 | 25.21 | 111.00 | 231.00 | 9.99 |
| Pentane | 25.50 | 3.67 | 29.84 | 127.50 | 270.80 | 11.83 |
| Hexane | 31.36 | 4.21 | 34.47 | 144.00 | 310.60 | 13.66 |
| Heptane | 37.23 | 4.74 | 39.11 | 160.50 | 350.40 | 15.50 |
| Octane | 43.11 | 5.28 | 43.74 | 177.00 | 390.20 | 17.34 |
| Nonane | 49.01 | 5.82 | 48.37 | 193.60 | 430.00 | 19.17 |
| Decane | 54.90 | 6.35 | 53.01 | 210.10 | 469.70 | 21.01 |
| Undecane | 60.81 | 6.89 | 57.64 | 226.60 | 509.50 | 22.85 |
| Dodecane | 66.71 | 7.42 | 62.27 | 243.10 | 549.30 | 24.28 |
| Tridecane | 72.62 | 7.96 | 66.90 | 259.60 | 589.10 | 26.52 |
| Tetradecane | 78.53 | 8.50 | 71.54 | 276.10 | 628.90 | 28.36 |
| Pentadecane | 84.45 | 9.03 | 76.17 | 292.60 | 668.70 | 30.19 |
| Hexadecane | 90.36 | 9.57 | 80.80 | 309.10 | 708.40 | 32.03 |
| Heptadecane | 96.28 | 10.10 | 85.44 | 325.60 | 748.20 | 33.87 |
| Octadecane | 102.19 | 10.64 | 90.07 | 342.10 | 788.00 | 35.70 |
| Nonadecane | 108.11 | 11.18 | 94.70 | 358.60 | 825.80 | 37.40 |

[a] Experimental data were taken from [30].

The correlation results for listed properties in Table 1 as well as corresponding statistical quantities are given in Eqs. 7–11. As obvious from these equations, for all five properties, the $R$ values are near to 1 with very small $s$. Moreover, the values of $R_{cv}$ and $s_{sv}$ are very close to $R$ and $s$, and $s_{sv}$ are slightly larger than $s$ indicating that the resulted models are statistically significant and validated for physico–chemical properties of alkanes. Further, there is a good correlativity between $M$–index and the molecular structure. Inspection of the results in Table 1 reveals that while the considered properties increase with the increase in the alkane size, the values of $M$–index increase as well.

$$\log p = 1.382 + 0.090M \tag{7}$$
$$R = 0.9999,\ s = 0.042,\ R_{cv} = 0.9999,\ s_{cv} = 0.049 \quad N = 19$$

$$MR = 10.050 + 0.783M \tag{8}$$
$$R = 0.9999,\ s = 0.354,\ R_{cv} = 0.9999,\ s_{cv} = 0.413 \quad N = 19$$

$$MV = 57 + 2.788M \tag{9}$$
$$R = 0.9999,\ s = 1.264,\ R_{cv} = 0.9999,\ s_{cv} = 1.477 \quad N = 19$$

$$PR = 101.100 + 6.714M \tag{10}$$
$$R = 0.9999,\ s = 3.055,\ R_{cv} = 0.9999,\ s_{cv} = 3.563 \quad N = 19$$

$$\alpha = 3.986 + 0.310M \tag{11}$$
$$R = 0.9998,\ s = 0.168,\ R_{cv} = 0.9998,\ s_{cv} = 0.189 \quad N = 19$$

**Table 2.** The comparison between correlation parameters for modeling physico–chemical properties of alkanes using *Sz*, *PI*, and $M$–indices.[a]

| | $R$ | | | $s$ | | |
|---|---|---|---|---|---|---|
| | *Sz* | *PI* | $M$ | *Sz* | *PI* | $M$ |
| $\log p$ | 0.8586 | 0.9397 | 0.9999 | 1.957 | 1.305 | 0.042 |
| $MR$ | 0.8586 | 0.9397 | 0.9999 | 16.914 | 11.283 | 0.354 |
| $MV$ | 0.8395 | 0.9402 | 0.9999 | 60.194 | 40.100 | 1.264 |
| $PR$ | 0.8586 | 0.9397 | 0.9999 | 145.251 | 96.895 | 3.055 |
| $\alpha$ | 0.8592 | 0.9401 | 0.9998 | 6.691 | 4.457 | 0.168 |

[a] The correlation parameters for *Sz*, *PI* were taken from [30] and those for $M$–index were computed in the present work.

The quality of the obtained correlations will be confirmed by comparison of our results with other indices. For this reason, we compared the correlation coefficients and the standard errors of the considered physico–chemical properties obtained by $M$–index with those previously reported using Szeged (*Sz*) and Padmakar–Ivan (*PI*) indices in Table 2 [30]. The quality parameters presented in Table 2 confirm the superiority of the $M$–index over *Sz* and *PI* for modeling property/activity of alkanes. Inspection of the reported data in Table 2 reveals that for all five properties, the models base on the $M$–index provides significantly less standard errors.

Now, we extend our QSPR models to normal (un–branched) cycloalkanes. Listed in Table 3 are the values of $M$–index and similar experimental physico–chemical properties for considered cycloalkanes [30]. The corresponding linear correlation results are as follows:

$$\log p = 0.148 + 0.090M \tag{12}$$

$$R = 1.0000,\ s = 0.012,\ R_{cv} = 1.0000,\ s_{cv} = 0.013 \quad N = 17$$

$$MR = 1.226 + 0.770M \tag{13}$$

$$R = 1.0000,\ s = 0.122,\ R_{cv} = 1.0000,\ s_{cv} = 0.133 \quad N = 17$$

$$MV = 4.754 + 2.963M \tag{14}$$

$$R = 1.0000,\ s = 0.452,\ R_{cv} = 1.0000,\ s_{cv} = 0.503 \quad N = 17$$

$$PR = 10.960 + 6.684M \tag{15}$$

$$R = 1.0000,\ s = 0.905,\ R_{cv} = 1.0000,\ s_{cv} = 0.994 \quad N = 17$$

$$\alpha = 0.494 + 0.305M \tag{16}$$

$$R = 1.0000,\ s = 0.045,\ R_{cv} = 1.0000,\ s_{cv} = 0.049 \quad N = 17$$

The Eqs.12–16 can outstandingly reproduce the physico–chemical properties of cycloalkanes which in turn imply that $M$–index can successfully be applied for cyclic structures as well as non-cyclic alkanes.

### 3.2 INORGANIC COMPOUNDS OF SILICON/TITANIUM HALIDES

To assess the applicability of $M$–index for inorganic compounds, we considered standard enthalpy of formation for a series of silicon/titanium halides with the general formula of $Si(Ti)X_m$ ($X$=$F$, $Cl$, $Br$, $I$ and $m$=1,2,3,4). The calculated $M$–index and the corresponding experimental standard enthalpy of formation for studied systems are given in Table 4 [31,32].

**Table 3.** The calculated $M$–index and the experimental values of five physico–chemical properties for un–branched cycloalkanes.[a]

| Compd. | $M$ | log $p$ | $MR$ | $MV$ | $PR$ | $\alpha$ |
|---|---|---|---|---|---|---|
| Cyclopropane | 16.28 | 1.61 | 13.83 | 53.20 | 120.10 | 5.48 |
| Cyclobutane | 21.96 | 2.14 | 18.44 | 70.90 | 160.10 | 7.31 |
| Cyclopentane | 28.38 | 2.68 | 23.05 | 88.70 | 200.20 | 9.14 |
| Cyclohexane | 34.20 | 3.22 | 27.67 | 106.40 | 240.20 | 10.96 |
| Cycloheptane | 40.43 | 3.75 | 32.28 | 124.10 | 280.20 | 12.79 |
| Cyclooctane | 46.31 | 4.29 | 36.69 | 141.90 | 320.30 | 14.62 |
| Cyclononane | 52.43 | 4.82 | 41.50 | 159.60 | 360.30 | 16.41 |
| Cyclodecane | 58.33 | 5.36 | 46.11 | 177.40 | 400.40 | 18.28 |
| Cycloundecane | 64.40 | 5.90 | 50.72 | 195.10 | 440.40 | 20.11 |
| Cyclododecane | 70.30 | 6.43 | 55.34 | 212.80 | 480.40 | 21.93 |
| Cyclotridecane | 76.34 | 6.97 | 59.95 | 230.60 | 520.50 | 23.76 |
| Cyclotetradecane | 82.25 | 7.50 | 64.54 | 248.30 | 560.50 | 25.59 |
| Cyclopentadecane | 88.26 | 8.04 | 69.17 | 266.00 | 600.60 | 27.42 |
| Cyclohexadecane | 94.18 | 8.58 | 73.78 | 283.80 | 640.60 | 29.25 |
| Cycloheptadecane | 100.05 | 9.11 | 78.39 | 301.50 | 680.60 | 31.07 |
| Cyclooctadecane | 106.09 | 9.65 | 83.01 | 319.30 | 720.70 | 32.90 |
| Cyclononadecane | 112.07 | 10.18 | 87.67 | 337.60 | 760.70 | 34.73 |

[a] Experimental data were taken from [30].

The linear correlation equations are given in Eqs. 17 and 18. For both $SiX_m$ and $TiX_m$ series, there are good linear correlations with $R$ equals to 0.9614 and 0.9631, respectively.

$$\Delta_f H^0_{Si} = 658.100 - 45.720M \tag{17}$$

$$R = 0.9531,\ s = 156,\ R_{cv} = 0.9374,\ s_{cv} = 160 \quad N = 16$$

$$\Delta_f H^0_{Ti} = 583.900 - 51.290M \tag{18}$$

$$R = 0.9586,\ s = 146,\ R_{cv} = 0.9489,\ s_{cv} = 146 \quad N = 16$$

**Table 4.** The calculated $M$–index and the experimental standard enthalpy of formation, $\Delta_f H^0$ (kJ/mol), for silicon/titanium halides.[a]

| Compd. | $M$ | $\Delta_f H^0$ | Compd. | $M$ | $\Delta_f H^0$ |
|---|---|---|---|---|---|
| SiF | 12.77 | -20.92 | TiF | 11.18 | -66.90 |
| $SiF_2$ | 23.27 | -589.94 | $TiF_2$ | 20.50 | -688.30 |
| $SiF_3$ | 34.03 | -999.98 | $TiF_3$ | 30.07 | -1188.70 |
| $SiF_4$ | 45.03 | -1625.90 | $TiF_4$ | 39.88 | -1551.40 |
| SiCl | 9.07 | 154.81 | TiCl | 8.27 | 154.40 |
| $SiCl_2$ | 16.12 | -167.78 | $TiCl_2$ | 14.88 | -282.40 |
| $SiCl_3$ | 23.42 | -334.72 | $TiCl_3$ | 21.73 | -539.30 |
| $SiCl_4$ | 30.97 | -662.75 | $TiCl_4$ | 28.84 | -763.20 |
| SiBr | 8.10 | 196.65 | TiBr | 7.48 | 212.50 |
| $SiBr_2$ | 14.26 | -46.02 | $TiBr_2$ | 13.38 | -179.10 |
| $SiBr_3$ | 20.67 | -158.99 | $TiBr_3$ | 19.52 | -374.90 |
| $SiBr_4$ | 27.33 | -415.47 | $TiBr_4$ | 25.91 | -550.20 |
| SiI | 7.11 | 259.41 | TiI | 6.63 | 274.10 |
| $SiI_2$ | 12.35 | 92.05 | $TiI_2$ | 11.75 | -57.70 |
| $SiI_3$ | 17.85 | 58.58 | $TiI_3$ | 17.11 | -149.80 |
| $SiI_4$ | 23.59 | -110.46 | $TiI_4$ | 22.72 | -287.00 |

[a] Experimental data were taken from [31,32].

Similarly, for the whole set containing all halides of Table 4 there is a good correlation result as well. By taking into account the 32 compounds as a whole, the obtained QSPR equation for the correlation between $M$–index and $\Delta_f H^0$ is as follow

$$\Delta_f H^0_{Si+Ti} = 601.200 - 47.380M \tag{19}$$

$$R = 0.9362,\ s = 176,\ R_{cv} = 0.9281,\ s_{cv} = 168 \quad N = 32$$

The overall indication of these results is that the $M$–index can be applied not only for organic compounds but also for inorganic compounds containing transition metals.

### 3.3 METHYL HALIDES

The usefulness of newly constructed topological indices was demonstrated by correlating standard enthalpy of formation of methyl halides. These sets of compounds contain only five atoms. For such small molecules, hydrogen atoms have great impact on their properties and they cannot be neglected. Thus, the hydrogen atoms must be treated as vertices in the molecular graph. Table 5 contains the calculated *M*–index and the experimental values of the standard enthalpy of formation for 41 methyl halides [33]. By using the linear regression analysis, we obtain

$$\begin{aligned}&\Delta_f H^0 = 1171 - 40.120M\\&R = 0.9940,\ s = 30.000,\qquad R_{cv} = 0.9933,\ s_{cv} = 30.660\quad N = 41\end{aligned} \tag{20}$$

Again, the obtained statistical quantities indicate that there is a good correlation between $M$–index and $\Delta_f H^0$ for methyl halides compounds.

### 3.4 SATURATED ALCOHOLS, KETONES, DIOLS, CARBOXYLIC ACIDS

In this part, we intend to study the applicability of our proposed index to predict relative toxic potency of aliphatic compounds. Here we consider the population growth inhibition of the ciliate *Tetrahymenapyri form* is to develop such a toxicity–based QSAR. *T. pyriformis* is one of the generally used ciliated protozoa [34,35] in which diverse endpoints can be used to originate the cytotoxic effects. The experimental $\log(1/IGC_{50})$ values for the four groups of aliphatic compounds [36] under consideration as well as the resulted QSAR models are provided in Table 6. The statistical parameters reported in Table 6 demonstrate very good consistency between *R* and $R_{cv}$. The small $s$ values indicate that the $M$–index can be successfully used as a way for quantifying toxicity of aliphatic compounds even when they have not exhibit a common skeleton requirement of QSAR analysis.

The suitable quality criteria to judge present results can be set up through the comparison with other theoretical predictions for the toxicity of these molecular sets. Roy et. al. [37] reported results for QSAR calculations on these aliphatic compounds using electrophilicity as a possible descriptor. Their statistical parameters for different molecular set(for instance; diols: *R*=0.899, *s* =0.486 and ketones: *R*=0.882, *s* =0.612) are clearly inferior with respect to present results (diols: *R*=0.9864, *s* =0.192 and ketones: *R*=0.9850, *s* =0.225, Table 6).

.

**Table 5.** The calculated $M$–index and the experimental standard enthalpy of formation, $\Delta_f H^0$ (kJ/mol), for methyl halides.[a]

| Compd. | $M$ | $\Delta_f H^0$ | Compd. | $M$ | $\Delta_f H^0$ |
|---|---|---|---|---|---|
| $CH_2F_2$ | 40.43 | -452.9 | $CH_3Br$ | 28.94 | -37.7 |
| $CCl_2F_2$ | 42.17 | -493.3 | $CBrClF_2$ | 41.04 | -471.5 |
| $CH_2Cl_2$ | 30.94 | -95.4 | $CH_2BrCl$ | 29.81 | -50.2 |
| $CHCl_2F$ | 36.56 | -284.9 | $CHBrClF$ | 35.43 | -295.0 |
| $CBr_2F_2$ | 39.91 | -429.7 | $CH_2BrI$ | 27.59 | 50.2 |
| $CBr_2Cl_2$ | 30.42 | -29.3 | $CHF_3$ | 46.05 | -693.3 |
| $CH_2Br_2$ | 28.68 | -14.8 | $CF_3I$ | 44.70 | -589.9 |
| $CHBr_2F$ | 34.30 | -223.4 | $CCl_3F$ | 37.43 | -284.9 |
| $CHBr_2Cl$ | 29.55 | -20.9 | $CHF_2Cl$ | 41.30 | -483.7 |
| $CBr_2ClF$ | 35.17 | -231.8 | $CF_3Cl$ | 46.92 | -707.9 |
| $CHBr_3$ | 28.42 | 16.7 | $CF_3Br$ | 45.79 | -648.9 |
| $CBr_3F$ | 34.04 | -190.0 | $CH_3F$ | 34.82 | -237.7 |
| $CBr_3Cl$ | 29.29 | 12.6 | $CH_2I_2$ | 26.49 | 118.4 |
| $CHI_3$ | 25.14 | 210.9 | $CH_2FCl$ | 35.69 | -264.4 |
| $CF_4$ | 51.67 | -933.0 | $CH_2ClI$ | 28.72 | 12.6 |
| $CCl_4$ | 32.68 | -95.8 | $CH_3I$ | 27.84 | 13.8 |
| $CI_4$ | 23.79 | 262.9 | $CH_2FBr$ | 34.56 | -252.7 |
| $CBr_4$ | 28.16 | 79.5 | $CH_4$ | 29.20 | -74.9 |
| $CHBrF_2$ | 40.17 | -463.6 | $CBrCl_3$ | 31.55 | -37.2 |
| $CHBrCl_2$ | 30.68 | -58.6 | $CHCl_3$ | 31.81 | -102.9 |
| $CBrCl_2F$ | 36.30 | -269.4 | | | |

[a] Experimental data were taken from [33].

An alternative manner for predicting the toxicity is utilizing topological parameters derived from the electron density, as previously done in our research group [12]. As shown in Ref. [12], predictions improve significantly with respect to the results obtained by Roy et al. [37]. However, despite its successful in predicting the toxicity, such approach is very computational demanding because it is necessary to perform quantum chemical calculation on each molecule and to derive the topological properties of the electron densities from the calculated wave functions. On the contrary, the most important advantage of the proposed $M$–index is its computational ease with no need to complicated calculations.

**Table 6.** The calculated $M$–index and the experimental values of the toxicity, $\log 1/IGC_{50}$, for aliphatic compounds.$^{a}$

| Compd. | $M$ | $\log 1/IGC_{50}$ | Correlation Equation $^{b}$ |
|---|---|---|---|
| *Saturated alcohols* | | | |
| 1-Propanol | 83.61 | -1.7464 | |
| 2-Propanol | 83.67 | -1.8819 | |
| 1-Butanol | 104.64 | -1.4306 | |
| (±)-2-Butanol | 104.74 | -1.5420 | |
| 2-Methyl-1-propanol | 104.87 | -1.3724 | |
| 2-Pentanol | 125.93 | -1.1596 | |
| 3-Pentanol | 125.97 | -1.2437 | |
| 3-Methyl-2-butanol | 126.26 | -0.9959 | |
| 2-Methyl-1-butanol | 126.16 | -0.9528 | |
| 3-Methyl-1-butanol | 126.13 | -1.0359 | $\log 1/IGC_{50} = -3.721 + 0.022M$ |
| 2,2-Dimethyl-1-propanol | 126.58 | -0.8702 | (21) |
| 2-Methyl-2-propanol | 104.99 | -1.7911 | $R = 0.9932,\ s = 0.164$, |
| 1-Hexanol | 147.09 | -0.3789 | $R_{cv} = 0.9914,\ s_{cv} = 0.176$ N = 21 |
| 3,3-Dimethyl-1-butanol | 148.06 | -0.7368 | |
| 1-Heptanol | 168.44 | 0.1050 | |
| 1-Octanol | 189.83 | 0.5827 | |
| 1-Nonanol | 211.26 | 0.8551 | |
| 1-Decano | 232.72 | 1.3354 | |
| 1-Undecanol | 254.20 | 1.9547 | |

| | | | |
|---|---|---|---|
| 1-Dodecanol | 275.70 | 2.1612 | |
| 1-Tridecanol | 297.22 | 2.4497 | |
| *Ketones* | | | |
| Acetone | 64.52 | -2.2036 | |
| 2-Butanone | 85.27 | -1.7457 | |
| 2-Pentanone | 106.29 | -1.2224 | |
| 3-Pentanone | 106.19 | -1.4561 | |
| 4-Methyl-2-pentanone | 127.75 | -1.2085 | |
| 2-Heptanone | 148.73 | -0.4872 | |
| 5-Methyl-2-hexanone | 149.08 | -0.6459 | $\log 1/IGC_{50} = -3.446 + 0.020M$ |
| 4-Heptanone | 148.51 | -0.6690 | (22) |
| 2-Octanone | 170.07 | -0.1455 | $R = 0.9850,\ s = 0.225$ |
| 2-Nonanone | 191.46 | 0.6598 | $R_{cv} = 0.9779,\ s_{cv} = 0.25 \quad \mathrm{N} = 15$ |
| 2-Decanone | 212.89 | 0.5822 | |
| 3-Decanone | 212.69 | 0.6265 | |
| 2-Undecanone | 234.35 | 1.5346 | |
| 2-Dodecanone | 255.84 | 1.6696 | |
| 7-Tridecanone | 276.90 | 1.5214 | |
| *Alcohols: diols* | | | |
| (±)-1,2-Butanediol | 118.23 | -2.0482 | |
| (±)-1,3-Butanediol | 118.22 | -2.3013 | |
| 1,4-Butanediol | 118.12 | -2.2365 | $\log 1/IGC_{50} = -4.709 + 0.021M$ |
| 1,2-Pentanediol | 139.46 | -1.6269 | (23) |
| 1,5-Pentanediol | 139.33 | -1.9344 | $R = 0.9864,\ s = 0.192$ |
| (±)-1,2-Hexanediol | 160.77 | -1.2669 | $R_{cv} = 0.9653,\ s_{cv} = 0.269 \quad \mathrm{N} = 9$ |
| 1,6-Hexanediol | 160.62 | -1.4946 | |
| 1,2-Decanediol | 246.46 | 0.7640 | |
| 1,10-Decanediol | 246.30 | 0.2240 | |
| *Carboxylic acids* | | | |

| | | | |
|---|---|---|---|
| Propanoic acid | 78.00 | -0.5123 | |
| Butyric acid | 98.91 | -0.5720 | |
| Valeric acid | 120.01 | -0.2674 | |
| Hexanoic acid | 141.23 | -0.2083 | |
| Heptanoic acid | 162.53 | -0.1126 | |
| Octanoic acid | 183.90 | 0.0807 | |
| Nonanoic acid | 181.25 | 0.3509 | |
| Decanoic acid | 226.75 | 0.5063 | $\log 1/IGC_{50} = -1.181 + 0.007M$ (24) |
| Undecanoic acid | 248.22 | 0.8983 | $R = 0.9429,\ s = 0.124$ |
| Iso-Butyric acid | 99.04 | -0.3334 | $R_{cv} = 0.9197,\ s_{cv} = 0.139$ $N = 20$ |
| Isovalerianic acid | 120.26 | -0.3415 | |
| Trimethylacetic acid | 120.52 | -0.2543 | |
| 3-Methylvaleric acid | 141.62 | -0.2331 | |
| 4-Methylvaleric acid | 141.55 | -0.2724 | |
| 2-Ethylbutyric acid | 141.52 | -0.1523 | |
| 2-Propylpentanoic acid | 184.33 | 0.0258 | |
| 2-Ethylhexanoic acid | 184.29 | 0.0756 | |
| Crotonic acid | 81.91 | -0.5448 | |
| trans-2-Pentenoic acid | 102.55 | -0.2774 | |
| trans-2-Hexenoic acid | 123.50 | -0.1279 | |

[a] Experimental data were taken from [36].

[b] For each series, the correlation equations and statistical quantities were presented in the last column.

### 3.5. KETONES AND ALDEHYDES

Molar refraction (MR) is a particularly useful physical parameter in chemistry, biological chemistry, and pharmaceutical science because it is closely related to the bulkiness and polarizability of a molecule. We have checked the ability of $M$–index to predict the molar refraction of a set containing 22 aldehydes and 24 ketones listed in Table 7 [38]. The relationship between $MR$ and $M$–index is give below

$$MR = 2.223 + 0.217M \tag{25}$$
$$R = 0.9997,\ s = 0.262,\ R_{cv} = 0.9997,\ s_{cv} = 0.266$$

This linear equation indicates an outstanding correlation with high $R$ and small $s$, implying the reliability of $M$–index for prediction of molar refraction.

**Table 7.** The calculated $M$–index and the experimental molar refraction for aldehydes and ketones.[a]

| Compd. | *M* | *MR* | Compd. | *M* | *MR* |
|---|---|---|---|---|---|
| Acetaldehyde | 44.24 | 11.5829 | 2-Butanone | 85.27 | 20.6039 |
| Propionaldehyde | 64.74 | 16.1632 | 2-Pentanone | 106.29 | 25.2926 |
| Butyl aldehyde | 85.60 | 20.8011 | 3-Pentanone | 106.19 | 25.2487 |
| 2-Methyl propanal | 87.31 | 20.8219 | 3-Methyl-2-butanone | 102.74 | 25.2603 |
| Pentaldehyde | 106.66 | 25.4983 | 2-Hexanone | 127.46 | 29.9308 |
| 2-Methyl butanal | 106.80 | 25.3943 | 3-Hexanone | 127.31 | 29.7251 |
| 3-Methyl butanal | 106.90 | 25.5327 | 3-Methyl-2-pentanone | 127.75 | 29.9453 |
| Hexanal | 127.85 | 30.9280 | 4-Methyl-2-pentanone | 127.75 | 29.9877 |
| 2-Methylpentanal | 128.02 | 29.8497 | 3,3-Dimethyl-2-butanone | 128.14 | 29.6748 |
| 2-Ethylbutanal | 128.07 | 29.9981 | 2-Heptanone | 148.73 | 34.5663 |
| 2,3-Dimethylbutanal | 128.36 | 30.0640 | 3-Heptanone | 148.56 | 34.4230 |
| Heptanal | 149.14 | 34.7004 | 4-Heptanone | 148.51 | 34.3083 |
| 2,2-Dimethylpentanal | 149.84 | 34.7537 | 5-Methyl-2-hexanone | 149.08 | 34.5773 |
| Octanal | 170.49 | 39.4396 | 2-Octanone | 170.07 | 39.1959 |
| 2-Ethylhexanal | 170.78 | 39.2395 | 5-Octanone | 169.81 | 39.0616 |
| 2-Ethyl-3- | 171.38 | 38.9423 | 6-Methyl-3- | 170.25 | 38.9478 |

| | | | | | |
|---|---|---|---|---|---|
| methylpentanal | | | heptanone | | |
| Nonanal | 191.89 | 44.2669 | 2-Nonanone | 191.46 | 43.3542 |
| 3,5,5-Trimethylhexanal | 193.64 | 43.9887 | 5-Nonanone | 191.15 | 43.8710 |
| Decanal | 213.33 | 48.6737 | 2,6-Dimethyl-4-heptanone | 191.88 | 43.8902 |
| 2-Methyldecanal | 235.00 | 53.0003 | 2-Decanone | 212.89 | 48.5304 |
| Dodecanal | 256.28 | 58.0913 | 2-Undecanone | 234.35 | 52.7129 |
| 2-Methylundecanal | 256.49 | 57.9284 | 6-Undecanone | 233.97 | 53.2109 |
| Acetone | 64.52 | 16.2963 | 2-Methyl-4-undecanone | 255.83 | 57.7027 |

[a]Experimental data were taken from [38].

### 3.6. ISOMERIC SYSTEMS

One of the main drawbacks of the most topological indices is their poor discrimination of isomers and the index has the same value for different isomeric compounds. It is well–known that this degeneracy increases when the number of atoms in the molecule increases, even for simple molecules such as alkanes. In previous studies the capacity of one index to discriminate isomers was measured by using a discrimination index, *D*, which has been calculated as the number of isomers having different values of the index divided by the total number of isomers [39].

Similar to other descriptors, the $M$–index introduced in the present work requires some modifications to be applicable for isomers. Theproposed $M$–index in Eq. 6 was constructed by parameters which are only dependent on the number of atoms, bond strength and the property of individual atoms. Thus, the calculated values of the $M$–index for the structural isomers of a specific molecular formula are very close together and cannot well discriminate the isomers. To increase the discriminative power of the $M$–index, we introduce a quantity in which the effect of different configurations due to structural isomers is taken into account. For the isomeric systems, we first calculate $q$ as

$$q_i = \sum_{d=1} \frac{n_i(d)}{d}, \tag{26}$$

where $d$ is the distance and $n_i(d)$ is the number of neighboring nodes in the distance $d$ of the vertex $i$. Finally, we defined the discrimination parameter, $\tilde{D}$ used in the present study as

$$\tilde{D} = \left( \frac{\sum_{i=1}^{N} q_i}{N} \right), \tag{27}$$

where $N$ is the total number of vertices in the molecular graph. Finally, the appropriate form of our proposed index designed for isomeric systems is $\tilde{D}^{\beta} M$, where $\beta$ is a free parameter dependent on the considered property.

The predictive ability of $M$–index modified by discrimination parameter has been checked within two sets of isomeric compounds. First, a large set of 88 aldehydes and ketones and their boiling points [40−42] (Table 8) was taken to construct model. For such a data set, the correlative model was obtained by setting $\beta = -1$ as following equation

$$Bp = -172.800 + 15.590 \tilde{D}^{-1} M \tag{28}$$

$$R = 0.9922,\ s = 7.557,\ R_{cv} = 0.9918,\ s_{cv} = 7.647 \quad N = 88.$$

**Table 8.** The calculated $M$–index and the experimental boiling points (°C) for aldehydes and ketones. [a]

| Compd. | $Bp$ | $\tilde{D}^{\beta} M$ | Compd. | $Bp$ | $\tilde{D}^{\beta} M$ |
|---|---|---|---|---|---|
| Acetaldehyde | 20.8 | 12.39 | 5-Methyl-2-hexanone | 144.0 | 20.02 |
| Propionaldehyde | 48.8 | 14.07 | 2-Methyl-3-hexanone | 135.0 | 20.10 |
| Butyl aldehyde | 75.7 | 15.69 | 4-Methyl-3-hexanone | 134.0 | 19.99 |
| 2-Methyl propanal | 64.4 | 15.87 | 5-Methyl-3-hexanone | 135.0 | 20.08 |
| Pentaldehyde | 103.0 | 17.24 | 2,2-Dimethyl-3-pentanone | 125.0 | 19.75 |
| 2-Methyl butanal | 92.5 | 17.09 | 2,4-Dimethyl-3-pentanone | 125.0 | 19.89 |
| 3-Methyl butanal | 92.5 | 17.05 | 4,4-Dimethyl-3-pentanone | 126.0 | 19.75 |
| 2,2-Dimethylpropanal | 77.5 | 16.88 | 2-Octanone | 172.5 | 21.59 |
| Hexanal | 128.0 | 18.71 | 3-Octanone | 167.5 | 21.65 |

| | | | | | |
|---|---|---|---|---|---|
| 2-Methylpentanal | 117.0 | 18.55 | 4-Octanone | 165.5 | 21.68 |
| 3-Methylpentanal | 118.0 | 18.46 | 2-Methyl-4-heptanone | 154.0 | 21.47 |
| 2-Ethylbutanal | 117.0 | 18.50 | 3-Methyl-4-heptanone | 153.0 | 21.37 |
| Heptanal | 152.8 | 20.13 | 3-Methyl-2-heptanone | 164.0 | 21.31 |
| 3-Methylhexanal | 143.0 | 19.85 | 6-Methyl-2-heptanone | 167.0 | 21.38 |
| 2,2-Dimethylpentanal | 126.5 | 19.66 | 2-Methyl-3-heptanone | 158.0 | 21.48 |
| Octanal | 171.0 | 21.50 | 3,3-Dimethyl-2-hexanone | 151.5 | 20.92 |
| 2-Ethylhexanal | 160.0 | 21.26 | 2,2-Dimethyl-3-hexanone | 146.0 | 21.03 |
| 2-Propylpentanal | 160.0 | 21.23 | 2,5-Dimethyl-3-hexanone | 147.5 | 21.27 |
| Nonanal | 191.0 | 22.83 | 4,4-Dimethyl-3-hexanone | 148.0 | 20.95 |
| 3,5,5-Trimethylhexanal | 170.5 | 21.96 | 2,2,4-Trimethyl-3-pentanone | 135.1 | 20.91 |
| Decanal | 208.5 | 24.12 | 2-Nonanone | 195.0 | 22.91 |
| Undecanal | 233.0 | 25.38 | 3-Nonanone | 190.0 | 23.28 |
| 2-Methyldecanal | 229.0 | 24.33 | 4-Nonanone | 187.5 | 23.31 |
| Dodecanal | 254.0 | 26.61 | 5-Nonanone | 188.4 | 23.02 |
| 2-Methylundecanal | 246.0 | 26.48 | 7-Methyl-3-octanone | 182.5 | 22.77 |
| Tridecanal | 267.0 | 27.82 | 3-Methyl-4-octanone | 174.0 | 22.70 |
| Tetradecanal | 287.0 | 29.00 | 7-Methyl-4-octanone | 178.0 | 22.80 |
| Pentadecanal | 304.0 | 30.17 | 2,6-Dimethyl-4-heptanone | 169.4 | 22.61 |
| Acetone | 56.2 | 14.18 | 3,5-Dimethyl-4-heptanone | 162.0 | 22.37 |
| 2-Butanone | 79.6 | 15.81 | 2,2,4,4-Tetramethyl-3-pentanone | 152.0 | 21.89 |
| 2-Pentanone | 102.0 | 17.35 | 2-Decanone | 210.0 | 24.20 |
| 3-Pentanone | 101.7 | 17.39 | 3-Decanone | 211.0 | 24.26 |
| 3-Methyl-2-butanone | 93.5 | 16.57 | 4-Decanone | 206.5 | 24.29 |
| 2-Hexanone | 127.6 | 18.82 | 2-Undecanone | 231.5 | 25.45 |
| 3-Hexanone | 123.5 | 18.87 | 3-Undecanone | 227.0 | 25.51 |

| | | | | | |
|---|---|---|---|---|---|
| 3-Methyl-2-pentanone | 118.0 | 18.57 | 5-Undecanone | 227.0 | 25.57 |
| 4-Methyl-2-pentanone | 117.0 | 18.61 | 6-Undecanone | 226.0 | 25.58 |
| 2-Methyl-3-pentanone | 115.0 | 18.75 | 2-Dodecanone | 246.5 | 26.68 |
| 3,3-Dimethyl-2-butanone | 106.0 | 18.33 | 2-Tridecanone | 263.0 | 27.88 |
| 2-Heptanone | 151.4 | 20.23 | 7-Tridecanone | 261.0 | 28.02 |
| 3-Heptanone | 147.0 | 20.29 | 2-Methyl-3-tridecanone | 267.0 | 28.99 |
| 4-Heptanone | 144.0 | 20.31 | 7-Ethyl-2-methyl-4-undecanone | 252.5 | 28.29 |
| 3-Methyl-2-hexanone | 143.5 | 19.95 | 2-Pentadecanone | 294.0 | 30.23 |
| 4-Methyl-2-hexanone | 139.0 | 19.93 | 8-Pentadecanone | 291.0 | 30.37 |

[a] Experimental data were taken from [40−42].

In this model, there is a high $R$(0.9922) and a small $s$ (7.557) value. The values of $R_{cv}$ (0.9918) and $s_{cv}$ (7.647) are very close to the values of $R$ and $s$, and $s_{cv}$ are only slightly larger than $s$. The correlation results and the cross-validation results demonstrate that the obtained model is statistically significant and validated.

In the second attempt, six physico–chemical properties of 77 hydrocarbons [43] (Table 9) including isomers have been studied. For heat capacity and boiling point $\beta$ is zero, while for density, refractive index, Gibbs free energy, and the standard enthalpy of formation the obtained values for $\beta$ are, respectively, 2, 2, 3.5, and −0.9. The correlation equations for these six properties are given below

$$C_p = 20 + 3.851M \tag{29}$$

$$R = 0.9899,\ s = 3.886,\ R_{cv} = 0.9894,\ s_{cv} = 3.926\quad N = 77$$

$$Bp = -66.810 + 4.068M \tag{30}$$

$$R = 0.9849,\ s = 5.035,\ R_{cv} = 0.9838,\ s_{cv} = 5.145\quad N = 77$$

$$\rho = 633.5 + 0.111\tilde{D}^2 M \tag{31}$$

$$R = 0.9495,\ s = 8.291,\ R_{cv} = 0.9460,\ s_{cv} = 8.455\quad N = 77$$

$$RI = 1.360 + 5.752e - 5\tilde{D}^2 M \tag{32}$$

$$R = 0.9584,\ s = 3.866e-3,\ R_{cv} = 0.9553,\ s_{cv} = 4.010e-3 \quad N = 77$$

$$\Delta G = -7.522 + 5.171e-3\tilde{D}^{3.5}M \tag{33}$$
$$R = 0.9610,\ s = 3.675,\ R_{cv} = 0.9586,\ s_{cv} = 3.735 \quad N = 77$$

$$\Delta_f H^0 = -24.180 + 4.305\tilde{D}^{-0.9}M \tag{34}$$
$$R = 0.9819,\ s = 0.923,\ R_{cv} = 0.9810,\ s_{cv} = 0.935 \quad N = 77$$

As can be seen, all equations have acceptable quality and the propose index used in these equations can explain more than 95% of the variance in the considered physico–chemical properties. The high correlation coefficients of cross validation show that the constructed models are statistically significant.

**Table 9.** The calculated $M$–index and the experimental heat capacity ($C_p$, J/mol.K), boiling point ($Bp$, °C), density ($\rho$, kg/m$^3$), refractive index ($RI$), Gibbs free energy ($\Delta G$, kJ/mol), and the standard enthalpy of formation ($\Delta_f H^0$, kJ/mol) for hydrocarbons.[a]

| Compd. | $C_p$ | $Bp$ | $\tilde{D}^{\beta}M$ | $\rho$ | $RI$ | $\tilde{D}^{\beta}M$ | $\Delta G$ | $\tilde{D}^{\beta}M$ | $\Delta_f H^0$ | $\tilde{D}^{\beta}M$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 3-Methylpentane | 140.88 | 63.28 | 31.63 | 659.76 | 1.3739 | 289.95 | -2.12 | 1527.59 | 26.32 | 11.67 |
| 2,2-Dimethylbutane | 142.26 | 49.74 | 31.98 | 644.46 | 1.3660 | 320.67 | -7.42 | 1807.01 | 25.40 | 11.33 |
| 2,3-Dimethylbutane | 140.21 | 57.99 | 31.83 | 657.02 | 1.3723 | 308.06 | -1.77 | 1690.50 | 24.77 | 11.46 |
| 3-Methylhexane | 164.50 | 91.85 | 37.53 | 682.88 | 1.3861 | 413.46 | 6.60 | 2500.07 | 30.71 | 12.75 |
| 3-Ethylpentane | 166.80 | 93.48 | 37.60 | 693.92 | 1.3911 | 423.81 | 12.70 | 2606.91 | 31.71 | 12.64 |
| 2,2-Dimethylpentane | 167.70 | 79.17 | 37.90 | 669.48 | 1.3800 | 451.77 | 2.10 | 2898.00 | 29.50 | 12.43 |
| 2,3-Dimethylpentane | 161.80 | 89.75 | 37.80 | 690.81 | 1.3895 | 444.39 | 7.60 | 2821.19 | 28.62 | 12.47 |
| 2,4-Dimethylpentane | 171.70 | 80.47 | 37.70 | 668.23 | 1.3788 | 430.98 | 4.90 | 2679.30 | 29.58 | 12.60 |
| 3,3-Dimethylpentane | 166.70 | 86.04 | 38.00 | 689.16 | 1.3884 | 465.55 | 4.80 | 3048.35 | 29.33 | 12.31 |
| 2,2,3-Trimethylbutane | 164.20 | 80.86 | 38.20 | 685.64 | 1.3869 | 487.29 | 6.30 | 3288.92 | 28.28 | 12.15 |
| N-Octane | 188.70 | 125.68 | 43.11 | 698.54 | 1.3951 | 508.91 | 17.67 | 3240.89 | 38.12 | 14.20 |
| 2-Methylheptane | 188.20 | 117.65 | 43.35 | 693.87 | 1.3926 | 538.66 | 13.37 | 3564.95 | 35.82 | 13.95 |
| 3-Methylheptane | 186.82 | 118.93 | 43.43 | 701.73 | 1.3961 | 552.51 | 13.79 | 3721.67 | 35.31 | 13.83 |
| 2,4-Dimethylhexane | 193.35 | 109.43 | 43.69 | 696.17 | 1.3929 | 586.01 | 13.07 | 4107.42 | 33.76 | 13.58 |
| 2,5-Dimethylhexane | 186.52 | 109.11 | 43.59 | 689.37 | 1.3900 | 570.22 | 11.40 | 3921.96 | 33.39 | 13.71 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3,3-Dimethylhexane | 191.96 | 111.97 | 43.96 | 705.95 | 1.3978 | 620.91 | 15.13 | 4523.98 | 33.43 | 13.35 |
| 3,4-Dimethylhexane | 182.72 | 117.73 | 43.81 | 715.15 | 1.4018 | 605.15 | 18.43 | 4336.01 | 32.47 | 13.44 |
| 3-Ethyl-2-Methylpentan | 193.05 | 115.66 | 43.83 | 715.20 | 1.4017 | 609.52 | 20.68 | 4389.42 | 34.31 | 13.41 |
| 2,2,3-Trimethylpentane | 186.77 | 109.84 | 44.23 | 712.03 | 1.4007 | 657.01 | 19.45 | 4971.27 | 32.13 | 13.13 |
| 2,3,3-Trimethylpentane | 188.20 | 114.77 | 44.28 | 722.30 | 1.4052 | 664.88 | 20.04 | 5071.69 | 32.17 | 13.08 |
| 2,3,4-Trimethylpentane | 192.72 | 113.47 | 44.03 | 715.09 | 1.4020 | 633.00 | 20.76 | 4673.57 | 32.55 | 13.27 |
| 2,2,3,3-Tetramethylbutane | 188.28 | 106.29 | 44.68 | 729.88 | 1.4057 | 714.87 | 24.04 | 5718.96 | 31.84 | 12.83 |
| 2-Methyloctane | 210.90 | 143.28 | 49.25 | 709.60 | 1.4008 | 689.31 | 21.60 | 4988.20 | 40.42 | 15.02 |
| 3-Methyloctane | 209.70 | 144.23 | 49.33 | 716.70 | 1.4040 | 706.26 | 22.00 | 5197.81 | 39.92 | 14.89 |
| 4-Ethylheptane | 214.30 | 141.20 | 49.49 | 722.30 | 1.4067 | 738.47 | 26.80 | 5606.79 | 40.50 | 14.66 |
| 2,2-Dimethylheptane | 212.40 | 132.82 | 49.72 | 706.60 | 1.3995 | 756.29 | 19.50 | 5824.82 | 38.83 | 14.61 |
| 2,3-Dimethylheptane | 207.70 | 140.50 | 49.66 | 722.00 | 1.4064 | 755.25 | 23.50 | 5816.87 | 37.82 | 14.59 |
| 2,4-Dimethylheptane | 217.10 | 133.20 | 49.63 | 711.50 | 1.4011 | 751.95 | 20.80 | 5774.96 | 38.16 | 14.60 |
| 2,5-Dimethylheptane | 208.20 | 136.00 | 49.58 | 713.60 | 1.4015 | 742.75 | 18.20 | 5655.51 | 37.53 | 14.67 |
| 2,6-Dimethylheptane | 210.40 | 135.22 | 49.49 | 704.50 | 1.3985 | 724.43 | 19.80 | 5421.28 | 37.99 | 14.79 |
| 3,3-Dimethylheptane | 214.00 | 137.02 | 49.89 | 721.60 | 1.4063 | 787.88 | 22.00 | 6241.86 | 38.20 | 14.41 |
| 3,4-Dimethylheptane | 206.80 | 140.40 | 49.76 | 727.50 | 1.4091 | 775.63 | 24.90 | 6084.78 | 37.02 | 14.46 |
| 3,5-Dimethylheptane | 214.60 | 135.70 | 49.69 | 716.60 | 1.4046 | 762.93 | 22.00 | 5917.94 | 38.07 | 14.54 |
| 3-Ethyl-3-Methylhexane | 214.10 | 140.60 | 50.06 | 736.00 | 1.4134 | 821.89 | 30.50 | 6703.42 | 37.36 | 14.21 |
| 4-Ethyl-2-Methylhexane | 219.70 | 133.80 | 49.72 | 724.20 | 1.4054 | 770.66 | 24.50 | 6020.30 | 39.25 | 14.48 |
| 2,2,4-Trimethylhexane | 210.70 | 129.91 | 50.09 | 711.80 | 1.4010 | 817.86 | 23.60 | 6643.10 | 36.61 | 14.25 |
| 2,3,3-Trimethylhexane | 213.30 | 137.69 | 50.26 | 733.50 | 1.4119 | 847.96 | 29.40 | 7058.75 | 36.28 | 14.09 |
| 2,3,4-Trimethylhexane | 214.00 | 138.96 | 50.06 | 735.10 | 1.4120 | 821.89 | 28.60 | 6703.42 | 36.86 | 14.21 |
| 2,3,5-Trimethylhexane | 212.50 | 131.36 | 49.92 | 717.90 | 1.4037 | 795.75 | 22.20 | 6348.35 | 36.02 | 14.36 |
| 2,4,4-Trimethylhexan | 213.50 | 130.66 | 50.16 | 720.05 | 1.4052 | 831.08 | 26.60 | 6824.87 | 36.44 | 14.18 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3-Ethyl-2,2-Dimethylpentan | 205.00 | 133.84 | 50.31 | 731.00 | 1.4101 | 857.89 | 37.50 | 7199.46 | 36.16 | 14.04 |
| 2,2,3,3-Tetramethylpentane | 213.34 | 140.29 | 50.81 | 752.97 | 1.4214 | 929.70 | 39.00 | 8225.65 | 35.86 | 13.73 |
| 2,2,3,4-Tetramethylpentane | 208.50 | 133.03 | 50.51 | 735.22 | 1.4125 | 884.63 | 36.70 | 7574.18 | 35.06 | 13.92 |
| 3-Ethyloctane | 235.80 | 166.50 | 55.37 | 735.90 | 1.4136 | 903.92 | 34.90 | 7341.39 | 45.31 | 15.76 |
| 4-Ethyloctane | 236.50 | 163.64 | 55.42 | 734.30 | 1.4131 | 917.34 | 33.40 | 7528.24 | 45.10 | 15.67 |
| 2,2-Dimethyloctane | 235.10 | 156.90 | 55.63 | 720.80 | 1.4060 | 928.65 | 27.70 | 7669.23 | 43.43 | 15.67 |
| 2,5-Dimethyloctane | 231.80 | 158.50 | 55.53 | 726.40 | 1.4089 | 924.58 | 26.90 | 7620.98 | 41.92 | 15.66 |
| 3,4-Dimethyloctane | 229.30 | 163.40 | 55.69 | 741.80 | 1.4159 | 956.90 | 33.00 | 8075.93 | 41.80 | 15.49 |
| 3,5-Dimethyloctane | 238.30 | 159.40 | 55.64 | 732.90 | 1.4115 | 948.37 | 29.10 | 7955.70 | 42.47 | 15.53 |
| 3,6-Dimethyloctane | 229.60 | 160.80 | 55.59 | 732.90 | 1.4115 | 935.26 | 28.90 | 7769.81 | 41.63 | 15.61 |
| 4,4-Dimethyloctane | 239.30 | 157.50 | 55.87 | 731.20 | 1.4122 | 983.33 | 31.90 | 8449.58 | 42.30 | 15.37 |
| 4,5-Dimethyloctane | 230.10 | 162.13 | 55.72 | 743.20 | 1.4167 | 965.17 | 35.30 | 8194.87 | 41.51 | 15.44 |
| 4-N-Propylheptane | 237.70 | 157.50 | 55.46 | 732.10 | 1.4113 | 927.67 | 38.20 | 7673.21 | 44.85 | 15.61 |
| 4-Isopropylheptane | 239.20 | 158.90 | 55.77 | 735.40 | 1.4132 | 979.10 | 37.90 | 8397.48 | 43.10 | 15.36 |
| 2-Methyl-3-Ethylheptane | 238.50 | 161.20 | 55.73 | 739.80 | 1.4151 | 967.48 | 35.70 | 8228.62 | 43.30 | 15.43 |
| 2-Methyl-4-Ethylheptane | 243.40 | 156.20 | 55.69 | 732.20 | 1.4114 | 962.17 | 31.60 | 8153.95 | 43.64 | 15.45 |
| 3-Methyl-4-Ethylheptane | 236.20 | 162.20 | 55.84 | 746.60 | 1.4183 | 992.87 | 36.90 | 8597.02 | 42.47 | 15.29 |
| 3-Methyl-5-Ethylheptane | 240.90 | 158.20 | 55.74 | 736.80 | 1.4141 | 970.77 | 33.10 | 8276.38 | 43.35 | 15.41 |
| 2,2,3-Trimethylheptane | 232.50 | 157.60 | 56.13 | 738.50 | 1.4145 | 1021.76 | 34.80 | 9004.99 | 41.30 | 15.21 |
| 2,3,3-Trimethylheptane | 235.10 | 160.20 | 56.21 | 748.80 | 1.4202 | 1039.31 | 37.30 | 9267.17 | 41.00 | 15.12 |
| 2,3,4-Trimethylheptane | 237.60 | 159.90 | 56.03 | 748.50 | 1.4195 | 1016.82 | 37.20 | 8940.44 | 40.96 | 15.20 |
| 2,3,5-Trimethylheptane | 233.90 | 160.70 | 55.94 | 754.50 | 1.4169 | 997.74 | 30.30 | 8659.63 | 40.12 | 15.30 |
| 2,3,6-Trimethylheptane | 228.50 | 156.00 | 55.82 | 734.70 | 1.4131 | 972.26 | 28.50 | 8289.11 | 39.75 | 15.43 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2,4,5-Trimethylheptane | 234.10 | 156.50 | 55.96 | 737.30 | 1.4160 | 1002.85 | 36.90 | 8734.94 | 39.98 | 15.27 |
| 3,3,5-Trimethylheptane | 234.10 | 155.68 | 56.19 | 739.00 | 1.4170 | 1038.98 | 34.10 | 9264.22 | 40.46 | 15.12 |
| 3,4,4-Trimethylheptane | 235.60 | 161.10 | 56.33 | 753.50 | 1.4235 | 1067.62 | 40.30 | 9697.26 | 40.08 | 14.99 |
| 3,4,5-Trimethylheptane | 235.10 | 162.50 | 56.11 | 751.90 | 1.4229 | 1034.31 | 39.70 | 9201.17 | 41.14 | 15.12 |
| 2-Methyl-3-Isopropylhexane | 231.80 | 166.70 | 56.09 | 743.60 | 1.4172 | 1033.98 | 46.80 | 9198.24 | 40.46 | 15.12 |
| 2,3-Dimethyl-3-Ethylhexane | 238.20 | 163.70 | 56.42 | 759.98 | 1.4247 | 1085.61 | 45.00 | 9974.15 | 40.71 | 14.91 |
| 2,3-Dimethyl-4-Ethylhexane | 243.00 | 160.90 | 56.14 | 751.60 | 1.4203 | 1042.95 | 42.10 | 9332.13 | 42.43 | 15.08 |
| 2,4-Dimethyl-4-Ethylhexane | 235.00 | 160.10 | 56.29 | 751.40 | 1.4202 | 1061.98 | 42.30 | 9612.80 | 40.29 | 15.01 |
| 3,3-Dimethyl-4-Ethylhexane | 228.20 | 162.90 | 56.44 | 759.80 | 1.4246 | 1090.98 | 50.00 | 10057.79 | 39.92 | 14.89 |
| 3,4-Dimethyl-4-Ethylhexane | 235.50 | 162.10 | 56.49 | 759.60 | 1.4244 | 1100.24 | 47.60 | 10200.89 | 40.42 | 14.85 |
| 2,2,3,3-Tetramethylhexane | 238.20 | 160.31 | 56.82 | 760.89 | 1.4260 | 1143.72 | 48.80 | 10869.42 | 40.00 | 14.71 |
| 2,3,3,4-Tetramethylhexan | 241.50 | 164.59 | 56.69 | 765.60 | 1.4298 | 1129.30 | 49.10 | 10648.72 | 40.04 | 14.75 |
| 2,3,4,4-Tetramethylhexane | 231.80 | 161.60 | 56.64 | 758.60 | 1.4267 | 1119.89 | 49.20 | 10500.92 | 38.87 | 14.79 |
| 2,3,4,5-Tetramethylhexane | 243.10 | 156.20 | 56.34 | 745.60 | 1.4204 | 1071.09 | 42.70 | 9751.18 | 40.71 | 14.97 |

[a] Experimental data were taken from [43].

## 4. CONCLUSIONS

We have employed the expanded form of the Wiener index to introduce a novel topological descriptor, named $M$–index, which includes both the atom parameter such as electronegativity and principle quantum number and the bond parameter. In fact, we modified the adjacency matrix for constructing the $M$–index in such a way that not only the characteristics of an individual atom but also the role of that atom in establishing the connection with other atoms in a molecular graph is considered. The proposed $M$–index was used to correlate with a wide range of properties in various data sets, including; logarithm of partition coefficient in octanol/water, molar refraction, molar volume, parachor, and polarizability for alkanes and cycloalkanes; standard enthalpy of formation for silicon, titanium, and methyl halides; toxicity of saturated alcohols, ketones, diols, and carboxylic acids; molar refraction, and boiling point of aldehydes and ketones, as well as heat capacity, boiling point, density, refractive index, Gibbs energy, and the standard enthalpy of formation of isomeric compounds. The predictive ability of the developed models has been assessed by leave–one-outcross–validation test. All the constructed models have favorable statistical parameters and demonstrate satisfactory predictability.

Finally, it is important to note that the $M$–index is easy to generate and the developed QSPR/QSAR models based on this index are linearly correlated. This is an interesting feature of the $M$–index when compared with quantum chemical descriptors which require vast computational cost and exhibit limitations for large size molecules.

## 5. SUPPLEMENTARY INFORMATION

Illustrative examples for computing M–index for the compounds of different tables are presented in Supplementary Information.

**ACKNOWLEDGMENTS.** Shiraz University is gratefully acknowledged for its support. P. Manshour would also like to acknowledge support of Persian Gulf University Research Council.

## REFERENCES

1. A.-L. Barabási, *Linked: The New Science of Networks*, Perseus Publishing, Cambridge, 2003.
2. A.T. Balaban, *Chemical Applications of Graph Theory*, Academic Press, London, 1976.
3. A.T. Balaban, From chemical graphs to 3D molecular modeling, in: A. T. Balaban (Ed.), *From Chemical Topology to Three–Dimensional Geometry*, Plenum Press, New York, 1997, pp. 1–24.
4. L. Euler, Solutio problematisadgeometriam situs pertinentis, *Commentarii Academiae Scientiarum Petropolitanae* **8** (1741) 128–140.
5. N. Trinajstić, *Chemical Graph Theory*, CRC Press, Florida, USA, 1992.
6. Q. Ivanciuc, QSAR comparative study of Wiener descriptors for weighted molecular graphs, *J. Chem. Inf. Comput. Sci.* **40** (2000) 1412–1422.
7. A. Mohajeri, M. Alipour, M. B. Ahmadi, A graph theory study on $(ZnS)_n$ (n= 3–10) nanoclusters, *Chem. Phys. Lett.* **503** (2011) 162–166.
8. A. Kurup, R. Garg, C. Hansch, Comparative QSAR study of tyrosine kinase inhibitors, *Chem. Rev.* **101** (2001) 2573–2600.

9. R. García–Domenech, J. Gálvez, J. V. de Julián–Ortiz, L. Pogliani, Some new trends in chemical graph theory, *Chem. Rev.***108** (2008) 1127–1169.
10. C. Cao, Y. Hua, Topological indices based on vertex, distance, and ring: On the boiling points of paraffins and cycloakanes, *J. Chem. Inf. Comput. Sci.* **41** (2001) 867–877.
11. H. Yuan, A. L. Parrill, QSAR development to describe HIV–1 integrase inhibition, *J. Mol. Struct.* (*THEOCHEM*) **529** (2000) 273–282.
12. A. Mohajeri, M. H. Dinpajooh, Structure–toxicity relationship for aliphatic compounds using quantum topological descriptors, *J. Mol. Struct.* (*THEOCHEM*) **855** (2008) 1–5.
13. B. Hemmateenejad, A. Mohajeri, Application of quantum topological molecular similarity descriptors in QSPR study of the O–methylation of substituted phenols, *J. Comput. Chem.* **29** (2008) 266–274.
14. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley VCH, Weinheim, 2000.
15. H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.
16. A. T. Balaban, Highly discriminating distance–based topological index, *Chem. Phys. Lett.* **89** (1982) 399–404.
17. M. Randić, Characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
18. H. Hosoya, Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.* **44** (1971) 2332–2339.
19. D. H. Rouvray, *The role of the topological distance matrix in chemistry*, in: N. Trinajstić (Ed.), Mathematics and Computational Concepts in Chemistry, Horwood, Chichester, 1986, pp. 295–306.
20. I. Gutman, J. H. Potgieter, Wiener index and intermolecular forces, *J. Serb. Chem. Soc.* **62** (1997) 185–192.
21. D. H. Rouvray, Should we have designs on topological indices? Chemical Applications of Topology and Graph Theory, *Stud. in Phys. & Theoret. Chem.*, Elsevier, Amsterdam, 1983, pp. 159–177.
22. I. Lukovits, Wiener–type graph invariants, in: M. V. Diudea (Ed.), QSPR/QSAR Studies by Molecular Descriptors, Nova Science, Huntington, 2001, pp. 31–38.
23. I. Gutman, J. Žerovnik, Corroborating a modification of the Wiener index, *Croat. Chem. Acta.* **75** (2002) 603–612.
24. O. Ivanciuc, T. Ivanciuc, A. T. Balaban, Design of topological indices, part 10.1: Parameters based on electronegativity and covalent radius for the computation of

molecular graph descriptors for heteroatom–containing molecules, *J. Chem. Inf. Comput. Sci.* **38** (1998) 395–401.

25. F. Yang, Z. D. Wang, Y. P. Huang, X. R. Ding, Modification of Wiener index and its application, *J. Chem. Inf. Comput. Sci.* **43** (2003) 753–756.
26. F. Yang, Z. D. Wang, Y. P. Huang, P. J. Zhou, Modification of the Wiener index 2, *J. Chem. Inf. Comput. Sci.* **43** (2003) 1337–1341.
27. F. Yang, Z. D. Wang, Y. P. Huang, H. L. Zhu, Novel topological index F based on incidence matrix, *J. Comput. Chem.* **24** (2003) 1812–1820.
28. F. Yang, Z. D. Wang, Y. P. Huang, Modification of the Wiener index 4, *J. Comput. Chem.* **25** (2004) 881–887.
29. S. S. Tratch, M. I. Stankevitch, N. S. Zefirov, Combinatorial models and algorithms in chemistry. The expanded Wiener number–A novel topological index, *J. Comput. Chem.* **11** (1990) 899–908.
30. P. V. Khadikar, S. Karmarkar, A novel PI index and its applications to QSPR/QSAR studies, *J. Chem. Inf. Comput. Sci.* **41** (2001) 934–949.
31. R. Walsh, Thermochemistry of Silicon–containing compounds, *J. Chem. Soc. Faraday Trans.* **79** (1983) 2233–2248.
32. X. Zhi-hong, W. Le-shan, *The Data Base of Inorganic Chemical Thermodynamics,* Science Press, Beijing, 1987.
33. J. G. Stark, H. G. Wallace, *Chemistry Data Book,* John Murray, London, 1982.
34. G. Gini, M. V. Craciun, C. König, Combining unsupervised and supervised artificial neural networks to predict aquatic toxicity, *J. Chem. Inf. Comp. Sci*. **44** (2004) 1897–1902.
35. Y. Xue, H. Li, C. Y. Ung, C. W. Yap, Y. Z. Chen, Classification of a diverse set of Tetrahymena pyriformistoxicity chemical compounds from molecular descriptors by statistical learning methods, *Chem. Res. Toxicol.* **19** (2006) 1030–1039.
36. T. W. Schultz, Tetratox: Tetrahymena pyriformis population growth impairment endpoint a surrogate for fish lethality, *Toxicol. Mech. Method* **7** (1997) 289–309.
37. D. R. Roy, R. Parthasarathi, B. Maiti, V. Subramanian, P. K. Chattaraj, Electrophilicity as a possible descriptor for toxicity prediction, *Bioorg. Med. Chem.* **13** (2005) 3405–3412.
38. C. L. Yaws, *Chemical Properties Handbook*, McGraw–Hill, New York, 1999.
39. E. Estrada, L. Rodriguez, A. Gutiétrez, Matrix algebraic manipulations of molecular graphs. 1. Distance and Vertex–Adjacency Matrices, *MATCH Commun. Math. Comput. Chem*. **35** (1997) 145–156.
40. R. C. Weast, *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, 1989.
41. D. R. Lide, G. W. A. Milne, *Handbook of Data on Common Organic Compounds,* CRC Press, Boca Raton, FL, 1995.

42. J. A. Dean, *Lange's Handbook of Chemistry*, McGraw–Hill, New York, 1999.
43. A. A. Gakh, E. G. Gakh, B. G. Sumpter, D. W. Noid, Neural network–graph theory approach to the prediction of the physical properties of organic compounds, *J. Chem. Inf. Comput. Sci.* **34** (1994) 832–839.

Iranian J. Math. Chem. 8(2) June (2017) 137−159

Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

# A New Two–Step Obrechkoff Method with Vanished Phase–Lag and some of its Derivatives for the Numerical Solution of Radial Schrödinger Equation and Related IVPs with Oscillating Solutions

**ALI SHOKRI AND MORTAZA TAHMOURASI**•

Faculty of Mathematical Science, University of Maragheh, Maragheh, Iran

## ARTICLE INFO

**Article History:**

Received 2 October 2016
Accepted 26 February 2017
Published online 22 March 2017
Academic Editor: Ivan Gutman

**Keywords:**

Schrödinger equation
Phase–lag
Ordinary differential equations
Symmetric multistep methods.

## ABSTRACT

A new two–step implicit linear Obrechkoff twelfth algebraic order method with vanished phase–lag and its first, second, third and fourth derivatives is constructed in this paper. The purpose of this paper is to develop an efficient algorithm for the approximate solution of the one–dimensional radial Schrödinger equation and related problems. This algorithm belongs in the category of the multistep methods. In order to produce an efficient multistep method the phase–lag property and its derivatives are used. An error analysis and a stability analysis are also investigated and a comparison with other methods is also studied. The efficiency of the new methodology is proved via theoretical analysis and numerical applications.

## 1. INTRODUCTION

The radial time–independent Schrödinger equation can be written as:

$$y''(x)=\left(\frac{l(l+1)}{x^2}+V(x)-E\right)y(x), \tag{1}$$

The boundary conditions are $y(0) = 0$, and a second boundary condition, for large values of x, determined by physical considerations. Large research on the algorithmic development of numerical methods for the solution of the Schrödinger equation has been done in the last decades. The aim and scope of this research is the construction of fast and reliable algorithms for the solution of the Schrödinger equation and related problems.

---

• Corresponding Author: (Email address: mortazatahmoras@gmail.com)

DOI: 10.22052/ijmc.2017.62671.1243

Mathematical models in theoretical physics and chemistry, material sciences, quantum mechanics and quantum chemistry, electronics etc. can be express via the above boundary value problem [1]. The numerical methods for the approximate solution of the Schrödinger equation and related problems can be divided into two main categories:

1. Methods with constant coefficients.
2. Methods with coefficients depending on the frequency of the problem.

The main result of this paper is the development of an efficient multistep method for the numerical solution of systems of ordinary differential equations with oscillating or periodical solutions. The reason of their efficiency, as the analysis proved, is that the phase–lag and its derivatives are eliminated. Another reason of the efficiency of the new obtained method is that it has high algebraic order. The purpose of this paper is to extend the methodology for the development of numerical methods for the approximate solution periodic initial–value problems. The new methodology is based on the requirement of the phase–lag and its derivatives vanishing. Based on this new methodology we will develop a method one will have phase–lag and its first, second, third and fourth derivatives vanishing. We will apply the new developed method on the numerical solution of the radial Schrödinger equation. We will study the efficiency of the new obtained methods via:

- A comparative error analysis,
- A comparative stability analysis and finally,
- The numerical results produced from the numerical solution of the radial Schrödinger with application to the specific potential.

More specifically, we will develop a family of implicit symmetric two–step Obrechkoff methods of twelfth algebraic order. The development of the new family of methods is based on the requirement of the phase–lag and its first, second, third and fourth derivatives vanishing. We will give a comparative error analysis and a comparative stability analysis in order to study the efficiency of new proposed method of the family. Finally, we will apply both methods to the resonance problem. This is one of the most difficult problems arising from the radial Schrödinger equation.

For several decades, there has been strong interest in searching for better numerical methods to integrate first order and second–order initial value problems, because these problems are usually encountered in celestial mechanics, quantum mechanical scattering theory, theoretical physics and chemistry, and electronics. Computational methods involving a parameter proposed by Gautschi [8], Jain et al. [13] and Steifel and Bettis [24] yield numerical solution of problems of class (1). Chawla et al. [3, 4], Ananthakrishnaiah [1], Shokri and et al. [17, 18,19], Dahlquist [5], Asadzadeh [2], Franco [6], Lambert and Watson [14], Simos and et al. [20, 21, 22], Saldanha and Achar [16], and Daele and Vanden Berghe [26] have developed methods to solve problems of class (2). Consider the class of Obrechkoff methods of the form

$$\sum_{j=0}^{k}\alpha_i y_{n-j+1} = \sum_{i=1}^{l} h^{2i}\sum_{j=0}^{k}\beta_i y^{(2i)}{}_{n-j+1,} \tag{2}$$

for the numerical integration of the problem (1). The method (2) is symmetric when $\alpha_j = \alpha_{k-j}, \beta_j = \beta_{k-j}, j = 0,1,2,k,$ and it is of order $q$ if the truncation error associated with the linear difference operator is given as

$$TE = C_{q+2}h^{q+2}y^{(q+2)}, \quad x_{n-k+1} < \eta < x_{n+1},$$

where $C_{q+2}$ is a constant dependent on *h*. We have organized the paper as follows: In Section 2 we present the theory of the new methodology. In Section 3 we present the development of the new method. A comparative error analysis is presented in Section 4. Finally, the numerical results are presented in Section 5.

## 2. PRELIMINARIES

In order to define the interval of periodicity of a method the periodic stability analysis of this method is very important. The interval of periodicity defines the step size which can be used in order the approximation of the solution of problems with high oscillatory or periodic solution to be of the same order as the algebraic order of the method. It can be seen that when we have a large interval of periodicity then we can have a large step size for the same accuracy. To investigate the stability properties of methods for solving the initial value problem (l), Lambert and Watson [14] introduced the scalar test equation. From the form (2) and without loss of generality we assume

$$y'' = -\omega^2 y, \quad \omega \in \mathbb{R}. \tag{3}$$

and the interval of periodicity, where w is defined as the frequency of the problem and may be a constant. When we apply a symmetric two–step method to the scalar test equation (3), we obtain a difference equation of the form

$$y_{n+1} - 2C(v)y_n + y_{n-1} = 0, \tag{4}$$

where $v = \omega h$, $h$ is the step length, $C(v) = B(v)/A(v)$ where $A(v)$ and $B(v)$ are polynomials in $v$ and $y_n$ is the computed approximation to $y(nh), n = 0,1,2,...$. The characteristic equation associated with (4) is

$$\zeta^2 - 2C(v)\zeta + 1 = 0. \tag{5}$$

We have the following definitions.

**Definition 2.1.** (See [25]). The method of Eq. (4) with the characteristic Eq. (5) is unconditionally stable if $|\zeta_1| \leq 1$ and $|\zeta_1| \leq 1$ for all values of $\omega h$.

Following Lambert and Watson [14], we say that the numerical method (4) has an interval of periodicity $\left(0, v_0^2\right)$, if for all $v \in \left(0, v_0^2\right)$, $\zeta_1$ and $\zeta_2$ satisfy $\zeta_1 = \exp(i\theta(v))$, and $\zeta_2 = \exp(-i\theta(v))$, where $\theta(v)$ is a real function of *v*. For any method corresponding to the characteristic equation (4) the phase–lag is defined as the leading term in the expansion of

$$t = v - \theta(v) = v - \cos^{-1}[C(v)]. \tag{6}$$

If the quantity $t = O\left(v^{q+1}\right)$ as $v \to 0$, the order of phase–lag is *q*.

**Definition 2.2.** Suppose (5) is the characteristic equation of (4), and

$$|C(v)| < 1, \ \forall v^2 \in \left(0, v_0^2\right).$$

Then the periodicity interval of the method is $\left(0, v_0^2\right)$.

**Definition 2.2.** The method (4) is said to be P–stable if its interval of periodicity is $(0, \infty)$ .

**Theorem 2.4.** (See Ibraheem and Simos [10]) The phase–lag of a symmetric two–step method with characteristic equation given by (5) is the leading term in the expansion of

$$\frac{[C(v) - \cos(v)]}{v^2}.$$

## 3. DEVELOPMENT AND ANALYSIS

From the form (2) and without loss of generality we assume

$$\alpha_j = \alpha_{m-j}, \ \beta_{ij} = \beta_{i,m-j}, \quad j = 0(1)\left\lfloor \frac{m}{2} \right\rfloor,$$

and we can write

$$y_{n+1} - 2y_n + y_{n-1} = \sum_{i=1}^{m} h^{2i}\left[\beta_{i0} y_{n+1}{}^{(2i)} + \beta_{i1} y_n{}^{(2i)} + \beta_{i0} y_{n-1}{}^{(2i)}\right] \tag{7}$$

When $m = 3$ we get

$$\begin{aligned} y_{n+1} - 2y_n + y_{n-1} = {} & h^2\left[\beta_{10}(y_{n+1}^{(2)} + y_{n-1}^{(2)}) + \beta_{11} y_n^{(2)}\right] \\ & + h^4\left[\beta_{20}(y_{n+1}^{(4)} + y_{n-1}^{(4)}) + \beta_{21} y_n^{(4)}\right] \\ & + h^6\left[\beta_{30}(y_{n+1}^{(6)} + y_{n-1}^{(6)}) + \beta_{31} y_n^{(6)}\right]. \end{aligned} \tag{8}$$

$M-3$ for method (8) is 11 so that if $P=-1$, $K=13$ we obtain classic method and the coefficients of this method are

$$\begin{aligned} &\beta_{10} = \frac{229}{7788}, \quad \beta_{11} = \frac{3665}{3894}, \quad \beta_{20} = -\frac{1}{2360}, \\ &\beta_{21} = \frac{711}{12980}, \quad \beta_{30} = \frac{127}{39251520}, \quad \beta_{31} = \frac{2923}{3925152}, \end{aligned} \tag{9}$$

where its phase–lag is given by

$$pl_{clas} = -\frac{45469}{3394722659328000} v^{12} + O\left(v^{14}\right),$$

and its local truncation error is given by

$$LTE_{clas} = -\frac{45469}{1697361329664000} y^{(14)} h^{14} + O(h^{16}).$$

If $P=6$, $K=-1$ then we obtain the method with zero phase–lag ($PL$), and the coefficients of this case are given in [16].

### 3.1. DEVELOPMENT

Application of the method (8) to the scalar test equation (3) leads to the difference equation (4) with $C(v^2)$ given by

$$C\left(v^2\right) = \frac{1 - \frac{1}{2}\beta_{11} v^2 + \frac{1}{2}\beta_{21} v^4 - \frac{1}{2}\beta_{31} v^6}{1 + \beta_{10} v^2 - \beta_{20} v^4 + \beta_{30} v^6}. \tag{10}$$

We require the above mentioned method to have the phase–lag and its derivatives vanished. Using the Eq. (10) and Theorem 2.4, and requiring the above method (8) to have the maximum algebraic order with five free parameters, the following relations are obtained:

$$\beta_{10} = \frac{1}{2} - \frac{1}{2}\beta_{11}.$$

So the phase–lag is equal to:

$$PL = \left( \frac{1 - \frac{1}{2}\beta_{11}v^2 + \frac{1}{2}\beta_{21}v^4 - \frac{1}{2}\beta_{31}v^6}{1 + \left(\frac{1}{2} - \frac{1}{2}\beta_{11}\right)v^2 - \beta_{20}v^4 + \beta_{30}v^6} - \cos(v) \right) v^{-2}.$$

We require the above mentioned method to have the phase–lag and some of its derivatives vanished. Hence we can write

$$PL^{(i)} = 0, \qquad i = 0,1,2,3,4.$$

Demanding the phase–lag and the first, second, third and fourth derivatives of the phase to vanish we can find of all coefficients. For small values of $|v|$ in the coefficients, are subject to heavy cancelations. In this case the following Taylor series expansions should be used:

$$\begin{aligned}\beta_{10} = {} & \frac{229}{7788} + \frac{45469}{262829424}v^2 - \frac{24889175}{21492613318176}v^4 - \frac{678202459751}{17615223060643802880}v^6 \\ & + \frac{35019787379195}{18245918507107093598315552}v^8 + \frac{51820722922675986337}{197923904285737562600127482880 0}v^{10} \\ & + \frac{640261689488085567511}{621981077741727283564653249884160}v^{12} \\ & + \frac{30583141728568702155177793466 9}{110739500591277795437521893413571142533120 00}v^{14} + \ldots\end{aligned}$$

$$\begin{aligned}\beta_{11} = {} & \frac{3665}{3894} - \frac{45469}{131414712}v^2 + \frac{24889175}{10746306659088}v^4 + \frac{678202459751}{8807611530321901440}v^6 \\ & - \frac{35019787379195}{9122959253553546799157 76}v^8 - \frac{51820722922675986337}{98961952142868781300063741440 0}v^{10} \\ & - \frac{640261689488085567511}{310990538870863641782326624942080}v^{12} \\ & - \frac{30583141728568702155177793466 9}{553697502956388977187609467067855712665600 0}v^{14} - \ldots\end{aligned}$$

$$\begin{aligned}\beta_{20} = {} & -\frac{1}{2360} - \frac{45469}{6021183168}v^2 - \frac{805517}{58616218140480}v^4 + \frac{441092244757}{13451624882673449472 0}v^6 \\ & + \frac{5333451495777199}{417997405799180689706864 64}v^8 + \frac{116638144035866913376 9}{352664411272768747905681696768000}v^{10} \\ & + \frac{731684285369995100804 33}{105284433341372381999762213753118720 0}v^{12} \\ & + \frac{373962299301113409721054638 71}{300427257584949808674999586437774343618560 00}v^{14} + \ldots\end{aligned}$$

$$\beta_{21} = \frac{711}{12980} - \frac{5228935}{33116507424}\nu^2 + \frac{47774789}{40298649971580}\nu^4 + \frac{4726497723443}{1479678737094079441920}\nu^6$$
$$- \frac{63080459663327759}{22989857318954937933877552}\nu^8 - \frac{233098982919042820567 3}{387930852400045622696249866444800}\nu^{10}$$
$$- \frac{1228101235325079961895 9}{629417808019074022824665408306688 00}\nu^{12}$$
$$- \frac{24686049449918639502839304 34259}{62789296835254510013074913565494837816279 0400}\nu^{14} - \ldots$$

$$\beta_{30} = \frac{127}{39251520} + \frac{45469}{305690837760}\nu^2 + \frac{5624977}{1289556799090560}\nu^4 + \frac{14566546103957}{14341501605681077667 8400}\nu^6$$
$$+ \frac{184980240090953987}{9195942927581975173551022 0800}\nu^8 + \frac{41313447360649468349}{119363333920001403775269226659840 00}\nu^{10}$$
$$+ \frac{2245777138282281077923249 3}{437772673833426364355011284785467637760 00}\nu^{12}$$
$$+ \frac{200122349577094995860851362 11}{321996394026946205195255967002537629827072 0000}\nu^{14} + \ldots$$

$$\beta_{31} = \frac{2923}{3925152} - \frac{14231797}{1986990445440}\nu^2 + \frac{26183459}{257911359818112}\nu^4 - \frac{255261508015541}{93219760436927004840 9600}\nu^6$$
$$+ \frac{6643202390908931}{9195942927581975173551022080}\nu^8 - \frac{768504540290345312837}{775861704800091245392499732889 6000}\nu^{10}$$
$$- \frac{281805583458416972116 3}{1152033352193227274618450749435441152}\nu^{12}$$
$$- \frac{3369063927072101706481851260 81}{697658853725050111256387928505498197958656 0000}\nu^{14} - \ldots$$

hence

$$PL_{New} = \frac{3111088424452198220240000327708014988665362375064863999939597}{37062592984163314261892997730058360207116126560610832607916100700873211067278295 0400000}\nu^{30},$$

and

$$LTE_{New} = -\frac{45469}{1697361329\,664000}\left(5\omega^8 y^{(6)} + 10\omega^4 y^{(10)} + 5\omega^2 y^{(12)} + y^{(14)} + 10\omega^6 y^{(8)} + \omega^{10} y^{(4)}\right)h^{14},$$

where $\nu = \omega h$, $\omega$ is the frequency and h is the step length. As $\nu \to 0$, the LTE of the method (8) with above derived coefficients, tends to

$$\frac{45469}{169736132966400}h^{14}y^{(14)} + O(h^{16}),$$

which agrees with the LTE of the three methods due to Wang [27], Simos [20] and Daele [26], Achar [1], as $H \to 0$. The characteristic equation

$$\Omega\left(s;v^2\right) = A(v)s^2 - 2B(v)s + A(v) = 0$$

has complex roots of unit magnitude when

$$\left|\cos(\theta(v))\right| = \left|\frac{B(v)}{A(v)}\right| < 1,$$

i.e. when $A(v)^2 \pm B(v) > 0$. Substituting for $A(v)$ and $B(v)$ for new method, the interval of periodicity of the classical Obrechkoff method, PL′ and PL″ methods [18], the new method when $v \to 0$ are obtained [0, 25.2004], [0,408.04], [0, 1428.84] and [0, 6593.44] respectively. The behaviors of the coefficients are given in Figures 1, 2 and 3.

## 4. COMPARATIVE ERROR ANALYSIS

We will study the following methods:

- The ten–step tenth algebraic order method developed by Quinlan and Tremaine [15] which is indicated as QT10.
- The twelve–step twelfth algebraic order method developed by Quinlan and Tremaine [15] which is indicated as QT12.
- The classical two–step method of the family of methods mentioned in Section 3 of this paper which is indicated as CL2.
- The classical ten–step method of the family of methods mentioned in paragraph 3 of [9] which is indicated as CL10
- The method with vanished phase–lag produced by Alolyan and Simos [10] which is indicated as PF.
- The ten–step predictor–corrector method produced by Shokri [17] which is indicated as PC.
- High phase–lag order trigonometrically fitted two–step Obrechkoff produced by Shokri [18] which is indicated as TFO.
- The method with vanished phase–lag and its first derivative produced by Alolyan and Simos [10] which is indicated as PFDF.
- The ten–step method with phase–lag and its first and second derivatives equal to zero produced by Alolyan and Simos [9] which is indicated as PFDF12.
- The ten–step method with phase–lag and its first, second and third derivatives equal to zero produced by Alolyan and Simos [9] which is indicated as PFDF123.
- The new developed two–step Obrechkoff method with vanished phase–lag and its first, second, third and fourth derivatives obtained in this paper which is indicated as new.

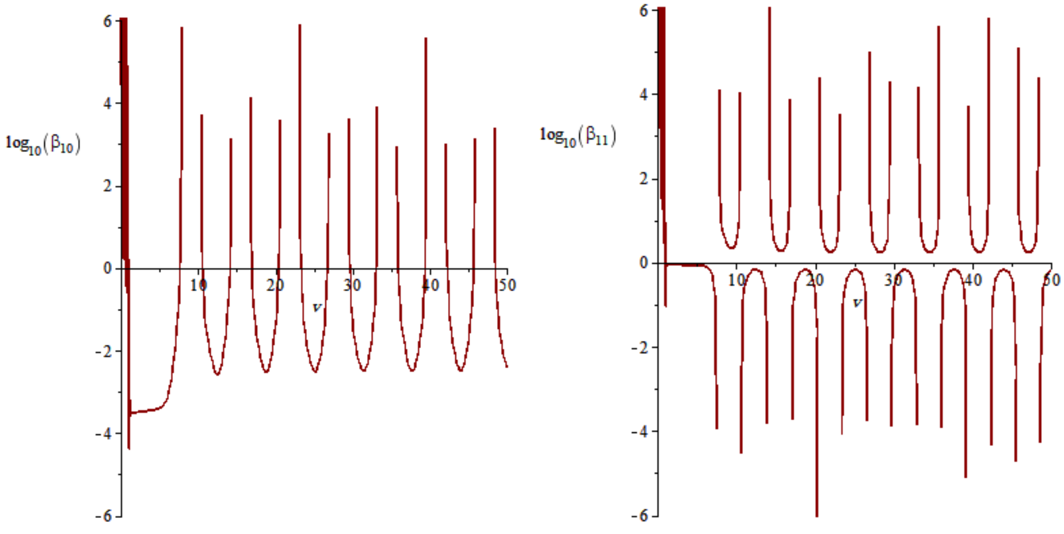


**Figure 1:** Behavior of the coefficients $\beta_{10}$ and $\beta_{11}$ in the new method.

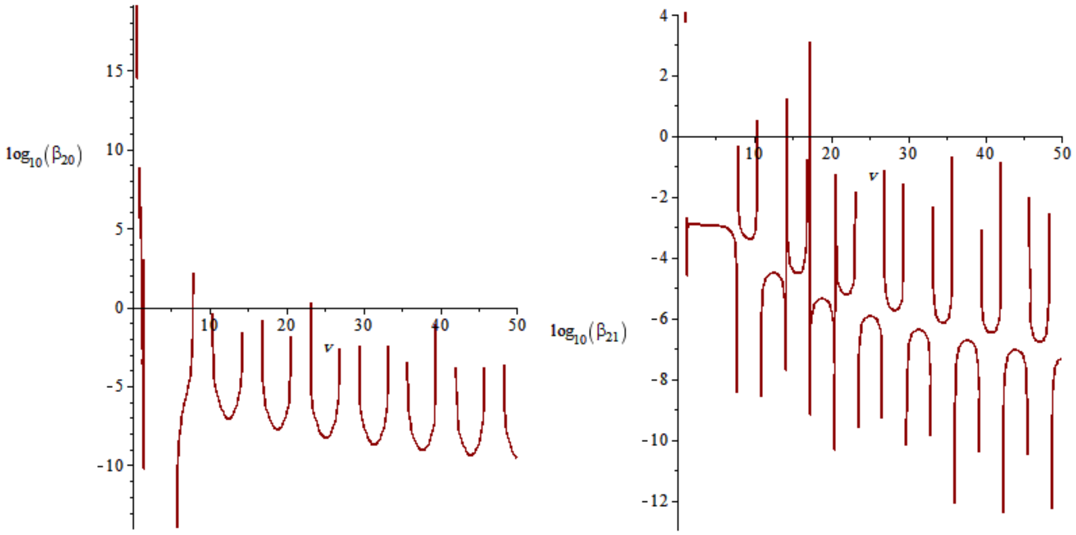


**Figure 2**: Behavior of the coefficients $\beta_{20}$ and $\beta_{21}$ in the new method.

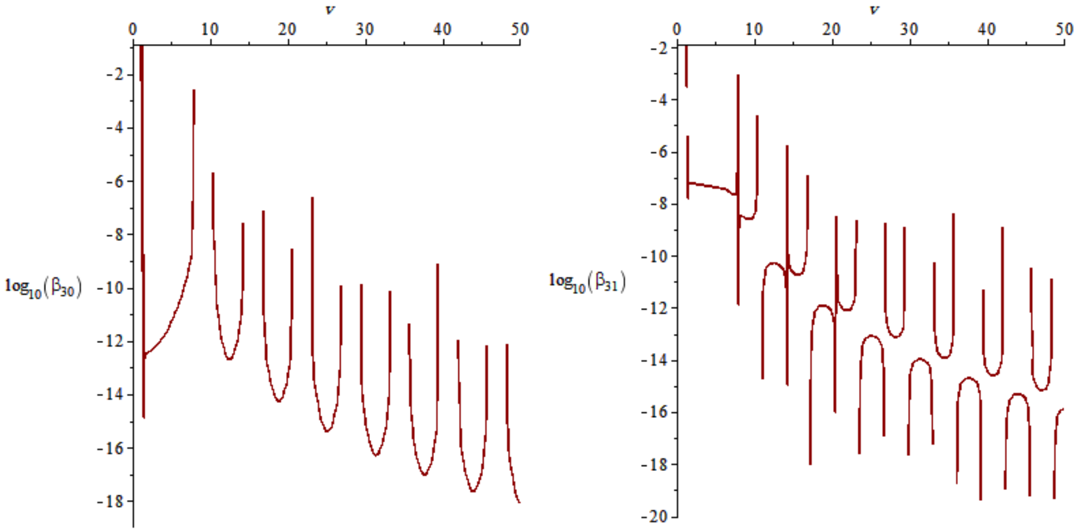


**Figure 3**: Behavior of the coefficients $\beta_{30}$ and $\beta_{31}$ in the new method.

From the above equations we have the following theorem:

**Theorem 4.1.** For the numerical solution of the time independent radial Schrödinger equation the new proposed method produced in this paper is the most accurate method, especially for large values of $|G| = |V_C - E|$.

**Proof.** The radial time independent Schrödinger equation is of the form

$$y'' = f(x)y(x) \tag{12}$$

Based on the paper of Ixaru and Rizea [12], the function *f*(*x*) can be written in the form $f(x) = g(x) + G$, where $g(x) = V(x) - V_C = g$, and $V_C$ is the constant approximation of the potential and $G = v^2 = V_C - E$. We express the derivatives $y_n^{(i)}$, $i = 2,3,4,...$ which are terms of the local truncation error formulae, in terms of Eq. (12). The expressions are presented as polynomials of $G$. Finally, we substitute the expressions of the derivatives, produced in the previous step, into the local truncation error formulae. We use the procedure mentioned above and the formulae:

$$\begin{aligned}
y_n^{(2)} &= \left(V(x)-V_C+G\right)y(x),\\
y_n^{(4)} &= \left(\frac{d^2}{dx^2}V(x)\right)y(x)+2\left(\frac{d}{dx}V(x)\right)\left(\frac{d}{dx}y(x)\right)+\left(V(x)-V_C+G\right)\left(\frac{d^2}{dx^2}y(x)\right),\\
y_n^{(6)} &= \left(\frac{d^4}{dx^4}V(x)\right)y(x)+4\left(\frac{d^3}{dx^3}V(x)\right)\left(\frac{d}{dx}y(x)\right)+3\left(\frac{d^2}{dx^2}V(x)\right)\left(\frac{d^2}{dx^2 y(x)}\right)\\
&\quad +4\left(\frac{d}{dx}V(x)\right)^2 y(x)+6\left(V(x)-V_C+G\right)\left(\frac{d}{dx}V(x)\right)\left(\frac{d}{dx}y(x)\right)\\
&\quad +4\left(V(x)-V_C+G\right)\left(\frac{d^2}{dx^2}V(x)\right)y(x)+\left(V(x)-V_C+G\right)^2\left(\frac{d^2}{dx^2}y(x)\right)
\end{aligned}\tag{13}$$

We consider two cases in terms of the value of $E$:

1. The energy is close to the potential, i.e. $G=V_C-E\approx 0$. So only the free terms of the polynomials in $G$ are considered. Thus for these values of $G$, the methods are of comparable accuracy. This is because the free terms of the polynomials in $G$, are the same for the cases of the classical method and of the new developed methods.
2. $G>>$ or $G<<0$. Then $|G|$ is a large number.

So, we have the following asymptotic expansions of the equations produced from the Local Truncation Errors and based on the above procedure:

a. The ten–step tenth algebraic order method developed by Quinlan and Tremaine [15], for the analysis of the local truncation error see [11]

$$LTE_{QT10}=h^{12}\left[-\frac{52559}{912384}y(x)G^6+\ldots\right].\tag{14}$$

b. The twelve–step twelfth algebraic order method developed by Quinlan and Tremaine [15], for the analysis of the local truncation error see [11]

$$LTE_{QT14}=h^{14}\left[-\frac{16301796103}{290594304000}y(x)G^7+\ldots\right].\tag{15}$$

c. The classical two–step Obrechkoff method with $m=3$ which is indicated as CL2

$$LTE_{CL2}=h^{14}\left[-\frac{45469}{1697361329664000}y(x)G^7+\ldots\right].\tag{16}$$

d. The classical ten–step method of the family 1, [9] which is indicated as CL10

$$LTE_{CL10}=h^{14}\left[-\frac{547336457}{373621248000}y(x)G^7+\ldots\right].\tag{17}$$

e. The method with vanished phase–lag produced by Alolyan and Simos [10] which is indicated as PF

$$LTE_{PF}=h^{14}\left[-\frac{547336457}{373621248000}g(x)y(x)G^{6}+\ldots\right]. \tag{18}$$

f. The ten–step predictor–corrector method produced by Shokri [17] which is indicated as PC

$$LTE_{PC}=h^{14}\left[-\frac{96506469327691}{47345284546560000}g(x)y(x)G^{6}+\ldots\right]. \tag{19}$$

g. High phase–lag order trigonometrically fitted two–step Obrechkoff PL″ produced by Shokri [18] which is indicated as TFO

$$LTE_{TFO}=h^{14}\left[-\frac{45469}{1697361329664000}g(x)y(x)G^{6}+\ldots\right]. \tag{20}$$

h. The method with vanished phase–lag and its first derivative produced by Alolyan and Simos [9] which is indicated as PFDF

$$LTE_{PFDF}=h^{14}\left[\left[\begin{array}{l}\left(-\frac{547336457}{17791488000}\left(\frac{d^{2}}{dx^{2}}g(x)\right)y(x)-\frac{547336457}{373621248000}\left(g(x)\right)^{2}y(x)\right)\\+\left(-\frac{547336457}{186810624000}\left(\frac{d}{dx}g(x)\right)\frac{d}{dx}y(x)\right)\end{array}\right]G^{5}+\ldots\right]. \tag{21}$$

i. The method with vanished phase–lag and its first and second derivatives produced by Alolyan and Simos [9] which is indicated as PFDF12.

$$LTE_{PFDF12}=h^{14}\left[-\frac{547336457}{17791488000}\left(\frac{d^{2}}{dx^{2}}g(x)\right)y(x)G^{5}+\ldots\right]. \tag{22}$$

j. The method with vanished phase–lag and its first, second and third derivatives produced by Alolyan and Simos [9] which is indicated as PFDF123.

$$LTE_{PFDF123}=h^{14}\left[\left[\begin{array}{l}\left(-\frac{547336457}{17791488000}\left(\frac{d^{4}}{dx^{4}}g(x)\right)y(x)-\frac{547336457}{373621248000}\left(\frac{d^{3}}{dx^{3}}g(x)\right)\frac{d}{dx}y(x)\right)\\+\left(-\frac{547336457}{31135104000}\left(\frac{d}{dx}g(x)\right)^{2}y(x)-\frac{547336457}{23351328000}g(x)y(x)\frac{d^{2}}{dx^{2}}g(x)\right)\end{array}\right]G^{4}+\ldots\right]. \tag{23}$$

k. The new two–step Obrechkoff method with phase–lag and its first, second, third and fourth derivatives equal to zero obtained in this paper which is indicated as new:

$$LTE_{New}=h^{14}\left[-\frac{45469}{106085083104000}\left(\frac{d^{4}}{dx^{4}}g(x)\right)y(x)G^{4}+\ldots\right]. \tag{24}$$

Based on the analysis presented above, we studied the interval ofperiodicity of the eight methods mentioned in the previousparagraph. The results are presented in Table 1.

| Method | Interval of periodicity |
|---|---|
| QT10 | (0,0.17) |
| QT12 | (0,0.046) |
| CL2 | (0,25.2004) |
| CL12 | (0,0.8) |
| PF(see [10]) | (0,1.2) |
| PC (see [17]) | (0,9.89) |
| TFO (see PL" in [18]) | (0,1428.84) |
| PFDF (see [9]) | (0,1.5) |
| PFDF12(see [9]) | (0,6.6) |
| PFDF (see [9]) | (0,3.6) |
| New method | (0,65559993.44) |

**Table 1:** Comparative interval of periodicity for the methods mentioned in Section 4.

Hence for the classical two–step Obrechkoff methods, the errorincreases as the seventh power of $G$. For the classical ten–stepmethods, the error increases as the seventh power of $G$. For themethod with vanished phase–lag produced by Alolyan and Simos [10], the error increases as the sixth power of $G$. Forten–step predictor–corrector method produced by Shokri [17], the error increases as the sixth power of $G$. For two–step twelfthorder Obrechkoff method produced by Shokri [18], the errorincreases as the sixth power of $G$. For twelfth order method with vanished phase–lag and its first derivative produced by Alolyan and Simos [9], the error increases as the fifth power of $G$. For ten-step twelfth order method with vanished phase–lag and its firstand second derivatives produced by Alolyan and Simos [9], theerror increases as the fifth power of $G$. For ten–step twelfthorder method with vanished phase–lag and its first, second and thirdderivatives produced by Alolyan and Simos [9], the errorincreases as the fourth power of $G$. For the new two–step Obrechkoff method with vanished phase–lag and its first, second,third and fourth derivatives obtained in this paper, the errorincreases as the fourth power of $G$ but it has lower coefficientsthan the method developed in [9]. So, for the numericalsolution of the time–independent radial Schrödinger equation thenew obtained two–step Obrechkoff method with vanished phase–lag andits derivatives is the most accurate ones, especially for largevalues of $|G| = |V_C - E|$. ■

**Remark 4.2.** In Figures 4, 5, we present the $s - v$ plane and behavior of stability polynomial (respectively) for the method developed in this paper ($s$ is frequency of test problem and $v$ is frequency of method).

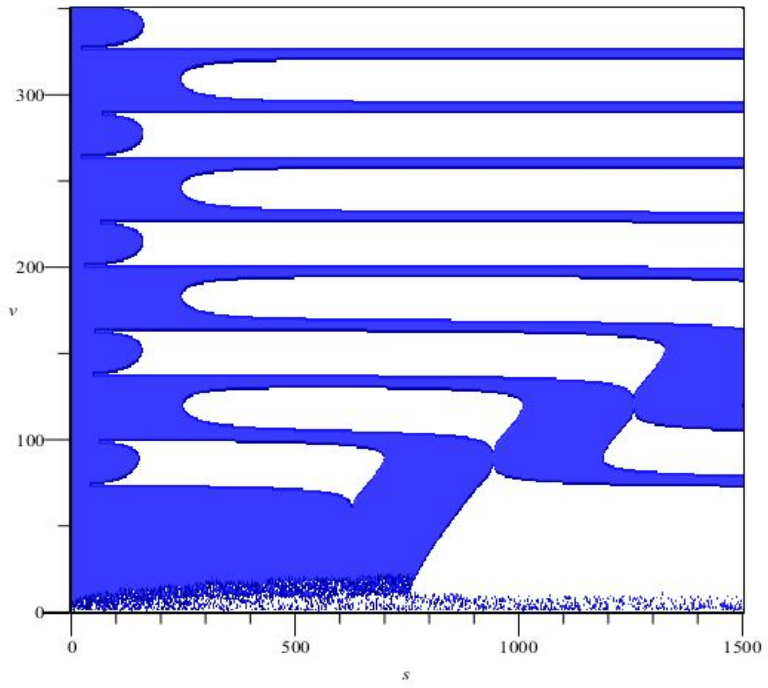


**Figure 4:** The stability region for the new method.

For the solution of the Schrödinger equation the frequency of the exponential fitting is equal to the frequency of the scalar test equation. So, it is necessary to observe the surroundings of the first diagonal of the $s - v$ plane.

## 5. NUMERICAL RESULTS

### 5.1 THE METHODS

We have used several multistep methods for the integration of the five test problems. These methods are

- The ten–step tenth algebraic order method developed by Quinlan and Tremaine [15] which is indicated as Method I.
- The twelve–step twelfth algebraic order method developed by Quinlan and Tremaine [15] which is indicated as Method II.
- The ten–step method with phase–lag and its first and second derivatives equal to zero obtained in [9] which is indicated as Method III.
- The ten–step method with phase–lag and its first, second and third derivatives equal to zero obtained in [9] which is indicated as Method IV.
- The ten–step predictor–corrector method produced by Shokri [17], which is indicated as Method V.

- The new method obtained in this paper which is indicated as VI.

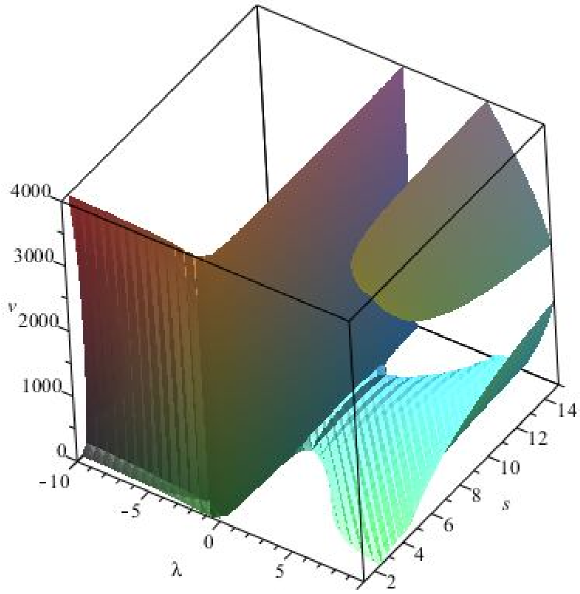


**Figure 5:** Behavior of the stability polynomial for the new method.

### 5.2. The Problems

The efficiency of the new symmetric two–step Obrechkoff method will be measured through the integration of five initial value problems with oscillating solution. In order to apply the new method to the radial Schrödinger equation the value of parameter $\omega$ is needed. For every problem of the one-dimensional Schrödinger equation given by (1) the parameter $\omega$ is given by

$$\omega = \sqrt{|q(x)|} = \sqrt{|V(x) - E|}\,, \qquad (25)$$

where $V(x)$ is the potential and $E$ is the energy.

**Example 5.1.** We consider the Schrödinger equation resonance problem. We will integrate problem (1) with $l = 0$ at the interval $[0,15]$ using the well–known Woods–Saxon potential

$$V(x) = \frac{u_0}{(1+q)} + \frac{u_1 q}{(1+q)^2}, \qquad q = \exp\left(\frac{x - x_0}{a}\right),$$

where $u_0 = 50$, $a = 0.6$, $x_0 = 7$, $u_1 = -\frac{u_0}{a}$. The behavior of the Woods–Saxonpotential is shown in Figure 6 and with boundary condition $y(0) = 0$. The potential $V(x)$ decays more quickly than $\frac{l(l+1)}{x^2}$, so for large $x$ (asymptotic region) the Schrödinger equation (1) becomes

$$y''=\left(\frac{l(l+1)}{x^2}+V(x)-E\right)y(x).$$

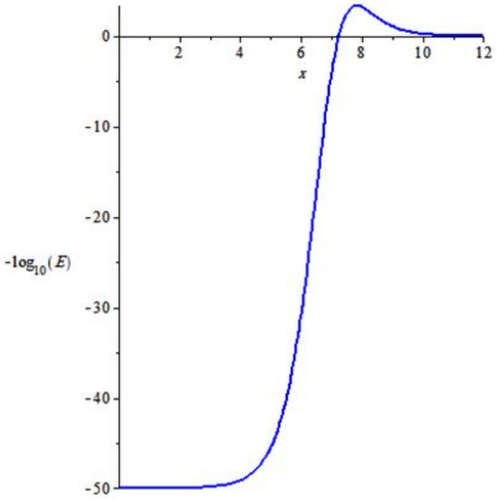


**Figure 6:** The Woods-Saxon potential.

The last equation has two linearly independent solutions $kxj_l(kx)$ and $kxn_l(kx)$, where $j_l$ and $n_l$ are the spherical Bessel and Neumann functions respectively. When $x\to\infty$ thesolution of Schrödinger has the asymptotic form

$$\begin{aligned} y(x) &\approx Akxj_l(kx)-Bkxn_l(kx) \\ &\approx D\left[\sin\left(kx-\frac{l\pi}{2}\right)+\tan(\delta_l)\cos\left(kx-\frac{l\pi}{2}\right)\right], \end{aligned}$$

where $\delta_l$ is called *scattering phase shift* and it iscalculated by the following expression:

$$\tan(\delta_l)=\frac{y(x_i)S(x_{i+1})-y(x_{i+1})S(x_i)}{y(x_{i+1})C(x_i)-y(x_i)C(x_{i+1})},$$

where $S(x)=kxj_l(kx)$, $C(x)=kxn_l(kx)$ and $x_i<x_{i+1}$ both belong to the asymptotic region. Given the energy we approximate thephase shift, the accurate value of which is π/2 for the above problem.

We will use for the energy the value $E=989.701916$. For some well–known potentials, such as the Woods–Saxon potential, the definition of parameter $\omega$ is not given as a function of $x$ but based on some critical points which have been defined from the study of the appropriate potential (see for details [12]). For the purpose of obtaining our numerical results it is appropriate to choose $\omega$ as follows (see for details [12]):

$$\omega = \begin{cases} \sqrt{E+50} & , \quad x \in [0,6.5], \\ \sqrt{E} & , \quad x \in [6.5,15]. \end{cases}$$

**Example 5.2.** The almost periodic orbital problem studied by Franco and Palacios [7], can be described by

$$y''+y = \varepsilon \exp(i\psi x), \qquad y(0)=1, \quad y'(0)=i, \quad y \in C$$

or equivalently by

$$\begin{cases} u''+u = \varepsilon \cos(i\psi x), \quad u(0)=1, \quad u'(0)=0 \\ v''+v = \varepsilon \sin(\psi x), \quad u(0)=0, \quad v'(0)=1 \end{cases}$$

where $\varepsilon = 0.001$ and $\psi = 0.01$. The theoretical solution of this problem is given by

$$y(x) = u(x) + iv(x), \qquad x \in R, \tag{26}$$

where

$$u(x) = \frac{1-\varepsilon-\psi^2}{1-\psi^2}\cos(x) + \frac{\varepsilon}{1-\psi^2}\cos(\psi x),$$

$$v(x) = \frac{1-\varepsilon\psi-\psi^2}{-\psi^2}\sin(x) + \frac{\varepsilon}{1-\psi^2}\sin(\psi x).$$

This system of equations has been solved for $x \in [0,1000\pi]$. For this problem we use $\omega = 1$.

**Example 5.3.** The almost periodic orbital problem studied by **Stiefel and Bettis** [24], can be described by

$$y''+y = 0.001\exp(ix), \qquad y(0)=1, \quad y'(0)=0.9995i, \quad y \in C$$

or equivalently by

$$\begin{cases} u''+u = 0.001\cos(\psi x), \quad u(0)=1, \quad u'(0)=0, \\ v''+v = 0.001\sin(\psi x), \quad u(0)=0, \quad v'(0)=0.9995i. \end{cases}$$

The theoretical solution of this problem is given by $y(x) = u(x) + iv(x),\ u, v \in R$ and

$$u(x) = \cos(x) + 0.0005\cos(x),$$

$$v(x) = \sin(x) + 0.0005x\cos(x).$$

This system of equations has been solved for $x \in [0,1000\pi]$. For this problem we use $\omega = 1$.

**Example 5.4.** (**Inhomogeneous Equation**) Consider the initial value problem

$$y'' = -100y + 99\sin(x), \qquad y(0)=1, \qquad y'(0)=11, \qquad t \in [0,1000\pi].$$

With the exact solution $y(t)=\sin(t)+\sin(10t)+\cos(10t)$. For this problem we use $\omega=1$.

**Example 5.5.** We consider the nonlinear undamped **Duffing equation**

$$y''=-y-y^3+B\cos(\omega x),\qquad y(0)=0.200426728067,\quad y'(0)=0, \tag{12}$$

where $B=0.002$, $\omega=1.01$ and $x\in\left[0,\frac{40.5\pi}{1.01}\right]$. We use the following exact solution for (27), from [23],

$$g(x)=\sum_{i=0}^{3}K_{2i+1}\cos((2i+1)\omega x),$$

where $\{K_1,K_3,K_5,K_7\}=\{0.200179477536, 0.246946143\times10^{-3}, 0.304016\times10^{-6}, 0.374\times10^{-9}\}$.

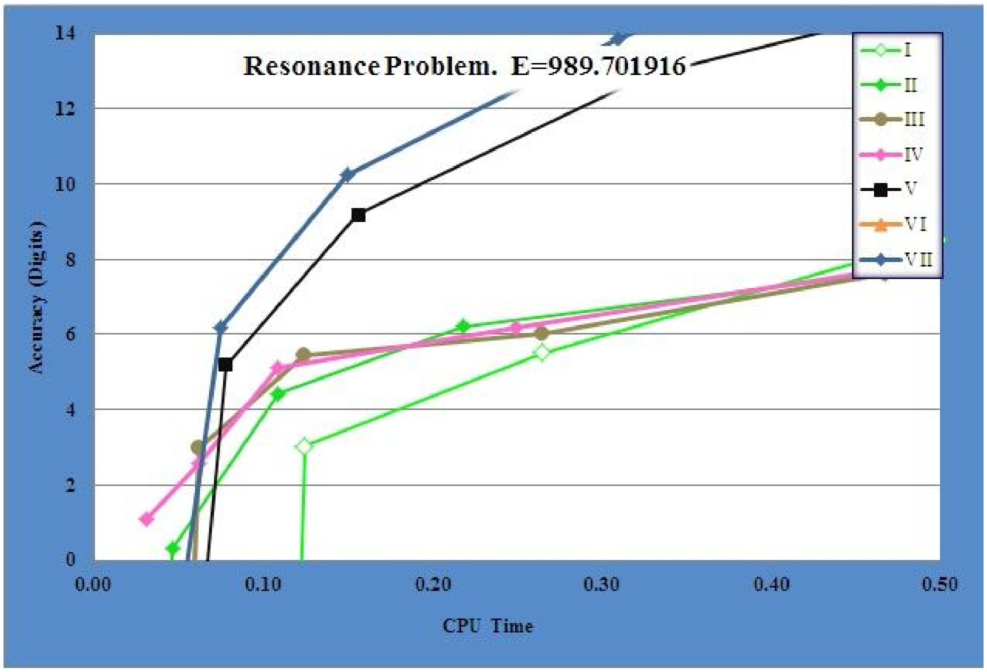


**Figure 7:** Efficiency for the resonance problem using $E$ = 989:701916.

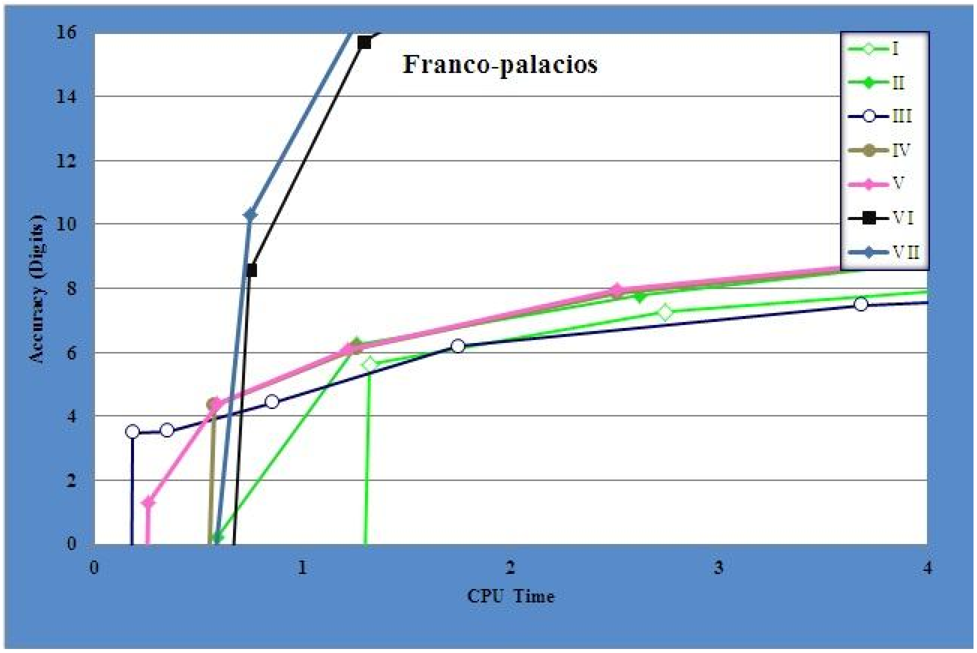


**Figure 8:** Efficiency for the Franco and Palacios equation.

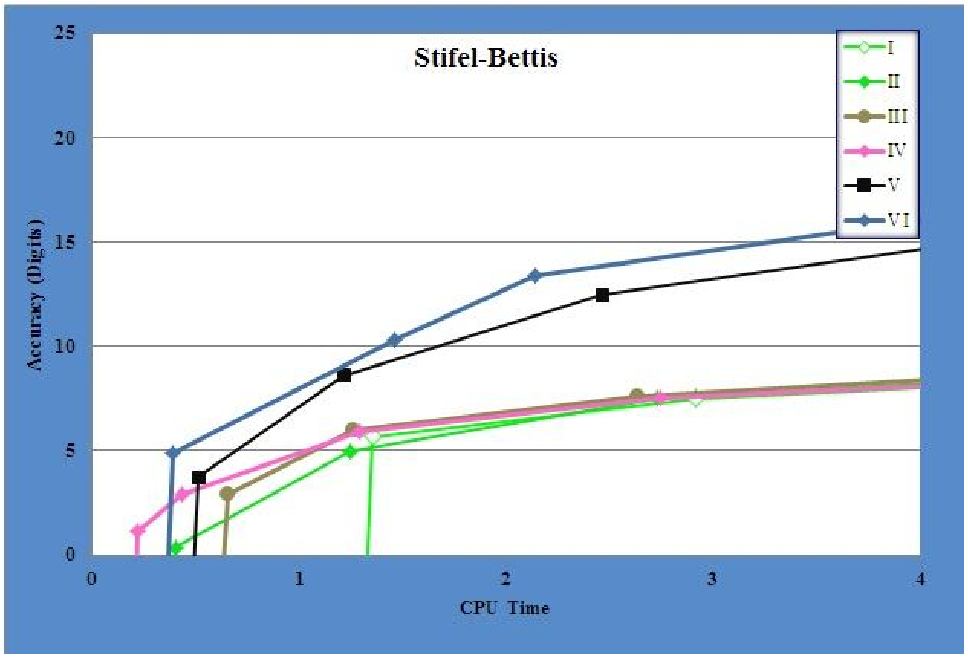


**Figure 9:** Efficiency for the orbital problem by Stiefel and Bettis.

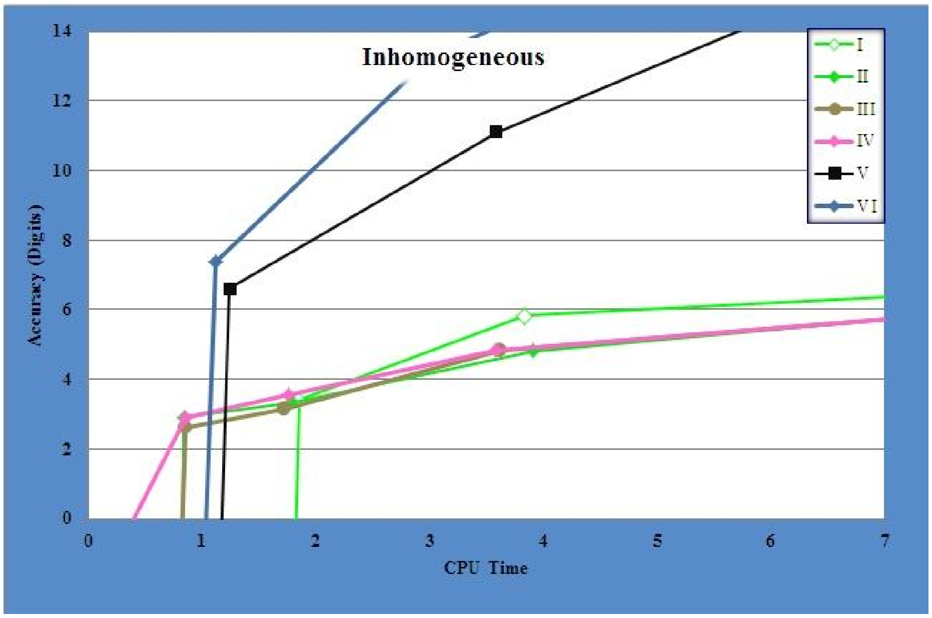


**Figure 10:** Efficiency for the inhomogeneous equation.

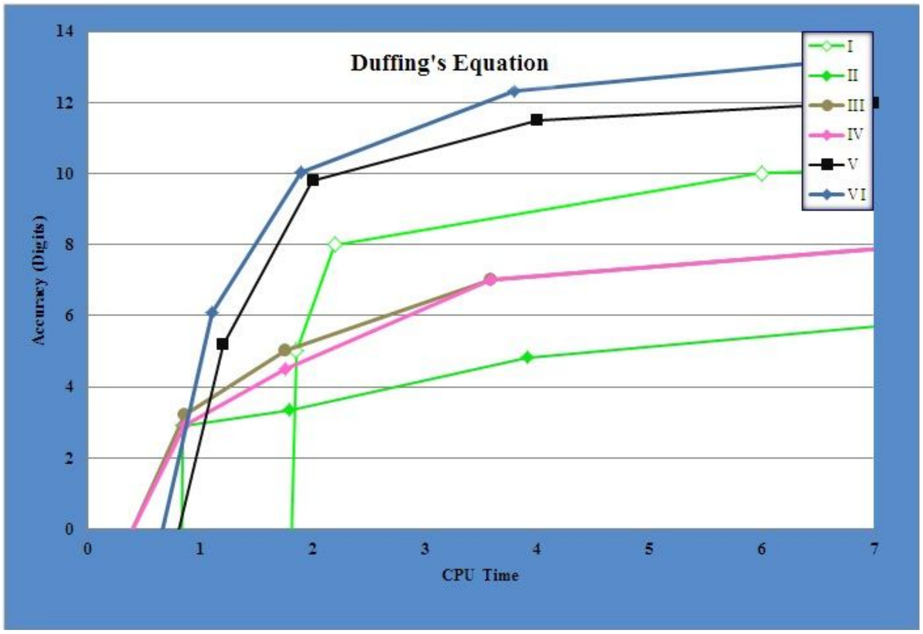


**Figure 11:** Efficiency for the Duffing Equation.

## 4. CONCLUSIONS

In Figure 7, we see the results for the resonance problem for energy $E = 989.495874$. In Figure 8, we see the results for the Franco–Palacios almost periodic problem, in Figure 9, the results for the Stiefel–Bettis almost periodic problem are present, in Figure 10, the results for the inhomogeneous equation are present and finally in Figure 11, we see the results for the Duffing equation.

Among all the methods used the new symmetric two–step Obrechkoff method with twelfth algebraic order and vanished some of its derivatives was the most efficient.

**Acknowledgments.** The authors wish to thank the anonymous referees for their careful reading of the manuscript and their fruitful comments and suggestions.

## REFERENCES

[1] U. Ananthakrishnaiah, P–stable Obrechkoff methods with minimal phase–lag for periodic initial value problems, *Math. Comput.* **49** (1987) 553–559.

[2] M. Asadzadeh, D. Rostamy and F. Zabihi, Discontinuous Galerkin and multiscale variational schemes for a coupled damped nonlinear system of Schrödinger equations, *J. Numer. Methods Partial Differential Equations* **29** (6) (2013) 1912–1945.

[3] M. M. Chawla, P. S. Rao, A Numerov–type method with minimal phase–lag for the integration of second order periodic initial value problems. II: Explicit method, *J. Comput. Appl. Math.* **15** (1986) 329–337.

[4] M. M. Chawla, P. S. Rao, An explicit sixth–order method with phase–lag of order eight for $y'' = f(t, y)$, *J. Comput. Appl. Math.* **17** (1987) 363–368.

[5] G. Dahlquist, On accuracy and unconditional stability of linear multistep methods for second order differential equations, *BIT* **18** (1978) 133–136.

[6] J. M. Franco, An explicit hybrid method of Numerov type for second–order periodic initial–value problems, *J. Comput. Appl. Math.* **59** (1995) 79–90.

[7] J. M. Franco, M. Palacios, High–order P–stable multistep methods, *J. Comput. Appl. Math.* **30** (1990) 1–10.

[8] W. Gautschi, Numerical integration of ordinary differential equations based on trigonometric polynomials, *Numer. Math.* **3** (1961) 381–397.

[9] A. Ibraheem, T. E. Simos, A family of high–order multistep methods with vanished phase–lag and its derivatives for the numerical solution of the Schrödinger equation, *Comput. Math. Appl.* **62** (2011) 3756–3774.

[10] A. Ibraheem, T. E. Simos, A family of ten–step methods with vanished phase–lag and its first derivative for the numerical solution of the Schrödinger equation, *J. Math. Chem.* **49** (2011) 1843–1888.

[11] A. Ibraheem, T. E. Simos, Mulitstep methods with vanished phase–lag and its first and second derivatives for the numerical integration of the Schrödinger equation, *J. Math. Chem.* **48** (2010) 1092–1143.

[12] L. Gr. Ixaru, M. Rizea, A Numerov–like scheme for the numerical solution of the Schrödinger equation in the deep continuum spectrum of energies, *Comput. Phys. Commun.* **19** (1) (1980) 23–27.

[13] M. K. Jain, R. K. Jain and U. Krishnaiah, Obrechkoff methods for periodic initial value problems of second order differential equations, *J. Math. Phys. Sci.* **15** (1981) 239–250.

[14] J. D. Lambert, I. A. Watson, Symmetric multistep methods for periodic initial value problems, *IMA J. Appl. Math.* **18** (1976) 189–202.

[15] G. D. Quinlan, S. Tremaine, Symmetric multistep methods for the numerical integration of planetary orbits, *Astron. J.* **100** (1990) 1694–1700.

[16] D. P. Sakas, T. E. Simos, Multiderivative methods of eighth algebraic order with minimal phase–lag for the numerical solution of the radial Schrödinger equation, *J. Comput. Appl. Math.* **175** (2005) 161–172.

[17] A. Shokri, H. Saadat, Trigonometrically fitted high–order predictor–corrector method with phase–lag of order infinity for the numerical solution of radial Schrödinger equation, *J. Math. Chem.* **52** (2014) 1870–1894.

[18] A. Shokri, H. Saadat, High phase–lag order trigonometrically fitted two–step Obrechkoff methods for the numerical solution of periodic initial value problems, *Numer. Algor.* **68** (2015) 337–354.

[19] A. Shokri, A. A. Shokri, Sh. Mostafavi, H. Saadat, Trigonometrically fitted two-step Obrechkoff methods for the numerical solution of periodic initial value problems, *Iranian J. Math. Chem.* **6** (2015) 145-161.

[20] T. E. Simos, A P-stable complete in phase Obrechkoff trigonometric fitted method for periodic initial value problems, *Proc. Roy. Soc. London Ser. A.* **441** (1993) 283–289.

[21] T. E. Simos, A two–step method with vanished phase–lag and its first two derivatives for the numerical solution of the Schrödinger equation, *J. Math. Chem.* **49** (2011) 2486–2518.

[22] T. E. Simos, Exponentially fitted multiderivative methods for the numerical solution of the Schrödinger equation, *J. Math. Chem.* **36** (2004) 13–27.

[23] T. E. Simos, Multiderivative methods for the numerical solution of the Schrödinger equation, *MATCH Commun. Math. Comput. Chem.* **50** (2004) 7–26.

[24] E. Steifel, D. G. Bettis, Stabilization of Cowells methods, *Numer. Math.* **13** (1969) 154–175.

[25] R. M. Thomas, Phase properties of high order, almost P–stable formulae, *BIT* **24** (1984) 225–238.

[26] M. Van Daele, G. Vanden Berghe, P-stable exponentially fitted Obrechkoff methods of arbitrary order for second order differential equations, *Numer. Algor*. **46** (2007) 333–350.

[27] Z. Wang, D. Zhao, Y. Dai and D. Wu, An improved trigonometrically fitted P–stable Obrechkoff method for periodic initial value problems, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **461** (2005) 1639–1658.

Iranian J. Math. Chem. 8(2) June (2017) 161−173

Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

# Optimal Control of Switched Systems by a Modified Pseudo Spectral Method

**HAMID REZA TABRIZIDOOZ[•], MARZIEH POURBABAEE AND MEHRNOOSH HEDAYATI**

Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Kashan, Kashan 87317–53153, Iran

**ARTICLE INFO**

**Article History:**

Received 18 January 2016
Accepted 16 April 2016
Published online 5 April 2017
Academic Editor: IVAN GUTMAN

**Keywords:**

Optimal control
Switched systems
Legendre pseudo spectral method

**ABSTRACT**

In the present paper, we develop a modified pseudo spectral scheme for solving an optimal control problem which is governed by a switched dynamical system. Many real–world processes such as chemical processes, automotive systems and manufacturing processes can be modeled as such systems. For this purpose, we replace the problem with an alternative optimal control problem in which the switching times appear as unknown parameters. Using the Legendre–Gauss–Lobatto quadrature and the corresponding differentiation matrix, the alternative problem is discretized to a nonlinear programming problem. At last, we examine three examples in order to illustrate the efficiency of the proposed method.

## 1. INTRODUCTION

It is well known that pseudospectral (PS) methods are powerful methods for the numerical solution of differential equations. In fact, they arose from spectral methods which were traditionally used to solve fluid dynamics problems [1, 2]. They can often achieve ten digits of accuracy where a finite difference scheme or a finite element method would get two or three [3]. The key point in PS methods is that they avoid the poor behavior of the classical polynomial interpolation methods by removing the restriction to equally spaced interpolation points.

The variational method of optimal control theory, which typically consists of the calculus of variations and Pontryagin’s methods, can be used to derive a set of necessary conditions that must be satisfied by an optimal control law and its associated state–control equations [4, 5]. These necessary conditions of optimality lead to a generally nonlinear

---

[•] Corresponding Author: (Email address: htabrizidooz@kashanu.ac.ir)

DOI: 10.22052/ijmc.2017.44718

two–point boundary value problem that must be solved to determine the explicit expression for the optimal control. Except in some special cases, the solution of this two–point boundary value problem is difficult and not practical to obtain.

Various alternative computational techniques for optimal control problems have been developed in the literature. The techniques are basically of three types: parameterization on both state and control [6, 7, 8], parameterization on control only [9, 10] and nonparameterization [11, 12, 13]. As a technique of the first type, PS methods can be interpreted as direct transcription methods for discretizing a continuous optimal control problem into a nonlinear programming (NLP) problem [14, 15, 16, 17, 18, 19]. The resulting NLP problem can be solved numerically by the well developed algorithms [20, 21].

Although PS methods enjoy many nice properties, but their use in solving problems with nonsmooth solutions or problems with switches may cause major difficulties. The reason lies in the famous Gibbs phenomenon which happens when a nonsmooth function is approximated by means of a finite number of smooth functions [2]. In [22], the authors developed the method of PS knotting in order to address this issue. In fact, they introduced the concepts of hard and soft knots to eliminate the mentioned difficulties.

The switched systems are a particular class of hybrid systems. The hybrid systems arise in varied contexts in chemical processes, automotive engine control, traffic control, and manufacturing processes, etc. The abundance of hybrid phenomena in many engineering systems in general, and in the chemical process industries in particular has fostered a large and growing body of research work in this area [23, 24, 25, 26, 27, 28, 29, 30]. In [31], the authors discussed important hybrid aspects of chemical processing plants. Recently, optimal control of switched systems arising in fermentation processes has been studied in [32]. A hybrid system consists of several subsystems and a switching law, where the switching law is determined by a switching sequence and a set of switching times. At each time instant, only one subsystem is active. A hybrid system can be described by a differential inclusion of the form

$$x(t) \in \{ f_v(t, x(t), u(t)): \quad v \in \{1,2,\ldots,M\}\}, \tag{1}$$

where $t \in [t_0, t_f]$, $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m$and for each$v \in \{1,2,\ldots,M\}$,$f_v$: $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, is continuously differentiable with respect to its arguments. A switching law $\sigma$ for system (1) is defined as $\sigma = ((t_0, i_0), (t_1, i_1), \ldots, (t_{K-1}, i_{K-1}))$, where $1 \le K < \infty$, $t_0 \le t_1 \le \cdots \le t_{K-1} \le t_f$, and $i_k \in \{1,2,\ldots,M\}$ for $k = 0,1,\ldots,K-1$. Note here $t_1,\ldots,t_{K-1}$ are the switching instants. An optimal control of such a system involves finding a control $u(t)$, and a switching law $\sigma$ such that the corresponding state trajectory subject to the dynamical system (1) departs from a given initial state and minimize a given cost functional. In [33], a method which is based on parameterization of the switching instants is proposed for this kind of optimal control problems.

In this paper, we investigate a modified Legendre PS scheme in order to explore accurate and efficient solutions of optimal control problems for switched systems. Here, we consider the optimal control problems in which a prespecified sequence of active subsystems is given. In order to explore numerical solutions of such problems, we need to seek the solutions of both the optimal switching instants and the optimal piecewise input. The rest of this paper is organized as follows. The problem statement is given in Section 2. In Section 3, we describe the preliminaries for subsequent development. The present method is proposed in Section 4. Then, three examples are provided in Section 5 to illustrate the efficiency of the proposed method. Conclusions are presented in Section 6.

## 2. PROBLEM STATEMENT

We consider switched systems defined on the fixed time interval $[t_0, t_f]$ with $K-1$ switches, consisting of the subsystems

$$x(t) = f_k(x(t), u(t)), \quad t \in [t_{k-1}, t_k), \quad k = 1,2,\ldots,K, \tag{2}$$

with initial conditions

$$x(t_0) = x_0, \tag{3}$$

where $x(t) = (x_1(t), \ldots, x_n(t)) \in \mathbb{R}^n$ is the state function and $u(t) = (u_1(t), \ldots, u_m(t)) \in \mathbb{R}^m$ is the corresponding control function. Also, $f_k: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ , $k = 1,2,\ldots,K$, are given functions. We assume that the switching sequence is preassigned, such that

$$t_0 \le t_1 \le \cdots \le t_{K-1} \le t_K = t_f, \tag{4}$$

where the switching times $t_1, \ldots, t_{K-1}$ are decision variables. Our objective is to find a piecewise continuous function $u(t)$ and switching instants $t_1, \ldots, t_{K-1}$ subject to the condition (4) for the switched system (2) and (3) such that the cost functional

$$J = \phi(x(t_f)) + \int_{t_0}^{t_f} g(x(t), u(t))dt \tag{5}$$

is minimized. It is noted that the considered problem is an optimal control problem in Bolza form. Also, the vector functions $f_k: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $k = 1,2,\ldots,K$, and the scalar functions $g: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $\phi: \mathbb{R}^n \to \mathbb{R}$, are assumed to be smooth with respect to all their arguments.

## 3. PRELIMINARIES

Let $\tau_0 < \tau_1 < \cdots < \tau_N$ be the Legendre–Gauss–Lobatto (LGL) nodes where $\tau_0 = -1$, $\tau_N = 1$ and $\tau_1, \ldots, \tau_{N-1}$ are the roots of $\dot{P}_N(\tau)$. Here $\dot{P}_N(\tau)$ is the derivative of the $N$–th order Legendre polynomial $P_N(\tau)$. In other words, the LGL points $\tau_0, \tau_1, \ldots, \tau_N$ are the $N+1$ roots of $(1-\tau^2)\dot{P}_N(\tau)$. The reader is referred to [1, 34] for details.

Let $h(t)$ be a continuous real function which is defined on $[-1,1]$. The Lagrange interpolating polynomial of degree $N$ interpolates the function $h(t)$ at the points $\tau_0, \tau_1, \ldots, \tau_N$, as

$$h(\tau) = \sum_{j=0}^{N} h(\tau_j) L_j(\tau). \tag{6}$$

Here for $j = 0,1,\ldots,N$, $L_j(\tau)$ denotes the Lagrange polynomial of degree $N$ corresponding to the point $\tau_j$, defined by

$$L_j(\tau) = \prod_{i=0, i\neq j}^{N} \frac{\tau - \tau_i}{\tau_j - \tau_i}.$$

Note that the Lagrange polynomials satisfy in the Kronecker property

$$L_j(\tau_i) = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases}.$$

In order to approximate the derivative of $h(t)$ at the points $\tau_i$, $i = 0,1,\ldots,N$, the interpolation formula (6) is differentiated yielding

$$\dot{h}(\tau_i) = \sum_{j=0}^{N} d_{ij} h(\tau_j), \tag{7}$$

where $d_{ij} = \dot{L}_j(\tau_i)$. The $(N+1)\times(N+1)$ matrix $D = [d_{ij}]$ is the so–called derivative matrix. According to [1]

$$d_{ij} = \begin{cases} \dfrac{P_N(\tau_i)}{P_N(\tau_j)} \cdot \dfrac{1}{\tau_i - \tau_j}, & i \neq j \\ \dfrac{-N(N+1)}{4}, & i = j = 0 \\ \dfrac{N(N+1)}{4}, & i = j = N \\ 0, & otherwise \end{cases}.$$

Furthermore, for approximating the definite integral of $h(t)$ on $[-1,1]$, the LGL quadrature rule is used. According to this quadrature rule, the definite integral is replaced by a summation, in which the values of $h(t)$ at the LGL points are utilized, as

$$\int_{-1}^{1} h(\tau)d\tau = \sum_{j=0}^{N} w_j h(\tau_j), \tag{8}$$

where $w_j$, $j = 0,1,\ldots,N$, are the LGL weights, corresponding to the LGL points $\tau_j$, $j = 0,1,\ldots,N$, given by

$$w_j = \frac{2}{N(N+1)} \cdot \frac{1}{\left[P_N(\tau_j)\right]^2}, \qquad j = 0,1,\ldots,N.$$

## 4. PROPOSED METHOD

We suppose that in the problem stated in the Eqs. (2)–(5), the switching sequence is preassigned and $t_1,\ldots,t_{K-1}$ are the corresponding unknown switching times for which the condition (4) holds.

We denote the restriction of vector functions $x(t)$ and $u(t)$ to the $k$–th subinterval $[t_{k-1},t_k)$ by $x^k(t)$ and $u^k(t)$, respectively. According to these notations, the dynamic subsystems in Eq. (2) are expressed as

$$x^k(t) = f_k(x^k(t),u^k(t)), \quad t_{k-1} \le t < t_k, \quad k = 1,\ldots,K, \tag{9}$$

$$x^k(t_{k-1}) = \lim_{t \to t_{k-1}^-} x^{k-1}(t), \qquad k = 2,\ldots,K. \tag{10}$$

Note that in Eq. (10), the continuity constraints are added in order to guarantee the continuity of state functions. Accordingly, the cost functional (5) reformulated as

$$J = \phi(x^K(t_f)) + \sum_{k=1}^{K} \int_{t_{k-1}}^{t_k} g(x^k(t),u^k(t))dt, \tag{11}$$

and the initial conditions (3) restated as

$$x^1(t_0) = x_0. \tag{12}$$

To apply the approximations described in the previous section, we must transfer each subinterval to the interval $[-1,1]$. For this purpose, we use the transformation formula $\tau = \dfrac{2t - (t_{k-1} + t_k)}{t_k - t_{k-1}}$ in the $k$–th subinterval $[t_{k-1},t_k)$. In this respect, the problem is restated in the following alternative form:

$$\min J = \phi(x^K(1)) + \sum_{k=1}^{K} \left(\frac{t_k - t_{k-1}}{2}\right) \int_{-1}^{1} g(x^k(\tau),u^k(\tau))d\tau \tag{13}$$

$$s.t. x^k(\tau) = \left(\frac{t_k - t_{k-1}}{2}\right) f_k(x^k(\tau),u^k(\tau)), \quad k = 1,\ldots,K, \tag{14}$$

$$x^k(-1) = x^{k-1}(1), \qquad k = 2,\ldots,K, \tag{15}$$

$$x^1(-1) = x_0. \tag{16}$$

The alternative problem (13)–(16) provides us with some advantage, namely that it no longer has varying switching instants. In fact, the switching instants are considered as parameters in the alternative problem.

It has to be noted that for $k = 1,\dots,K$, the components of vector functions $x^k(\tau)$ and $u^k(\tau)$ are smooth on $[-1,1]$ and then can be expanded in terms of Lagrange basis functions according to Eq. (6). Therefore, using the formula (8), the performance index $J$ in Eq. (13) is approximated as

$$J = \phi(X_N^{(K)}) + \sum_{k=1}^{K}\left(\frac{t_k - t_{k-1}}{2}\right)\sum_{j=0}^{N} g(X_j^{(k)}, U_j^{(k)}) w_j, \tag{17}$$

where $X_j^{(k)}$ and $U_j^{(k)}$ are vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and defined by

$$X_j^{(k)} = x^k(\tau_j), U_j^{(k)} = u^k(\tau_j), \quad j = 0,1,\dots,N, \quad k = 1,\dots,K.$$

Also, using the formula (7), the alternative dynamical systems (14) are approximated by

$$DX^{(k)} - \left(\frac{t_k - t_{k-1}}{2}\right)F^{(k)} = 0, \quad k = 1,\dots,K, \tag{18}$$

where $X^{(k)}$ and $F^{(k)}$ are $(N+1)\times n$ matrices, respectively, defined by

$$X^{(k)} = \begin{bmatrix} X_0^{(k)} \\ X_1^{(k)} \\ \vdots \\ X_N^{(k)} \end{bmatrix}, \quad F^{(k)} = \begin{bmatrix} f_k(X_0^{(k)}, U_0^{(k)}) \\ f_k(X_1^{(k)}, U_1^{(k)}) \\ \vdots \\ f_k(X_N^{(k)}, U_N^{(k)}) \end{bmatrix}.$$

Furthermore, the continuity constraints (15) and the initial conditions (16), respectively, are stated as

$$X_0^{(k)} - X_N^{(k-1)} = 0, \quad k = 2,\dots,K, \tag{19}$$

and

$$X_0^{(1)} = x_0. \tag{20}$$

We also assume that no two endpoints of subintervals coincide. Then, for a small given $\varepsilon > 0$, we add the extra constraints

$$t_k - t_{k-1} > \varepsilon, \quad k = 1,\dots,K. \tag{21}$$

In summary, the alternative optimal control (13)–(16) is discretized to the following NLP problem: Find vectors $X_j^{(k)}$, $U_j^{(k)}$, $j = 0,1,\dots,N$, $k = 1,\dots,K$ and the parameters $t_k$, $k = 1,\dots,K-1$ to minimize the expression (17) subject to the constraints (18)–(21).

The relations between the solutions of obtained NLP problem and the solutions of alternative problem (13)–(16) are given by

$$x^{(k)}(\tau)=\sum_{j=0}^{N}X_j^{(k)}L_j(\tau), \qquad k=1,\ldots,K,$$

and

$$u^{(k)}(\tau)=\sum_{j=0}^{N}U_j^{(k)}L_j(\tau), \qquad k=1,\ldots,K.$$

## 5. ILLUSTRATIVE EXAMPLES

In this section, we consider three examples to illustrate the efficiency of proposed method. Here, we consider the numerical examples given in [33]. According to the present method, each example in modeled using the mathematical software package Maple 17 and the resulting NLP problems are solved by the command NLPSolve.

**Example 1.** Consider a switched system consisting of nonlinear subsystems

$$\begin{aligned}
&subsystem1: \begin{cases}\dot{x}_1(t) &= x_1(t)+u(t)\sin x_1(t)\\ \dot{x}_2(t) &= -x_2(t)-u(t)\cos x_2(t)\end{cases},\\
&subsystem2: \begin{cases}\dot{x}_1(t) &= x_2(t)+u(t)\sin x_2(t)\\ \dot{x}_2(t) &= -x_1(t)-u(t)\cos x_1(t)\end{cases},\\
&subsystem3: \begin{cases}\dot{x}_1(t) &= -x_1(t)-u(t)\sin x_1(t)\\ \dot{x}_2(t) &= x_2(t)+u(t)\cos x_2(t)\end{cases}.
\end{aligned}$$

Assume that $t_0=0$, $t_f=3$ and the system switches at $t=t_1$ from subsystem 1 to 2 and at $t=t_2$ from subsystem 2 to 3 ($0\le t_1\le t_2\le 3$). The initial conditions are $x_1(0)=2$ and $x_2(0)=3$. We want to find optimal switching instants $t_1$, $t_2$ and an optimal input $u(t)$ such that the cost functional

$$J=\frac{1}{2}(x_1(3)-1)^2+\frac{1}{2}(x_2(3)+1)^2+\frac{1}{2}\int_0^3[(x_1(t)-1)^2+(x_2(t)+1)^2+u^2(t)]dt$$

is minimized.

In Table 1, we listed the results of optimal switching instants $t_1$, $t_2$ and optimal cost $J$ obtained by the present method with $K=3$ and different values of $N$. In the last of Table 1, we reported the CPU time (seconds) for the computations of the corresponding results. Also, in Figure 1, we plot the graphs of optimal control and the corresponding state trajectory obtained by the present method with $K=3$ and $N=9$.column

**Table 1.** The results of optimal switching instants $t_1$, $t_2$ and optimal cost $J$ obtained by the present method with $K = 3$ and different values of $N$, for Example 1.

| | $t_1$ | $t_2$ | $J$ | CPU time (seconds) |
|---|---|---|---|---|
| $K = 3$ $N = 6$ | 0.22451889 | 1.01940266 | 5.44119735 | 4.04 |
| $N = 8$ | 0.22452199 | 1.02006802 | 5.44100709 | 5.03 |
| $N = 10$ | 0.22451866 | 1.02002342 | 5.44097522 | 6.20 |
| $N = 12$ | 0.22451838 | 1.02002491 | 5.44097350 | 8.15 |
| $N = 14$ | 0.22451835 | 1.02002485 | 5.44097350 | 9.96 |

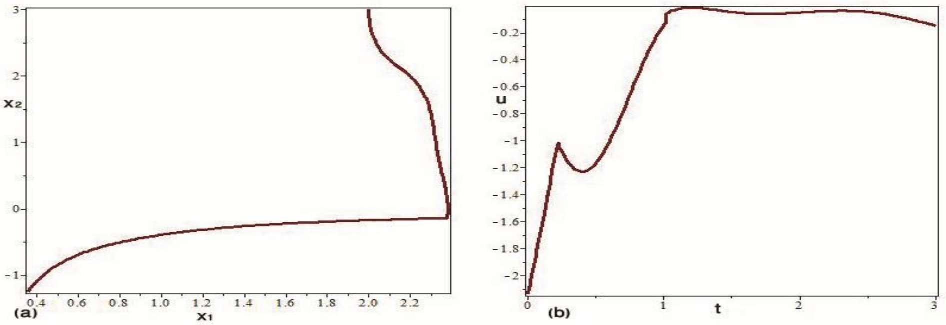


**Figure 1:** The graphs of (a) state trajectory and (b) optimal control obtained by the present method with $K = 3$ and $N = 9$, for Example 1.

**Example 2.** Consider a switched system consisting

$$
\begin{aligned}
\mathit{subsystem}\,1: &\quad \begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix} = \begin{bmatrix}0.6 & 1.2\\ -0.8 & 3.4\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix} + \begin{bmatrix}1\\ 1\end{bmatrix}u(t),\\
\mathit{subsystem}\,2: &\quad \begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix} = \begin{bmatrix}4 & 3\\ -1 & 0\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix} + \begin{bmatrix}2\\ -1\end{bmatrix}u(t).
\end{aligned}
$$

Assume that $t_0 = 0$, $t_f = 2$ and the system switches once at $t = t_1$ ($0 \le t_1 \le 2$) from subsystem 1 to 2. The initial conditions are $x_1(0) = 0$ and $x_2(0) = 2$. We want to find an optimal switching instant $t_1$ and an optimal input $u(t)$ such that the cost functional

$$J = \frac{1}{2}(x_1(2) - 4)^2 + \frac{1}{2}(x_2(2) - 2)^2 + \frac{1}{2}\int_0^2 [(x_2(t) - 2)^2 + u^2(t)]dt$$

is minimized.

We applied the proposed method to solve this example. In Table 2, we reported the results of $t_1$ and $J$ obtained by the present method with $K=2$ and different values of $N$. Also in Figure 2, we plot the graphs of optimal control and the corresponding state trajectory with $K=2$ and $N=9$.

**Table 2.** The results of optimal switching instant $t_1$ and optimal cost $J$ obtained by the present method with $K=2$ and different values of $N$, for Example 2.

| | $t_1$ | $J$ | CPU time (seconds) |
|---|---|---|---|
| $K=2$ $N=6$ | 0.19007133 | 9.78402619 | 2.62 |
| $N=8$ | 0.18967215 | 9.76657993 | 3.04 |
| $N=10$ | 0.18967109 | 9.76654884 | 3.46 |
| $N=12$ | 0.18967110 | 9.76654882 | 4.42 |
| $N=14$ | 0.18967107 | 9.76654882 | 5.60 |

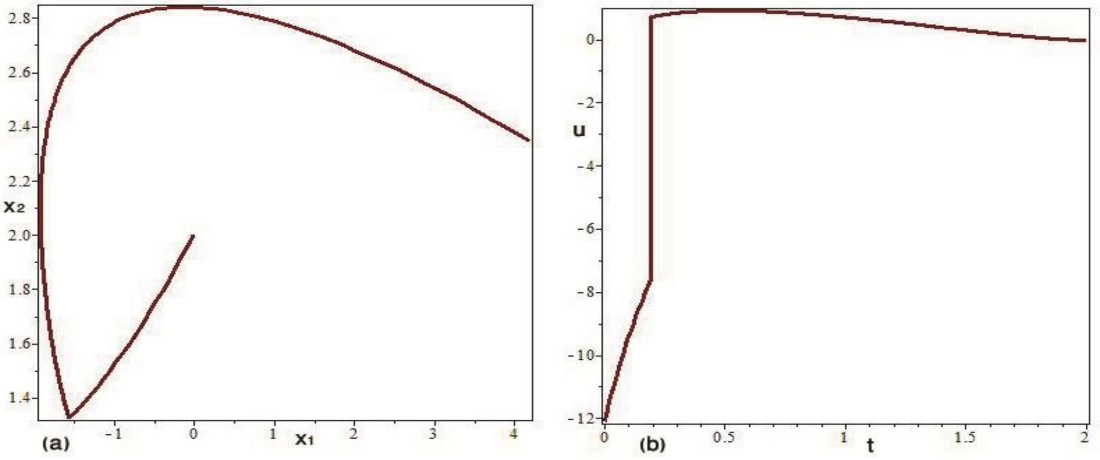


**Figure 2:** The graphs of (a) state trajectory and (b) optimal control obtained by the present method with $K=2$ and $N=9$, for Example 2.

**Example 3.** Consider a switched system with internally forced switching only consisting of

$$subsystem1: \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 1.5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t),$$

$$subsystem2: \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.866 \\ 0.866 & -0.5 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u(t).$$

Assume that $t_0 = 0$, $t_f = 2$ and the system state starts at $x_1(0) = 1$, $x_2(0) = 1$, following subsystem 1 (subsystem 1 is active for $c(x_1(t), x_2(t)) = x_1(t) + x_2(t) - 7 \le 0$ and subsystem 2 is active for $c(x_1(t), x_2(t)) \ge 0$). Assume that upon intersecting the hyper surface $c(x_1, x_2) = 0$, the system switches from subsystem 1 to 2. Also, assume there is only one switching which takes place at time $t_1$ $(0 \le t_1 \le 2)$. We want to find an optimal input $u(t)$ such that the cost functional

$$J = \frac{1}{2}(x_1(2) - 10)^2 + \frac{1}{2}(x_2(2) - 6)^2 + \frac{1}{2}\int_0^2 u^2(t)dt$$

is minimized.

Note that we have not considered state constraints in the subsystems of our problem modeled by Eqs. (2)–(5). For this reason, we state our technique in order to approximate state constraints. By setting $K = 2$, according to the proposed method, we have two sets of state functions values: $X_j^{(1)}$, $j = 0,1,\dots,N$, are the values of state functions in subsystem 1, and $X_j^{(2)}$, $j = 0,1,\dots,N$, are the values of state functions in subsystem 2. According to this, we obtain the constraints $c(X_j^{(1)}) \le 0$, $j = 0,1,\dots,N$, in subsystem 1, and $-c(X_j^{(2)}) \le 0$, $j = 0,1,\dots,N$, in subsystem 2. These new inequality constraints must be added to Eqs. (18)–(21).

In Table 3, we listed the results of $t_1$ and $J$ obtained by the present method with $K = 2$ and different values of $N$. Also in Figure 3, we plot the graphs of optimal control and the corresponding state trajectory with $K = 2$ and $N = 9$.

**Table 3**. The results of optimal switching instant $t_1$ and optimal cost $J$ obtained by the present method with $K = 2$ and different values of $N$, for Example 3.

| | $t_1$ | $J$ | CPU time (seconds) |
|---|---|---|---|
| $K = 2$ $N = 6$ | 1.16328653 | 0.11315919 | 5.78 |
| $N = 8$ | 1.16293205 | 0.11309541 | 6.00 |
| $N = 10$ | 1.16278027 | 0.11306590 | 6.46 |
| $N = 12$ | 1.16270441 | 0.11305120 | 6.84 |
| $N = 14$ | 1.16266144 | 0.11304283 | 7.42 |

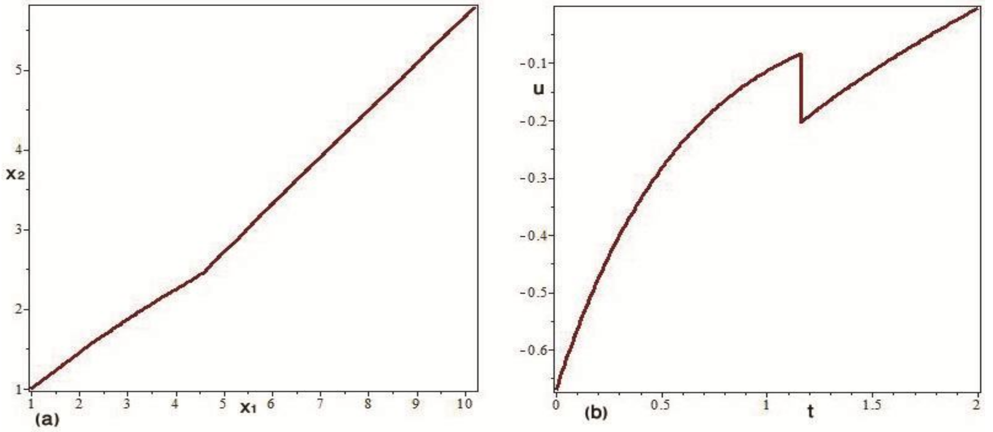


**Figure 3:** The graphs of (a) state trajectory and (b) optimal control obtained by the present method with $K = 2$ and $N = 9$, for Example 3.

## 6. CONCLUSION

In this paper, we have considered a class of optimal control problems governed by switched systems. Such systems arise in varied contexts in chemical processes, automotive engine control, traffic control, and manufacturing processes, etc. We have proposed a modified Legendre pseudospectral scheme in order to explore accurate solutions. For this purpose, we have restated the problem in form of an alternative problem in which the switching instants are considered as parameters. Then, we can solve the obtained NLP problem using existing subroutines. Three numerical examples considered in order to show the validity and applicability of the proposed method.

## REFERENCES

1. C. Canuto, M. Y. Hussaini, A. Quarteroni and T. A. Zang, *Spectral Methods: Fundamentals in Single Domains*, Springer–Verlag, Berlin, 2006.
2. B. Fornberg, *A Practical Guide to Pseudospectral Methods*, Cambridge University Press, 1996.
3. L. N. Trefethen, *Spectral Methods in MATLAB*, SIAM, Philadelphia, 2000.
4. D. E. Kirk, *Optimal Control Theory*, Prentice–Hall, Englewood Cliffs, New Jersey, 1970.
5. L. S. Pontryagin, V. Boltyanskii, R. Gamkrelidze and E. Mischenko, *The Mathematical Theory of Optimal Processes*, Interscience Publishers, New York, 1962.
6. C. P. Neuman and A. Sen, A suboptimal control algorithm for constrained problems using cubic splines, *Automatica* **9** (1973) 601–613.

7. J. Vlassenbroeck, A Chebyshev polynomial method for optimal control with state constraint, *Automatica* **24** (1988) 499–506.
8. J. Vlassenbroeck and R. Van Doreen, A Chebyshev technique for solving nonlinear optimal control problems, *IEEE Trans. Automat. Control* **33** (1988) 333–340.
9. C. J. Goh and K. L. Teo, Control parametrization: a unified approach to optimal control problems with general constraint, *Automatica* **24** (1988) 3–18.
10. O. Rosen and R. Luus, Evaluation of gradients for piecewise constraint optimal control, *Computers and Chemical Engineering* **15** (1991) 273–281.
11. W. W. Hager, Multiplier methods for nonlinear optimal control, *SIAM J. Numer. Anal*. **27** (1990) 1061–1080.
12. D. H. Jacobson and M. M. Lele, A Transformation technique for optimal control problems with a state variable inequality constraints, *IEEE Trans. Automat. Control* AC–14 (1969) 457–464.
13. E. Polak, T. H. Yang and D. Q. Mayne, A method of centers based on barrier function methods for solving optimal control problems with continuum state and control constraints, *SIAM J. Control Optim*. **31** (1993) 159–179.
14. G. N. Elnagar, M. A. Kazemi and M. Razzaghi, The pseudospectral Legendre method for discretizing optimal control problems, *IEEE Trans. Automat. Control* **40** (1995) 1793–1796.
15. G. N. Elnagar and M. A. Kazemi, Pseudospectral Chebyshev optimal control of constrained nonlinear dynamical systems, *Comput. Optim. Appl*. **11** (1998) 195–217.
16. F. Fahroo and I. M. Ross, Costate estimation by a Legendre pseudospectral method, *J. Guid. Control Dyn.* **24** (2001) 270–277.
17. G. T. Huntington, D. A. Benson and A. V. Rao, Post–optimality evaluation and analysis of a formation flying problem via a Gauss pseudospectral method, 2005, AAS paper no. 05–339.
18. Q. Gong, W. Kang and I. M. Ross, A pseudospectral method for the optimal control of constrained feedback linearizable systems, *IEEE Trans. Automat. Control* **51** (2006) 1115–1129.
19. M. Shamsi, A modified pseudospectral scheme for accurate solution of Bang–Bang optimal control problems, *Optimal Control Appl. Methods* **32** (2011) 668–680.
20. R. Fletcher, *Practical Methods of Optimization*, John Wiely, Chichester, 1987.
21. J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer, New York, 1999.
22. I. M. Ross and F. Fahroo, Pseudospectral knotting methods for solving optimal control problems, *J. Guid. Control Dyn.* **27** (2004) 397–405.

23. I. E. Grossmann, S. A. Van Den Heever and I. Harjukoski, Discrete optimization methods and their role in the integration of planning and scheduling, *AIChE Symposium Series* **98** (2002) 150–168.
24. N. H. El–Farra, P. Mhaskar and P. D. Christofides, Feedback control of switched nonlinear systems using multiple Lyapunov functions, in Proceedings of American Control Conference, pages 3496–3502, Arlington, VA, 2001.
25. R. A. Decarlo, M. S. Branicky, S. Petterson and B. Lennartson, Perspectives and results on the stability and stabilizability of hybrid systems, *Proceedings of the IEEE* **88** (2000) 1069–1082.
26. A. Bemporad and M. Morari, Control of systems integrating logic, dynamics and constraints, *Automatica* **35** (1999) 407–427.
27. B. Hu, X. Xu, P. J. Antsaklis and A. N. Michel, Robust stabilizing control law for a class of second–order switched systems, *Systems Control Lett.* **38** (1999) 197–207.
28. N. H. El–Farra, P. Mhaskar and P. D. Christofides, Output feedback control of switched nonlinear systems using multiple Lyapunov functions, *Systems Control Lett.* **54** (2005) 1163–1182.
29. R. Ghosh and C. Tomlin, Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta–Notch protein signalling, *Syst. Biol.* **1** (2004) 170–183.
30. P. G. Howlett, P. J. Pudney and X. Vu, Local energy minimization in optimal train control, *Automatica* **45** (2009) 2692–2698.
31. S. Engell, S. Kowalewski, C. Schulz and O. Stursberg, Continuous–discrete interactions in chemical processing plants, *Proceedings of the IEEE* **88** (2000) 1050–1068.
32. C. Liu and Z. Gong, *Optimal Control of Switched Systems Arising in Fermentation Processes*, Springer–Verlag, Berlin Heidelberg, 2014.
33. X. Xu and P.J. Antsaklis, Optimal control of switched systems based on parameterization of the switching instants, *IEEE Trans. Automat. control* **49** (2004) 2–16.
34. P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, New York, 1984.

Iranian J. Math. Chem. 8(2) June (2017) 175–180

Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

# Computing Szeged Index of Graphs on Triples

M. R. DARAFSHEH[a] , R. MODABERNIA[b] AND M. NAMDARI[b,•]

[a] School of Mathematics, College of Science, University of Tehran, Tehran, Iran
[b]Department of Mathematics, Shahid Chamran University of Ahvaz, Ahvaz, Iran

**ARTICLE INFO**

**Article History:**

Received 20 October 2016
Accepted 3 April 2017
Published online April 5 2017
Academic Editor: Hassan Yousefi-Azari

**Keywords:**

Szeged index
Intersection graph
Automorphism of graph

**ABSTRACT**

Let G=(V,E) be a simple connected graph with vertex set V and edge set E. The Szeged index of G is defined by $Sz(G) = \sum_{e=uv\in E} n_u(e|G)n_v(e|G)$, where $n_u(e|G)$ is the number of vertices of $G$ closer to u than v and $n_v(e|G)$ can be defined in a similar way. Let S be a set of size $n \geq 8$ and V be the set of all subsets of S of size 3. We define three types of intersection graphs with vertex set V. These graphs are denoted by $G_i(n)$, i=0,1,2 and we will find their Szeged indices.

## 1. INTRODUCTION

Let $G = (V,E)$ be a simple graph with vertex set $V$ and edge set $E$. An automorphism of $G$ is a one–to–one mapping $\sigma : V \rightarrow V$ that preserves adjacency of vertices in $G$. The distance between two vertices $u$ and $v$ is the length of a shortest path from $u$ to $v$ and is denoted by $d(u,v)$. A function $f$ from the set of all graphs into real numbers is called a *graph invariant* if and only if $G \cong H$ implies that $f(G) = f(H)$. A graph invariant is said to be *distance−based* if it can be can defined by distance function $d(-,-)$. A graph invariant applicable in chemistry is called a *topological index*.

In recent research in mathematical chemistry, distance–based graph invariants are of particular interest. One of the oldest descriptors concerned with the molecular graph is the *Wiener index*, which was proposed by Wiener [8]. The definition of theWiener index in terms of distances between vertices of a graph is due to Hosoya ]6].

The *Szeged index* [4,5,7] is a topological index closely related to the the Wiener index and coincides with the Wiener index in the case when the graph is a tree. For the

---

•Corresponding Author (Email: namdari@ipm.ir)

DOI: 10.22052/ijmc.2017.80007.1275

basic definition of the Szeged index of graph, let $G = (V,E)$ be a connected simple graph. Let $e = uv$ be an edge of $G$. We define two subsets of vertices of $G$ as follows:

$$N_u(e|G) = \{w \in V \mid d(w,u) < d(w,v)\}$$
$$N_v(e|G) = \{w \in V \mid d(w,v) < d(w,u)\}$$

Let $n_u(e|G) = |N_u(e|G)|$ and $n_v(e|G) = |N_v(e|G)|$. The Szeged index of the graph $G$ is defined by the following formula:

$$Sz(G) = \sum\nolimits_{e=uv\in E} n_u(e|G)n_v(e|G)$$

We see that the Szeged index is a sum of edge–contribution for the edge $e = uv$ of the graph $G$, we set $sz(e) = n_u(e|G)n_v(e|G)$ , hence $Sz(G) = \sum\nolimits_{e\in E} sz(e)$.

Let $\Gamma$ denote the automorphism group of the graph $G$. Then $\Gamma$ acts as a permutation group on the vertex set $V$ of $G$. If $e = uv$ is an edge of $G$ and $\sigma \in \Gamma$, then by defining $e^\sigma = u^\sigma v^\sigma$, we observe that $\Gamma$ acts on the set $E$ of edges of $G$. If $\Gamma$ acts transitively on $V$, then $G$ is called a vertex–transitive graph and if it acts transitively on $E$, then $G$ is called an edge–transitive graph. We refer the reader to the book ]2] for further reading about permutation groups.

In [1], the case of edge–transitive graph is studied. In this case, the edge–distribution at each edge is the same, i.e., $sz(e) = sz(e')$ for all edges $e$ and $e'$ of $G$ holds, hence $Sz(G) = |E|\, sz(e)$ for a single edge of $G$ holds. The above situation is also studied in [9].

## 2. PRELIMINARY RESULTS

In this paper we are concerned with the graphs on triples. Let $S$ be a set of size $n$ where $n$ is a suitable natural number. Let $V$ be the set of all the 3–element subsets of $S$. The graph $G_i, i = 0, 1, 2,$ called intersection graphs, are defined as $G = (V, E_i)$, where $V$ is the set of vertices of $G$ and two vertices are joined by an edge if and only if they intersect in $i$ elements. It is clear that $|V| = \binom{n}{3}$ and the size of each $E_i; i = 0,1,2,$ is $\binom{n-3}{3}, 3\binom{n-3}{2}$ and $3(n-3)$ respectively, it is worth mentioning that the Weiner indices of the graphs $G_i; i = 0,1,2,$ were computed in [3].

**Lemma 2.1.** Each of the graphs $G_i; i = 0,1,2,$ is edge–transitive.

**Proof.** By [3], the automorphism graph of each graph $G_i; i = 0,1,2,$ has a subgroup isomorphic to the symmetric group $S_n$. Let $e = uv$ and $e' = u'v'$ be two edges of $G_i; i = 0,1,2.$ Then $|u \cap v| = i = |u' \cap v'|$.

**Case 1**. $i = 0.$ In this case we may take $u = \{1,2,3\}, v = \{4,5,6\},$ $u' = \{1',2',3'\}, v' = \{4',5',6'\}$ where $\{1',2',\ldots,6'\} \subseteq \{1,2,\cdots,6\}$ .The permutation $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1' & 2' & 3' & 4' & 5' & 6' \end{pmatrix} \in S_n$ take $e$ to $e'$.

**Case 2**. $i = 1.$ In this case we may take $u$ = {1,2,3}, $v$ = {1,4,5}, $u'$ = {1′, 2′, 3′}, $v'$= {1′, 4′, 5′} and choose $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1' & 2' & 3' & 4' & 5' \end{pmatrix} \in S_n$ which takes e to $e'$ .

**Case 3**. $i = 2.$ In this case we may choose $u$ = {1,2,3}, $v$ = {1,2,4}, $u'$ = {1′, 2′, 3′}, $v'$= {1′, 2′, 4′} and in this case $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1' & 2' & 3' & 4' \end{pmatrix} \in S_n$ takes e to $e'$ . ■

We have the following result from [3] that will be used.

**Result 2.1.** Let $u$ and $v$ be two vertices of $G_i; i = 0,1,2.$ Then $d(u,v) \leq 2$ unless $i = 2$ where $d(u,v) = 3$ also occurs.

## 3. COMPUTATION OF THE SZEGED INDEX

Now because of Lemma 2.1, we have $Sz(G_i) = |E_i| sz(e)$, $i$ = 0, 1, 2, where $sz(e) = n_u(e|G_i)n_v(e|G_i)$. By definition we have $n_u(e|G_i) = |\{ w \in V \mid d(w,u) < d(w,v)\}|$. By the above result $d(w,v)$ = 0, 1, 2 in the case $G_1$ and $G_2$.

**Case 1**. $d(w,v) = 0$ is impossible.

**Case 2.** If $d(w,v) = 1,$ then $d(w,v) = 0$ implying $w = u$.

**Case3.** If $d(w,v) = 2,$ then $d(w,v) = 0$ or 1. If $w = u,$ then $d(u,v) = 1$ a contradiction, hence $d(w,u) = 1.$ We conclude that

$$n_u(e \mid G) = 1 + |\{v \neq w \in V \mid d(w,u) = 1\}|.$$

By symmetry we have $n_v(e \mid G) = n_u(e \mid G).$

**Corollary 3.1.** The Szeged index of $G_0$ and $G_1$ are as follows:

$$Sz(G_0)=\binom{n-3}{3}\left(1+3\binom{n-6}{2}+3\binom{n-6}{1}\right)^2$$

$$Sz(G_1)=3\binom{n-3}{2}\left(1+2\binom{n-5}{2}+2\binom{n-5}{1}\right)^2$$

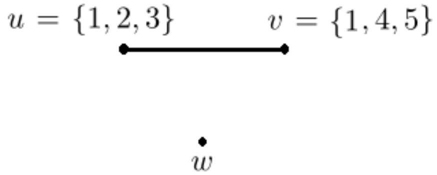
**Proof**. According to what we proved earlier $Sz(G_i)=|E_i|\,sz(e),$ where $e=uv$ is a fixed edge of $G_i, i=0,1.$ But

$$sz(e)=n_u(e\,|\,G)\,n_v(e\,|\,G)=n_u(e\,|\,G)^2$$
$$=\left(1+|\{v\neq w\in V\,|\,d(w,u)=1\}|\right)^2$$

Therefore we must find the number of vertices $w\neq v$ of $V$ with distance 1 from $u$.

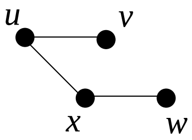
**Case 1.** $i=0$. In this case we may take $u=\{1,2,3\}$ and $v=\{4,5,6\}$, the vertex w should be of distance 2 from v, hence should meet v and $w\cap u=\phi$. If w meets v in one element we have 3/2(n − 6)(n − 7) choices for it and if it meets v in 2 elements again we have $3(n-6)$ choices for it and the formula for $Sz(G_0)$ is obtained as above.

**Case 2**. $i=1$. In this case we may choose $u=\{1,2,3\}, v=\{1,4,5\}$. we have $d(w,v)=2$, hence $w\cap v=\phi$ or $|w\cap v|=2$, but $|w\cap u|=1$.

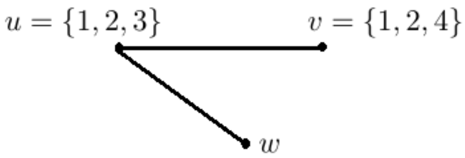


If $w\cap v=\phi$, then we have $(n-5)(n-6)$ choices for W. If $|w\cap v|=2$, then if $1\in w$, we must have $w=\{1,4,x\}$ or $w=\{1,5,y\}$, hence the number of choices for w is $2(n-5)$ .For $1\notin w$ we don't obtain a possibility for w. Therefore $Sz(G_1)$ is as above. ■

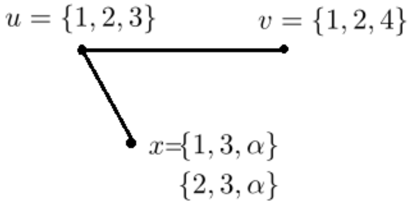
To calculate the Szeged index of $G_2$ we must calculate the size of the set $N_u(e\,|\,G)=\{w\in V\,|\,d(w,u)<d(w,v)\}$ .In this case $d(w,v)=3$ may occur and $d(w,u)=1$ or 2. If $d(w,u)=1$, then $d(w,u)=2$, a contradiction. Therefore $d(w,u)=2$, i.e there is a vertex $x$ such that $d(w,x)=1$. If we set $A_1=\{v\neq w\in V\,|\,d(w,u)=1\}$ and $A_2=\{w\in V\,|\,d(w,u)=2\}$ then we must find the sizes of $A_1$ and $A_2$.



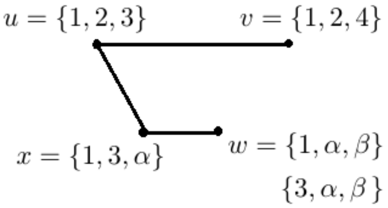
Let $u=\{1,2,3\}$, $v=\{1,2,4\}$ and find $|A_1|$.



In this case $d(w,v)=2,$ hence $|w\cap v|\neq 2.$ If $w\cap v=\phi,$ then there is no possibility for $w$. If $|w\cap v|=1,$ then $w=\{1,3,x\},\{2,3,x\},$ and hence the following corollary is proved. There are 2($n$ – 4) possibilities for $w$ and $|A_1|$ = 2($n$ – 4). To find $|A_2|$ we may assume again $u=\{1,2,3\},v=\{1,2,4\}$.



The number of vertices $x$ is 2($n$ – 4). Now having chosen $x$ the number of $w$ with distance 1 from $x$ is 2($n$ – 5).



**Corollary 3.2.** For the Szeged index of $G_2$ we have

$$Sz(G_2)=3(n-3)\left(1+2\binom{n-4}{1}+4\binom{n-4}{1}\binom{n-5}{1}\right)^2.$$

## REFERENCES

1. M. R. Darafsheh, Computation of topological indices of some graphs, *Acta. Appl. Math.* **110** (2010) 1225–1235.
2. J. D. Dixon and B. Mortimer, *Permutation Groups*, Springer–Verley, NewYork, 1996.

3. M. Ghorbani, Computing the Wiener index of graphs on triples, *Creat. Math. Inform.* **24** (2015) no.1 49–52.
4. I. Gutman, A formula for the Wiener number of trees and its extension to graphs containing cyclic, *Graph Theory Notes NY.* **27** (1994) 9–15.
5. I. Gutman and A. A. Dobrynin, The Szeged index-asuccess story, *Graph Theory Notes NY.* **34** (1998) 37–44.
6. H. Hosoya, Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons, *Bull. Chem. Soc. Japan*. **44** (1971) 2332–2339.
7. P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. Dobrinin, I. Gutman and G. Domotor, The Szeged index and analogy with the Wiener index, *J. Chem. Inf. Comput. Sci.* **35** (1995) 547–550.
8. H. Wiener, Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.* **69** (1947) 17–20.
9. J. Zerovnik, Szeged index of symmetric graphs, *J. Chem. Inf. Comput. Sci*. **39** (1999) 77–80.

Iranian J. Math. Chem. 8(2) June (2017) 181–198

Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

# Nordhaus–Gaddum type results for the Harary index of graphs

ZHAO WANG[1], YAPING MAO[2,•], XIA WANG[2] AND CHUNXIA WANG[2]

[1]School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China

[2]Department of Mathematics, Qinghai Normal University, Xining, Qinghai 810008, China

**ARTICLE INFO**

**Article History:**

Received 16 November 2016
Accepted 7 April 2017
Published online April 7 2017
Academic Editor: Tomislav Doslić

**Keywords:**

distance
Steiner distance
Harary index
k-center Steiner Harary index

**ABSTRACT**

The Harary index $H(G)$ of a connected graph $G$is defined as $H(G) = \sum_{u,v\in V(G)} \frac{1}{d_G(u,v)}$, where $d_G(u,v)$ is the distance between vertices $u$ and $v$ of $G$. The Steiner distance in agraph, introduced by Chartrand et al. in 1989, is a natural generalization of the concept of classical graph distance. For a connected graph $G$ of order at least 2 and $S \subseteq V(G)$, the Steiner distance $d_G(S)$ of the vertices of $S$ is the minimum size of a connected subgraph whose vertex set contains $S$. Recently, Furtula, Gutman, and Katanić introduced the concept of Steiner Harary index and gave its chemical applications. The k-center Steiner Harary index $SH_k(G)$ of $G$ is defined by $SH_k(G) = \sum_{S\subseteq V(G),|S|=k} \frac{1}{d_G(S)}$. In this paper, we get the sharp upper and lower bounds for $SH_k(G) + SH_k(\overline{G})$ and $SH_k(G) \cdot SH_k(\overline{G})$, valid for any connected graph $G$ whose complement $\overline{G}$ is also connected.

## 1. INTRODUCTION

All graphs in this paper are assumed to be undirected, finite and simple and connected. We refer to [5] for graph theoretical notation and terminology not specified here. For a graph $G$, let $V(G), E(G)$ and $e(G) = |E(G)|$denote the set of vertices, the set of edges and the sizeof $G$, respectively.

---

• Corresponding author (Email: maoyaping@ymail.com)

DOI: 10.22052/ijmc.2017.67735.1254

If $S$ is a vertex-subset of a graph $G$, the subgraph of $G$ induced by $S$ is denoted by $G[S]$. We denote by $E_G[X,Y]$ the set of edges of $G$with one end in $X$ and the other in $Y$. If $X = \{x\}$, we simply write $E_G[x,Y]$ for $E_G[\{x\},Y]$.

The connectivity of a graph $G$, written $\kappa(G)$, is the order of a minimum vertex-subset $S \subseteq V(G)$ such that $G - S$ is disconnected or has only one vertex. Thus, if $G$ is connected, then $\kappa(G) \geq 1$; if $G$ has cut vertices, then $\kappa(G) = 1$.

The introduction is divided into the three subsections, in order to state the motivations and results of this paper.

### 1.1 DISTANCE AND ITS GENERALIZATION

Distance is one of the basic concepts of graph theory [6]. If $G$ is a connected graph and $u, v \in V(G)$, then the distance $d(u,v)$ between $u$and $v$ is the length of a shortest path connecting $u$ and $v$.

The distance between two vertices $u$and $v$ in a connected graph $G$ also equals the minimum size of a connected subgraph of $G$ containing both $u$ and $v$. This observation suggests a generalization of the distance concept. The Steiner distance of a graph, introduced by Chartrand et al. in 1989 [8], is a natural generalization of the classical graph distance. For a graph $G(V,E)$ and a set $S \subseteq V(G)$ of at least two vertices, an S-Steiner tree or a Steiner tree connecting $S$ (or simply, an S-tree) is a subgraph $T(V^{'},\mathrm{E}')$ of $G$ that is a tree with $S \subseteq V'$. Then the Steiner distance $d_G(S)$ of the vertices of $S$ (or simply the distance of $S$) is the minimum size of all connected subgraphs whose vertex sets contain $S$. Observe that $d_G(S) = \min\{e(T)|S \subseteq V(T)\}$, where $T$ is subtree of $G$. Furthermore, if $S = \{u,v\}$, then $d_G(S)$ coincides with the classical distance between $u$and $v$.

**Observation 1.1** Let $G$ be a connected graph of order $n$ and $k$ be an integer, $2 \leq k \leq n$. If $S \subseteq V(G)$ and $|S| = k$, then $k - 1 \leq d_G(S) \leq n - 1$.

The average Steiner distance $\mu_k(G)$ of a graph $G$, introduced by Dankelmann et al. [9, 10], is defined as the average of the Steiner distances of all k-subsets of $V(G)$, i.e.,

$$\mu_k(G) = \binom{n}{k}^{-1} \sum_{\substack{S \subseteq V(G) \\ |S|=k}} d_G(S)\,. \tag{1.1}$$

Let $n$ and $k$ be integers such that $2 \leq k \leq n$. The Steiner k-eccentricity $e_k(v)$ of a vertex $v$ of $G$ is defined by $e_k(v) = max\{d(S)|S \subseteq V(G), |S| = k, v \in S\}$. The Steiner k-radius of $G$ is $srad_k(G) = \min\{e_k(v)|v \in V(G)\}$, whereas the Steiner k-diameter of $G$ is $sdiam_k(G) = max\{e_k(v)|v \in V(G)\}$. Note that for any vertex $v$ of any connected graph

$G$, $e_2(v) = e(v)$, and in addition $srad_2(G) = rad(G)$ and $sdiam_2(G) = diam(G)$. For more details on Steiner distance, we refer to [3, 7, 8, 9, 10, 17, 25, 29].

Mao [25] obtained the following results. By $\Delta(G)$ we denote the greatest degree of a vertex of $G$.

**Lemma 1.1** [25] Let $G$ be a connected graph with connected complement $\overline{G}$. If $sdiam_k(G) \geq 2k$, then $sdiam_k(\overline{G}) \leq k$.

**Lemma 1.2** [25] Let $G$ be a connected graph of order $n$. Then $sdiam_3(G) = 2$ if and only if $0 \leq \Delta(\overline{G}) \leq 1$.

**Lemma 1.3** [25] Let $n, k$ be integers such that $2 \leq k \leq n$, and let $G$ be a connected graph of order $n$. If $sdiam_k(G) = k - 1$, then $0 \leq \Delta(\overline{G}) \leq k - 2$.

**Lemma 1.4** [25] Let $G$ be a connected graph of order $n$ with connected complement. Let $k$ be an integer such that $3 \leq k \leq n$. Let $x = 0$ if $n \geq 2k - 2$ and $x = 1$ if $n < 2k - 2$. Then

(1) $2k - 1 - x \leq sdiam_k(G) + sdiam_k(\overline{G}) \leq max\{n + k - 1, 4k - 2\}$;

(2) $(k - 1)(k - x) \leq sdiam_k(G) \cdot sdiam_k(\overline{G}) \leq max\{k(n - 1), (2k - 1)^2\}$.

**Lemma 1.5** [25] Let $G$ be a graph. Then $sdiam_{n-1}(G) = n - 2$ if and only if $G$ is 2-connected.

The following corollary is immediate from the above lemmas.

**Corollary 1.1** [28] Let $G$ and $\overline{G}$ be connected graphs. If $sdiam_3(G) \geq 6$, then $sdiam_3(\overline{G}) = 3$.

### 1.2 WIENER INDEX AND ITS GENERALIZATION

The *Wiener index* is defined as the sum of ordinary distances of all pairs of vertices of the underlying graph, i.e., as $W(G) = \sum_{u,v \in V(G)} d(u, v)$ and its mathematical theory is nowadays well elaborated. For details see the surveys [13, 34].

Li et al. [22] generalized the concept of Wiener index using Steiner distance, by defining the Steiner k-Wiener index $SW_k(G)$ of the connected graph $G$ as

$$SW_k(G) = \sum_{\substack{S \subseteq V(G) \\ |S| = k}} d_G(S) .$$

However, with regard to this definition, one should bear in mind Eq. (1.1), and the references [9, 10].

For $k = 2$, the Steiner Wiener index coincides with the ordinary Wiener index. It is usual to consider $SW_k$ for $2 \leq k \leq n - 1$, but the above definition implies $SW_1(G) = 0$ and $SW_n(G) = n - 1$.

An application in chemistry of the Steiner Wiener index was reported in [18]. Expressions for $SW_k$ for some special graphs were reported in [22]. Li et al. [22] also gave sharp upper and lower bounds on $SW_k$, and established some of its properties in the case of trees. For more details on the Steiner Wiener index, we refer to [18, 22, 23, 27].

### 1.3 HARARY INDEX AND ITS GENERALIZATION

The *Harary index* $H(G)$ of $G$is defined by $H(G) = \sum_{u,v\in V(G)} \frac{1}{d_G(u,v)}$. For more details on the Harary index, we refer to [4, 21, 24, 33].

Furtula et al. [15] introduced the concept of Steiner Harary index. The *Steiner Harary k-index* or *k-center Steiner Harary index* $SH_k(G)$ of $G$ is defined as

$$SH_k(G) = \sum_{\substack{S\subseteq V(G) \\ |S|=k}} \frac{1}{d_G(S)}.$$

For $k = 2$, the above defined Steiner Harary index coincides with the ordinary Harary index. It is usual to consider $SH_k$ for $2 \leq k \leq n - 1$, but the above definition implies $SH_1(G) = 0$ and $SH_n(G) = \frac{1}{n-1}$.

The following results will be needed later.

**Lemma 1.6** [26] Let $T$ be a tree of order $n$, and let $k$ be an integer such that $2 \leq k \leq n$. Then

$$n \sum_{k-1\leq t\leq n-1} \frac{1}{t}\binom{t-1}{k-2} - \binom{n-1}{k-1} \leq SH_k(T) \leq \frac{kn-n+k}{k^2(k-1)}\binom{n-1}{k-1}.$$

Moreover, among all trees of order $n$, the star $S_n$ maximizes the Steiner Harary k-index whereas the path $P_n$ minimizes the Steiner Harary k-index.

**Lemma 1.7** [26] Let $P_n$ be the path of order $n$ $(n \geq 3)$, and let $k$ be an integer such that $2 \leq k \leq n$. Then

$$SH_k(P_n) = n \sum_{k-1 \le t \le n-1} \frac{1}{t}\binom{t-1}{k-2} - \binom{n-1}{k-1}.$$

## 2. MAIN RESULTS

Let $f(G)$ be a graph invariant and $n$a positive integer, $n \geq 2$. The *Nordhaus–Gaddum Problem* is to determine sharp bounds for $f(G) + f(\overline{G})$ and $f(G) \cdot f(\overline{G})$, as $G$ ranges over the class of all graphs of order $n$, and to characterize the extremal graphs, i.e., graphs that achieve the bounds. Nordhaus–Gaddum type relations have received wide attention; see the recent survey [2] by Aouchiche and Hansen.

Denote by $\mathcal{G}(n)$ the class of connected graphs of order $n$ whose complements are also connected. In the studies of Nordhaus–Gaddum–type relations it must be assumed that $f(G)$ and $f(\overline{G})$ exist. Therefore, such relations are examined in the case of Wiener and Steiner Wiener indices, one must restrict the consideration to the class $\mathcal{G}(n), n \geq 2$.

Mao et al. [28] studied the Nordhaus-Gaddum type results for the Wiener index. In this paper, we investigate the analogous problem for the Steiner Harary index. Our basic idea is from [28].

### 2.1 RESULTS PERTAINING TO GENERAL k

For general $k$, we obtain the following result:

**Theorem 2.1** Let $G \in \mathcal{G}(n)$ and let $k$ be an integer such that $3 \leq k \leq n$. Then:

(1) $\binom{n}{k}\frac{2k-2}{max\{k(n-1),(2k-1)^2\}} \leq SH_k(G) + SH_k(\overline{G}) \leq \frac{(n+k-2)\binom{n}{k}}{(k-1)^2}$.

(2) $\frac{1}{max\{k(n-1),(2k-1)^2\}}\binom{n}{k}^2 \leq SH_k(G) \cdot SH_k(\overline{G}) \leq \frac{1}{(k-1)^2}\binom{n}{k}^2$.

Moreover, the lower bounds are sharp.

**Proof.** Proof of part (1):

For any $S \subseteq V(G)$ and $|S| = k$, from the definition of Steiner diameter, we have $d_G(S) + d_{\overline{G}}(S) \leq \max\{n + k - 2, 2k - 2\} = n + k - 2$. Then

$$SH_k(G) + SH_k(\overline{G}) = \sum_{S\subseteq V(G)} \frac{1}{d_G(S)} + \sum_{S\subseteq V(\overline{G})} \frac{1}{d_{\overline{G}}(S)} = \sum_{S\subseteq V(G)} \frac{d_G(S)+d_{\overline{G}}(S)}{d_G(S)d_{\overline{G}}(S)}$$
$$\leq \frac{(n+k-2)\binom{n}{k}}{(k-1)^2}.$$

By the same reason, Lemma 1.4 implies

$$SH_k(G) + SH_k(\overline{G}) = \sum_{S\subseteq V(G)} \frac{d_G(S)+d_{\overline{G}}(S)}{d_G(S)d_{\overline{G}}(S)} \geq \binom{n}{k} \frac{2k-2}{max\{k(n-1),(2k-1)^2\}}.$$

Proof of part (2):

For any $S' \subseteq V(G)$, $|S'| = k$ and any $S'' \subseteq V(\bar{G})$, $|S''| = k$, from the definition of Steiner diameter and Lemma 1.4, we have $d_G(S') \cdot d_{\bar{G}}(S'') \leq \max\{k(n-1),(2k-1)^2\}$. Then

$$SH_k(G) \cdot SH_k(\overline{G}) = \sum_{S'\subseteq V(G)} \frac{1}{d_G(S')} \cdot \sum_{S''\subseteq V(\overline{G})} \frac{1}{d_{\overline{G}}(S'')} = \sum_{S'\subseteq V(G), S''\subseteq V(\overline{G})} \frac{1}{d_G(S')} \cdot \frac{1}{d_{\overline{G}}(S'')}$$
$$\geq \frac{1}{\max\{k(n-1),\ (2k-1)^2\}} \binom{n}{k}^2.$$

For any $S' \subseteq V(G)$, $|S'| = k$ and any $S'' \subseteq V(\overline{G})$, $|S''| = k$, from the definition of Steiner diameter and Lemma 1.4, we have $d_G(S') \cdot d_{\overline{G}}(S'') \geq (k-1)^2$. Then

$$SH_k(G) \cdot SH_k(\overline{G}) = \sum_{S'\subseteq V(G)} \frac{1}{d_G(S')} \cdot \sum_{S''\subseteq V(\overline{G})} \frac{1}{d_{\overline{G}}(S'')} = \sum_{S'\subseteq V(G), S''\subseteq V(\overline{G})} \frac{1}{d_G(S')} \cdot \frac{1}{d_{\overline{G}}(S'')}$$
$$\leq \frac{1}{(k-1)^2} \binom{n}{k}^2,$$

as desired.

## 3. FOR SOME k

For $k = n, n-1, 3$, we can improve the results in Theorem 2.1.

### 3.1 THE CASE k = n, n– 1

For $k = n$, the following result is immediate.

**Observation 3.1** Let $G \in \mathcal{G}(n)$. Then

(1) $SH_n(G) + SH_n(\overline{G}) = \frac{2}{n-1}$;

(2) $SH_n(G) \cdot SH_n(\overline{G}) = \frac{1}{(n-1)^2}$.

Akiyama and Harary [1] characterized the graphs for which both $G$ and $\overline{G}$ are connected.

**Lemma 3.1** [1] Let $G$ be graph with $n$ vertices and maximal vertex degree $\Delta(G)$. Then $\kappa(G) = \kappa(\overline{G}) = 1$ if and only if $G$ satisfies the following conditions.

- i. $\kappa(G) = 1$ and $\Delta(G) = n - 2$;
- ii. $\kappa(G) = 1, \Delta(G) \leq n - 3$, and $G$ has a cut vertex $v$ with pendent edge $uv$, such that $G - u$ contains a spanning complete bipartite subgraph.

For $k = n - 1$, we have the following result:

**Proposition 3.1** Let $G$ be a graph of order $n$ $(n \geq 5)$.

1. If $G$ and $\overline{G}$ are both 2-connected, then $SH_{n-1}(G) + SH_{n-1}(\overline{G}) = \frac{2n}{n-2}$ and $SH_{n-1}(G) \cdot SH_{n-1}(\overline{G}) = \frac{n^2}{(n-2)^2}$.
2. If $\kappa(G) = 1$ and $\overline{G}$ is 2-connected, then $SH_{n-1}(G) + SH_{n-1}(\overline{G}) = \frac{p}{n-1} + \frac{2n-p}{n-2}$ and $SH_{n-1}(G) \cdot SH_{n-1}(\overline{G}) = \frac{pn}{(n-1)(n-2)} + \frac{n(n-p)}{(n-2)^2}$, where $p$ is the number of cut vertices in $G$.
3. If $\kappa(G) = \kappa(\overline{G}) = 1$, $\Delta(G) \leq n - 3$, and $G$ has a cut vertex $v$ with pendent edge $uv$ such that $G - u$ contains a spanning complete bipartite subgraph, and $\Delta(\overline{G}) \leq n - 3$ and $\overline{G}$ has a cut vertex $q$ with pendent edge $pq$ such that $G - p$ contains a spanning complete bipartite subgraph, then $SH_{n-1}(G) + SH_{n-1}(\overline{G}) = \frac{2n^2-2n-2}{(n-1)(n-2)}$ and $SH_{n-1}(G) \cdot SH_{n-1}(\overline{G}) = \frac{(n^2-n-1)^2}{(n-1)^2(n-2)^2}$.
4. If $\kappa(G) = \kappa(\overline{G}) = 1$, $\Delta(\overline{G}) = n - 2$, $\Delta(G) \leq n - 3$ and $G$ has a cut vertex $v$ with pendent edge $uv$ such that $G - u$ contains a spanning complete bipartite subgraph, then $SH_{n-1}(G) + SH_{n-1}(\overline{G}) = \frac{2n^2-2n-2}{(n-1)(n-2)}$ or $SH_{n-1}(G) +$

$SH_{n-1}(\overline{G}) = \frac{2n^2-2n-3}{(n-1)(n-2)}$ and $SH_{n-1}(G) \cdot SH_{n-1}(\overline{G}) = \frac{(n^2-n-1)^2}{(n-1)^2(n-2)^2}$ or $SH_{n-1}(G) \cdot SH_{n-1}(\overline{G}) = \frac{(n^2-n-1)(n+1)}{(n-1)^2(n-2)}$.

5. If $\kappa(G) = \kappa(\overline{G}) = 1$, $\Delta(G) = \Delta(\overline{G}) = n-2$, then $\frac{2(n+1)}{n-1} \leq SH_{n-1}(G) + SH_{n-1}(\overline{G}) \leq \frac{2n^2-2n-2}{(n-1)(n-2)}$ and $\frac{(n+1)^2}{(n-1)^2} \leq SH_{n-1}(G) \cdot SH_{n-1}(\overline{G}) \leq \frac{(n^2-n-1)^2}{(n-1)^2(n-2)^2}$.

**Proof.** (1): From Lemma 1.5, if $G$ and $\overline{G}$ are both connected, then $d_G(S) = n-2$ and $d_{\overline{G}}(S) = n-2$ for any $S \subseteq V(G)$ and $|S| = n-1$. Therefore, $SH_{n-1}(G) + SH_{n-1}\overline{G} = \frac{2n}{n-2}$ and $SH_{n-1}(G) \cdot SH_{n-1}(\overline{G}) = \frac{n^2}{(n-2)^2}$.

(2): Since $\overline{G}$ is 2-connected, it follows that $d_{\overline{G}}(S) = n-2$ for any $S \subseteq V(G)$ and $|S| = n-1$, and hence $SH_{n-1}(\overline{G}) = \frac{n}{n-2}$. Note that $\kappa(G) = 1$ and there are exactly $p$ cut vertices in $G$. For any $S \subseteq V(G)$ and $|S| = n-1$, if the unique vertex in $V(G) \setminus S$ is a cut vertex, then $d_G(S) = n-1$. If the unique vertex in $V(G) \setminus S$ is not a cut vertex, then $d_G(S) = n-2$. Therefore, we have $SH_{n-1}(G) = \frac{p}{n-1} + \frac{n-p}{n-2}$, and hence $SH_{n-1}(G) + SH_{n-1}(\overline{G}) = \frac{p}{n-1} + \frac{2n-p}{n-2}$ and $SH_{n-1}(G) \cdot SH_{n-1}(\overline{G}) = \frac{pn}{(n-1)(n-2)} + \frac{n(n-p)}{(n-2)^2}$, where $p$ is the number of cut vertices in $G$.

(3), (4), (5): We have $\kappa(G) = \kappa(\overline{G}) = 1$. By condition $(i)$ of Lemma 3.1, since $\Delta(G) = n-2$, there is a vertex of degree $n-2$, say $x$. Let the set of first neighbors of $x$ be $N_G(x) = \{y_1, y_2, \cdots, y_{n-2}\}$. Let $V(G) \setminus (\{x\} \cup N_G(x)) = \{z\}$. Since $zx \notin E(G)$, there must exist a vertex in $N_G(x)$, say $y_1$, such that $zy_1 \in E(G)$, since $G$ is connected. Since $x, y_1$ may be the cut vertices in $G$, it follows that there are one or two cut vertices in $G$. So

$$SH_{n-1}(G) = \frac{1}{n-1} + \frac{n-1}{n-2} = \frac{n^2-n-1}{(n-1)(n-2)} \text{ or } SH_{n-1}(G) = \frac{2}{n-1} + \frac{n-2}{n-2} = \frac{n+1}{n-1}.$$

By condition $(ii)$ of Lemma 3.1, since $\Delta(G) \leq n-3$ and $G$ has a cut vertex $v$ with pendent edge $uv$ such that $G-u$ contains a spanning complete bipartite subgraph, it follows that $v$ is the unique cut vertex. So $SH_{n-1}(G) = \frac{1}{n-1} + \frac{n-1}{n-2} = \frac{n^2-n-1}{(n-1)(n-2)}$. From this argument, (3), (4), (5) are true.

### 3.2 THE CASE $k = 3$

The following lemmas and corollaries will be used later.

**Lemma 3.2** [28] Let $T$ be a tree of order $n$, and let $k$ be an integer such that $3 \leq k \leq n$. Then there exist at least $(n - k + 1)$ subsets of $V(T)$ for which the Steiner k-distance is equal to $k - 1$.

**Corollary 3.1** [28] Let $G$ be a connected graph of order $n$, and let $k$ be an integer such that $3 \leq k \leq n$. Then there exist at least $(n - k + 1)$ subsets of $V(G)$ whose Steiner k-distance is $k - 1$.

**Lemma 3.3** [28] Let $T$ be a tree of order $n$, and let $k$ be an integer such that $3 \leq k \leq n - 1$. Then there exist at least $(n - k)$ subsets of $V(T)$ whose Steiner k-distance is $k$.

In this section, we focus our attention on the case $k = 3$. For $k = 3$ and $n \geq 10$, from Theorem 2.1, we have $\binom{n}{3}\frac{4}{3(n-1)} \leq SH_3(G) + SH_3(\overline{G}) \leq \frac{(n+1)\binom{n}{3}}{4}$ and $\frac{1}{3(n-1)}\binom{n}{3}^2 \leq SH_3(G) \cdot SH_3(\overline{G}) \leq \frac{1}{4}\binom{n}{3}^2$.

We improve these bounds and prove the following result.

**Theorem 3.1** Let $G \in \mathcal{G}(n)$ with $n \geq 4$. Then

1. $\frac{5}{6}\binom{n}{3} \geq SH_3(G) + SH_3(\overline{G}) \geq$
$$\begin{cases} \frac{7}{10}\binom{n}{3} + \frac{11}{60}n - \frac{1}{2} & if\ n = 6{,}7\ and\ sdiam_3(G) = 5 \\ & or\ n = 6{,}7\ and\ sdiam_3(\bar{G}) = 5 \\ \frac{1}{2}\binom{n-3}{3} - \sum_{i=2}^{n-1}\frac{n}{i} + \frac{7n^2-23n+20}{6} & otherwise. \end{cases}$$
2. $\frac{25}{144}\left[\binom{n}{3}^2\right] \geq SH_3(G) \cdot SH_3(\overline{G}) \geq \left[\frac{1}{n-1}\binom{n}{3} + \frac{(n-3)(n-2)}{2(n-1)}\right]\left[\frac{1}{2}\binom{n}{3} - \frac{(n-3)(n-2)}{2(n-1)}\right]$.

Moreover, the bounds are sharp.

We first need the following lemma.

**Lemma 3.4** [28] Let $G$ be a connected graph. If $sdiam_3(G) = 5$, then $sdiam_3(\overline{G}) \leq 4$.

**Lemma 3.5** Let $G \in \mathcal{G}(n)$. Then

$$SH_3(G) + SH_3(\overline{G}) \leq \frac{5}{6}\binom{n}{3} \tag{3.1}$$

$$SH_3(G) \cdot SH_3(\overline{G}) \leq \frac{25}{144}\binom{n}{3}^2 \tag{3.2}$$

and

$$SH_3(G)\cdot SH_3(\overline{G}) \geq \left[\frac{1}{n-1}\binom{n}{3}+\frac{(n-3)(n-2)}{2(n-1)}\right]\left[\frac{1}{2}\binom{n}{3}-\frac{(n-3)(n-2)}{2(n-1)}\right]. \tag{3.3}$$

Moreover, the bounds are sharp.

**Proof.** (1) For any $S\subseteq V(G)$ and $|S|=3$, $G[S]\cong K_3$ or $G[S]\cong P_3$ or $G[S]\cong K_2\cup K_1$ or $G[S]\cong 3K_1$. If $G[S]\cong K_3$ or $G[S]\cong P_3$, then $d_G(S)=2$. If $G[S]\cong K_2\cup K_1$ or $G[S]\cong 3K_1$, then $d_G(S)\geq 3$. Let $S_1,S_2,\cdots,S_{\binom{n}{3}}$ be all the 3-subsets of $V(G)$. Without loss of generality, let $S_1,S_2,\cdots,S_x$ be all the 3-subsets of $V(G)$ such that $G[S_i]\cong K_3$ or $G[S_i]\cong P_3$, where $1\leq i\leq x$. Therefore, $d_G(S_i)=2$ and $d_{\overline{G}}(S_i)\geq 3$ for each $i\ (1\leq i\leq x)$. Furthermore, for any $S_j\ (x+1\leq j\leq \binom{n}{3})$, we have

$$SH_3(G)\leq \frac{x}{2}+\frac{[\binom{n}{3}-x]}{3}=\frac{1}{3}\binom{n}{3}+\frac{x}{6}$$

$$SH_3(\overline{G})\leq \frac{x}{3}+\frac{[\binom{n}{3}-x]}{2}=\frac{1}{2}\binom{n}{3}-\frac{x}{6}$$

$$SH_3(G)\geq \frac{x}{2}+\frac{[\binom{n}{3}-x]}{n-1}=\frac{1}{n-1}\binom{n}{3}+\frac{(n-3)x}{2(n-1)}.$$

and

$$SH_3(\overline{G})\geq \frac{x}{n-1}+\frac{[\binom{n}{3}-x]}{2}=\frac{1}{2}\binom{n}{3}-\frac{(n-3)x}{2(n-1)}.$$

implying inequality (3.1).

By Corollary 3.1, there exist at least $n-2$ subsets of $V(G)$ whose Steiner 3-distances are equal to 2. The same is true for $\overline{G}$. Therefore, $n-2\leq x\leq \binom{n}{3}-n+2$, and hence

$$\begin{aligned} SH_3(G)\cdot SH_3(\overline{G}) &\leq \left[\frac{1}{3}\binom{n}{3}+\frac{x}{6}\right]\left[\frac{1}{2}\binom{n}{3}-\frac{x}{6}\right]\\ &=\frac{1}{6}\binom{n}{3}^2+\frac{x}{36}\binom{n}{3}-\frac{x^2}{36}\\ &\leq \frac{1}{36}\left[6\binom{n}{3}^2+\frac{1}{4}\binom{n}{3}^2\right]\\ &=\frac{25}{144}\left[\binom{n}{3}^2\right] \end{aligned}$$

i.e., inequality (3.2) holds.

$$\begin{aligned}\mathrm{S}H_3(G)\cdot\mathrm{S}H_3(\overline{G}) &\geq \left[\frac{1}{n-1}\binom{n}{3}+\frac{(n-3)x}{2(n-1)}\right]\left[\frac{1}{2}\binom{n}{3}-\frac{(n-3)x}{2(n-1)}\right]\\ &= \frac{1}{2(n-1)}\binom{n}{3}^2+\frac{(n-3)^2x}{4(n-1)^2}\binom{n}{3}-\frac{(n-3)^2x^2}{4(n-1)^2}\\ &\geq \left[\frac{1}{n-1}\binom{n}{3}+\frac{(n-3)(n-2)}{2(n-1)}\right]\left[\frac{1}{2}\binom{n}{3}-\frac{(n-3)(n-2)}{2(n-1)}\right]\end{aligned}$$

i.e., inequality (3.3) holds.

The sharpness of the above bounds is illustrated by the following example.

**Example 3.2** Let $G \cong P_4$. Then $\overline{G} \cong P_4$. By Lemma 1.7, $SH_3(G) = SH_3(\overline{G}) = \frac{5}{3}$, and hence $\mathrm{S}H_3(G) + \mathrm{S}H_3(\overline{G}) = \frac{10}{3} = \frac{5}{6}\binom{n}{3}$ and $\mathrm{S}H_3(G)\cdot\mathrm{S}H_3(\overline{G}) = \frac{25}{9} = \frac{25}{144}\left[\binom{n}{3}^2\right] = \left[\frac{1}{n-1}\binom{n}{3}+\frac{(n-3)(n-2)}{2(n-1)}\right]\left[\frac{1}{2}\binom{n}{3}-\frac{(n-3)(n-2)}{2(n-1)}\right]$, which confirms that the lower and upper bounds are sharp.

Let $S^*$ be a tree obtained from a star of order $n-2$ and a path of length 2 by identifying the center of the star and a vertex of degree one in the path. Then $\overline{S^*}$ is a graph obtained from a clique of order $n-1$ by deleting an edge $uv$ and then adding an pendent edge at $v$.

**Observation 3.2**

(1) $SH_3(S^*) = \frac{13}{12}\binom{n-3}{2} + \frac{1}{3}\binom{n-3}{3} + \frac{7}{6}n - 3$;

(2) $SH_3(\overline{S^*}) = \frac{4}{3}\binom{n-3}{2} + \frac{1}{2}\binom{n-3}{3} + \frac{4}{3}n - \frac{11}{3}$.

**Proof.** From the structure of $S^*$and $\overline{S^*}$, we conclude

$$\begin{aligned}\mathrm{S}H_3(S^*) &= \frac{1}{4}\binom{n-3}{2}+\frac{1}{2}\left[\binom{n-3}{2}+(n-3)+1\right]\\ &\quad+\frac{1}{3}\left[\binom{n-3}{2}+\binom{n-3}{3}+2(n-3)\right]\\ &= \frac{13}{12}\binom{n-3}{2}+\frac{1}{3}\binom{n-3}{3}+\frac{7}{6}n-3\end{aligned}$$

and

$$\begin{aligned} SH_3(\overline{S^*}) &= \frac{1}{2}\left[2\binom{n-3}{2} + 2(n-3) + \binom{n-3}{3}\right] + \frac{1}{3}\left[\binom{n-3}{2} + (n-2)\right] \\ &= \frac{4}{3}\binom{n-3}{2} + \frac{1}{2}\binom{n-3}{3} + \frac{4}{3}n - \frac{11}{3}. \end{aligned}$$

In order to show the sharpness of the above bounds, we consider the following example.

**Example 3.3** Let $S^*$ be the same tree as before. From Observation 3.2, we have

$$SH_3(S^*) + SH_3(\overline{S^*}) = \frac{29}{12}\binom{n-3}{2} + \frac{5}{6}\binom{n-3}{3} + \frac{15}{6}n - \frac{20}{3}$$

and

$$\begin{aligned} SH_3(S^*) \cdot SH_3(\overline{S^*}) = {} & \frac{52}{36}\binom{n-3}{2}^2 + \frac{1}{6}\binom{n-3}{3}^2 + \frac{71}{72}\binom{n-3}{2}\binom{n-3}{3} \\ & + \left(\frac{27}{9}n - \frac{287}{36}\right)\binom{n-3}{2} + \left(\frac{37}{36}n - \frac{49}{18}\right)\binom{n-3}{3} \\ & + \left(\frac{4}{3}n - \frac{11}{3}\right)\left(\frac{7}{6}n - 3\right). \end{aligned}$$

The following lemmas are preparations for deducing an upper bound on $SH_3(G) + SH_3(\overline{G})$.

**Lemma 3.6** Let $G$ be a connected graph of order $n$, and let $T$ be a spanning tree of $G$. If $sdiam_3(\overline{G}) = 3$, then

$$SH_3(T) + SH_3(\overline{T}) \leq SH_3(G) + SH_3(\overline{G}).$$

**Proof.** Note that $\overline{G}$ is a spanning subgraph of $\overline{T}$. It suffices to prove that

$$SH_3(\overline{T}) - SH_3(\overline{G}) \leq SH_3(G) - SH_3(T).$$

Since $sdiam_3(\overline{G}) = 3$, it follows that $d_{\overline{G}}(S) = 2$ or $d_{\overline{G}}(S) = 3$ for any $S \subseteq V(G)$ and $|S| = 3$. Since $\overline{G}$ is a spanning subgraph of $\overline{T}$ and $sdiam_3(\overline{G}) = 3$, it follows that $sdiam_3(\overline{T}) \leq 3$, and hence $d_{\overline{T}}(S) = 2$ or $d_{\overline{T}}(S) = 3$ for any $S \subseteq V(T)$ and $|S| = 3$. Then $0 \leq \frac{1}{d_{\overline{T}}(S)} - \frac{1}{d_{\overline{G}}(S)} \leq \frac{1}{6}$. We claim that $\frac{1}{d_{\overline{T}}(S)} - \frac{1}{d_{\overline{G}}(S)} \leq \frac{1}{d_G(S)} - \frac{1}{d_T(S)}$ for $S \subseteq V(T)$ and

$|S| = 3$. Because $\overline{G}$ is a spanning subgraph of $\overline{T}$, $\frac{1}{d_{\overline{G}}(S)} \leq \frac{1}{d_{\overline{T}}(S)}$ for any $S \subseteq V(T)$ and $|S| = 3$. Similarly, since $T$ is a spanning subgraph of $G$, $\frac{1}{d_T(S)} \leq \frac{1}{d_G(S)}$ for any $S \subseteq V(T)$ and $|S| = 3$. If $\frac{1}{d_{\overline{T}}(S)} - \frac{1}{d_{\overline{G}}(S)} = 0$, then $\frac{1}{d_{\overline{T}}(S)} - \frac{1}{d_{\overline{G}}(S)} = 0 \leq \frac{1}{d_G(S)} - \frac{1}{d_T(S)}$, as desired. If $\frac{1}{d_{\overline{T}}(S)} - \frac{1}{d_{\overline{G}}(S)} = \frac{1}{6}$, then $d_{\overline{G}}(S) = 3$ and $d_{\overline{T}}(S) = 2$, and hence $d_G(S) = 2$ and $d_T(S) \geq 3$. Therefore, $\frac{1}{d_G(S)} - \frac{1}{d_T(S)} \geq \frac{1}{6} = \frac{1}{d_{\overline{T}}(S)} - \frac{1}{d_{\overline{G}}(S)}$, as desired. The result follows from the arbitrariness of $S$ and the definition of Steiner Wiener index.

**Lemma 3.7** Let $T$ be a tree of order $n$, different from the star $S_n$. Let $S^*$ be the tree same as in Observation 3.2. If $sdiam_3(\overline{G}) = 3$, then

$$\mathrm{S}H_3(P_n) + SH_3(\overline{S^*}) \leq \mathrm{S}H_3(T) + SH_3(\overline{T}).$$

**Proof.** Note first that the complements of all trees, except of the star, are connected. Therefore, $SH_3(\overline{T})$ in Lemma 3.7 is always well defined.

By Lemma 1.6 and 1.7, $\mathrm{S}H_3(P_n) \leq SH_3(T)$. It suffices to prove $SH_3(\overline{S^*}) \leq SH_3(\overline{T})$. Since $sdiam_3(\overline{G}) \leq 3$, it follows that $sdiam_3(\overline{T}) \leq 3$. For any $S \subseteq V(T)$ and $|S| = 3$, if $T[S]$ is not connected, then $d_{\overline{T}}(S) = 2$. If $T[S]$ is connected, then $d_{\overline{T}}(S) \geq 3$. So if we want to obtain the minimum value of $SH_3(\overline{T})$ for a tree $T$, then we need to find as less as possible 3-subsets of $V(T)$ whose induced subgraphs in $\overline{T}$ are disconnected. Since the complement of $S_n$ is not connected, it follows that $\overline{S^*}$ is our desired. So $SH_3(\overline{S^*}) \leq SH_3(\overline{T})$, and hence $SH_3(P_n) + SH_3(\overline{S^*}) \leq SH_3(T) + SH_3(\overline{T})$.

We are now in the position to complete the proof of Theorem 3.1. This will be achieved by combining Lemmas 3.5 and 3.8.

Let $G \in \mathcal{G}(n)$. If $n = 6, 7$ and $sdiam_3(G) = 5$, then the validity of Theorem 3.1 can be verified by direct checking.

**Lemma 3.8** Let $G \in \mathcal{G}(n)$. Let $n \geq 8$, or $n \leq 5$, or $n = 6, 7$ and $sdiam_3(G) \neq 5$, or $n = 6, 7$ and $sdiam_3(\overline{G}) \neq 5$. Then the lower bounds in parts (1) and (2) of Theorem 3.1 are obeyed. Moreover, these bounds are sharp.

**Proof.** We need to separately examine three cases.

Case 1. $sdiam_3(G) \geq 6$ or $sdiam_3(\overline{G}) \geq 6$. Without loss of generality, let $sdiam_3(G) \geq 6$. From Corollary 1.1 it is known that $sdiam_3(\overline{G}) = 3$, and hence $SH_3(G) + SH_3(\overline{G}) \geq SH_3(P_n) + SH_3(\overline{S^*})$. By Lemma 1.7, $SH_3(P_n) = \frac{(n+1)(n-2)}{2} - \sum_{i=2}^{n-1} \frac{n}{i}$. Note that $\overline{S^*}$ is a graph obtained from a clique of order $n-1$ by deleting an edge $uv$ and then adding a pendent edge at $v$. Then $SH_3(\overline{S^*}) = \frac{4}{3}\binom{n-3}{2} + \frac{1}{2}\binom{n-3}{3} + \frac{4}{3}n - \frac{11}{3}$, and hence $SH_3(G) + SH_3(\overline{G}) \geq \frac{(n+1)(n-2)}{2} - \sum_{i=2}^{n-1} \frac{n}{i} + \frac{4}{3}\binom{n-3}{2} + \frac{1}{2}\binom{n-3}{3} + \frac{4}{3}n - \frac{11}{3} = \frac{1}{2}\binom{n-3}{3} - \sum_{i=2}^{n-1} \frac{n}{i} + \frac{7n^2-23n+20}{6}$.

Case 2. $sdiam_3(G) = 5$or$sdiam_3(\overline{G}) = 5$. In view of Lemma 3.4, we can assume that $sdiam_3(G) = 5$ and $sdiam_3(\overline{G}) \leq 4$. Let $S_1, S_{2,\dots,}S_{\binom{n}{3}}$ be all the 3-subsets of $V(G)$. Without loss of generality, assume that $S_1, S_{2,\dots,}S_x$ are the 3-subsets of $V(G)$ for which $G[S_i] \cong K_3$ or $G[S_i] \cong P_3$, where $1 \leq i \leq x$.

For each $i$ $(1 \leq i \leq x)$, $d_G(S_i) = 2$. For any $S_j$ $(x + 1 \leq j \leq \binom{n}{3})$, $G[S_j] \cong K_2 \cup K_1$ or $G[S_j] \cong 3K_1$. Since $G$ is connected, it follows that there exists a spanning tree, say $T$. By Lemmas 3.2 and 3.3, there exist at least $(n-3)$ subsets of $V(T)$ whose Steiner 3-distance is 3, and there exist at least $(n-2)$ subsets of $V(T)$ whose Steiner 3-distance is 2. Therefore, there exist at least $(2n-5)$ subsets of $V(G)$ whose Steiner 3-distance is at most 3. Without loss of generality, let $d_G(S_j) = 3$for $S_j$ $(x + 1 \leq j \leq 2n - 5)$. Then $d_G(S_j) \leq 5$ and $d_{\overline{G}}(S_j) = 2$ for each j $(2n - 4 \leq j \leq \binom{n}{3})$. For each $i$ $(1 \leq i \leq x)$, $d_G(S_i) = 2$. By Lemma 3.3, there exist at least $(n-3)$ subsets of $V(\overline{G})$ whose Steiner 3-distance is 3. Then there exist at most $x - (n-3)$ subsets of $V(\overline{G})$ whose Steiner 3-distance is 4. If $x \leq 2n - 5$, then $SH_3(G) \geq \frac{1}{2}x + \frac{1}{3}(2n - 5 - x) + \frac{1}{5}\left[\binom{n}{3} - 2n + 5\right]$ and $SW_3(\overline{G}) \geq \frac{1}{3}(n-3) + \frac{1}{4}(x - n + 3) + \frac{1}{2}\left[\binom{n}{3} - x\right]$, and hence $SH_3(G) + SH_3(\overline{G}) \geq \frac{7}{10}\binom{n}{3} - \frac{1}{12}x + \frac{7}{20}n - \frac{11}{12} \geq \frac{7}{10}\binom{n}{3} + \frac{11}{60}n - \frac{1}{2}$. If $x \geq 2n - 5$, then $SH_3(G) \geq \frac{1}{2}x + \frac{1}{5}\left[\binom{n}{3} - x\right]$ and $SH_3(\overline{G}) \geq \frac{1}{3}(n-3) + \frac{1}{4}(x - n + 3) + \frac{1}{2}\left[\binom{n}{3} - x\right]$, and hence $SH_3(G) + SH_3(\overline{G}) \geq \frac{7}{10}\binom{n}{3} + \frac{1}{20}x + \frac{1}{12}n - \frac{1}{4} \geq \frac{7}{10}\binom{n}{3} + \frac{11}{60}n - \frac{1}{2}$.

Case 3. $sdiam_3(G) \leq 4$ and $sdiam_3(\overline{G}) \leq 4$. Let $S_1, S_{2,\dots,}S_{\binom{n}{3}}$ be the 3-subsets of $V(G)$. Without loss of generality, let $S_1, S_{2,\dots,}S_x$ be the 3-subsets of $V(G)$ for which $G[S_i] \cong K_3$ or

$G[S_i] \cong P_3$, where $1 \le i \le x$. For each $i$ $(1 \le i \le x)$, $d_G(S_i) = 2$. For any $S_j$ $(x + 1 \le j \le \binom{n}{3})$, $G[S_j] \cong K_2 \cup K_1$ or $G[S_j] \cong 3K_1$.Since $G$ is connected, there exists a spanning tree, say $T$. By Lemmas 3.2 and 3.3, there exist at least $(n - 3)$ subsets of $V(T)$ whose Steiner 3-distance is equal to 3, and there exist at least $(n-2)$ subsets of $V(T)$ whose Steiner 3-distance is 2. Therefore, there exist at least $(2n-5)$ subsets of $V(G)$ whose Steiner 3-distance is at most 3. Without loss of generality, let $d_G\ (S_j) = 3$ for $S_j$ $(x + 1 \le j \le 2n - 5)$. Then $d_G\ (S_j) \le 4$ and $d_{\overline{G}}(S_j) = 2$ for each $j$ $\left(2n - 4 \le j \le \binom{n}{3}\right)$. For each $i$ $(1 \le i \le x)$, $d_G\ (S_i) = 2$. By Lemma 3.3, there exist at least $(n-3)$ subsets of $V(\overline{G})$ whose Steiner 3-distance in $\overline{G}$ is 3. Then there exist at most $x - (n-3)$ subsets of $V(\overline{G})$ whose Steiner 3-distance in $\overline{G}$ is 4. If $x \le 2n - 5$, then $SH_3(G) \ge \frac{1}{2}x + \frac{1}{3}(2n - 5 - x) + \frac{1}{4}\left[\binom{n}{3} - 2n + 5\right]$ and $SH_3(\overline{G}) \ge \frac{1}{3}(n-3) + \frac{1}{4}(x - n + 3) + \frac{1}{2}\left[\binom{n}{3} - x\right]$. Thus

$$SH_3(G) + SH_3(\overline{G}) \ge \frac{3}{4}\binom{n}{3} - \frac{1}{12}x + \frac{1}{4}n - \frac{2}{3} \ge \frac{3}{4}\binom{n}{3} + \frac{1}{12}n - \frac{3}{12}.$$

If $x \ge 2n - 5$, then $SH_3(G) \ge \frac{1}{2}x + \frac{1}{4}\left[\binom{n}{3} - x\right]$ and $SH_3(\overline{G}) \ge \frac{1}{3}(n-3) + \frac{1}{4}(x - n + 3) + \frac{1}{2}\left[\binom{n}{3} - x\right]$. Thus $SH_3(G) + SH_3(\overline{G}) \ge \frac{3}{4}\binom{n}{3} + \frac{1}{12}n - \frac{3}{12}$.

For $n \ge 6$, one can check that $\frac{1}{2}\binom{n-3}{3} - \sum_{i=2}^{n-1}\frac{n}{i} + \frac{7n^2 - 23n + 20}{6} \le \frac{3}{4}\binom{n}{3} + \frac{1}{12}n - \frac{3}{12}$ and $\frac{7}{10}\binom{n}{3} + \frac{11}{60}n - \frac{1}{2} \le \frac{3}{4}\binom{n}{3} + \frac{1}{12}n - \frac{3}{12}$. So we only need to consider the lower bounds in Cases 1 and 2.

From the above argument, we conclude the following:

1. For $n \ge 8$, $\frac{1}{2}\binom{n-3}{3} - \sum_{i=2}^{n-1}\frac{n}{i} + \frac{7n^2 - 23n + 20}{6} \le \frac{7}{10}\binom{n}{3} + \frac{11}{60}n - \frac{1}{2}$ and $SH_3(G) + SH_3(\overline{G}) \ge \frac{1}{2}\binom{n-3}{3} - \sum_{i=2}^{n-1}\frac{n}{i} + \frac{7n^2 - 23n + 20}{6}$.
2. For $n \le 5$, the lower bound in Case 2 does not exist. Then $SH_3(G) + SH_3(\overline{G}) \ge \frac{1}{2}\binom{n-3}{3} - \sum_{i=2}^{n-1}\frac{n}{i} + \frac{7n^2 - 23n + 20}{6}$.
3. If $n = 6, 7$, $\mathrm{sdiam}_3(G) \ne 5$, and $\mathrm{sdiam}_3(\overline{G}) \ne 5$, then $SH_3(G) + SH_3(\overline{G}) \ge \frac{1}{2}\binom{n-3}{3} - \sum_{i=2}^{n-1}\frac{n}{i} + \frac{7n^2 - 23n + 20}{6}$.

4. If $n = 6, 7$ and $\mathrm{sdiam}_3(G) = 5$, or $n = 6, 7$ and $\mathrm{sdiam}_3(\overline{G}) = 5$, then

$$SH_3(G) + SH_3(\overline{G}) \geq \frac{7}{10}\binom{n}{3} + \frac{11}{60}n - \frac{1}{2}.$$

This completes the proof.

In order to demonstrate the sharpness of the above bounds, we point out the following example.

**Example 3.4** Let $G \cong P_4$. Then $\overline{G} \cong P_4$. By Lemma 1.1, $SH_3(G) = SH_3(\overline{G}) = \frac{5}{3}$ , and hence $SH_3(G) + SH_3(\overline{G}) = \frac{10}{3} = \frac{1}{2}\binom{n-3}{3} - \sum_{i=2}^{n-1}\frac{n}{i} + \frac{7n^2-23n+20}{6}$ and $SH_3(G) \cdot SH_3(\overline{G}) = \frac{25}{9} = \left[\frac{1}{n-1}\binom{n}{3} + \frac{(n-3)(n-2)}{2(n-1)}\right]\left[\frac{1}{2}\binom{n}{3} - \frac{(n-3)(n-2)}{2(n-1)}\right]$, which implies that the upper and lower bounds are sharp.

**Acknowledgment.** The authors are very grateful to the referees for their valuable comments and suggestions, which improved the presentation of this paper. This work was supported by the National Science Foundation of China (Nos. 11601254, 11551001, 11161037, 11461054) and the Science Found of Qinghai Province (Nos. 2016-ZJ-948Q, and 2014-ZJ-907).

## REFERENCES

1. J. Akiyama, F. Harary, A graph and its complement with specified properties, *Int. J. Math. Math. Sci.* **2** (1979) 223–228.
2. M. Aouchiche, P. Hansen, A survey of Nordhaus–Gaddum type relations, *Discrete Appl. Math.* **161** (2013) 466–546.
3. P. Ali, P. Dankelmann, S. Mukwembi, Upper bounds on the Steiner diameter of a graph, *Discrete Appl. Math.* **160** (2012) 1845–1850.
4. A. T. Balaban, The Harary index of a graph, *MATCH Commun. Math. Comput. Chem.* **75** (2016) 243–245.
5. J. A. Bondy, U. S. R. Murty, *Graph Theory*, Springer, 2008.
6. F. Buckley, F. Harary, *Distance in Graphs*, Addison–Wesley, Redwood, 1990.

7. J. Cáceresa, A. Márquezb, M. L. Puertasa, Steiner distance and convexity in graphs, *Eur. J. Combin.* **29** (2008) 726–736.
8. G. Chartrand, O. R. Oellermann, S. Tian, H. B. Zou, Steiner distance in graphs, *Časopis Pest. Mat.* **114** (1989) 399–410.
9. P. Dankelmann, O. R. Oellermann, H. C. Swart, The average Steiner distance of a graph, *J. Graph Theory* **22** (1996) 15–22.
10. P. Dankelmann, H. C. Swart, O. R. Oellermann, On the average Steiner distance of graphs with prescribed properties, *Discrete Appl. Math.* **79** (1997) 91–103.
11. P. Dankelmann, R. Entringer, Average distance, minimum degree, and spanning trees, *J. Graph Theory* **33** (2000) 1–13.
12. P. Dankelmann, O. R. Oellermann, H. C. Swart, The average Steiner distance of a graph, *J. Graph Theory* **22** (1996) 15–22.
13. A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: theory and application, *Acta Appl. Math.* **66** (2001) 211–249.
14. R. C. Entringer, D. E. Jackson, D. A. Snyder, Distance in graphs, *Czech. Math. J.* **26** (1976), 283–296.
15. B. Furtula, I. Gutman, V. Katanić, Three-center Harary index and its applications, *Iranian J. Math. Chem.* **7** (1) (2016) 61–68.
16. M. R. Garey, D. S. Johnson, *Computers and Intractability – A Guide to the Theory of NP−Completeness*, Freeman, San Francisco, 1979, pp. 208–209.
17. W. Goddard, O. R. Oellermann, Distance in graphs, in: M. Dehmer (Ed.), *Structural Analysis of Complex Networks*, Birkhäuser, Dordrecht, 2011, pp. 49–72.
18. I. Gutman, B. Furtula, X. Li, Multicenter Wiener indices and their applications, *J. Serb. Chem. Soc.* **80** (2015) 1009–1017.
19. I. Gutman, S. Klavžar, B. Mohar (Eds.), Fifty years of the Wiener index, *MATCH Commun. Math. Comput. Chem.* **35** (1997) 1–159.
20. I. Gutman, O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986.
21. X. Li, Y. Fan, The connectivity and the Harary index of a graph, *Discrete Appl. Math.* **181** (2015) 167–173.
22. X. Li, Y. Mao, I. Gutman, The Steiner Wiener index of a graph, *Discuss. Math. Graph Theory* **36** (2) (2016) 455–465.

23. X. Li, Y. Mao, I. Gutman, Inverse problem on the Steiner Wiener index, *Discuss. Math. Graph Theory*, in press.
24. B. Lučić, A. Miličević, S. Nikolić, N. Trinajstić, Harary index-twelve years later, *Croat. Chem. Acta* **75** (2002) 847–868.
25. Y. Mao, The Steiner diameter of a graph, *Bull. Iranian Math. Soc.* **43** (2) (2017) 439−454.
26. Y. Mao, Steiner Harary index, *Kragujevac J. Math.* **42** (1) (2018) 29−39.
27. Y. Mao, Z. Wang, I. Gutman, Steiner Wiener index of graph products, *Trans. Combin.* **5** (3) (2016) 39–50.
28. Y. Mao, Z. Wang, I. Gutman, H. Li, Nordhaus-Gaddum-type results for the Steiner Wiener index of graphs, *Discrete Appl. Math.* **219** (2017) 167−175.
29. O. R. Oellermann, S. Tian, Steiner centers in graphs, *J. Graph Theory* **14** (1990) 585–597.
30. D. H. Rouvray, Harry in the limelight: The life and times of Harry Wiener, in: D. H. Rouvray, R. B. King (Eds.), *Topology in Chemistry — Discrete Mathematics of Molecules*, Horwood, Chichester, 2002, pp. 1–15.
31. D. H. Rouvray, The rich legacy of half century of the Wiener index, in: D. H. Rouvray, R. B. King (Eds.), *Topology in Chemistry — Discrete Mathematics of Molecules*, Horwood, Chichester, 2002, pp. 16–37.
32. H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.
33. K. Xu, K. C. Das, N. Trinajstić, *The Harary Index of a Graph*, Springer, Heidelberg, 2015.
34. K. Xu, M. Liu, K. C. Das, I. Gutman, B. Furtula, A survey on graphs extremal with respect to distance–based topological indices, *MATCH Commun. Math. Comput. Chem.***71** (2014) 461–508.

Iranian J. Math. Chem. 8(2) June (2017) 199−220

Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

# Determination of Critical Properties of Alkanes Derivatives using Multiple Linear Regression

**ESMAT MOHAMMADINASAB**•

Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran

**ARTICLE INFO**

**Article History:**
Received: 27 July 2016
Accepted: 14 October 2016

Academic Editor: Roberto Todeschini

**Keywords:**
Alkanes
MLR
Critical properties
QSPR

**ABSTRACT**

This study presents some mathematical methods for estimating the critical properties of 40 different types of alkanes and their derivatives including critical temperature, critical pressure and critical volume. This algorithm used QSPR modeling based on graph theory, several structural indices, and geometric descriptors of chemical compounds. Multiple linear regression was used to estimate the correlation between these critical properties and molecular descriptors using proper coefficients. To achieve this aim, the most appropriate molecular descriptors were chosen from among 11 structural and geometric descriptors in order to determine the critical properties of the intended molecules. The results showed that among all the proposed models to predict critical temperature, pressure and volume, a model including the combination of such descriptors as hyper-Wiener, Platt, MinZL is the most appropriate one.

## 1. INTRODUCTION

Alkanes are considered as one of the most significant aliphatic hydrocarbons used in such various industries as food, pharmaceutical, petrochemical and oil [1]. Since these chemicals are present in many refining processes (crude oil), it is quite essential to take their various physical and chemical properties– especially critical properties– into consideration. Critical point is a state in which there is no boundary between the two phases of a matter. This state occurs for every matter at the presence of a certain amount of temperature, pressure and combination. Theoretically, it is possible to estimate most of the thermodynamic properties

---

• Corresponding Author: (Email address: e-mohammadinasab@iau-arak.ac.ir)

DOI: 10.22052/ijmc.2017.58461.1225

of chemicals using their critical properties. Practically, a large amount of theoretical correlations are based on these properties [2–4].

The initial methods for estimating critical properties were experimental and were used for hydrocarbon systems. Due to the fact that experimental values of critical properties are not available for heavy alkanes, it is important to take advantage of appropriate methods for estimating these properties. In this study, the model of multiple linear regression was applied for the first time to find out the most appropriate molecular descriptors in order to estimate critical temperature, critical pressure and critical volume of alkanes and their derivatives [5]. The independent variables in multiple regression model can be obtained through the use of various methods [6–7]. In addition, the graph theory provides us with a suitable means for calculating topological descriptors which function as independent variables [8–9].

The graph theory has a long history in mathematics and its application dates back to about 200 years ago. In 1730, Euler was recognized as the father of graph theory after publishing the “Seven Bridges of Königsberg”. This theory is one the most practical branches of mathematics in other disciplines. It has a wide range of applications in such disciplines as biology, chemistry, nanotechnology, operational research and engineering [10].

Chemical Graph Theory is one of the branches of mathematical chemistry which is typically related to theoretical chemistry [11]. According to this theory, a graph indicates a set of elements of a group and their interrelationships. In chemical graphs, the hydrogen atoms are ignored since these atoms do not usually play a significant role in determining the molecular structure. After drawing the chemical graph for a molecule, it would not be difficult to extract topological indices–which are some constant numbers–for that graph. Mathematicians call such constant numbers topological indices. These indices include the structure, size, molecular polymerization, number of atoms and types of freaking molecular atoms. The concept topological indices were initially expressed in 1947. Wiener and Platt were the first to develop graph theory–based quantitative topological variables in 1947 known as Wiener index and Platt index, respectively, and reported Quantitative Structure Property Relationship (QSPR) models on boiling points of hydrocarbones. At that time, this concept was most often used for physical properties such as alkanes and paraffin’s boiling points. QSPRs have provided a valuable approach in research into physico–chemical properties of organic compounds [12–13]. In 1994, Ivan Gutman paid much more attention to these issues, specifically the distances and weighted functions, in his paper entitled “on the sum of all distances in composite graphs”. In theoretical chemistry, these indices help to predict the chemical and pharmaceutical properties of materials. QSPR is a model that relates the predictor variables of a molecule to its physico–chemical properties. The essential problem in the development of a suitable correlation between chemical structures and properties can be imputed to the quantitative nature of chemical structures. Graph theory was successfully applied through the translation of chemical structures into characteristic numerical descriptors by resorting to graph invariants.

Hosseini and Shafiei proposed QSPR model for the prediction of gas heat capacity of benzene derivatives through the use of topological indices. The best model was obtained as follows: $C_v = -84.569 + 43.970{}^1\chi - 2.298W + 1.463Sz$. This means that ${}^1\chi$, W, Sz descriptors play an important role in affecting heat capacity ($C_v$) of benzene derivatives [14].

QSPR modeling produces predictive model derived from application of statistical tools correlating physico–chemical properties in QSPR models of chemicals with descriptors representative of molecular structure [15–17].

In a nutshell, the aim of present research is to investigate the relationship between all critical properties (as dependent variables) and 2–dimentional and 3–dimentional descriptors (as independent variables) using QSPR and multiple linear regression (MLR) methods for estimating the critical properties of 40 different types of alkanes and their derivatives including critical temperature ($T_c$), critical pressure ($P_c$) and critical volume ($V_c$).

## 2. TOPOLOGICAL INDICES

Considering the research studies in which several two–dimensional indices (topological) were used, the current paper makes an attempt to investigate several three–dimensional (geometrical) indices as molecular descriptors and their application for prediction of critical properties of alkanes [16–17]. As a matter of fact, critical properties are sensitive to molecular geometry; hence, some of the geometric descriptors were employed as independent variables in this research. The statistical formulas used in this regard are presented below.

### 2.1 WIENER INDEX

In 1947, Harold Wiener [18] introduced one of the first molecular descriptors of topological nature for acyclic saturated hydrocarbons. The Wiener index W(G) of a graph G is defined as the half of the sum of the distances between every pair of vertices of G, $D_{ij}$, is the distance of two vertices i and j in the graph G).

$$W = \frac{1}{2}\sum_{i=}^{n}\sum_{j=1}^{n} D_{ij} \tag{1}$$

### 2.2 HYPER WIENER INDEX

Another related distance–based structural descriptor of the graph G is the hyper–Wiener index [19], WW(G) [5]. WW(G) index is introduced as:

$$WW(G) = \sum d(U,V)^2 + \sum d(U,V)/2 \tag{2}$$

where d(U, V) denotes the distance between the vertices U and V in the graph G and the summations run over all (unordered) pairs of vertices of G.

### 2.3 RANDIĆ INDEX

In 1975, Milan Randić a Croatian–American scientist introduced the Randić index [20–23], the first connectivity index. The Randić index of a chemical graph is the sum of all the bonds contributions:

$$\chi = \sum (\frac{1}{d_i d_j})^{\frac{1}{2}} \tag{3}$$

where $d_i$ and $d_j$ are the degrees of the vertices representing atoms "i", "j".

### 2.4 BALABAN INDEX

Defined by the Romanian chemist Alexandru T. Balaban in 1982, Balaban index is: J=J(G) of a Graph G on n node and m edges and $D_i$ are the sum of all entries in the $^i$th row (or column) of graph distance matrix [24–25]:

$$J = \frac{m}{\mu + 1} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[ (D_i)(D_j) \right]^{-0/5} \tag{4}$$

where $\mu = m - n + 1$ is the cyclomatic number.

### 2.5 HARARY INDEX

The Harary index of a graph G was defined from the inverse of the squared elements of the distance matrix according to the expression [26]:

$$\mathrm{H} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left( D_{ij} \right)^{-2} \tag{5}$$

where $D^{-2}$ is the matrix whose elements are the squares of the reciprocal distances.

### 2.6 GEOMETRIC INDICES

Geometric analysis provides characteristic values related to the geometrical structure of a molecule such as minimal and maximal z length, minimal and maximal projection area, force field energies or van der Waals volume.

## 3. COMPUTATIONAL METHODS

To analyze the relationship between critical properties such as temperature ($T_c$), pressure ($P_c$) and volume ($V_c$) of alkanes derivatives in contrast with molecular descriptors, the research data were collected in two stages as follow:

First, the structure and existing laboratory quantities (experimental) of 40 different types of alkanes and their derivatives in the present investigation were taken from National Institute of Standard and Technology chemistry webbook and were listed in Table 1.

Second, the values of Randić ($\chi$), Harary (H), Balaban (J), Wiener (W), Platt (Platt) and hyper–Wiener (WW) topological indices were calculated by formulas 1–5 using graph theory for 40 different types of alkanes derivatives, and the values of geometry descriptors such as the minimal projection area (MinPA/A$^{°2}$), the maximal projection area (MaxPA/A$^{°2}$), the minimal z length (MinZL/A$^{°}$), the maximal z length (MaxZL/A$^{°}$), the van der Waals volume (V/A$^{°3}$) were taken for 40 compounds of mentioned training set from the book and webbook [27].

Third, the relationships between $T_c$, $P_c$, $V_c$ with all the used molecular indices were investigated for 40 different types of alkanes derivatives using excel software and relevant equations were extracted.

Fourth, the estimation of critical properties was performed by SPSS software version 16 with MLR method and backward procedure. According to the important determining factors of this method such as correlation coefficient, square correlation coefficient, adjust square correlation coefficient, Fisher statistics, Durbin Watson, the best topological indices were determined for estimating the properties.

The linear regression model is a statistical means for analyzing the correlation between an independent variable and a dependent variable. Now, if we increase the number of independent variables to more than one, the regression model is called multiple regression model [28]. The simple linear regression equation is indicated as y=a+bx, while the multiple regression equation is introduced as:

$$y = b_0 + b_1 x_1 + \cdots + b_k x_k + e \tag{6}$$

where, y: dependent variable; $b_0$: regression constant; $b_1$: regression coefficient for first independent variable $x_1$; $b_k$: regression coefficient for k$^{th}$ independent variable $x_k$; e: the observed amount of error.

The simple regression model is linear since the increase of a descriptor x value causes the increase of dependent variable y (if the coefficient $b_i$ is positive). One of the assumptions behind the multiple regression model is that there is a linear correlation (a straight line) between the independent variables and dependent one. Several independent variables form a model which predicts the amount of dependent variable.

This research paper attempts to scrutinize the correlation between critical properties of alkanes derivatives and molecular descriptors through the use of MLR method.

## 4. RESULTS AND DISCUSSION

The experimental data of $T_c$, $P_c$ and $V_c$ of alkanes and their derivatives were shown in Table 1:

**Table 1.** Used compound, experimental data of critical temperature ($T_c$/K), critical pressure ($P_c$/Pa), critical volume ($V_c$/m$^3$) of alkanes derivatives.

| **No** | **Alkane** | **$T_c$/K** | **$P_c\times10^{-6}$/Pa** | **$V_c\times10^4$/m** |
|---|---|---|---|---|
| 1 | Ethane | 305.3 | 4.9 | 1.47 |
| 2 | Propane | 369.9 | 4.25 | 2.0 |
| 3 | n-Butane | 425.1 | 3.80 | 2.55 |
| 4 | 2-Methylpropane | 407.7 | 3.65 | 2.59 |
| 5 | n-Pentane | 469.8 | 3.36 | 3.11 |
| 6 | 2-Methylbutane | 461.0 | 3.38 | 3.06 |
| 7 | 2,2-Dimethylpropane | 433.8 | 3.20 | 3.07 |
| 8 | n-Hexane | 507.6 | 3.02 | 3.68 |
| 9 | 2-Methylpentane | 497.8 | 3.03 | 3.68 |
| 10 | 3-Methylpentane | 504.0 | 3.11 | 3.68 |
| 11 | 2,2-Dimethylbutane | 489.0 | 3.10 | 3.58 |
| 12 | 2,2-Dimethylbutane | 500.1 | 3.15 | 3.61 |
| 13 | n-Heptane | 540.0 | 2.74 | 4.28 |
| 14 | 2-Methylhexane | 530.5 | 2.74 | 4.21 |
| 15 | 3-Methylhexane | 535.2 | 2.81 | 4.04 |
| 16 | 2,2-Dimethylpentane | 520.5 | 2.77 | 4.16 |
| 17 | 2,3-Dimethylpentane | 537.3 | 2.91 | 3.93 |
| 18 | 2,4-Dimethylpentane | 519.8 | 2.74 | 4.18 |
| 19 | 3,3-Dimethylpentane | 536.4 | 2.95 | 4.14 |
| 20 | 2,2,3-Trimethylbutane | 531.1 | 2.95 | 3.98 |

**Table 1.** (Continued).

| No | Alkane | $T_c$/K | $P_c \times 10^{-6}$/ Pa | $V_c \times 10^4/m^3$ |
|---|---|---|---|---|
| 21 | 3-ethylpentane | 540.6 | 2.89 | 4.16 |
| 22 | n-octane | 568.9 | 2.49 | 4.92 |
| 23 | 2,2,3-trimethylpentane | 563.5 | 2.73 | 4.36 |
| 24 | 2,3,3-trimethylpentane | 573.5 | 2.82 | 4.55 |
| 25 | 2,2,4-Trimethylpentane | 543.9 | 2.57 | 4.68 |
| 26 | 2,2-Dimethylhexane | 549.8 | 2.53 | 4.78 |
| 27 | 3,3-Dimethylhexane | 562.0 | 2.65 | 4.43 |
| 28 | 3-Methyl-3-ethylpentane | 576.5 | 2.81 | 4.55 |
| 29 | 2,3,4-Trimethylpentane | 566.4 | 2.73 | 4.6 |
| 30 | 2,3-Dimethylhexane | 563.5 | 2.63 | 4.68 |
| 31 | 2-Methyl-3-ethylpentane | 567.1 | 2.70 | 4.42 |
| 32 | 3,4-Dimethylhexane | 568.8 | 2.69 | 4.66 |
| 33 | 2,4-Dimethylhexane | 553.5 | 2.56 | 4.72 |
| 34 | 2,5-Dimethylhexane | 550.0 | 2.49 | 4.82 |
| 35 | 2-Methylheptane | 559.7 | 2.50 | 4.88 |
| 36 | 3-Methylheptane | 563.6 | 2.55 | 4.64 |
| 37 | 4-Methylheptane | 561.7 | 2.54 | 4.76 |
| 38 | 3-Ethylhexane | 565.5 | 2.61 | 4.55 |
| 39 | n-Nonane | 595.0 | 2.30 | 5.55 |
| 40 | n-Decane | 617.8 | 2.11 | 6.24 |

The values of used topological indices of 40 different types of alkanes and their derivatives were calculated by formula 1–5, and the values of the geometry descriptors of all the mentioned compounds were taken from the book and webbook [27].

In the first section, in order to apply simple linear regression method the relationship between critical properties of the used compound with all used indices was investigated using excel software (see equations 7–39).

The following equations indicated the relationship between $T_c$ and the values of molecular indices.

| **Number** | **Equation** | **$R^2$** |
|---|---|---|
| (7) | $T_C=0.2974\ \text{Platt} + 7.0538$ | 0.5219 |
| (8) | $T_C=0.0609\ X + 1.9244$ | 0.8051 |
| (9) | $T_C=0.0255J + 2.353$ | 0.294 2 |
| (10) | $T_C=0.3181H + 5.2614$ | 0.8193 |
| (11) | $T_C=2.349\ W + 5.4462$ | 0.7733 |
| (12) | $T_C=5.9864\ WW + 9.9962$ | 0.6133 |
| (13) | $T_C=2.1433\ V+84.807$ | 0.7969 |
| (14) | $T_C=0.2754\ \text{Min PA}+21.597$ | 0.4973 |
| (15) | $T_C=0.1149\ \text{Min Z L}+7.0897$ | 0.4415 |
| (16) | $T_C=0.6768\ \text{Max PA}+29.759$ | 0.7547 |
| (17) | $T_C=0.0113\ \text{Max Z L} +5.5147$ | 0.0718 |

According to equations (7 to 17) and the square correlation coefficients ($R^2$), it can be inferred that there was better correlations between $T_c$ and H>X>V of this type of alkanes, respectively.

Also, research results indicated that the correlation coefficients values of equations (8), (10), and (13) are very close to each other and there is a significant distinction between these values and other regression coefficients. On the other hand, the correlation coefficients of the equations (7), (9), (11), (14), (15), (16), and (17) demonstrate that there is not a strong correlation between Tc and J, Platt, W, WW, MinPA, MinZL, MaxPA, MaxZL descriptors. Consequently, the descriptors MaxZL and J which possess a lower correlation compared with other descriptors were not used for predicting $T_c$ of alkanes using the MLR method.

According to the square correlation coefficient of equations (18–28), the following rank was found among $P_c$ and molecular descriptors: V>X>MaxPA>H:

| **Number** | **Equation** | **$R^2$** |
|---|---|---|
| (18) | $P_C= -86352\ \text{Platt} +4\times10^6$ | 0.6427 |
| (19) | $P_C= -615873\ X +5\times10^6$ | 0.8877 |
| (20) | $P_C=-645673\ J + 5\times10^6$ | 0.4705 |
| (21) | $P_C=-116735\ H + 4\times10^6$ | 0.8562 |
| (22) | $P_C=-13768\ W + 4\times10^6$ | 0.6881 |

| | | |
|---|---|---|
| (23) | $P_C = -4210.3\ WW + 3\times10^6$ | 0.5269 |
| (24) | $P_C = -17730\ V + 5\times10^6$ | 0.9219 |
| (25) | $P_C = -81662\ MinPA + 5\times10^6$ | 0.5172 |
| (26) | $P_C = -198960\ MinZL + 5\times10^6$ | 0.6020 |
| (27) | $P_C = -52914\ MaxPA + 5\times10^6$ | 0.8645 |
| (28) | $P_C = -370030\ MaxZL + 5\times10^6$ | 0.1238 |

Therefore, the descriptors MaxZL and J which showed a weak correlation were ignored and the descriptors X, H, V, MaxPA which had a higher correlation were preserved for prediction of alkanes $P_c$ through the use of MLR method. Also, a linear relationship between $V_c$ and X, V, H, MaxPA of this class alkanes was obtained. In accordance with the equations (29–39) it was observed that the Randic index and Volume geometry descriptor had the highest linear relationship with Vc, (R2 > 0.97).

According to the $R^2$ values of equations (30), (32), (35), and (38) the following rank was found among $V_c$ and efficient molecular descriptors: X>V >H >MaxPA.

Also, the descriptors Platt, X, H, W, WW, V, MinPA, MinZL, MaxPA which illustrated a correlation coefficient above 0.5 were used for prediction of alkanes $V_c$ using the MLR method, and the descriptors J, MaxZL which possessed a fairly weak correlation were removed.

| **Number** | **Equation** | **$R^2$** |
|---|---|---|
| (29) | $V_C = 41667\ Platt - 3.9284$ | 0.6240 |
| (30) | $V_C = 8563X - 0.3378$ | 0.9706 |
| (31) | $V_C = 3704.6\ J + 1.3583$ | 0.3767 |
| (32) | $V_C = 43537\ H - 6.0622$ | 0.9345 |
| (33) | $V_C = 320744\ W - 77.865$ | 0.8780 |
| (34) | $V_C = 835066\ WW - 229.55$ | 0.7267 |
| (35) | $V_C = 303037\ V + 4.5371$ | 0.9700 |
| (36) | $V_C = 35070 MinPA + 12.867$ | 0.4912 |
| (37) | $V_C = 18439\ MinZL + 1.8871$ | 0.6926 |
| (38) | $V_C = 96265\ MaxPA + 4.1775$ | 0.9297 |
| (39) | $V_C = 1757.2\ MaxZL + 5.026$ | 0.1058 |

In addition, the autocorrelation of descriptors used in the selected model was tested. If the regression coefficients of the diagrams indicating interrelationship between the independent variables were above 0.9, one of the independent variables was ignored.

Following MLR guidelines, the experimental critical properties, i.e. $T_c$, $P_c$, $V_c$ were selected as the dependent variables and the suitable molecular descriptors– as the independent variables– were chosen on SPSS software and backward procedure.

Then, different models were examined and the best model was defined using correlation coefficient (Pearson's r), determination coefficient, Std. Error of estimate, mean square, the Fischer statistic, sum of squares of residual and specifically Fisher statistic and the associated significance values (see Table 2).

**Table 2.** Property, Equations, R, $R^2$, $R^2_{Adjust}$,. RMSE, F statistic, SS, SSE and Sig for estimating of $T_c$, $P_c$, $V_c$.

| **Mod.** | **Prop.** | **Equation** | **R** | **$R^2$** | **$R^2_{Adjust}$** | **RMSE** | **F** | **SS** | **SSE** |
|---|---|---|---|---|---|---|---|---|---|
| 40 | $T_c$ | $_c$=8.75 Platt – 0.113 WW+ 19.995 MinZL+ 32.130 | 0.959 | 0.920 | 0.913 | 1.8509 E1 | 137.877 | 141718.56 | 12334.32 |
| 41 | $P_c$ | $P_c$= -68615.237 Platt+1413.541 WW – 194862.228 MinZL+ 5494998.974 | 0.961 | 0.923 | 0.917 | 1.4933 E5 | 144.619 | 9.67E12 | 8.028 E11 |
| 42 | $V_C$ | $V_c$=1.006E-5 Platt +1.931E-7 WW+ 2.123E-5 MinZL +5.526 E-5 | 0.986 | 0.972 | 0.970 | 1.5879 E5 | 417.097 | 3.15E-7 | 9.076 E-9 |

## 4.1 STATISTICAL PARAMETERS

**4.1.1.** Significance Level (sig): A coefficient used in the statistical method is significance level. The more the significance level is close to zero, the smaller the significance level and the linear model will be more meaningful. Therefore, the higher the Fisher statistic, the lesser significance level. As it's seen in Table 2, the best three descriptors, as predictors of $T_c$, $P_c$ and $V_C$ in terms of non–standardized coefficients, are represented using the models (40), (41) and (42), respectively.

**4.1.2.** Correlation Coefficients (R): It indicates the correlation between two variables. Statistically, the higher correlation between variables X and Y, the more accurate the prediction will be. R=0.959 in equation (40) illustrates a strong correlation between $T_c$ and Platt, WW, MinZL descriptors using the MLR method.

**4.1.3.** Determination Coefficient ($R^2$): For example, the value of $R^2$=0.972 in equation (42) illustrates that 97.2% of variation is residing in the residual meaning that the fitted line or model is very good.

**4.1.4**. Adjusted Determination Coefficient ($R^2_{Adjust}$): the percentage of adjusted determination coefficient does not represent the influence of all the independent variables, but it only illustrates the real influence of applied independent variables on the dependent variable. Thus, the high value of $R^2_{Adjust}$ (%97.0) can be used to explain the values of the $V_{c(exp)}$ variations in terms of the values of Platt, WW, MinZL independent variables.

**4.1.5.** Also, adjusted determination coefficient $R^2_{adjust}$ indicates the percentage of dependent variable that is justified by the independent variable. The small differences between $R^2_{adjust}$ and $R^2$ indicates that independent variables added to the model have been chosen more appropriately. The slight difference between the above amounts in the proposed model verifies the precision and accuracy of the model for predicting the critical properties. So, in accordance with the unstandardized coefficients, the models (40), (41) and (42) were determined for estimation of $T_c$, $P_c$ and $V_c$, respectively.

**4.1.6.** If the standard deviation of a set of data is close to zero, it indicates that the data are close to the average and have low dispersion.

**4.1.7**. Standard Error of Estimate (STD) or RMSE is used to indicate the spread of values in a distribution. It is a standard method for determining the normal, above–normal and below–normal values. It measures the error rate between the two datasets. RMSE usually compares a predicted value and an observed value.

Finally, the comparison between equations and the values of statistical coefficients showed the best models for predicting $T_c$, $P_c$ and $V_c$ of alkanes using the MLR method which are summarized as follow:

$T_c$=8.75 Platt – 0.113 WW+ 19.995 MinZL+ 232.130 ; DW=2.01

$P_c$=−68615.237 Platt+1413.541 WW – 194862.228 MinZL+ 5494998.974; DW=1.55

$V_c$=1.006E−5 Platt +1.931E-7 WW+ 2.123E-5 MinZL +5.526 E-5; DW=1.85

**4.1.8.** Standard Coefficient β: The values of standard coefficients of β related to effective descriptors used for predicting $T_c$ and $V_c$ in the final equations using MLR method were obtained as follow:

**Table 3.** The standard coefficients β values of Platt, WW, MinZL

| Descriptor/ $T_c$ | **B** | Descriptor / $P_c$ | **β** | Descriptor / $V_c$ | **β** |
|---|---|---|---|---|---|
| Platt | 0.670 | Platt | -0.637 | Platt | 0.531 |
| WW | -0.160 | WW | 0.244 | WW | 0.189 |
| MinZL | 0.643 | MinZL | -0.760 | MinZL | 0.470 |

The standard correlation coefficient β value is a measure of how strongly each predictor variable influences the dependent variable. For example, the standard coefficients β=0.670, 0.531 for the Platt variable which are used for predicting $T_c$ and $V_c$, respectively, illustrate that compared to WW and MinZL predictors, the Platt index has the strongest influence on Tc and $V_c$. Similar to above explanations, the correlation coefficient β=-0.760 reveals that the descriptor MinZL has the highest influence on dependent variable $P_c$ than Platt index. Table 4 indicates the definitive values of $T_{c(pred)}$, $P_{c(pred)}$, $V_{c(pred)}$ of alkanes and their residuals using the equations (40), (41), (42) and MLR method.

**Table 4.** The values of predicted critical properties and residuals of alkanes derivatives.

| **No** | **$T_{c\ Pred}$/K** | **Res($T_c$)/K** | **$P_{c\ Pred}$ ×$10^{-6}$/Pa** | **Res($P_c$)/Pa** | **$V_{c\ Pred}$×$10^4$/m$^3$** | **Res ($V_c$) /m$^3$** |
|---|---|---|---|---|---|---|
| **1** | 330.59 | −25.293 | 4.53574 | 364258.3 | 1.60 | −1.31E-05 |
| **2** | 379.43 | −9.535 | 4.09433 | 155665.5 | 2.15 | −1.48E-05 |
| **3** | 422.80 | 2.298 | 3.70817 | 91824.6 | 2.65 | −1.05E-05 |
| **4** | 418.04 | −10.346 | 3.78689 | −136899 | 2.61 | −2.05E-06 |
| **5** | 466.63 | 3.160 | 3.32056 | 39437.23 | 3.20 | −8.85E-06 |
| **6** | 455.33 | 5.663 | 3.46183 | −81833.6 | 3.07 | −1.20E-06 |
| **7** | 453.63 | −19.824 | 3.54328 | −343284 | 3.07 | 4.18E-07 |
| **8** | 500.99 | 6.614 | 3.03014 | −10149.5 | 3.69 | −8.11E-07 |
| **9** | 499.84 | −2.046 | 3.07081 | -35818.8 | 3.65 | 2.61E-06 |
| **10** | 484.50 | 19.499 | 3.21910 | -109106 | 3.48 | 2.02E-05 |
| **11** | 510.43 | −21.433 | 3.02988 | 70110.85 | 3.75 | −1.73E-05 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **12** | 489.79 | 10.305 | 3.19864 | −48641 | 3.53 | 8.19E-06 |
| **13** | 538.36 | 1.637 | 2.71680 | 23192.17 | 4.28 | 4.36E-07 |
| **14** | 535.29 | −4.799 | 2.77432 | −34327.9 | 4.20 | 7.80E-07 |
| **15** | 517.92 | 17.280 | 2.94087 | −130879 | 3.99 | 5.05E-06 |
| **16** | 549.01 | −28.515 | 2.69977 | 70223.33 | 4.30 | −1.44E-05 |
| **17** | 531.62 | 5.673 | 2.83559 | 74406.4 | 4.10 | −1.70E-05 |
| **18** | 515.73 | 4.0718 | 2.99304 | −253049 | 3.96 | 2.23E-05 |
| **19** | 534.92 | 1.478 | 2.83461 | 115385 | 4.13 | 1.11E-06 |
| **20** | 542.61 | −11.514 | 2.79076 | 159233.3 | 4.20 | −2.24E-05 |
| **21** | 500.54 | 40.060 | 3.10743 | −217431 | 3.78 | 3.83E-05 |
| **22** | 572.58 | −3.680 | 2.44304 | 46954.71 | 4.92 | 2.74E-07 |
| **23** | 572.12 | −8.621 | 2.55097 | 179022.4 | 4.68 | −3.17E-05 |
| **24** | 563.77 | 9.725 | 2.63106 | 188937.2 | 4.58 | −2.56E-06 |
| **25** | 579.56 | −35.665 | 2.48220 | 87799.26 | 4.79 | −1.13E-05 |
| **26** | 583.38 | −33.576 | 2.41864 | 111356.9 | 4.89 | −1.07E-05 |
| **27** | 570.81 | −8.811 | 2.53545 | 114551.2 | 4.70 | −2.67E-05 |
| **28** | 547.68 | 28.816 | 2.75675 | 53246.74 | 4.41 | 1.39E-05 |
| **29** | 563.03 | 3.371 | 2.60846 | 121533.7 | 4.59 | 1.33E-06 |
| **30** | 569.55 | −6.052 | 2.51816 | 111837 | 4.71 | −2.63E-06 |
| **31** | 526.94 | 40.157 | 2.92901 | −229019 | 4.21 | 2.10E-05 |
| **32** | 548.57 | 20.227 | 2.71979 | −29789.5 | 4.46 | 2.05E-05 |
| **33** | 549.50 | 3.994 | 2.71478 | −154782 | 4.51 | 2.14E-05 |
| **34** | 563.72 | −13.722 | 2.58063 | −90630.5 | 4.70 | 1.19E-05 |
| **35** | 569.70 | −10.007 | 2.49651 | 3483.528 | 4.82 | 5.61E-06 |
| **36** | 557.60 | 5.997 | 2.60976 | −59768.3 | 4.65 | −8.46E-07 |
| **37** | 556.17 | 5.532 | 2.62218 | −82186.8 | 4.62 | 1.42E-05 |
| **38** | 532.46 | 33.0340 | 2.84845 | −238459 | 4.32 | 2.31E-05 |
| **39** | 602.54 | −7.535 | 2.22211 | 77881.14 | 5.63 | −7.63E-06 |
| **40** | 625.41 | −7.615 | 2.08428 | 25712 | 6.40 | −1.61E-05 |

Figures 1, 2 and 3, show that there was a high linear correlation between the experimental and obtained critical properties and the estimated critical properties using the models. Figure 1 shows a high linear correlation ( $R^2$=0.9199) between the experimental and the obtained Tc using the equation (40). This diagram illustrates the values of $T_{c(pred)}$ variations obtained from equation (40) using a MLR method in terms of the $T_{c(exp)}$. The value of $R^2$=0.9199, in this diagram, indicates the fact that 91.99% of the $T_{c(pred)}$ variations can appropriately be determined in terms of one unit variation in $T_{c(exp)}$. Figure 2, shows a high linear correlation ( $R^2$=0.9234) between the experimental and the obtained $P_c$ using the equation (41). In this diagram, the high correlation between $P_{c(exp)}$ and $P_{c(pred)}$ was obtained using the MLR. The obtained value of 0.9234 for $R^2$ indicates that %92.34 of the $P_{c(pred)}$ variations can be determined in terms of one unit variation in $P_{c(exp)}$.

According to Figure 3, the indicative equation relationship between $V_{c(pred)}$ obtained from MLR model (42) was calculated as Y=X−1E-18. The value of 0.972 for $R^2$ shows that %97.2 of the $V_{c(pred)}$ variations can be determined in terms of one unit variation in $V_{c(exp)}$.

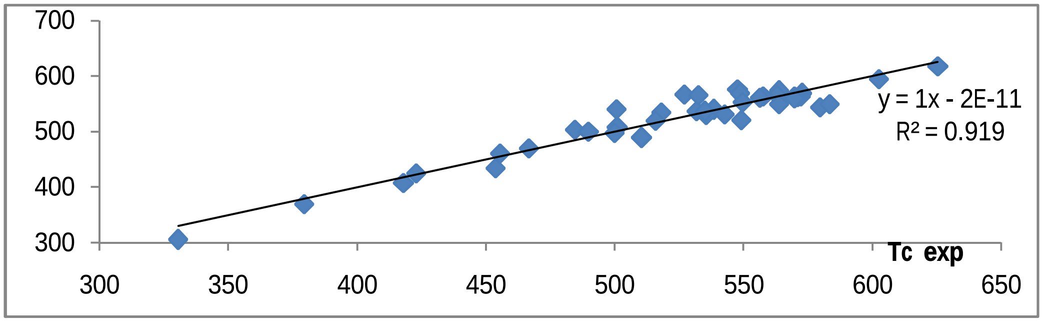


**Figure 1**. The plot of observed $T_c$ vs Predicted $T_c$.

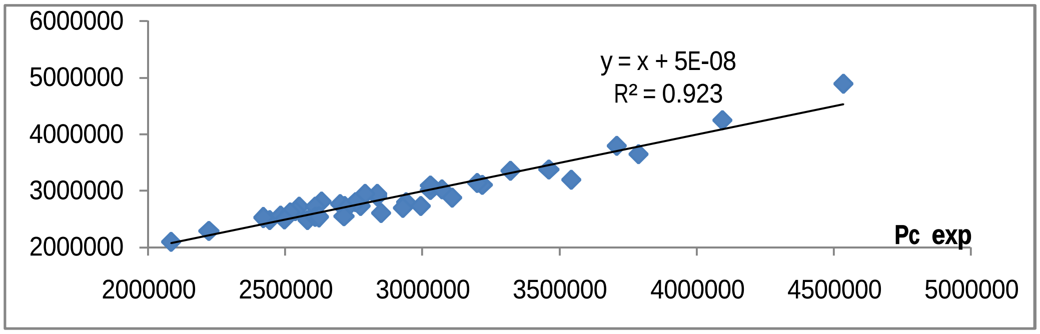


**Figure 2**. The plot of observed $P_c$ vs Predicted $P_c$.

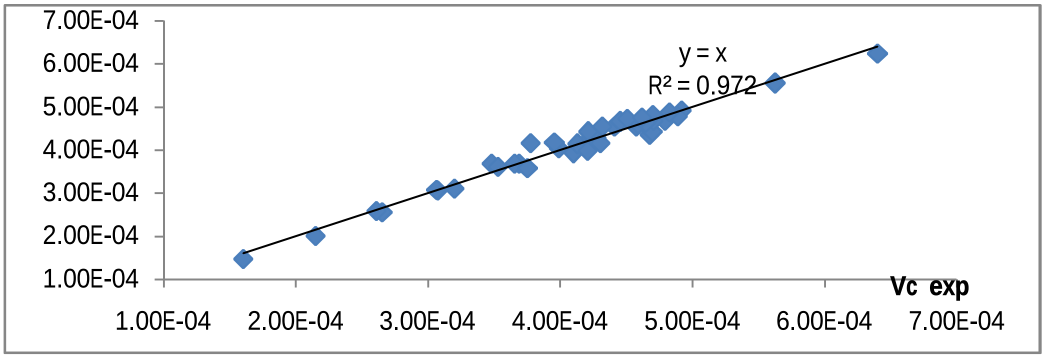


**Figure 3**. The plot of observed $V_c$ vs Predicted $V_c$.

It is obviously determined that the predicted values are so close to the experimental ones. So, it's inferred that the proposed patterns in these models have been selected correctly for determining critical properties of the studied molecules. The residual values are shown at a fairly random pattern (see Figures. 4, 5 and 6). Residuals are used to assess the normality of assumption. Figures 4, 5 and 6 show that the errors around x–axis have almost a uniform distribution. This proves the suitability of the selected pattern for proposed critical properties of alkanes.

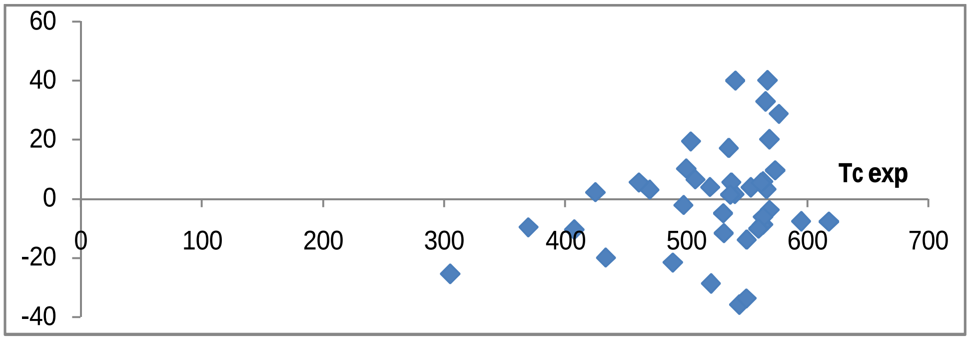


**Figure 4**. The plot of experimental $T_c$ vs the residuals.

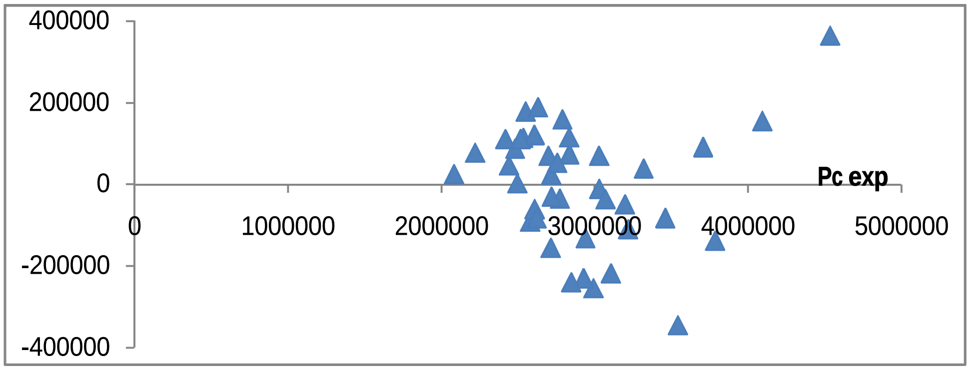


**Figure 5**. The plot of experimental $P_c$ vs the residuals.

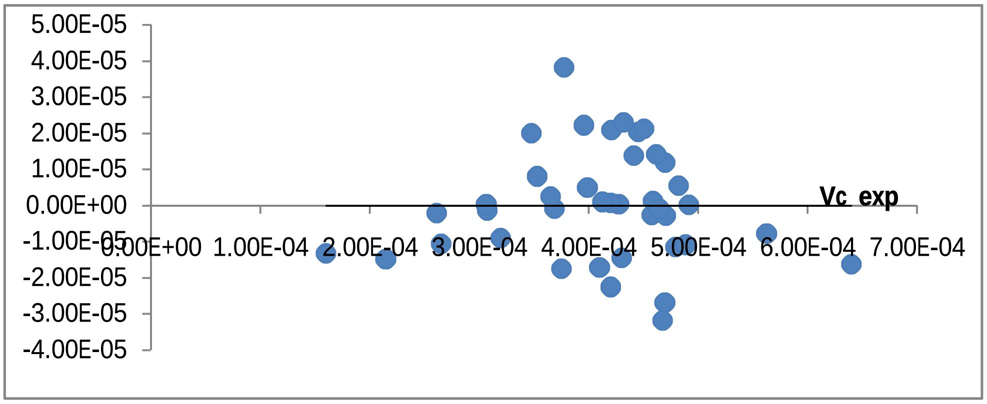


**Figure 6.** The plot of experimental $V_c$ vs the residuals.

## 5. QSAR MODEL VALIDATION

Typically, there are numerous methods for validation of QSAR models. Various statistical tests and coefficients can be used for validation of applied algorithms which, in the following, the most significant ones are represented. The statistical tests and coefficients used for estimation of $T_C$, $P_C$ and $V_C$ are as follow:

### 5.1. LEAVE–ONE–OUT CROSS–VALIDATION (LOOCV)

To determine the LOOCV, at first, a molecule from training set of 40 primary alkanes was removed. Then, QSPR was examined on the 39 remaining molecules. Considering the amounts of RSS and TSS, the amount of $Q^2$ was calculated based on the formula (43) and this cycle was repeated resulting in elimination of 25% of alkanes which were being studied leading to examining and calculating $Q^2$ for the remaining alkanes. Finally, the mean value of $Q^2$ was compared with $R^2$ in the final graphs, the results were shown in Table 5.

$$Q^2 = 1 - \frac{\sum (Y_{\exp(train)} - Y_{pred(train)})^2}{\sum (Y_{\exp(train)} - \bar{Y}_{Training})^2} = 1 - \frac{RSS}{TSS} \tag{43}$$

where, RSS refers to the residual sum of squares and TSS represents the total sum of square. This formula is the most widely used measure of the ability of a QSPR model to reproduce the data in the training set. We have computed the values of $Q^2$ (Eq. 43) using %25 of training set randomly. The values of $Q^2$ are defined as positive and less than one.

The small differences between mean $Q^2$ values of $T_c$, $P_c$, $V_c$ are equal to 0.9295, 0.9286, 0.9761, respectively, and the $R^2$ values of them verify the precision and accuracy of the model for predicting the critical properties.

**Table 5.** The values of $Q^2$ LOO of $T_c$, $P_c$, $V_c$.

| No. | $Q^2$LOO($T_c$) | $Q^2$LOO($P_c$) | $Q^2$LOO($V_c$) | Number | $Q^2$LOO($T_c$) | $Q^2$LOO($P_c$) | $Q^2$LOO($V_c$) |
|---|---|---|---|---|---|---|---|
| 1 | 0.9206 | 0.9250 | 0.9720 | 6 | 0.9350 | 0.9295 | 0.9755 |
| 2 | 0.9209 | 0.9252 | 0.9729 | 7 | 0.9353 | 0.9307 | 0.9757 |
| 3 | 0.9216 | 0.9254 | 0.9744 | 8 | 0.9355 | 0.9316 | 0.9802 |
| 4 | 0.9268 | 0.9259 | 0.9745 | 9 | 0.9359 | 0.9339 | 0.9803 |
| 5 | 0.9277 | 0.9283 | 0.9754 | 10 | 0.9360 | 0.9351 | 0.9805 |

### 5.2. MULTICOLLINEARITY TEST

Multicollinearity test was performed to avoid habits in the decision making process regarding the partial effect of independent variables on the dependent variable. A good regression model is a model in which there is not a high correlation between the independent variables. Multicollinearity test is performed through the use of SPSS software and the value of variance inflation factor (VIF) to avoid linear correlation between the independent variables. If the VIF value line is a number between 1 and 10, then there is no multicollinearity, and if VIF<1 or >10, then there is multicollinearity. In all our final models, the multicollinearity did not exist, because the values of correlations between independent variables are not close to one, and VIF values line between the numbers 1 to 10. The analysis of VIF values for all the descriptors indicated that the best models for predicting $T_c$, $P_c$ and $V_c$ values are: Platt, WW, and MinZL.

### 5.3. TEST FOR AUTOCORRELATION USING THE DURBIN–WATSON STATISTIC

From a statistical regression analysis lens of view, Durbin–Watson (DW) statistic is a number to examine autocorrelation in the residuals. The DW values 2.01, 1.55, 1.85 in final models are considered acceptable indicating that there is poor correlation between the errors and the independence of residuals. These numerical values indicate that our final models are perfect.

### 5.4. SKEWNESS AND KURTOSIS TEST

The normality of residuals represents whether the distribution function is symmetrical or asymmetrical. For a completely symmetrical distribution, the skewness and kurtosis are equal to zero. In a non–symmetrical distribution, when most of the scores "scrunch up" towards a few high scores it is positively skewed, and when most of the scores cluster towards a few low scores it is negatively skewed. Generally, if the skewness and kurtosis are placed at an interval between [2,-2], the data follow a normal distribution. The observed values for residuals skewness are 0.094, 0.076, 0.060, and the observed values for residuals kurtosis of variables $T_c$, $P_c$ and $V_c$, are 0.624, 0.500, -0.578, respectively. These indicate the normality of them.

### 5.5. APPLICABILITY DOMAIN

The applicability domain (AD) of QSAR model was used to verify the prediction reliability, identify the problematic compounds and predict the compounds with acceptable activity that fall within this domain. The most common methods used for determination of the AD of QSAR models have been described by Gramatica that used the leverage values for each compound. The leverage approach allows the determination of the position of new chemical in the QSAR model, *i.e.*

whether a new chemical will lie within the structural model domain or outside of it. The leverage approach along with the Williams Plot are used to determine the applicability domain in all QSAR models.

To construct the William Plot, the leverage $h_i$ for each chemical compound– in which QSAR model was used to predict its property– was calculated according to the following equation:

$$h_i=x^T(X^TX)^{-1}x \tag{44}$$

where, $x$ refers to the descriptor vector of the considered compound and $X$ represents the descriptor matrix derived from the training set descriptor values. The warning leverage ($h^*$) was determined as:

$$h^*=3(p+1)/n \tag{45}$$

where n is the number of training compounds and p is the number of predictor variables.
In this research, in each of the three models, the descriptor vector x includes the Platt, WW, MinZL descriptors and X is Platt, WW, MinZL descriptors matrix related to training set of alkanes. (The leverage values are shown in Table 6). Also, according to equation (45), the value of $h^*$ is equal to 0.3 (n=40 and p=3). Then, the defined applicability domain (AD) was visualized using a Williams plot, the plot of the standardized residuals versus the leverage values ($h$). A compound with $h_i > h^*$ seriously influences the regression performance and may be excluded from the applicability domain (See Figs. 7, 8, 9). The results indicated that among 40 different types of alkanes, there is just one outlier.

**Table 6.** The leverage values of used alkanes

| **Alkane** | $h_i$ | **Alkane** | $h_i$ | **Alkane** | $h_i$ | **Alkane** | $h_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.0836 | 11 | 0.0494 | 21 | 0.0234 | 31 | 0.0444 |
| 2 | 0.1047 | 12 | 0.0363 | 22 | 0.0946 | 32 | 0.0350 |
| 3 | 0.1146 | 13 | 0.0832 | 23 | 0.0952 | 33 | 0.0363 |
| 4 | 0.0569 | 14 | 0.0545 | 24 | 0.1050 | 34 | 0.0358 |
| 5 | 0.1193 | 15 | 0.0303 | 25 | 0.0864 | 35 | 0.0483 |
| 6 | 0.0558 | 16 | 0.0443 | 26 | 0.0458 | 36 | 0.0383 |
| 7 | 0.0453 | 17 | 0.0338 | 27 | 0.0512 | 37 | 0.0364 |
| 8 | 0.0979 | 18 | 0.0285 | 28 | 0.0673 | 38 | 0.0287 |
| 9 | 0.0556 | 19 | 0.0493 | 29 | 0.0560 | 39 | 0.2072 |
| 10 | 0.0417 | 20 | 0.0843 | 30 | 0.0329 | 40 | 0.5754 |

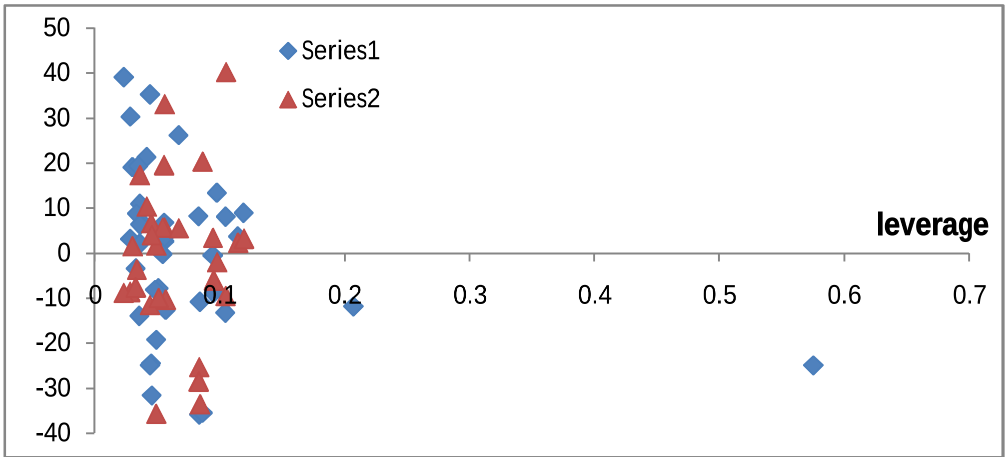


**Figure 7**. Williams plot of residual $T_c$ vs leverage.

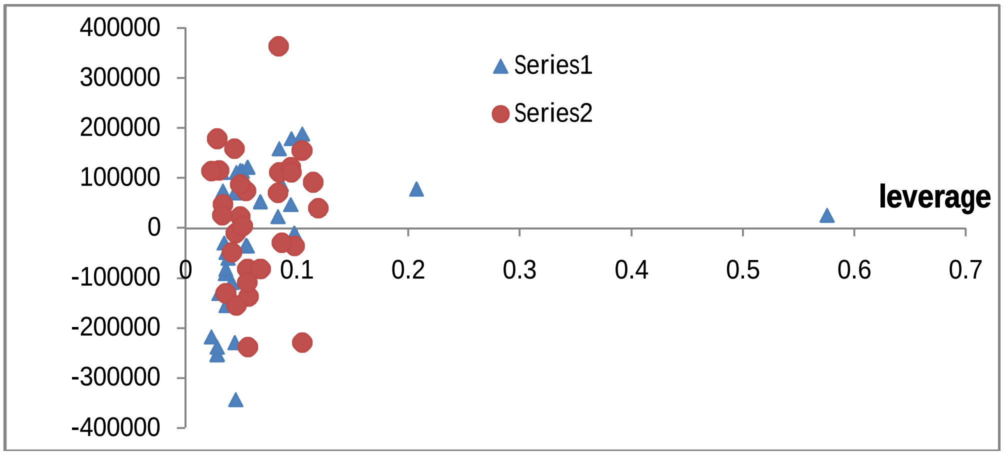


**Figure 8.** Williams plot of residual $P_c$ vs leverage.

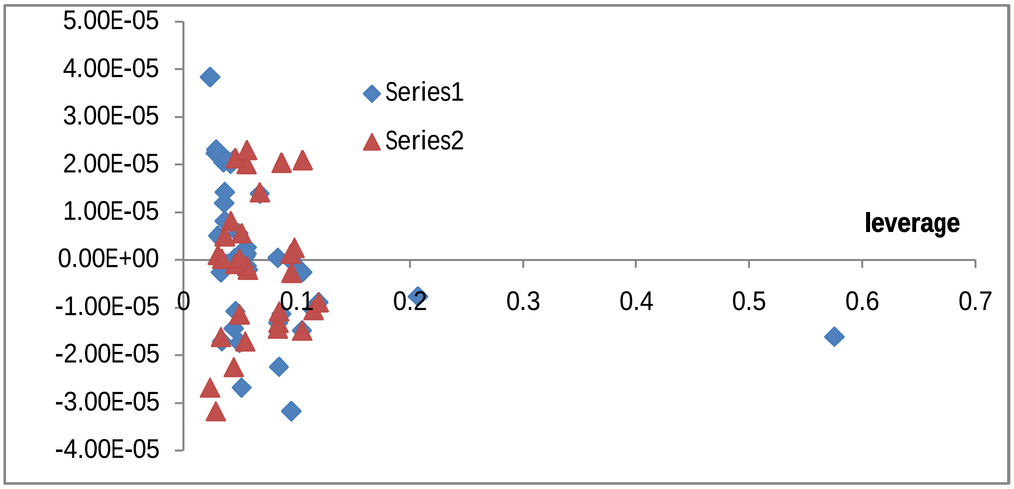


**Figure 9.** Williams plot of residual $V_c$ vs leverage.

Thus, the analyses of various statistical coefficients, tables, diagrams and QSPR model validation through the use of MLR method show that they possess the necessary and sufficient validity for predicting the critical properties of alkanes.

## 6. Conclusion

The afore–mentioned computational methods involve methods which all focus on the molecular structures and properties. The underlying concept of these methods is based on the fact that the molecular and geometric structures are responsible for all the physical and chemical properties of molecules including t. The results of the present study indicate that the simple linear regression model with dispersion coefficient (alone) is not sufficient for determining the critical properties of alkanes. However, the multiple linear and regression model benefiting from various descriptors, factors and efficient coefficients can suggest the best algorithm for determining these properties. It was also witnessed that among the proposed models to predict the critical properties, the model including a combination of descriptors Hyper–Wiener, Platt, and MinZL is the most appropriate one. And the last but not least, this was the first time that the relationship between critical properties with molecular descriptors of alkanes and their derivatives was investigated through the use of SPSS software and MLR method.

## REFERENCES

1. R. T. Morison and R. Neilson Boyd, *Organic Chemistry*, Allyn & Bacon, 2003.
2. H. Wiener, Correlation of heats of isomerization and differences in heats of vaporization of isomers, among the paraffin hydrocarbons, *J. Am. Chem. Soc.* **69** (1947) 2636–2638.
3. A. A. Gakh, E. G. Gakh, B. G. Sumpter and D. W. Noid, Neural network–graph theory approach to the prediction of the physical properties of organic compounds, *J. Chem. Inf. Comput. Sci.* **34** (1994) 832–839.
4. O. Ivanciuć, The neural network MolNet prediction of alkane enthalpies, *Anal. Chem. Acta.* **384** (1999) 271–284.
5. D. C. Montgomery, E. A. Peck and G. G. Vining, *Introduction to Linear Regression Analysis*, John Wiley & Sons, Inc, 2006.
6. R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors, Methods and Principles in Medicinal Chemistry,* Wiley–VCH Verlag GmbH, 2008.
7. I. Gutman and B. Furtula (eds), *Novel Molecular Structure Descriptors–Theory and Applications I and II*, University of Kragujevac and Faculty of Science Kragujevac, 2010.

8. I. Gutman, A formula for the Wiener number of trees and its extension to graphs containing cycles, *Graph Theory Notes N. Y.* **27** (1994) 9−15.
9. R. B. King, *Chemical Applications of Topology and Graph Theory*, Elsevier, Amsterdam, 1983.
10. I. Gutman and O. E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer–Verlag, Berlin, 1986.
11. M. Randić, *Chemical Graph Theory–Facts and Fiction*, NISCAIR–CSIR, India, 2003.
12. M. Randić, Quantitative Structure–property relationship: boiling points of planar 1009–benzenoids, *New. J. Chem.* **20** (1996) 1001–1009.
13. M. Shamsipur, B. Hemmateenejad and M. Akhond, Highly Correlating Distance/Connectivity–Based Topological Indices. 1: QSPR Studies of Alkanes, *Bull. Korean Chem. Soc*. 25 (2004) 253–259.
14. H. Hosseini and F. Shafiei, Quantitative Structure Property Relationship Models for the Prediction of Gas Heat Capacity of Benzene Derivatives Using Topological Indices, *MATCH Commun. Math. Comput. Chem*. **75** (2016) 583–592.
15. M. Goodarzi and E. Mohammadinasab, Theoretical investigation of relationship between quantum chemical descriptors, topological indices, energy and electric moments of zig–zag polyhex carbon nanotubes $TUHC_6[2p,q]$ with various circumference [2p] and fixed lengths, *Fullerenes, Nanotubes Carbon Nanostructures.* **21** (2013) 102–112.
16. A. Alaghebandi and F. Shafiei, QSPR modeling of heat capacity, thermal energy and entropy of aliphatic aldehydes by using topological indices and MLR method, *Iranian J. Math. Chem.* **7** (2016) 235–251.
17. M. Pashm Forush, F. Shafie and F. Dialamehpour, QSPR study on benzene derivatives to some physico chemical properties by using topological indices, *Iranian J. Math. Chem*. **7**(1) (2016) 93–110.
18. H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc*. **69** (1947) 17–20.
19. G. Cash, S. Klavžar and M Petkovsek, Three Methods for Calculation of the Hyper−Wiener Index of Molecular Graphs, *J. Chem. Inf. Comput. Sci*. **42** (2002) 571–576.
20. X. Li and Y. Shi, A survey on the Randić index, *MATCH Commun. Math. Comput. Chem.* **59** (2008) 127–156.
21. M. Randić, Characterization of atoms, molecules and classes of molecules based on paths. enumerations, *MATCH Commun. Math. Comput. Chem*. **7** (1979) 5–64.
22. B. Liu and I. Gutman, On general Randić indices, *MATCH Commun. Math. Comput. Chem.* **58** (2007) 147–154.

23. M. Randić, Charactrization of molecular branching, *J. Am. Chem. Soc*. **97** (1975) 6609–6615.
24. A. T. Balaban and T. S. Balaban, New Vertex Invariant and topological indices of chemical graphs based on information on distance, *J. Math. Chem*. **8** (1991) 383–397.
25. A. T. Balaban, Topological index based on topological distances in molecular graph, *Pure Appl. Chem.* **55** (1983) 199–206.
26. K. C. Das, B. Zhou and N. Trinajstić, Bounds on Harary index, *J. Math. Chem*. (2009) 1377–1393.
27. www.chemicalize.org
28. M. Randić and S. C. Basak, Multiple regression analysis with optimal molecular descriptors, *SAR QSAR Environ. Res.* **11** (2000) 1–23.
29. G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, Hoboken, NJ: Wiley–Interscience, 2003.
30. K. Roy and I. Mitra, On various metrics used fo r validation of predictive QSAR models with applications in virtual screening and focused library design, *Comb. Chem. High Throughput Screen.* **14** (2011) 450–474.

Iranian J. Math. Chem. 8(2) June (2017) 221 – 229

Iranian Journal of Mathematical Chemistry

Journal homepage: ijmc.kashanu.ac.ir

# *Some Relations between Kekulé Structure and Morgan−Voyce Polynomials*

**İNCI GÜLTEKIN[1,•] AND BÜNYAMIN ŞAHIN[2]**

[1]Department of Mathematics, Faculty of Science, Atatürk University, 25240 Erzurum, Turkey
[2]Department of Elemantary Mathematics Education, Faculty of Education, Bayburt University, 69000 Bayburt, Turkey

**ARTICLE INFO**

**Article History:**
Received 2 March 2016
Accepted 9 May 2016
Published online 11 April 2017
Academic Editor: Michel Marie Deza

**Keywords:**
Kekulé structure
Hosoya index
Morgan−Voyce polynomials
Caterpillar trees

**ABSTRACT**

In this paper, Kekulé structures of benzenoid chains are considered. It has been shown that the coefficients of a $B_n(x)$ Morgan Voyce polynomial equal to the number of $k$–matchings $(m(G,k))$ of a path graph which has $N = 2n + 1$ points. Furtermore, two relations are obtained between regularly zig–zag non-branched catacondensed benzenoid chains and Morgan–Voyce polynomials and between regularly zig−zag non branched catacondensed benzenoid chains and their corresponding caterpillar trees.

## 1. INTRODUCTION

A benzenoid system is obtained by using the regular hexagons consecutively so that two hexagons are either disjoint or have a common edge [1]. An example of benzenoid chain is illustrated in Figure 1.

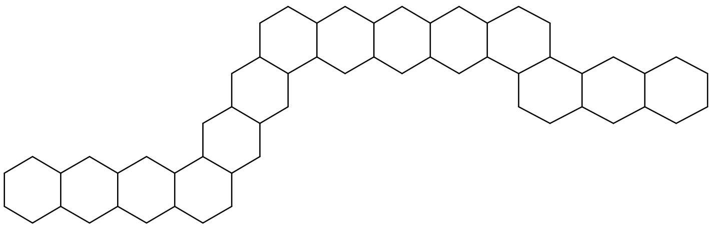


**Figure 1**. A Benzenoid Chain.

In connection with the benzenoid chains the $LA$–sequence is defined as an ordered $h$–tuple $(h > 1)$ of the symbols $L$ and $A$. The $i$–th symbol is $L$ if the $i$–th hexagon is of

---

[•] Corresponding Author (Email: shnbnymn25@gmail.com)

DOI: 10.22052/ijmc.2017.49481.1177

mode$L_1$ or $L_2$. The $i$–th symbol is $A$ if the $i$–th hexagon is of mode $A$. The definition of $L_1$, $L_2$ and $A$modes of hexagons is clear from Figure 2.

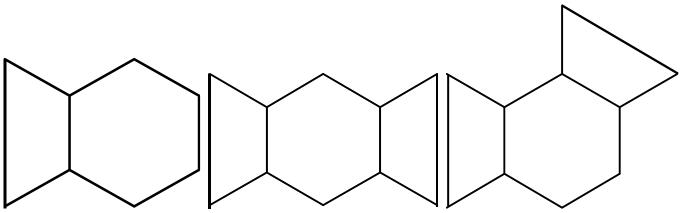


**Figure 2.** Illustration of $L_1$, $L_2$ and $A$ modes of hexagons, respectively.

For instance, the $LA$–sequence of the benzenoid chain in Figure 1 is $LLLALLALLLAALL$ or, in the abbreviated form $L^3AL^2AL^3A^2L^2$. Each perfect matching of a benzenoid system (if any exists) represents a Kekulé structure. The number of Kekulé structures of benzenoid chains is called its“$K$ number”. The $K$–number of a benzenoid chain is calculated by its $LA$–sequence.

Balaban and Tomescu coined the term isoarithmicity for the benzenoid chains which their K numbers are same [2]. It is denoted by $\langle x_1, x_2, \ldots, x_n\rangle$ the class of isoarithmic benzenoid chains with the $LA$–sequence

$$L^{x_1}AL^{x_2}A \ldots AL^{x_n}$$

where $n \geq 1$, and $x_1 \geq 1,\ x_n \geq 1,\ x_i \geq 0$ for $i = 2,3, \ldots, n-1$. For example isoarithmic class of the benzenoid chain which is depicted in Figure 1 is $\langle 3,2,3,0,2\rangle$.

Every benzenoid chain can be represented in this form. It is denoted by $K_n\langle x_1, x_2, \ldots, x_n\rangle$ the number of Kekulé structures of the chain $\langle x_1, x_2, \ldots, x_n\rangle$.It is defined for the initial terms of the K numbers such that ([1]) $K_0 = 1, K_1\langle x_1\rangle = 1 + x_1$.

**Theorem 1.** If $n \geq 2$ then for arbitrary $x_1 \geq 1,\ x_n \geq 1,\ x_i \geq 0,\ (i = 2,3, \ldots, n-1)$, the following recurrence relation holds [1]

$$K_n\langle x_1, x_2, \ldots, x_n\rangle = (x_n + 1)K_{n-1}\langle x_1, x_2, \ldots, x_{n-1}\rangle + K_{n-2}\langle x_1, x_2, \ldots, x_{n-2}\rangle.$$

## 2. THE HOSOYA INDEX AND MORGAN–VOYCE POLYNOMIALS

The Hosoya or $Z$–index was defined by Hosoya in 1971 [3] and the Hosoya index of a graph $G$ is denoted by $Z(G)$. The $Z(G)$, is the total number of $k$–matchings which are the number of $k$ choosing from a graph $G$ such that the $k$ lines are non–adjacent where $N$ is the number of points.

**Definition 1**. The number of $k$–matchings is denoted by $m(G,k)$ and the $Z(G)$ is defined as $Z(G) = \sum_{k=0}^{\lfloor N/2\rfloor} m(G,k)$ such that $m(G,0) = 1$ for any graph $G$.

**Theorem 2**. The number of $k$–matchings of the path graph is calculated by the following equation [4]

$$m(G,k) = \binom{N-k}{k}, \text{ for } 0 \leq k \leq \lfloor N/2 \rfloor.$$

Relations between topological indices and some orthogonal polynomials for example Hermite, Laguerre and Chebyshev polynomials were found by Hosoya ([5]). Another relation between the sextet polynomial of a hexagonal chain and the matching polynomial of a caterpillar tree was discovered by Gutman [6]. As a result of this paper, it has been shown that the $K$–number of a hexagonal chain is equal to the Hosoya index of the corresponding caterpillar [7]. For instance, corresponding caterpillar tree of the hexagonal chain which is depicted in Figure 1 is on the below.

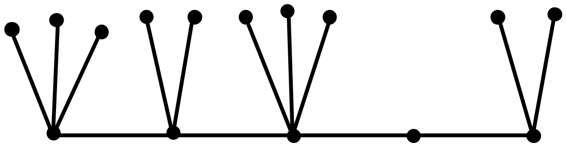


**Figure 3**. The hexagonal chain in Figure 1 has 14 hexagons and the corresponding caterpillar tree has 14 edges.

The caterpillar tree of the hexagonal chain in Figure 3 is $C_5(4,3,4,1,3)$.

**Definition 2.** The Morgan–Voyce polynomials $B_n(x)$ is defined by [8] as

$$B_n(x) = \sum_{i=0}^{n} \binom{n+i+1}{n-i} x^i$$

and the first five Morgan–Voyce polynomials are found from this equation like that

$$B_0(x) = 1$$
$$B_1(x) = x + 2$$
$$B_2(x) = x^2 + 4x + 3$$
$$B_3(x) = x^3 + 6x^2 + 10x + 4$$
$$B_4(x) = x^4 + 8x^3 + 21x^2 + 20x + 5\ .$$

## 3. REGULARLY ZIG–ZAG NON–BRANCHED CATACONDENSED BENZENOIDS

The Kekulé number of regularly zig–zag non-branched cata condensed benzenoids was found by He, He and Xie [9] by Peak–Valley matrix.

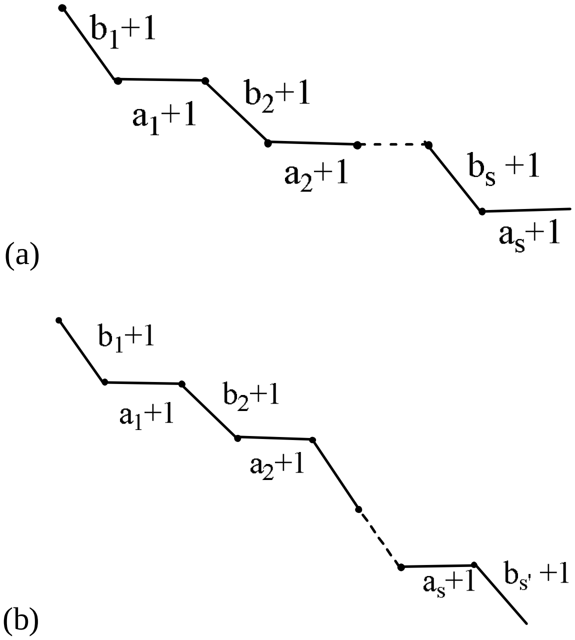


**Figure 4**. Dualist graph of a general non-branched cata–condensed benzenoids.

In Figure 4, $a_i \in \quad (i = 1,2, \dots s)$ and $b_i \in \quad (i = 1,2, \dots s')$ where $s' = s$ for Figure 4(a) and $s' = s + 1$ for Figure 4(b). $a_i + 1$ and $b_i + 1$ represent the numbers of linearly condensed six–membered rings horizontally and diagonally, respectively. For the benzenoid shown in Figure 4(a) and 4(b), the Peak–Valley matrix is as follows.

$$A_n = \begin{bmatrix} t_1 & 1 & 0 & & & \\ 1 & t_2 & 1 & & 0 & \\ 0 & 1 & t_3 & & & \\ & & & \ddots & 1 & 0 \\ & 0 & & 1 & t_{\ \ -1} & 1 \\ & & & 0 & 1 & t \end{bmatrix}$$

where $t_i = \begin{cases} b_{k+1} + 2, & if\ i = \sum_{j=0}^{k} a_j + 1 \\ 2, & if\ \ i \neq \sum_{j=0}^{k} a_j + 1 \end{cases}$, $k = 1,2, \dots, s$; $i = 1,2, \dots, \quad$. Here is the number of peaks (or valleys) in a graph $G$. The Kekulé number of a graph $G$ is shown by $K_n(G)(\ n = 1, \dots, \quad)$.

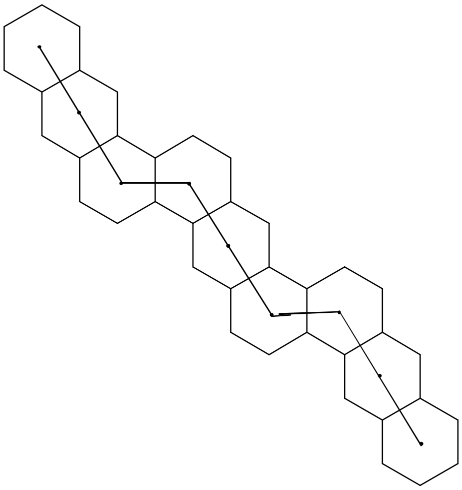


**Figure 5**. Simple binary regularly cata–condensed benzenoids.

**Lemma 1**. From Figure 5, the $K$–number of the graph $G$ is calculated by the following tri–diagonal determinantal expression[9]:

$$K_{\mathrm{n}}(G) = detA_n = \begin{vmatrix} b+2 & 1 & 0 & & & \\ 1 & b+2 & 1 & & 0 & \\ 0 & 1 & b+2 & & & \\ & & & \ddots & 1 & 0 \\ & 0 & & 1 & b+2 & 1 \\ & & & 0 & 1 & b+2 \end{vmatrix}.$$

The order of the above determinant is $s + 1$, where $s$ is the repeat times of horizontal linear segments on the graph$G$.

## 4. CONTINUANTS AND CATERPILLAR TREES

**Lemma 2.** If $H$ is a hexagonal chain whose $LA$–sequence is $L^{x_1}AL^{x_2}A \dots L^{x_{n-1}}AL^{x_n}$, then the number $K(H)$ of its Kekulé structures is equal to the $Z$–index of the caterpillar tree $C_n(x_1, x_2, \dots, x_n)$[7].

If it is written $C(H)$ for caterpillar tree of a $H$ hexagonal chain, Lemma 2 is equivalent to the equality $K(H) = \mathrm{Z}(C(H))$.

**Definition 3.**The continuants (or continuant polynomials) are introduced by Euler [10] as $L_n(x_1, x_2, \dots, x_n) = x_n L_{n-1}(x_1, x_2, \dots, x_{n-1}) + L_{n-2}(x_1, x_2, \dots, x_{n-2})$ with initial conditions $L_0() = 1$, $L_1(x_1) = x_1$ and $L_2(x_1, x_2) = x_1 x_2 + 1$.

From this it is shown that the $Z$–index of the caterpillar trees coincides with Euler 's continuant like the following lemma.

**Lemma 3.** $Z\big(C_n(x_1, x_2, \dots, x_n)\big) = L_n(x_1, x_2, \dots, x_n)$[7].

## 5. MAIN RESULTS

**Theorem 3.** The coefficients of a $B_n(x)$ Morgan–Voyce polynomial are equal to the number of $k$–matchings ($m(G,k)$) of a path graph which has $N = 2n + 1$ points.

**Proof.** We denote the coefficients of Morgan–Voyce polynomials with

$$C\big(B_n(x)\big) = \binom{n+i+1}{n-i}$$

such that $0 \le i \le n$ and we take the point number of the path graph $N = 2n + 1$. The number of $k$–matchings of a path graph for $0 \le k \le \lfloor N/2 \rfloor$ is

$$m(G,k) = \binom{N-k}{k}$$

and $\lfloor N/2 \rfloor = \lfloor (2n+1)/2 \rfloor = n$ by the definition of the Hosoya index. Now we demonstrate the coefficients of the Morgan–Voyce polynomials in combinatorial form with respectively for $0 \le i \le n$

$$C\big(B_n(x)\big) = \binom{n+1}{n}, \binom{n+2}{n-1}, \dots, \binom{2n}{1}, \binom{2n+1}{0}$$

and $m(G,k) = \binom{N-k}{k}$ for $0 \le k \le \lfloor N/2 \rfloor = n$ with respectively

$$m(G,k) = \binom{2n+1}{0}, \binom{2n}{1}, \dots, \binom{n+2}{n-1}, \binom{n+1}{n}.$$

It is clear that $C\big(B_n(x)\big)$ and $m(G,k)$ are same in reverse order. From this we say for every $n^{th}$ degree Morgan–Voyce polynomial there is a path graph ($P_N$) which has $N = 2n + 1$ points such that the coefficients of the Morgan–Voyce polynomials equal to the number of $k$–matchings of $P_N$.

**Example 1**. We show an application of the previous theorem for the first three Morgan–Voyce polynomials. For $B_0(x)$, $C\big(B_0(x)\big) = 1$ equals to $m(G,k)$ for $N = 2 \times 0 + 1 = 1$. For $B_1(x)$, $C\big(B_1(x)\big) = 1, 2$ equal to $m(G,k)$ for $N = 2 \times 1 + 1 = 3$. For $B_2(x)$,$C\big(B_2(x)\big) = 1, 4, 3$ equal to $m(G,k)$ for $N = 2 \times 2 + 1 = 5$.

**Lemma 4.** If $b_1 + 1 = b_2 + 1 = \cdots = b_s + 1 = b + 1$ (numbers of the regular hexagons on diagonal wise are same) like in Figure 5 and we take $x$ instead of $b_i$, then (the right equation is used to express many properties of the Morgan–Voyce polynomials like in [8])

$$K_n(G) = detA_n = B_n(x).$$

**Proof.**

$$\begin{array}{llll}
K_1(G) = \begin{vmatrix} x+2 \end{vmatrix} & = x+2 & = B_1(x) & \\
K_2(G) = \begin{vmatrix} x+2 & 1 \\ 1 & x+2 \end{vmatrix} & = (x+2)(x+2)-1 & = x^2+4x+3 & = B_2(x) \\
K_3(G) = \begin{vmatrix} x+2 & 1 & 0 \\ 1 & x+2 & 1 \\ 0 & 1 & x+2 \end{vmatrix} & = x^3+6x^2+10x+4 & = B_3(x) &
\end{array}$$

and by the determinant of the tri–diagonal matrix in Lemma 1,

$$K_n(G) = B_n(x) = (x+2)B_{n-1}(x) - B_{n-2}(x).$$

In Lemma 1, the $(n)$ indice on the notatin $K_n$ is the number of the repetition of the diagonal hexagons. We also take the number of the hexagons $b_i + 1$ on diagonal wise like the previous lemma. For Figure 5, $b_1 + 1 = b_2 + 1 = \cdots = b_s + 1 = b + 1$ and its corresponding caterpillar tree is $C_{2n}(b+1, 1, b, 1, \ldots, b, 1)$.

There is a relation between the $K$–number of the hexagonal chain in Figure 5 and $Z$–index of its corresponding caterpillar tree as noted in the next theorem.

**Theorem 4.** $K_n(G) = Z\big(C_{2n}(G)\big)$.

**Proof.** Induct on $n$. For $n = 1$, $K_1(G) = Z\big(C_2(b+1, 1)\big) = b + 2$, as desired. We assume that the equality is true for $n \leq k$ and we will show that it is true for $n = k + 1$. This means

$$K_{k+1}(G) = Z\big(C_{2k+2}(b+1, 1, b, 1, \ldots, b, 1)\big).$$

By assumption

$$K_k(G) = Z\big(C_{2k}(b+1, 1, b, 1, \ldots, b, 1)\big)$$

and

$$K_{k-1}(G) = Z\big(C_{2k-2}(b+1, 1, b, 1, \ldots, b, 1)\big).$$

By Lemma 1,

$$\begin{aligned}
K_{k+1}(G) &= (b+2)K_k(G) - K_{k-1}(G) \\
&= (b+2)Z(C_{2k}(G)) - Z(C_{2k-2}(G)) \\
&= bZ(C_{2k}(G)) + 2[Z(C_{2k-1}(G)) + Z(C_{2k-2}(G))] - Z(C_{2k-2}(G)) \\
&= bZ(C_{2k}(G)) + Z(C_{2k-1}(G)) + Z(C_{2k-1}(G)) + Z(C_{2k-2}(G)) \\
&= Z(C_{2k+1}(G)) + Z(C_{2k}(G)) = Z(C_{2k+2}(G))
\end{aligned}$$

This complete the proof.

**Example 2.** We calculate the Kekulé number of simple binary regularly catacondensed benzenoid in Figure 5 by two ways mentioned in the Theorem 4. The matrix form of $K$–number of the chain shown in Figure 5 is

$$K_3(G) = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix}$$

and $K_3(G) = detA = 56$. Now we use the corresponding caterpillar tree of the hexagonal chain as the follows:

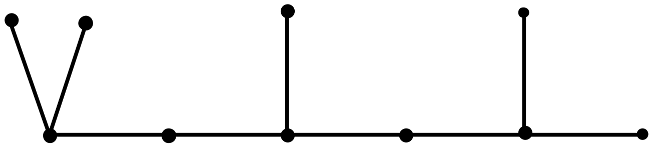


**Figure 6**. The hexagonal chain in Figure 5 has 9 hexagons and the corresponding caterpillar tree has 9 edges.

This caterpillar tree is denoted by $C_6(3,1,2,1,2,1)$ and $Z(C_6(3,1,2,1,2,1) = 56$. So that $K_3(G) = Z(C_6(3,1,2,1,2,1)$.

## ACKNOWLEDGEMENT

B.ŞAHİN thanks to the Scientific and Technological Research Council of Turkey (TUBITAK) for support.

## REFERENCES

1. R. Tošić, I. Stojmenović, Chemical graphs, Kekulé structures and Fibonacci numbers, *Zb. Rad. Prirod.–Mat. Fak. Ser. Mat.* **25** (2) (1995) 179–195.
2. A. T. Balaban, I. Tomescu, Algebratic expressions for the number of Kekulé structure of isoarithmic cata–condensed benzenoid polycyclic hydrocarbons, *MATCH Commun. Math. Comput. Chem.* **14** (1983) 155–182.
3. H. Hosoya, Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, *Bull. Chem. Soc. Jpn.* **44** (1971) 2332–2339.
4. H. Hosoya, Topological index and Fibonacci numbers with relation to chemistry, *Fibonacci Quart.***11** (1973) 255–269.
5. H. Hosoya, Graphical and combinatorial aspects of some orthogonal polynomials, *Natur. Sci. Rep. Ochanomizu Univ.* **32** (2) (1981) 127–138.

6. I. Gutman, Topological properties of benzenoid systems. An identity for the sextet polynomial, *Theor. Chim. Acta* **45** (1977) 309–315.
7. H. Hosoya, I. Gutman, Kekulé structures of hexagonal chains–some unusual connections, *J. Math. Chem.* **44** (2008) 559–568.
8. T. Koshy, *Fibonacci and Lucas numbers with applications*, Pure and Applied Mathematics (New York), Wiley–Interscience, New York, 2001.
9. W. J. He, W. C. He, S. L. Xie, Algebratic expressions for Kekulé structure counts of nonbranched cata–condensed benzenoid, *Discrete Appl. Math.* **35** (1992) 91–106.
10. R. L. Graham, D. E. Knuth, O. Patashnik, *Concrete Mathematics. A Foundation for Computer Science*, Addison–Wesley, Reading, 1989.

# ABSTRACTS
# IN
# PERSIAN

# A novel topological descriptor based on the expanded wiener index: applications to QSPR/QSAR studies

AFSHAN MOHAJERI[1], POUYA MANSHOUR[2], AND MAHBOUBEH MOUSAEE[1]

[1]Department of Chemistry, College of Sciences, Shiraz University, Shiraz 71454, Iran.
[2]Department of Physics, Persian Gulf University, Bushehr 75169, Iran.

# ارائه توصیف‌گر توپولوژیکی نوین بر پایه شاخص تعمیم‌یافته وینر: کاربردهایی در مطالعاتQSPR/QSAR

ادیتور رابط : علیرضا اشرفی

## چکیده

در این مقاله، یک شاخص جدید توپولوژیکی به نام شاخص M بر پایه شکل تعمیم‌یافته ماتریس وینر معرفی می‌شود. برای ساخت این شاخص، خصوصیات اتمی و برهم‌کنش بین راس‌ها در یک ملکول در نظر گرفته می‌شوند. مفید بودن شاخص M به وسیله مدل‌های QSPR/QSAR گوناگون برای خصوصیات مختلف فیزیکی-شیمیایی و فعالیت‌های زیستی تعداد زیادی از ترکیبات مختلف نشان داده می‌شود. به علاوه، کاربردی بودن شاخص پیشنهادی بین ترکیبات ایزومری نیز بررسی شده است. در هر مورد، پایداری مدل بدست آمده به وسیله‌ی آزمون اعتبار سنجی متقابل تایید می‌شود. نتایج مطالعه‌ی پیش رو نشان می‌دهد که شاخص M یک راه امید بخش برای توسعه‌ی مدل‌های شدیدا همبسته QSPR/QSAR را فراهم می‌آورد. از سوی دیگر، شاخص M به سادگی تولید شده و توسعه مدل‌های QSPR بر پایه‌ی این شاخص به صورت خطی همبسته هستند. این یک خاصیت جالب شاخص M در مقایسه با توصیف کننده‌های کوانتومی‌شیمیایی است که نیازمند هزینه‌های محاسباتی وسیع بوده و محدودیت‌هایی را برای ملکول‌های بزرگ از خود نشان می‌دهند.

**لغات کلیدی**: شاخص توپولوژیکی، نظریه‌ی گراف، شاخص وینر تعمیم‌یافته، QSPR، QSAR.

# A new two-step Obrechkoff method with vanished phase-lag and some of its derivatives for the numerical solution of radial Schrödinger equation and related IVPs with oscillating solutions

ALI SHOKRI AND MORTAZA TAHMOURASI*

Faculty of Mathematical Science, University of Maragheh, Maragheh, Iran

# یک روش دوگامی ابرشکوف جدید با صفر کردن فاز تاخیری و برخی از مشتق‌های آن برای حل عددی معادله شرودینگر شعاعی و مسائل مقدار اولیه وابسته به آن با جوابهای نوسانی

ادیتور رابط : ایوان گوتمن

## چکیده

در این مقاله، یک روش جدید صریح خطی دوگامی ابرشکوف از مرتبه جبری دوازدهم با صفر کردن فاز تاخیری و مشتق‌های اول، دوم، سوم و چهارم را ارائه می‌دهیم. هدف این مقاله توسعه الگوریتم کارا برای حل تقریبی معادله شرودینگر شعاعی تک-بعدی و مسائل وابسته می‌باشد. این الگوریتم در رسته روش‌های چندگامی قرار دارد. برای تهیه روش چندگامی کارا، خاصیت فاز تاخیری و مشتق‌های آن مورد استفاده قرار می‌گیرند. تحلیل خطا و تحلیل پایداری نیز بررسی، و مقایسه آن با روش‌های دیگر مورد مطالعه قرار می‌گیرد. دقت روش جدید با استفاده از تحلیل نظری و کاربردهای عددی اثبات می‌شود.

**لغات کلیدی**: معادله شرودینگر، فاز تاخیری، معادلات دیفرانسیل معمولی، روش‌های چندگامی متقارن.

# *Optimal Control of Switched Systems by a Modified Pseudo Spectral Method*

**HAMID REZA TABRIZIDOOZ, MARZIEH POURBABAEE AND MEHRNOOSH HEDAYATI**

Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Kashan, Kashan, I.R. Iran

# کنترل بهینه‌ی سیستم‌های سوییچ با استفاده از یک روش شبه‌طیفی بهبودیافته

ادیتور رابط : علیرضا اشرفی

## چکیده

در این مقاله، یک طرح شبه‌طیفی بهبودیافته به منظور حل مساله‌ی کنترل بهینه‌ای که توسط یک سیستم دینامیکی سوییچ کنترل می‌شود، ارایه می‌گردد. بسیاری از فرایندهای واقعی مانند فرایندهای شیمیایی، فرایندهای مربوط به سیستم‌های اتومبیل و فرایندهای صنعتی را می‌توان به صورت این سیستم‌ها مدل‌سازی کرد. برای این منظور، مساله‌ی اصلی با یک مساله‌ی کنترل بهینه‌ی دیگر، که در آن زمان‌های سوییچ به عنوان پارامترهای مجهول ظاهر شده‌اند، جایگزین می‌شود. سپس با استفاده از روش مربع‌سازی لژاندر-گاوس-لوباتو و ماتریس مشتق متناظر با آن، مساله‌ی جایگزین گسسته‌سازی گردیده و به یک مساله‌ی برنامه‌ریزی غیرخطی تبدیل می‌شود. در آخر، برای نشان دادن کارایی روش، سه مثال مورد آزمایش قرار می‌گیرد.

**لغات کلیدی**: کنترل بهینه، سیستم‌های سوییچ، روش شبه‌طیفی لژاندر.

# Computing Szeged Index of Graphs on Triples

**M. R. DARAFSHEH [a], R. MODABERNIA[b]AND M. NAMDARI[b]**

[a] School of Mathematics, College of Science, University of Tehran, Tehran, Iran
[b]Department of Mathematics, ShahidChamran University of Ahvaz, Ahvaz, Iran

# محاسبه اندیس سگد گرافها روی سه‌گانه‌ها

ادیتور رابط : علیرضا اشرفی

## چکیده

فرض کنیم G=(V,E) یک گراف همبند ساده با مجموعه‌ی اندیس V و مجموعه‌ی یال E باشد. اندیس سگد G بصورت $Sz(G) = \sum_{e=uv\in E} n_u(e|G)n_v(e|G)$ تعریف می‌شود که $n_u(e|G)$ تعداد راس‌های G نزدیکتر به u نسبت به v و نزدیکتر به v نسبت به u است و $n_v(e|G)$ به‌صورت مشابه تعریف می‌شود. گیریم S یک مجموعه با اندازه $n \geq 8$ و V مجموعه‌ی همه‌ی زیرمجموعه‌های S با اندازه‌ی 3 باشد. سه نوع از گراف‌های اشتراک با مجموعه‌ی اندیس V را تعریف می‌کنیم. این گراف‌ها را با $G_i(n),\ i = 0,1,2$ نمایش داده و اندیس‌های سگد آن‌ها را بدست می‌آوریم.

**لغات کلیدی**: اندیس سگد، گراف اشتراک، خودریختی گراف.

# *Nordhaus-Gaddum Type Results for the Harary Index Of Graphs*

**ZHAO WANG[1], YAPING MAO[2], XIA WANG[2] AND CHUNXIA WANG[2]**

[1]School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China
[2]Department of Mathematics, Qinghai Normal University, Xining, Qinghai 810008, China

# نتایج نوع نوردهاوس-گادم برای شاخص هراری گراف‌ها

ادیتور رابط : تومیسلاو داسلیک

## چکیده

شاخص هراری $H(G)$ از یک گراف همبند $G$ به صورت $H(G)=\sum_{u,v\in V(G)}\frac{1}{d_G(u,v)}$ تعریف می‌شود که در آن $d_G(u,v)$ ، فاصله‌ی بین رئوس $u,v$ از $G$ است. فاصله‌ی استینر در گراف، که توسط چارترند و همکاران در سال *1989* معرفی شد، بعنوان تعمیم طبیعی مفهوم فاصله‌ی گراف کلاسیک معرفی شده است. برای یک گراف همبند G از مرتبه حداقل $n\geq 2$ و $S\dot{\subseteq} V(G)$، *فاصله‌ی استینر* $d_G(S)$ برای رئوس $S$، کمترین اندازه‌ی یک زیرگراف همبند است که مجموعه ی رئوس آن، شامل $S$ است. اخیرا، فورتولا، گوتمن و کاتانیک مفهوم شاخص هراری استینر را معرفی کرده‌اند و کاربردهای شیمی آن را بیان نموده‌اند. شاخص هراری استینر $k$-مرکز $SH_k(G)$ از G، به صورت $SH_k(G)=\sum_{s\subseteq V(G),|s|=k}\frac{1}{d_G(s)}$ تعریف شده است. دراین مقاله، ما کران های بالایی و پایینی تیز را برای $SH_k(G)+SH_k(\overline{G})$ و $SH_k(G).SH_k(\overline{G})$ به دست آوردیم که برای هر گراف همبند G که مکمل آن $\overline{G}$ نیز همبند است، قابل قبول است.

**لغات کلیدی**: فاصله، فاصله‌ی استینر، شاخص هراری، شاخص هراری استینر $k$- مرکز

# Determination of Critical Properties of Alkanes Derivatives using Multiple Linear Regression

**ESMAT MOHAMMADINASAB**

Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran

# تعیین خواص بحرانی مشتقات آلکان‌ها با استفاده از رگرسیون خطی چندگانه

ادیتور رابط : علیرضا اشرفی

## چکیده

در این پژوهش، برخی روش‌های محاسباتی ریاضی برای تخمین دما و فشار و حجم بحرانی 40 نوع مختلف از آلکانها و مشتقاتشان ارائه شده است. در این مدل، رابطه کمی ساختار-خاصیت بر اساس نظریه گراف با برخی شاخصهای ساختاری و هندسی ترکیبات شیمیایی به کار گرفته شده است. برای بررسی همبستگی میان خواص بحرانی و توصیف‌گرهای مولکولی، مدل رگرسیون خطی چندگانه با کمک ضرایب مناسب بکار گرفته شده است. برای این منظور، از بین یازده توصیف‌گر ساختاری و هندسی مورد مطالعه، مناسب‌ترین آنها برای تعیین خواص بحرانی آلکانها انتخاب شدند. نتایج نشان دادند که از بین مدلهای پیشنهادی برای پیش بینی دما و فشار و حجم بحرانی آلکانهای مورد مطالعه، مدل شامل ترکیبی از توصیف گرهای hyper-Wiener، Platt، MinZL مناسب‌ترین مدل می باشد.

**لغات کلیدی:** آلکانها، MLR، خواص بحرانی، QSPR

# *Some Relation between Kekule Structure and Morgan-Voyce Polynomials*

**INCI GULTEKIN[1] AND BUNYAMIN SAHIN[2]**

[1]Department of Mathematics, Faculty of Science, Ataturk University, 25240 Erzurum, Turkey
[2]Department of Elemantary Mathematics Education, Faculty of Education, Bayburt University, 69000 Bayburt, Turkey

# برخی از روابط میان ساختار ککول و چندجملهایهای مورگان-وویس

ادیتور رابط: میشل ماری دزا

## چکیده

در این مقاله، ساختارهای ککول از زنجیرهای بنزوئیدی مورد توجه قرار گرفته‌اند. نشان داده شده است که ضرایب چندجمله‌ای مورگان-وویس $B_n(x)$، با تعداد $k$--جفتهای $(m(G,K))$ از یک گراف مسیر که $N=2n+1$ رأس دارد، برابر است. بعلاوه، دو رابطه بین زنجیرهای به‌طور منظم بنزوئیدی کاملا متراکم غیر شاخه‌ای مورب و چندجمله‌ایهای مورگان-وویس، و زنجیرهای به‌طور منظم بنزوئیدی کاملا متراکم غیرشاخه ای مورب و درختان کاترپیلار متناظر به دست آمده است.

**لغات کلیدی**: ساختار ککول، شاخص هوسویا، چندجمله‌ای‌های مورگان-وویس، درختان کاترپیلار

این نشریه طبق مجوز شماره 89/3/11/104372 مورخه 89/11/27 دارای اعتبار **علمی -پژوهشی** از وزارت علوم ، تحقیقات و فناوری می باشد. همچنین این مجله در پایگاه اطلاعاتی ISC (Islamic World Science Citation Center) وابسته به وزارت علوم ، تحقیقات و فناوری نمایه می شود.

# IRANIAN JOURNAL OF MATHEMATICAL CHEMISTRY

## CONTENTS

June 2017

ISSN
2228-6489 ( Print Version )
2008-9015 ( Online Version )