# DISCRETE AND CONTINUOUS MODELS AND APPLIED COMPUTATIONAL SCIENCE

**Volume 32 Number 4 (2024)**

**Founded in 1993**

**Founder: PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA NAMED AFTER PATRICE LUMUMBA**

**DOI: 10.22363/2658-4670-2024-32-4**

Edition registered by the Federal Service for Supervision of Communications, Information Technology and Mass Media

**Registration Certificate:** ПИ № ФС 77-76317, 19.07.2019

ISSN 2658-7149 (Online); 2658-4670 (Print)
4 issues per year.
Language: English.

## Publisher

Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University).

## Indexed by

- Scopus (https://www.scopus.com),
- Ulrich's Periodicals Directory (http://www.ulrichsweb.com),
- Directory of Open Access Journals (DOAJ) (https://doaj.org),
- Russian Index of Science Citation (https://elibrary.ru),
- CyberLeninka (https://cyberleninka.ru).

## Aim and Scope

Discrete and Continuous Models and Applied Computational Science arose in 2019 as a continuation of RUDN Journal of Mathematics, Information Sciences and Physics. RUDN Journal of Mathematics, Information Sciences and Physics arose in 2006 as a merger and continuation of the series "Physics", "Mathematics", "Applied Mathematics and Computer Science", "Applied Mathematics and Computer Mathematics".

Discussed issues affecting modern problems of physics, mathematics, queuing theory, the Teletraffic theory, computer science, software and databases development.

It's an international journal regarding both the editorial board and contributing authors as well as research and topics of publications. Its authors are leading researchers possessing PhD and PhDr degrees, and PhD and MA students from Russia and abroad. Articles are indexed in the Russian and foreign databases. Each paper is reviewed by at least two reviewers, the composition of which includes PhDs, are well known in their circles. Author's part of the magazine includes both young scientists, graduate students and talented students, who publish their works, and famous giants of world science.

The Journal is published in accordance with the policies of COPE (Committee on Publication Ethics). The editors are open to thematic issue initiatives with guest editors. Further information regarding notes for contributors, subscription, and back volumes is available at http://journals.rudn.ru/miph

E-mail: miphj@rudn.ru, dcm@sci.pfu.edu.ru

# Editorial board



---


**Computer Design:** *Anna V. Korolkova*, *Dmitry S. Kulyabov*

**English Text Editors:** *Nikolay E. Nikolaev*, *Ivan S. Zaryadov*, *Konstantin P. Lovetskiy*

**Address of editorial board:**
3 Ordzhonikidze St, 115419 Moscow, Russia
+7 (495) 955-07-16, e-mail: publishing@rudn.ru

**Editorial office:**
+7 (495) 952-02-50, miphj@rudn.ru, dcm@rudn.su,
site: http://journals.rudn.ru/miph

---

Paper size 70×108/16. Offset paper. Offset printing. Typeface “Adobe Source”.
Conventional printed sheet 8.92. Printing run 500 copies. Open price. The order 1714.
PEOPLES’ FRIENDSHIP UNIVERSITY OF RUSSIA NAMED AFTER PATRICE LUMUMBA
6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

**Printed at RUDN Publishing House:**
3 Ordzhonikidze St, Moscow, 115419, Russian Federation,
+7 (495) 955-08-61; e-mail: publishing@rudn.ru

**Discrete & Continuous Models**
**& Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4)**
http://journals.rudn.ru/miph

# Contents

## Editorial

*Kulyabov, D. S., Sevastianov, L. A.* IMRAD structure . . . . . 355

## Computer science

*Nikolaev, D. I., Beschastnyi, V. A., Gaidamaka, Y. V.* Two-queue polling system as a model of an integrated access and backhaul network node in half-duplex mode . . . . . 362

*Kiselev, G. A., Lubysheva, Y. M., Weizenfeld, D. A.* MMEmAsis: multimodal emotion and sentiment analysis . . . . . 370

*Voronina, N. M., Rozhkova, S. V.* Asymptotic diffusion analysis of RQ system M/M/1 with unreliable server . . . . . 380

## Modeling and Simulation

*Kroytor, O. K., Malykh, M. D., Sevastianov, L. A.* On the problem of normal modes of a waveguide 395

*Malykh, M. D., Malyshev, K. Yu.* On summation of Fourier series in finite form . . . . . 406

*Lovetskiy, K. P., Malykh, M. D., Sevastianov, L. A., Sergeev, S. V.* Solving a two-point second-order LODE problem by constructing a complete system of solutions using a modified Chebyshev collocation method . . . . . 414

*Zhanlav, T., Otgondorj, K.* Development and adaptation of higher-order iterative methods in $R^n$ with specific rules . . . . . 425

## Physics and Astronomy

*Rybakov, Y. P., Umar, M.* Superconductivity and special symmetry of twisted tri-layer graphene in chiral model . . . . . 445

**Discrete & Continuous Models**
**& Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 355–361**
http://journals.rudn.ru/miph

*Editorial*

DOI: 10.22363/2658-4670-2024-32-4-355-361

EDN: EKXUGR

# IMRAD structure

Dmitry S. Kulyabov[1, 2], Leonid A. Sevastianov[1, 2]

[1] RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

[2] Joint Institute for Nuclear Research, 6 Joliot-Curie St, Dubna, 141980, Russian Federation

**Abstract.** We describe introduced in the journal the rubric system.We describe the general structure of an IMRAD research publication. The IMRAD structure for a research article is described in detail.

**Key words and phrases:** IMRAD, research article

**For citation:** Kulyabov, D. S., Sevastianov, L. A. IMRAD structure. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 355–361. doi: 10.22363/2658-4670-2024-32-4-355-361. edn: EKXUGR (2024).

## 1. General structure of the IMRAD paper

IMRAD is an abbreviation of *introduction, materials and methods, results and discussion*. The structure of IMRAD is described in ANSI Z39.16-1972 (Preparation of Scientific Papers for Written or Oral Presentation). In the 1970s, IMRAD became the de facto standard for the design of scientific articles [1].

The structure of an article according to IMRAD should look as follows.

- *Introduction*.
  - Why the study was done.
  - What was researched.
  - Purpose of the study.
  - What hypotheses were tested.
- *Methods* (aka *Materials and Methods*, *Theoretical Framework*).
  - When, where, and how the research was conducted.
  - What materials were used.
- *Results*.
  - What answer was found.
  - Whether the hypothesis was tested correctly.
- *Discussion*.
  - What the answer implies and why it matters.
  - How it fits in with what other researchers have found.
  - What are the prospects for research.
- Also included in the article structure are *Title*, *Annotation*, *Keywords*.

## 2. Different IMRAD structure options

The IMRAD structure is not a dogma, but a guide to action. Therefore, for different types of articles and for different scientific fields, some modifications may be made to this structure.

- The *Materials and Methods* section may be replaced by the *Theory* section.
- The *Results* and *Discussion* sections can be combined into one section.
- The *Conclusions* section can be included as the last part of the *Discussion* section.
- Only the main aspects can be given in the article, all additional aspects are listed as *Supplemental Materials* (on the journal's website).
- Review articles do not have a *Results and Discussion* section.
- Grant information may be included in the *Funding* section.

## 3. Structure of a research paper based on IMRAD

### 3.1. Title

- The title of the article should describe the content of the article as accurately as possible.
- The title of the article should be understandable for both humans and tools for processing and analyzing scientific information.
- The title of the article—this is what is looked at first.

Guidelines for writing the title of the article:

- the title should contain as few words as possible ($7 \pm 2$ words) [2];
- is assumed that in our journal, the title of the article should not exceed 12 words;
- the title should accurately and specifically describe the content of the article;
- the title should not contain abbreviations, formulas and jargonisms;
- the title should not use abbreviations such as "Some notes on", "Observations on", "Investigations on", "Study of", "Effect of", etc;
- the title should not be flashy, as in newspapers;
- the title should communicate the subject of the study, not the results.

### 3.2. Authors

- This section is organized according to the rules of the author list.
- These rules may vary from one field of science to another.

### 3.3. Keywords

- Keywords should not duplicate terms specified in the title of the article.
- Keywords are used to analyze articles by automatic tools of abstract databases.

### 3.4. Abstract

- The main purpose of this section is to give the reader a comprehensive idea of what the article is about, so that he or she can decide whether or not to read it in its entirety.
- The abstract summarizes the main quantitative results of the study and the conclusions drawn from the work.

## 3.5. Introduction

- The introduction does not repeat the abstract; its purpose is to introduce the topic of the publication.
- The introduction aims to immerse the reader in the context of the research.
- The introduction provides a brief overview of the literature most relevant to the topic.
- The introduction answers the questions *what we are researching* and *why we are researching*.

In the introduction, the following objectives need to be addressed:

- to justify the relevance of the research (make a historical excursion, highlight the most significant works in the field);
- to reflect the most relevant recent achievements in the chosen field and the current state of the field;
- to introduce the necessary abbreviations and definitions;
- to formulate the prerequisites for the formation of the hypothesis to test which the article is devoted to;
- to define the specific problem, the solution to which the study is devoted;
- to formulate the goals and objectives of the work;
- to put the questions that this paper answers.

It is recommended to place service sections within the introduction: *Structure of the paper* and *Notations and conventions*.

### 3.5.1. Structure of the paper

In this section, the authors briefly summarize the structure of the article.

### 3.5.2. Notations and conventions

In this block the authors explain special terms, reveal abbreviations.

## 3.6. Materials and Methods

The section may also be called *Theoretical Basis*. The main objective of this section is to enable other scientists to reproduce the work done in the paper. It specifies the preparation mechanism, the equipment used, the algorithm for constructing the experimental process in chronological order, the laboratory procedures, the reagents and materials used, the laboratory objects, the methodology for processing the experimental results, and the research software used to write the paper.

Guidelines for preparing the section:

- avoid ambiguity in abbreviations or names;
- write all quantitative characteristics in standard international units of measurement (the choice of measurement system can be justified in the Introduction);
- explain each step of the study;
- explain all methods used;
- avoid irrelevant and unnecessary information that is not relevant to the results of the paper.

## 3.7. Results

In the section, authors presents the main results of the research findings.

- If data is obtained in the paper, the section presents a report on this data.
- It is recommended to use data visualization in the form of figures, tables, and charts.

Guidelines for preparing the section:

- try to present results clearly and concisely;
- not to give large amounts of data;
- reduce the data to statistically analyzable summary forms and present them in tables or graphs along with the necessary statistical information;
- do not repeat in the text the data presented in tables and figures;
- include only tables and figures that are necessary, understandable and should be reproduced;
- do not display the same data in tables and graphs at the same time.

### 3.8. Discussion

- The section summarizes the relationship between the findings and the answer to the questions posed in the Introduction.
- The section summarizes and shows the relationship between the results and conclusions.
- In the section, the authors justify their hypotheses.
- The section may include a theoretical justification of the findings.
- If there are deviations in the process of experimentation or doubts of the authors of the article, they are indicated here.
- The section indicates the consistency or discrepancy between the results of research published on the subject previously.
- In the section, you can emphasize the merits of your research and what could be improved in further research on the topic.
- This section demonstrates the significance of the work presented and its expected impact on the development of future research trajectories (can be included in the Conclusion section).
- Unfortunately, inconsistency between stated aims and discussion is a common problem in many manuscripts.

Guidelines for preparing the section:

- do not repeat what has already been said in the literature review;
- relate the results to the issues that were outlined in the Introduction;
- show whether the results and interpretations are consistent with current knowledge of the subject, i.e., previously published work;
- explain the theoretical background of the observed results;
- the significance of the results;
- suggest directions for future research;
- discuss only those results that were presented in the study;
- do not make generalizations or assumptions that are not justified by the findings;
- formulate conclusions with evidence for each conclusion.

### 3.9. Conclusion

- Summary of the analysis of the results of the study and perspectives.
- Should summarize the outcome of the study.
- The answers to each of the questions presented in the introduction should be noted.

### 3.10. Ethical clauses

The classic structure of this section usually included only an Acknowledgements section. This section thanked individuals and organizations that had supported the research in various ways. In addition, funding information was also provided in this section. Private or corporate investors or grant funds, laboratories that were used to conduct certain blocks of work, collaborators indirectly involved in this block of work and therefore not reflected in the section *Authors*, and technical specialists who assisted in conducting the measurements were mentioned. This section also reflected the genesis of the work, e.g., that the work was done as part of the dissertation.

Influenced by the COPE ethics committee [3], a more complex structure is now used for this section, and it is divided into several thematic subsections. For more information, see the related article [4].

### 3.11. References

- The formatting of the reference list should be done according to the style (in our journal we are trying to solve this issue with the help of automatic generation of bibliography).
- In the reference list, only those sources are listed that are referenced in the text of the article.
- For a scientific article usually use a minimum of 20 sources.
- If you cite information but do not provide a reference, it can be considered plagiarism.
- You can also link to sites with the date of visit (but do not abuse it).

### 3.12. Appendices

The section may also be called *Supplementary information*.

- The section provides details of the elements of the study for those who later wish to replicate the work.
- The section may include details of data processing, design of the experiment, and characterization of the equipment, objects, and materials used.
- The section provides any additional information that is relevant to the paper but of secondary importance.
- Appendices usually contain information that is necessary to fully explain and understand the results, but are too bulky and complex to be included in the main body of the paper.

## 4. Conclusion

The editors expect authors to adhere to the recommended structure of the article.

**Author Contributions:** The contributions of the authors are equal. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** No new data were created or analysed during this study. Data sharing is not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wu, J. Improving the writing of research papers: IMRAD and beyond. *Landscape Ecology* **26,** 1345–1349. doi:10.1007/s10980-011-9674-3 (Nov. 2011).

2. Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* **63,** 81–97. doi:10.1037/h0043158. eprint: 13310704 (Mar. 1956).
3. *COPE Council. COPE Discussion Document: Authorship* 2nd ed. doi:10.24318/cope.2019.3.3 (Sept. 2019).
4. Kulyabov, D. S. & Sevastianov, L. A. Author's ethics. *Discrete and Continuous Models and Applied Computational Science* **32,** 135–139. doi:10.22363/2658-4670-2024-32-2-135–139 (2024).

## Information about the authors

**Dmitry S. Kulyabov**—Professor, Doctor of Sciences in Physics and Mathematics, Professor of Department of Probability Theory and Cyber Security of RUDN University; Senior Researcher of Laboratory of Information Technologies, Joint Institute for Nuclear Research (e-mail: kulyabov-ds@rudn.ru, phone: +7 (495) 952-02-50, ORCID: 0000-0002-0877-7063, ResearcherID: I-3183-2013, Scopus Author ID: 35194130800)

**Leonid A. Sevastianov**—Professor, Doctor of Sciences in Physics and Mathematics, Professor of Department of Computational Mathematics and Artificial Intelligence of RUDN University (e-mail: sevastianov-la@rudn.ru, phone: +7 (495) 955-07-83, ORCID: 0000-0002-1856-4643, ResearcherID: B-8497-2016, Scopus Author ID: 8783969400)

DOI: 10.22363/2658-4670-2024-32-4-355-361 EDN: EKXUGR

# Структура IMRAD

Д. С. Кулябов[1,2], Л. А. Севастьянов[1,2]

[1] Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

[2] Объединённый институт ядерных исследований, ул. Жолио-Кюри, д. 6, Дубна, 141980, Российская Федерация

**Аннотация.** Описывается общая структура научной публикации IMRAD. Подробно описывается структура IMRAD для исследовательской статьи.

**Ключевые слова:** IMRAD, исследовательская статья

**Discrete & Continuous Models**
**& Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 362–369**
http://journals.rudn.ru/miph

*Research article*

UDC 519.872:519.217
PACS 02.50.Ey, 02.50.Fz
DOI: 10.22363/2658-4670-2024-32-4-362-369 EDN: DRHDFU

# Two-queue polling system as a model of an integrated access and backhaul network node in half-duplex mode

Dmitry I. Nikolaev[1], Vitalii A. Beschastnyi[1], Yuliya V. Gaidamaka[1,2]

[1] RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

[2] Institute of Informatics Problems, 44 Vavilova St, bldg 2, Moscow 119333, Russian Federation

(received: September 3, 2024; revised: September 25, 2024; accepted: September 28, 2024)

**Abstract.** Integrated Access and Backhaul (IAB) technology facilitates the establishment of a compact network by utilizing repeater nodes rather than fully equipped base stations, which subsequently minimizes the expenses associated with the transition towards next-generation networks. The majority of studies focusing on IAB networks rely on simulation tools and the creation of discrete-time models. This paper introduces a mathematical model for the boundary node in an IAB network functioning in half-duplex mode. The proposed model is structured as a polling service system with a dual-queue setup, represented as a random process in continuous time, and is examined through the lens of queueing theory, integral transforms, and generating functions (GF). As a result, analytical expressions were obtained for the GF, marginal distribution, as well as the mean and variance of the number of requests in the queues, which correspond to packets pending transmission by the relay node via access and backhaul channels.

**Key words and phrases:** polling, queuing system, integrated access and backhaul, half-duplex

**For citation:** Nikolaev, D. I., Beschastnyi, V. A., Gaidamaka, Y. V. Two-queue polling system as a model of an integrated access and backhaul network node in half-duplex mode. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 362–369. doi: 10.22363/2658-4670-2024-32-4-362-369. edn: DRHDFU (2024).

## 1. Introduction

To simplify and reduce the cost of deploying dense 5G networks, standardizing organizations have proposed various technologies, one of which is Integrated Access and Backhaul (IAB) [1]. This technology enables telecom operators to seamlessly transition to 5G-compliant networks by utilizing cost-effective relay nodes that implement wireless relay instead of fully-equipped base stations. By implementing a network with IAB technology, consisting of backbone and relay nodes, operators can meet the limitations of 5G standards and have the flexibility to upgrade relay nodes with access to the backbone network in the future, ultimately enhancing the quality of service for users.

Integrated Access and Backhaul is one of the approved objectives of the 17th Release of the 3GPP (3rd Generation Partnership Project) [2]. In IAB, a small number of backbone base stations (BSs) are connected to the existing fibre-optic network infrastructure. The remaining BSs transmit backhaul traffic over wireless channels [3].

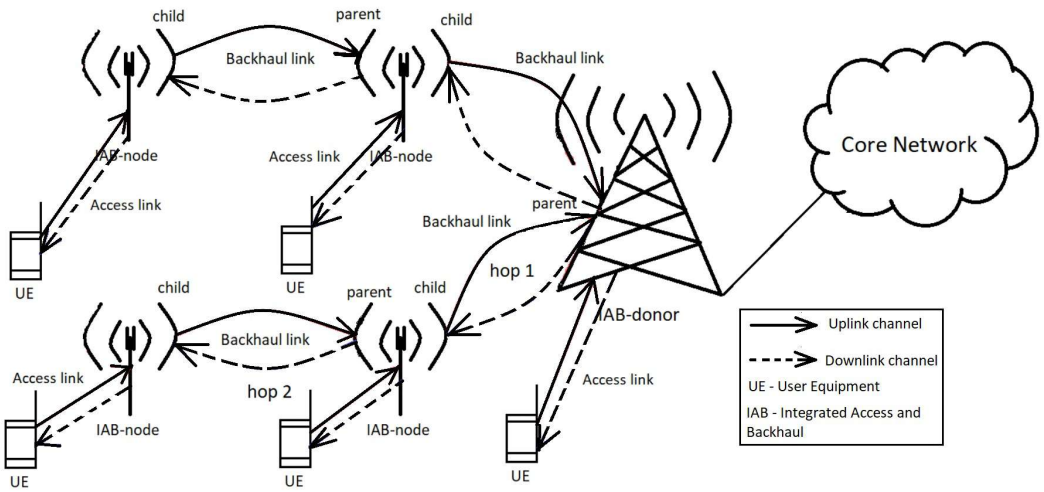


Figure 1. IAB network fragment in the form of a spanning tree

In comparison to LTE-Advanced, IAB is a more advanced solution that supports multi-hop, dynamic resource multiplexing, and plug-and-play design, significantly reducing the complexity of network deployment. Given the aforementioned benefits of IAB technology, the design of an efficient and high-performance 5G/6G network incorporating this technology has become a pressing research topic. As such, further exploration and utilization of IAB within the context of 5G/6G networks holds immense potential for enhancing network capabilities and improving overall user experience [4].

The IAB technology, along with its characteristics and operational mechanisms, has been investigated from multiple perspectives. Research efforts have addressed challenges such as routing in multi-hop networks [5, 6], the selection of optimal network topology [7], and efficient resource allocation [8]. In addition, advanced beamforming techniques have been explored [9], while the development of data channel management policies for latency control [10] and the establishment of network stability conditions that maximize throughput have also been studied [11, 12]. Furthermore, frequency reuse using graph coloring methods has been investigated [13]. Moreover, one study constructs a mathematical model of the IAB edge node as a Markov process and analyzes packet transmission delays [14], while another work develops a simulation model of the IAB edge node [15]. Complementary research includes the construction of mathematical models for IAB networks incorporating blockage effects [16], mean and the formulation of queuing system models to represent the number of users at an IAB node [17].

Figure 1 shows an example of the IAB network topology in the form of a spanning tree, with the IAB-donor reference base station (BS) located at the root vertex. The remaining IAB nodes in the network are branch vertices and leaf vertices. The focus of this study is on the IAB boundary node, which corresponds to the leaf vertex in the tree. The subject of the study is the packet flow that passes through this node.

Due to the separation of downlink and uplink channels, a mathematical model has been proposed in the form of a polling service system [18–23]. Data packets will be associated with requests, and the IAB boundary node will correspond to a server. Downlink traffic from the parent node to the current node and from the current node to user equipments (UEs) will be directed to queue $Q_1$ for receiving and servicing requests. Uplink traffic from UEs to the current node and from the current node to the parent node will correspond for receiving and servicing requests in queue $Q_2$ (see Table 1).

Table 1

The correspondence between the objects and processes of the technical system and the queueing system

| Technical system | IAB-node | Data packets | Downlink from the parent node | Uplink to the current node | Downlink from the current node | Uplink to the parent node |
|---|---|---|---|---|---|---|
| Queueing system | Server | Requests | Receipt of requests in the queue $Q_1$ | Receipt of requests in the queue $Q_2$ | Servicing of requests in the queue $Q_1$ | Servicing of requests in the queue $Q_2$ |

Considering the limitations imposed by the half-duplex data transmission mode, we can divide the operation of the IAB boundary node into phases shown in Table 2. For simplicity, let's combine the first two phases of request receipt into one. We then arrive at a polling service system with two queues, where requests are received during the switching of devices at the end of each service cycle. That is, the switching time between queues within the service cycle is zero, and applications are received exclusively during the above-described period. The characteristics of this model will be studied in the next section.

Table 2

Phases of operation of the IAB network boundary node

| | $Q_1$ | $Q_2$ |
|---|---|---|
| Downlink from the parent node | + | 0 |
| Uplink to the current node | 0 | + |
| Downlink from the current node | − | 0 |
| Uplink to the parent node | 0 | − |

## 2. Mathematical model

We will now delve deeper into the details of the $M_2|GI_2|1$ polling service system that was introduced in the preceding section. In order to study the system, we must make the assumption that it is operating in a stationary mode. Within this context, we will use $X_i^j$ to denote the number of requests in the queue $Q_j$ at any given time $Q_i$, $i,\ j = 1,\ 2$ [24].

Additionally, we make use of the notation $A_i(t)$ to represent the number of requests received in the $i$-th queue during time $t$. It is important to note that our system consists of 2 Poisson input flows, each with their own parameter, denoted by $\lambda_i$. Furthermore, the service time for a given request in queue $Q_i$ is denoted as $b_{ik}$, with $k$ representing the $k$-th request. It is also stated that these service times are independent and equally distributed with a cumulative distribution function (CDF) of $B_i(t)$.

Moving on to the half-duplex aspect of the system, we introduce the random variable $s_0$, which represents the switching time of the server. Its distribution is given by the CDF $S(t)$, and it has raw moments of arbitrary order $s_0^{(n)} = \int_0^\infty t^n \,\mathrm{d}(S(t))$, with $n \geq 1$. Finally, we arrive at the expressions for

$X_i^j$ for our system, which can be written as follows:

$$X_i^j = \begin{cases} 0, & j < i, \\ A_j(s_0), & j \geq i, \end{cases} \quad \Leftrightarrow \quad X_i^j = \begin{cases} 0, & (j=1) \wedge (i=2), \\ A_1(s_0), & (j=1) \wedge (i=1), \\ A_2(s_0), & j=2. \end{cases} \tag{1}$$

For these values, we have $p_i(n_1, n_2)$ — the probability distribution that at any moment of servicing the $i$-th queue $Q_i$ $j$-th queue $Q_j$ contains $n_j$ applications, $n_j \geq 0$, $i, j = 1, 2$.

The generating functions (GFs) of random variables $(X_i^1, X_i^2, \dots, X_i^K)$, $i = 1, \dots, K$ are expressed according to the following lemma.

**Lemma 1.** *GF of random variables* $(X_i^1, X_i^2)$, $i = 1, 2$ *have the following form*

$$P_i(\mathbf{z}) = P_i(z_1, z_2) = \tilde{S}\left(\sum_{j=i}^{2}(\lambda_j(1 - z_j))\right), \quad i = 1, 2, \tag{2}$$

*where* $\tilde{S}(w)$*—Laplace-Stieltjes Transform (LST) of RV* $s_0 \sim S(t)$.

Substituting the value of **1** into the variable **z** in the derivatives of (2), we obtain the values of the mean and variance of the number of requests in the queues.

**Theorem 1.** *For a polling system* $M_2|GI_2|1$ *with state-dependent input flows and switching time* $s_0$ *distributed according to the CDF* $S(t)$*, the mean* $\overline{N}_i(j)$ *and the variance* $\mathrm{Var}\left(X_i^j\right)$ *of the number of requests in queue* $Q_j$ *at the time of servicing queue* $Q_i$*,* $i, j = 1, 2$*, are expressed by the following formulas:*

$$\overline{N}_i(j) = \begin{cases} 0, & j < i, \\ \overline{s_0}\lambda_j, & j \geq i, \end{cases} \quad \mathrm{Var}\left(X_i^j\right) = \begin{cases} 0, & j < i, \\ s_0^{(2)}\lambda_j^2 + \overline{s_0}\lambda_j(1 - \overline{s_0}\lambda_j), & j \geq i, \end{cases} \tag{3}$$

*If the switching time is exponentially distributed with parameter* $s$ $(S(t) = 1 - e^{-st}$*,* $t \geq 0)$*, then formulas*(3) *are transformed to the form* (4).

$$\overline{N}_i(j) = \begin{cases} 0, & j < i, \\ \dfrac{\lambda_j}{s}, & j \geq i, \end{cases} \quad \mathrm{Var}\left(X_i^j\right) = \begin{cases} 0, & j < i, \\ \dfrac{\lambda_j^2}{s^2} + \dfrac{\lambda_j}{s}, & j \geq i, \end{cases} \tag{4}$$

*where* $\mathrm{Var}(\cdot)$ *— Variance of RV.*

Substituting the value of **0** into the variable **z** in the derivatives of (2), we obtain the probability distribution of number of requests in queues.

**Theorem 2.** *For a polling system* $M_2|GI_2|1$ *with state-dependent input flows and switching time* $s_0$ *distributed according to the exponential law* $S(t) = 1 - e^{-st}$*,* $t \geq 0$*, the distributions of the number of requests in queue* $Q_j$ *at the moment of servicing queue* $Q_i$*,* $i, j = 1, 2$*, are expressed by the following formulas:*

$$p_i(n_1, n_2) = \begin{cases} 0, & (n_1 \geq 1) \wedge (i = 2), \\ \dfrac{s\lambda_2^{n_2}}{(s + \lambda_2)^{n_2+1}}, & (n_1 = 0) \wedge (i = 2), \\ \dfrac{s\lambda_1^{n_1}\lambda_2^{n_2}(n_1 + n_2)!}{(n_1)!(n_2)!(s + \lambda_1 + \lambda_2)^{n_1+n_2+1}}, & i = 1, \end{cases} \tag{5}$$

*where* $n_k = 0, \dots, \infty$ *— the number of requests in queue* $Q_k$*,* $k = 1, 2$.

## 3. Conclusion

In the transition to the next generation of networks, integrated access and backhaul (IAB) technology is considered a key technology. However, due to limitations imposed by the half-duplex mode of data transmission, it is necessary to build adequate models of how IAB networks operate.

In this paper, a model of the boundary node of an IAB network in the form of a queueing polling system was constructed. We also derived analytical expressions for the generating functions, marginal distribution, raw, and central moments of the number of requests (packets) in queues.

These results allow us to estimate the probability and conditions of overloads at the IAB boundary node. For future research, we plan to build and analyze an energy-efficient model of the entire IAB network. This will be done for both the topology presented in this paper and for more complex topologies where there are multiple routes from the reference base station to user devices.

**Author Contributions:** Conceptualization, Dmitry I. Nikolaev and Yuliya V. Gaidamaka; methodology, Yuliya V. Gaidamaka; validation, Dmitry I. Nikolaev, Vitalii A. Beschastnyi, Yuliya V. Gaidamaka ; formal analysis, Dmitry I. Nikolaev; investigation, Dmitry I. Nikolaev; resources, Vitalii A. Beschastnyi and Yuliya V. Gaidamaka; writing—original draft preparation, Dmitry I. Nikolaev; writing—review and editing, Dmitry I. Nikolaev, Vitalii A. Beschastnyi, Yuliya V. Gaidamaka; supervision, Yuliya V. Gaidamaka; project administration, Vitalii A. Beschastnyi and Yuliya V. Gaidamaka; funding acquisition, Vitalii A. Beschastnyi All authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The reported study was funded by RSF, project number 23-79-01140, https://rscf.ru/en/project/23-79-01140/.

**Data Availability Statement:** Data sharing is not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. *3GPP: Study on Integrated Access and Backhaul. Technical report (TR) 38.874 v16.0.0 (2018)*
2. *3GPP: Integrated Access and Backhaul (IAB) radio transmission and reception. Technical Specification (TS) 38.174 v17.2.0 (2022)*
3. Polese, M., Giordani, M., Zugno, T., Roy, A., Goyal, S., Castor, D. & Zorzi, M. Integrated Access and Backhaul in 5G mmWave Networks: Potential and Challenges. *IEEE Communications Magazine* **58,** 62–68. doi:10.1109/MCOM.001.1900346 (Mar. 2020).
4. Sadovaya, Y., Molchanov, D., Mao, W., Orhan, O., Yeh, S.-p., Nikopour, H., Talwur, S. & Andreev, S. Integrated access and backhaul in millimeter-wave cellular: Benefits and challenges. *IEEE Communications Magazine* **60,** 81–86 (2022).
5. Gomez-Cuba, F. & Zorzi, M. Optimal link scheduling in millimeter wave multi-hop networks with MU-MIMO radios. *IEEE Transactions on Wireless Communications* **19,** 1839–1854 (2020).
6. Alghafari, H. & Sayad Haghighi, M. Decentralized Joint Resource Allocation and Path Selection in Multi-hop Integrated Access Backhaul 5G Networks. *Computer Networks,* 108837. doi:10.1016/j.comnet.2022.108837 (Feb. 2022).
7. Madapatha, C. *et al.* On topology optimization and routing in integrated access and backhaul networks: A genetic algorithm based approach. *IEEE Open Journal of the Communications Society* **2,** 2273–2291 (2021).

8. Tafintsev, N., Moltchanov, D., Yeh, S.-p., Nikopour, H., Mao, W., Orhan, O., Talwar, S., Valkama, M. & Andreev, S. *Joint Path Selection and Resource Allocation in Multi-Hop mmWave-based IAB Systems* in *ICC 2023 - IEEE International Conference on Communications* (2023), 4194–4199. doi:10.1109/ICC45041.2023.10279180.
9. Jayasinghe, P., Tölli, A., Kaleva, J. & Latva-Aho, M. Traffic Aware Beamformer Design for Flexible TDD-Based Integrated Access and Backhaul. *IEEE Access* **8,** 205534–205549. doi:10.1109/ACCESS.2020.3037814 (2020).
10. Yarkina, N., Moltchanov, D. & Koucheryavy, Y. Counter Waves Link Activation Policy for Latency Control in In-Band IAB Systems. *IEEE Communications Letters* **27,** 3108–3112. doi:10.1109/LCOMM.2023.3313233 (2023).
11. Neely, M. *Stochastic Network Optimization with Application to Communication and Queueing Systems* doi:10.2200/S00271ED1V01Y201006CNT007 (2010).
12. Tassiulas, L. & Ephremides, A. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control* **37,** 1936–1948. doi:10.1109/9.182479 (1992).
13. Silard, M., Fabian, P., Papadopoulos, G. Z. & Savelli, P. *Frequency Reuse in IAB-based 5G Networks using Graph Coloring Methods* in *2022 Global Information Infrastructure and Networking Symposium (GIIS)* (Argostoli, Greece, 2022), 104–110. doi:10.1109/GIIS56506.2022.9937005.
14. Nikolaev, D. & Gaidamaka, Y. *Leaf Node Polling Model Analysis in an Integrated Access and Backhaul Network* in *Information Technologies and Mathematical Modelling. Queueing Theory and Applications* (eds Dudin, A., Nazarov, A. & Moiseev, A.) (Springer Nature Switzerland, Cham, 2024), 106–117. doi:10.1007/978-3-031-65385-8_8.
15. Feoktistov, V., Nikolaev, D., Gaidamaka, Y. & Samouylov, K. *Analysis of Probabilistic Characteristics in the Integrated Access and Backhaul System* in *Distributed Computer and Communication Networks: Control, Computation, Communications* (eds Vishnevskiy, V. M., Samouylov, K. E. & Kozyrev, D. V.) (Springer Nature Switzerland, Cham, 2024), 277–290. doi:10.1007/978-3-031-50482-2_22.
16. Khayrov, E. & Koucheryavy, Y. Packet Level Performance of 5G NR System Under Blockage and Micromobility Impairments. *IEEE Access* **11,** 90383–90395. doi:10.1109/ACCESS.2023.3307021 (2023).
17. Salimzyanov, R. & Moiseev, A. *Local balance equation for the probability distribution of the number of customers in the IAB network* in *SUITMM, Omsk* (2023), 284–289.
18. Rykov, V. On analysis of periodic polling systems. *Autom. Remote Control* **70,** 997–1018. doi:10.1134/S0005117909060071 (2009).
19. Takagi, H. *Analysis of polling systems* p. 175. 175 pp. (MIT Press, 1986).
20. Takagi, H. & Kleinrock, L. *A tutorial on the analysis of polling systems* p. 172. 172 pp. (UCLA Computer Science Department, 1985).
21. Takagi, H. Mean message waiting times in symmetric multiqueue systems with cyclic service. *Performance Evaluation* **5,** 271–277 (1985).
22. Zaripova, E. *Metody analiza pokazateley effektivnosti telekommunikatsionnoy seti serverov protokola ustanovleniya sessiy [Methods of analyzing the efficiency indicators of the telecommunication network of session establishment protocol servers]* p. 18. PhD thesis (RUDN, Moscow, 2015), 18.
23. Ge, J., Bao, L., Ding, H. & Ding, X. *Performance Analysis of the First-order Characteristics of Two-level Priority Polling System Based on Parallel Gated and Exhaustive Services Mode* in *2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT)* (2021), 10–13. doi:10.1109/ICEICT53123.2021.9531122.
24. Vishnevsky, V. & Semenova, O. *Sistemy pollinga: Teoriya i primenenie v shirokopolosnykh besprovodnykh setyakh [Polling systems. Theory and applications for broadband wireless networks]* p. 312. 312 pp. (Tekhnosfera, Moscow, 2012).

## Information about the authors

**Dmitry I. Nikolaev**—Master's degree student of Department of Probability Theory and Cybersecurity of RUDN University (e-mail: nikolaev-di@rudn.ru, phone: +7 (919) 964-08-44, ORCID: 0009-0001-5885-8566, ResearcherID: LEM-8084-2024, Scopus Author ID: 58971412800)

**Vitalii A. Beschastnyi**—Candidate of Physical and Mathematical Sciences, assistant professor of Department of Probability Theory and Cybersecurity of RUDN University (e-mail: beschastnyy-va@rudn.ru, ORCID: 0000-0003-1373-4014, ResearcherID: AAH-7080-2019, Scopus Author ID: 57192573001)

**Yuliya V. Gaidamaka**—Doctor of Physical and Mathematical Sciences, full professor of Department of Probability Theory and Cybersecurity of RUDN University; Senior Researcher of Institute of Informatics Problems of Federal Research Center "Computer Science and Control" Russian Academy of Sciences (e-mail: gaydamaka-yuv@rudn.ru, ORCID: 0000-0003-2655-4805, ResearcherID: E-5620-2014, Scopus Author ID: 35106984800)

УДК 519.872:519.217
PACS 02.50.Ey, 02.50.Fz
DOI: 10.22363/2658-4670-2024-32-4-362-369 EDN: DRHDFU

# Система поллинга с двумя очередями как модель узла сети интегрированного доступа и транзита в полудуплексном режиме

Д. И. Николаев[1], В. А. Бесчастный[1], Ю. В. Гайдамака[1, 2]

[1] Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

[2] Институт проблем информатики, Федеральный исследовательский центр «Информатика и управление» РАН, ул. Вавилова, д. 44, корп. 2, Москва, 119333, Российская Федерация

**Аннотация.** Технология интегрированного доступа и транзита (Integrated Access and Backhaul, IAB) позволяет создать компактную сеть за счёт использования узлов ретрансляторов вместо полностью оборудованных базовых станций, что впоследствии минимизирует расходы, связанные с переходом к сетям следующего поколения. Большая часть работ, посвящённых сетям IAB, опираются на инструменты имитационного моделирования и создание моделей, функционирующих в дискретном времени. В данной работе представлена математическая модель граничного узла в сети IAB с полудуплексным режимом передачи данных. Предлагаемая модель конструируется как система поллинга с двумя очередями в непрерывном времени и анализируется с помощью аппарата теории массового обслуживания, интегральных преобразований и производящих функций (ПФ). В результате получены аналитические выражения для ПФ, вероятностных распределений, а также средних и дисперсий числа заявок в очередях, которые соответствуют пакетам, ожидающим своей передачи на ретрансляционном узле по каналам доступа и транзита.

**Ключевые слова:** поллинг, система массового обслуживания, интегрированный доступ и транзит, полудуплекс

**Discrete & Continuous Models**
**& Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 370–379**
http://journals.rudn.ru/miph

*Research article*

UDC 004.891.2
PACS 03B70, 68T27, 68W01
DOI: 10.22363/2658-4670-2024-32-4-370-379 EDN: EPGKRU

# MMEmAsis: multimodal emotion and sentiment analysis

Gleb A. Kiselev[1,2], Yaroslava M. Lubysheva[1], Daniil A. Weizenfeld[1,2]

[1] RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

[2] Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44 Vavilova St, bldg 2, Moscow 119333, Russian Federation

(received: July 24, 2024; revised: August 20, 2024; accepted: September 20, 2024)

**Abstract.** The paper presents a new multimodal approach to analyzing the psycho-emotional state of a person using nonlinear classifiers. The main modalities are the subject's speech data and video data of facial expressions. Speech is digitized and transcribed using the Scribe library, and then mood cues are extracted using the Titanis sentiment analyzer from the FRC CSC RAS. For visual analysis, two different approaches were implemented: a pre-trained ResNet model for direct sentiment classification from facial expressions, and a deep learning model that integrates ResNet with a graph-based deep neural network for facial recognition. Both approaches have faced challenges related to environmental factors affecting the stability of results. The second approach demonstrated greater flexibility with adjustable classification vocabularies, which facilitated post-deployment calibration. Integration of text and visual data has significantly improved the accuracy and reliability of the analysis of a person's psycho-emotional state

**Key words and phrases:** dataset, emotion analysis, multimodal data mining, artificial intelligence, machine learning, deep learning, neuroscience data mining

**For citation:** Kiselev, G. A., Lubysheva, Y. M., Weizenfeld, D. A. MMEmAsis: multimodal emotion and sentiment analysis. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 370–379. doi: 10.22363/2658-4670-2024-32-4-370-379. edn: EPGKRU (2024).

## 1. Introduction

Automatic detection and identification of signs of psychoemotional states are among the topical applied directions of engineering and artificial intelligence technologies development. Such systems make it possible to automate the process of controlling the actions of both individuals and groups of people, including in places of increased danger by timely informing the controlling services.

In recent years, in the field of recognizing the psycho-emotional state of users, the importance of automatic multimodal recognition has been increasing, providing the next level after syntax and semantics analysis, word search from emotion dictionaries. Automatic multimodal recognition techniques allow to increase the amount of information processed, which has a positive impact on the accuracy of emotion recognition. Addition of video and audio modalities allows to operate also on the analysis of users' gestures, their facial expressions, sequences of reactive movements, to analyze the timbre, volume of the voice, to find hidden artifacts in it. These factors significantly complement classical textual methods of analyzing the emotional state of users and allow to create applied actual systems.

Operational recognition of emotional states as an applied task of artificial intelligence technology is in demand in many fields. Risk analysis of employee behavior allows the employer to optimally plan the company's business processes and predict the personal efficiency of employees and the team as a whole. Monitoring of an employee's condition allows to take timely measures to stabilize it at the individual level or to solve general organizational problems. The library can be used in modeling the psychological climate of teams.

To recognize target psycho-emotional states, basic methods of emotion recognition supplemented with behavioral models can be used. Most existing developments (FaceReader by Dutch company Noldus, EmoDetect, etc.) are based on the theory of basic emotions, where the classes are 6 emotions: joy, surprise, sadness, disgust, anger, fear. In the present project, a complex psychophysiological state will be revealed not only on the basis of mimic signs, but also by analyzing the subject's movements and speech. This way of analysis is chosen on the basis of ideas about behavioral approach and its differences from the discrete model of emotions. An example of a discrete model system is the development of [1], which uses human skeletal landmarks to analyze movements and identify the six emotions mentioned above. In addition to analyzing movements and facial expression multimodal methods (video, audio, text) are used to recognize emotions, as in [2–4].

Subjective psychological experience is inevitably accompanied by physiological changes necessary to organize a particular behavior. Emotion allows rapid organization of responses of separated physiological systems, including facial expressions, somatic muscle tone, acoustic characteristics of the speech signal, autonomic nervous and endocrine systems, to prepare the organism for adaptive behavior [5–7].

## 2. Modality overview

Analyzing human emotion is a complex process with step-by-step extraction of feature space and its analysis.

Analysis of facial features. Mimics are coordinated movements of facial muscles. Certain facial expressions that occur to communicate one's state to others (expression of emotions) are closely related to the psychophysiological state. The mimic expression of basic emotions is very similar across cultures, but is often masked depending on certain cultural attitudes, partially discordant with subjective experiences and physiological indicators, justifying validation within a specific culture.

Analysis of gestures and posture. The need to analyze human gestures and posture is due to two main factors. Human posture, as well as facial expressions, is an important means of expressing emotions. The analysis of posture allows to reveal not only obvious psychophysiological states, but also more subtle non-verbal signals reflecting tension, fatigue or stress, which may not be explicitly expressed through facial expressions.

Speech analysis. The need for speech analysis stems from the increased accuracy of emotion recognition in identifying features such as acoustic and tempo-dynamic characteristics of speech.

Assessing the dynamics of posture change. One important quantitative measure is the change in posture over time. The dynamics of body movements offer a rich source of emotional information that cannot be obtained from static postures alone. The way a person moves from one pose to another, the speed and fluidity of movements can indicate specific emotions with greater clarity and nuance. Information from a person's face, voice, and posture is interrelated with the person's movements, reinforcing the emotions expressed in the face and voice. Certain emotions are closely related to specific temporal movement patterns. For example, sudden, jerky movements may indicate surprise or fear, while slow, jerky movements may signal fatigue and depression. Capturing these dynamics is critical for accurate emotion recognition.

Understanding the context of a movement sequence can greatly influence its emotional interpretation. By analyzing dynamics, the context and progression of emotional states can be better understood, leading to more accurate recognition. Some emotions are expressed through subtle changes in movement dynamics that might be missed if only static postures were analyzed. Evaluating the dynamics allows these subtle signals to be detected. In using applications, understanding the dynamics of body movement can lead to more immersive and responsive experiences. This allows systems to respond not only to the fact of movement, but also to its emotional content. Analyzing body movement dynamics can also help in predicting future actions and emotional states, which is valuable in the fields of safety, health, and education [8]. Analysis of dynamics (both pose and facial expressions) can eliminate artifacts associated with an individual's habitual postures and expressive expressions. While the analysis of static images can be distorted by facial or body features, analyzing the changes that occur significantly increases the reliability of the data obtained.

## 3. Methods

**Algorithm 1** Algorithmic representation of approach 1

**Require:** $i$ ▷ Input image
**Require:** $a$ ▷ Input audio (last $n$ seconds)
**Require:** $detector$ ▷ Face detector and cropper module
**Require:** $fer$ ▷ Face expression classifier model
**Require:** $transcriber$ ▷ Text transcriber module (Pisets)
**Require:** $ta$ ▷ Text analyzer module (Titanis)
**Require:** $\vec{bias}$ ▷ Bias for calibrating modality result weight in final classification
**while** $i$ **do** ▷ While the video stream supplies the image
  $\vec{faces} \leftarrow detector(i)$ ▷ Detecting all faces on image
  $\vec{logits}_{text} \leftarrow ta(transcriber(a))$ ▷ Classifying text sentiment
  **if** $|\vec{faces}| > 0$ **then** ▷ If we have detected at least one face
    **for** $face$ in $\vec{faces}$ **do** ▷ Iterating over detected faces
      $\vec{logits}_{face} \leftarrow fer(face)$ ▷ Classifying face expression
      **if** $|\vec{faces}| = 1$ **then** ▷ If we have detected only one face, combine the classifications
        $result \leftarrow argmax(\vec{logits}_{face} + \vec{logits}_{text}\vec{bias})$
      **else** ▷ Else, displaying each face label separately
        $result \leftarrow argmax(\vec{logits}_{face})$
      **end if**
      display $result$
    **end for**
    **if** $|\vec{faces}| > 1$ **then** ▷ If we have more than one face, displaying text sentiment separately
      display $\vec{logits}_{text}$
    **end if**
  **end if**
**end while**

Analytical review of methods of complex multimodal analysis of human psycho-emotional state, as well as multimodal datasets used for emotion detection, has shown that most of the existing datasets designed for training neural network models are based on the selection of a set of individual emotions

**Algorithm 2** Algorithmic representation of approach 2

**Require:** $i$ ▷ Input image
**Require:** $a$ ▷ Input audio (last $n$ seconds)
**Require:** $detector$ ▷ Face detector and cropper module
**Require:** $au$ ▷ Face action unit detector model
**Require:** $dict_{au}$ ▷ Label dict for action units combinations
**Require:** $transcriber$ ▷ Text transcriber module (Pisets)
**Require:** $ta$ ▷ Text analyzer module (Titanis)
**Require:** $\vec{bias}$ ▷ Bias for calibrating modality result weight in final classification

**while** $i$ **do**
  $\vec{faces} \leftarrow detector(i)$
  $\vec{logits}_{text} \leftarrow ta(transcriber(a))$
  **if** $|\vec{faces}| > 0$ **then**
    **for** $face$ in $\vec{faces}$ **do**
      $\vec{labels}_{face} \leftarrow au(face)$ ▷ Get action unit labels
      $\vec{logits}_{face} \leftarrow dict_{au}[\vec{labels}_{face}]$ ▷ Convert to logits using label dict
      **if** $|\vec{faces}| = 1$ **then**
        $result \leftarrow argmax(\vec{logits}_{face} + \vec{logits}_{text}\vec{bias})$
      **else**
        $result \leftarrow argmax(\vec{logits}_{face})$
      **end if**
      display $result$
    **end for**
    **if** $|\vec{faces}| > 1$ **then**
      display $\vec{logits}_{text}$
    **end if**
  **end if**
**end while**

[3, 4, 9–15]. Also, one of the disadvantages of existing databases is the frequent use of static images as material for processing, which gives a large error due to the individual characteristics of the subjects. We propose a dynamic option for processing and continuous feature extraction, which will improve the validity and predictive power of the data.

There are no open non-commercial datasets of this kind in Russia; the possibility of using foreign datasets may be limited by the cultural context, which determines the peculiarities of speech and rules of emotion expression. Four types of expression are distinguished in the literature: expression of an existing emotion according to its intensity; aggravation (amplification) of an emotion, masking (reduction or suppression) of an emotion, and distortion—expression of another emotional state. Which emotions are "allowed" or "forbidden" for expression, depends significantly on cultural attitudes and can affect markers of psychophysiological state. Thus, the proposed solution, implemented on the Russian sample, has a significant novelty. The achievability of the task is ensured by the fact that the team of authors is experienced in data collection, processing and utilization.

To identify the subject's emotional state, it is proposed to use an approach in which the dynamics of changes in the subject's facial expressions, posture, and voice (prosodic and temporal characteristics of speech) are considered as a marker of specific psychophysiological states. In these visually and auditorily registered characteristics, specific features (markers) associated with target states

will be identified with the help of self-learning neural networks. The input data will be the signal from a surveillance camera that allows recording both video and audio data of the subjects. The use of dynamic features will allow us to overcome the limitations associated with the individual characteristics of different people (stable features that, when analyzing static data, can be confused with expressive manifestations, e.g., constitutionally lowered corners of the lips—interpreted as a depressive state).

The analysis of behavioral sequences has been widely developed in the ethological approach, including human ethology [16]. Features of changes in mimicry and posture that are not characteristic of a conditionally healthy population are used within the framework of this approach in psychiatric practice, showing good criteria for differential diagnosis [17], which allows us to predict the possibility of applying them to the tasks of monitoring pronounced changes in psychophysiological states in conditionally healthy individuals.

Facial recognition technologies will enable a future strategy for tracking emergent change to be applied to people moving between different observation points by being able to identify the same person and collect consistent information about them.

In the first stage of the empirical study, in order to identify groups of people for whom certain states are characteristic (i.e., we can expect their manifestation in a wide range of conditions), standardized psychodiagnostic techniques, which will allow us to compare the results obtained with population norms. A survey method will also be used to screen out those who do not fit the criteria for participation in the study. The method of completing the methods and questionnaire online with initial automatic processing of the results will be used.

It is also planned to apply a psychodiagnostic method and an interview method immediately prior to the video recording, which will make it possible to control the current state of the subjects. Methods of induction of the appearance of specific psychophysiological states will be used during videoregistration. Selection of methods common for all groups of respondents and methods specific for each group will make it possible to achieve a higher probability of occurrence of the necessary states. Artifact control methods such as counterbalancing of influences (applying them in a random order for different subjects) and introduction of neutral stimuli (so that when the next stimuli are presented, the subject can get out of the previous emotional state) will be used. In all experimental sequences, the subject will say something aloud (his own or a suggested text), which will allow the extraction of speech characteristics.

The actual occurrence in each case of the expected psychoemotional states will be validated using the heart rate variability index, which is a good indicator of their occurrence, but in comparison with visual and auditory markers cannot be so easily (without a device attached to the human body) used as an independent monitoring tool in the conditions of everyday human activity [18].

Experts will be recruited to extract episodes of target psychophysiological states, accompanied by visual and voice changes, by partitioning and filtering the recordings. Candidates of extracted markers and combinations of markers could be, for example:

1. a combined lowering of the shoulders, the appearance of a transverse crease between the eyebrows, the lowering of the corners of the mouth, a decrease in the volume of the voice, and the appearance of pauses while the depressive state intensifies;
2. combined raising of shoulders and elbows, raising of eyebrows, appearance of transverse folds on the bridge of the nose, appearance of a specific sign “square mouth”, increase in the volume and tempo of the voice, etc. when the aggressive state increases.

The application of an integrated approach will increase the validity of detectable signs in terms of assessing the state of a person.

Table 1
Most frequent face expression decisions generated by model for one person showing happiness in different environments

| Method | Environment | Top Label | Frequency |
|---|---|---|---|
| 1 | A | Happiness | 75.4% |
| 1 | B | Contempt | 65.1% |
| 1 | C | Contempt | 71.8% |
| 1 | D | Happiness | 51.4% |
| 1 | E | Happiness | 55.3% |
| 2 | A | Happiness | 65.7% |
| 2 | B | Happiness | 65.1% |
| 2 | C | Contempt | 66.9% |
| 2 | D | Disgust | 59.8% |
| 2 | E | Contempt | 57.2% |

## 4. Prototype

Our prototype leverages both visual and textual data to enhance sentiment analysis capabilities. Textual data is generated through audio transcription using the Pisets library [19]. Subsequently, sentiment features are extracted from the transcribed text using the FRC CSC RAS Titanis sentiment analyzer, which provides detailed sentiment feature labels.

For the visual analysis component, we implemented and benchmarked two distinct approaches to sentiment recognition. The first approach employs a stand-alone ResNet model designed to directly classify sentiment based on cropped face images. Specifically, we utilized the model [20], which performs single-label classification to identify one of eight possible emotions from facial expressions. This approach is described by algorithm 1.

The second approach focuses on facial action unit (AU) recognition. For this, we employed a sophisticated deep learning model proposed by [21], which integrates a ResNet with a graph-based deep neural network (DNN) to achieve multi-label classification of facial action units. These AU labels are then translated into final sentiment labels using classification dictionaries. This method allows for the identification of specific sentiments, such as frustration or intoxication, in addition to the same eight emotions used in the first approach. To facilitate fine-tuning, a calibration tool was developed. This tool enables users to display chosen emotions or sentiments via a user interface, aiding in the adjustment of classification dictionaries. This approach is described by algorithm 2.

In both approaches, initial face boundary detection is essential for cropping the face from the image. We used the FaceTorch utility [22], which incorporates the RetinaFace model by [23] for accurate face localization. All the models we used were pre-trained by their papers' authors.

## 5. Results

Both visual sentiment recognition approaches encountered similar practical challenges. Environmental factors, such as background and lighting, significantly influenced the consistency of

Table 2
Most consistent action unit labels detected by model from the second approach. The shown expression is 'Happiness' in all cases. The labels in bold are least related to the facial expression shown, while the remaining labels are also not the main ones, indicating this expression only indirectly

| Environment | Top Final Label | Top Action Unit Labels | | | |
|---|---|---|---|---|---|
| | | lips part | **nose wrinkler** | left upper lip raiser | cheek raiser |
| A | Happiness | 80.9% | 0.0% | 95.2% | 40.2% |
| B | Happiness | 24.0% | 10.2% | 89.8% | 47.4% |
| C | Contempt | 33.3% | 45.6% | 74.1% | 10.0% |
| D | Disgust | 68.3% | 70.1% | 44.2% | 32.0% |
| E | Contempt | 30.1% | 66.6% | 33.8% | 5.9% |

results. Table 1 illustrates the classification outcomes for a single person displaying a 'Happiness' emotion across various environments, differing in lighting conditions, capturing devices, and backgrounds. The first approach frequently misclassified 'Happiness' as 'Contempt' in some environments. Conversely, the second approach produced varying results, with certain facial action units being detected consistently, irrespective of the actual facial expression, as shown in Table 2.

## 6. Conclusions

In summary, the first approach, utilizing a straightforward model architecture, requires extensive fine-tuning for application-specific environments and lacks manual adjustability of model outputs. In contrast, the second approach offers greater flexibility by outputting detected facial actions that can be mapped to final sentiment labels. This flexibility allows for both model and classification dictionary adjustments, facilitating easier post-deployment calibration and modification of the sentiment set.

For the textual analysis, it complements the visual results, contributing to the final sentiment label determination for the individual in the image. The integration of textual and visual data enhances the accuracy and robustness of the sentiment analysis in our prototype.

**Author Contributions:** Conceptualization, Gleb A. Kiselev and Yaroslava M. Lubysheva; methodology, Gleb A. Kiselev and Yaroslava M. Lubysheva; software, Daniil A. Weizenfeld; investigation, Daniil A. Weizenfeld, Yaroslava M. Lubysheva and Gleb A. Kiselev; data curation, Gleb A. Kiselev, Yaroslava M. Lubysheva and Daniil A. Weizenfeld; writing—original draft preparation, Yaroslava M. Lubysheva and Daniil A. Weizenfeld; writing—review and editing, Gleb A. Kiselev; visualization, Daniil A. Weizenfeld; supervision, Gleb A. Kiselev; project administration, Gleb A. Kiselev; funding acquisition, Gleb A. Kiselev. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper has been supported by the RUDN University Strategic Academic Leadership Program.

**Data Availability Statement:** Data sharing is not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; in the decision to publish the results.

## References


1. Piana, S., Staglianò, A., Odone, F., Verri, A. & Camurri, A. *Real-time Automatic Emotion Recognition from Body Gestures* 2014. doi:10.48550/arXiv.1402.5047.
2. Hu, G., Lin, T., Zhao, Y., Lu, G., Wu, Y. & Li, Y. UniMSE: Towards Unified Multimodal Sentiment Analysis and Emotion Recognition. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing.* doi:10.48550/arXiv.2211.11256 (2022).
3. Zhao, J., Zhang, T., Hu, J., Liu, Y., Jin, Q., Wang, X. & Li, H. *M3ED: Multi-modal Multi-scene Multi-label Emotional Dialogue Database* in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Association for Computational Linguistics, Dublin, Ireland, May, 2022, 2022), 5699–5710. doi:10.18653/v1/2022.acl-long.391.
4. Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E. & Mihalcea, R. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.* doi:10.48550/arXiv.1810.02508 (2018).
5. Ekman, P. Emotion: common characteristics and individual differences. *Lecture presented at 8th World Congress of I.O.P. Tampere Finland* (1996).
6. Levenson, R. W. The intrapersonal functions of emotion. *Cognition & Emotion* **13,** 481–504 (1999).
7. Keltner, D. & Gross, J. Functional accounts of emotions. *Cognition & Emotion* **13,** 467–480 (1999).
8. Ferdous, A., Bari, A. & Gavrilova, M. Emotion Recognition From Body Movement. *IEEE Access.* doi:10.1109/ACCESS.2019.2963113 (Dec. 2019).
9. Zadeh, A., Liang, P., Poria, S., Cambria, E. & Morency, L.-P. *Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph* in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (July 2018), 2236–2246. doi:10.18653/v1/P18-1208.
10. Busso, C., Bulut, M. & Lee, C. e. a. IEMOCAP: interactive emotional dyadic motion capture database. *Lang Resources & Evaluation* **42,** 335–359. doi:10.1007/s10579-008-9076-6 (2008).
11. Kossaifi, J. *et al.* SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13.** doi:10.1109/TPAMI.2019.2944808 (Oct. 2019).
12. O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., Bölte, S., Baron-Cohen, S. & Lundqvist, D. The EU-Emotion Stimulus Set: A validation study. *Behav Res Methods* **48,** 567–576. doi:10.3758/s13428-015-0601-4 (2016).
13. Soleymani, M., Lichtenauer, J., Pun, T. & Pantic, M. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing* **3,** 42–55. doi:10.1109/T-AFFC.2011.25 (2012).
14. Chou, H. C., Lin, W. C., Chang, L. C., Li, C. C., Ma, H. P. & Lee, C. C. *NNIME: The NTHU-NTUA Chinese interactive multimodal emotion corpus* in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (2017), 292–298. doi:10.1109/ACII.2017.8273615.
15. Ringeval, F., Sonderegger, A., Sauer, J. & Lalanne, D. *Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions* in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (2013), 1–8. doi:10.1109/FG.2013.6553805.
16. Reznikova, J. I. *Intelligence and language in animals and humans* 253 pp. (Yurayt, 2016).
17. Samokhvalov, V. P., Kornetov, A. N., Korobov, A. A. & Kornetov, N. A. *Ethology in psychiatry* 217 pp. (Health, 1990).
18. Gullett, N., Zajkowska, Z., Walsh, A., Harper, R. & Mondelli, V. Heart rate variability (HRV) as a way to understand associations between the autonomic nervous system (ANS) and affective

states: A critical review of the literature. *International Journal of Psychophysiology* **192,** 35–42. doi:10.1016/j.ijpsycho.2023.08.001 (2023).

19. Bondarenko, I. *Pisets: A Python library and service for automatic speech recognition and transcribing in Russian and English* https://github.com/bond005/pisets.
20. Savchenko, A. V. *Facial expression and attributes recognition based on multi-task learning of lightweight neural networks* in *2021 IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY)* (2021), 119–124.
21. Luo, C., Song, S., Xie, W., Shen, L. & Gunes, H. Learning multi-dimensional edge feature-based au relation graph for facial action unit recognition. *arXiv preprint arXiv:2205.01782* (2022).
22. Gajarsky, T. *Facetorch: A Python library for analysing faces using PyTorch* https://github.com/tomas-gajarsky/facetorch.
23. Deng, J., Guo, J., Ververas, E., Kotsia, I. & Zafeiriou, S. *Retinaface: Single-shot multi-level face localisation in the wild* in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), 5203–5212.

## Information about the authors

**Gleb A. Kiselev**—Candidate of Technical Sciences, Senior Lecturer at the Department of Mathematical Modeling and Artificial Intelligence of RUDN University; Researcher of Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (e-mail: kiselev@isa.ru, phone: +7 (906) 799-33-29, ORCID: 00000-0001-9231-8662, ResearcherID: Y-6971-2018, Scopus Author ID: 57195683637)

**Yaroslava M. Lubysheva**—Master's degree student of Department of Mathematical Modeling and Artificial Intelligence of RUDN University (e-mail: gorbunova_y_m@mail.ru, phone: +7 (977) 942-66-51, ORCID: 0000-0001-6280-6040)

**Daniil A. Weizenfeld**—Master's degree student of Department of Mechanics and Control Processes of RUDN University (e-mail: veicenfeld@isa.ru, phone: +7 (903) 123-76-05, ORCID: 0000-0002-2787-0714)

УДК 004.891.2
PACS 03B70, 68T27, 68W01
DOI: 10.22363/2658-4670-2024-32-4-370-379 EDN: EPGKRU

# ММЕмАсис: мультимодальный метод оценки психофизиологического состояния человека.

Г. А. Киселёв[1,2], Я. М. Лубышева[1], Д. А. Вейценфельд[1,2]

[1] Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

[2] Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ул. Вавилова, д. 44, корп. 2, Москва, 119333, Российская Федерация

**Аннотация.** В статье представлен новый мультимодальный подход анализа психоэмоционального состояния человека с помощью нелинейных классификаторов. Основными модальностями являются данные речи испытуемого и видеоданные мимики. Речь оцифровывается и транскрибируется библиотекой Писец, признаки настроения извлекаются системой Titanis от ФИЦ ИУ РАН. Для визуального анализа были реализованы два различных подхода: дообученная модель ResNet для прямой классификации настроений по выражениям лица и модель глубокого обучения, интегрирующая ResNet с основанной на графах глубокой нейронной сетью для распознавания мимических признаков. Оба подхода сталкивались с трудностями, связанными с факторами окружающей среды, влияющими на стабильность результатов. Второй подход продемонстрировал бóльшую гибкость благодаря регулируемым словарям классификации, что облегчало калибровку после развёртывания. Интеграция текстовых и визуальных данных значительно улучшила точность и надёжность анализа психоэмоционального состояния человека.

**Ключевые слова:** набор данных, анализ эмоций, мультимодальный анализ данных, искусственный интеллект, машинное обучение, глубокое обучение, анализ нейрофизиологических данных

**Discrete & Continuous Models**
**& Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 380–394**
http://journals.rudn.ru/miph

*Research article*

UDC 519.872:519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-4-380-394 EDN: DDNAPR

# Asymptotic diffusion analysis of RQ system M/M/1 with unreliable server

Nataliya M. Voronina[1], Svetlana V. Rozhkova[1, 2]

[1] National Research Tomsk Polytechnic University, 30 Lenina Ave, Tomsk, 634050, Russian Federation

[2] National Research Tomsk State University, 36 Lenina Ave, Tomsk, 634050, Russian Federation

(received: July 18, 2024; revised: August 20, 2024; accepted: September 10, 2024)

**Abstract.** The paper considers a single-line retrial queueing system with an unreliable server. Queuing systems are called unreliable if their servers may fail from time to time and require restoration (repair), only after which they can resume servicing customers. The input of the system is a simple Poisson flow of customers. The service time and uptime of the server are distributed exponentially. An incoming customer try to get service. The server can be free, busy or under repair. The customer is serviced immediately if the server is free. If it is busy or under repair, the customer goes into orbit. And after a random time it tries to get service again. The study is carried out by the method of asymptotically diffusion analysis under the condition of a large delay of requests in orbit. In this work, the transfer coefficient and diffusion coefficient were found and a diffusion approximation was constructed.

**Key words and phrases:** retrial queuing system, asymptotic diffusion method, unreliable device

**For citation:** Voronina, N. M., Rozhkova, S. V. Asymptotic diffusion analysis of RQ system M/M/1 with unreliable server. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 380–394. doi: 10.22363/2658-4670-2024-32-4-380-394. edn: DDNAPR (2024).

## 1. Introduction

Queuing systems with repeated requests are quite often used in various areas of telecommunications. Modern information processing systems often encounter unstable operating conditions, such as overloads, failures, and resource limitations. Under these conditions, conventional retrial queuing (RQ) systems may not be able to process all incoming requests, resulting in lost information and poor performance [1–4].

Repetitive request systems offer a solution to this problem by providing a mechanism for processing requests that cannot be fulfilled immediately. Instead of discarding such requests, they are resubmitted to the queue after a certain time, increasing the likelihood of successful completion of service. The most complete and detailed description of RQ systems and their detailed comparison with classical queuing systems was reflected in [5–7].

There are different types of unreliability. For example, the works [8–10] consider the unreliability of the server as a breakdown. The authors in [11–14] consider an unreliable server with collisions or conflicts during simultaneous access to the server.

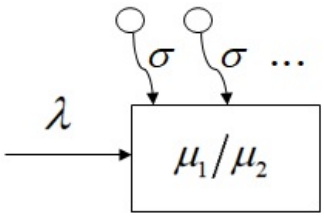


Figure 1. Model of retrial queueing system M/M/1 with unreliable server

This problem is especially relevant when it comes to unreliable servers that can fail due to software errors, hardware malfunctions or external factors. Server failures can lead to data loss, interruption of services, and decreased performance.

If the server fails while servicing the request, it goes to repair. A request under maintenance goes into orbit and awaits recovery of the server. A fairly large number of works are devoted to systems with unreliable server [15–20].

To understand the behavior of systems with repeated requests and evaluate their performance, it is necessary to use analytical methods.

In this paper, we consider a single-line queuing system with an unreliable server. We will conduct the study using the method of asymptotic diffusion analysis. It has been proven that the accuracy of the diffusion approximation exceeds the accuracy of the Gaussian approximation calculated in [21].

## 2. System description

Any data network, having generated customers, sends them to a shared resource (server). If the server is free, then the customer is served. If the server fails while servicing a customer, it is sent for repair, and the customers go into orbit.

Let's consider an RQ system with an unreliable server, the input of which receives a simple flow of customers with parameter $\lambda$. The request is served by the server at a random time, distributed according to an exponential law with the parameter $\mu_1$. An unreliable server can be in one of the following states: idle, busy, or under repair. If the server is idle and an entry customer is received, the server immediately begins servicing the incoming customer. If a customer arrives at a time when the server is busy, then the received customer goes into orbit and waits for the opportunity to occupy the server at the next attempt.

After a random delay, a customer with intensity $\sigma$ again contacts the device with an attempt to capture it (see Fig. 1). The server's uptime is distributed according to an exponential law with parameter $\gamma_1$ if the server is idle, and with parameter $\gamma_2$, if the server is busy. As soon as a breakdown occurs, the server is sent for repair. All incoming customers go into orbit. The recovery time after repair is distributed exponentially with the parameter $\mu_2$.

The goal of the work is to study such a system, as well as to find its main characteristics.

## 3. Kolmogorov equations

Let us denote by $P\{i(t) = i,\ k(t) = k,\ n(t) = n\} = P(k, i, t)$—the probability that at a given time $t$ the server is in state $k$ and in the orbit of $i$ customers. The probability distribution $P(k, i, t)$ satisfies the

following system of equations:

$$\begin{aligned}
P_0(i,t+\Delta t) =&(1-\lambda\Delta t)(1-i\sigma\Delta t)(1-\gamma_1\Delta t)P_0(i,t)+\mu_1\Delta tP_1(i,t)+\\
&+\mu_2\Delta tP_2(i,t)+o(\Delta t),\\
P_1(i,t+\Delta t) =&(1-\lambda\Delta t)(1-\mu_1\Delta t)(1-\gamma_2\Delta t)P_1(i,t)+\lambda\Delta tP_0(i,t)+\\
&+\sigma(i+1)\Delta tP_0(i+1,t)+\lambda P_1(i-1,t)+o(\Delta t),\\
P_2(i,t+\Delta t) =&(1-\lambda\Delta t)(1-\mu_2\Delta t)P_2(i,t)+\gamma_1\Delta tP_0(i,t)+\\
&+\gamma_2\Delta tP_1(i-1,t)+\lambda\Delta tP_2(i-1,t)+o(\Delta t).
\end{aligned}$$

Let's create a system of Kolmogorov differential equations:

$$\begin{cases}
\dfrac{\partial P_0(i,t)}{\partial t} =-(\lambda+i\sigma+\gamma_1)P_0(i,t)+\mu_1P_1(i,t)+\mu_2P_2(i,t),\\
\dfrac{\partial P_1(i,t)}{\partial t} =-(\lambda+\mu_1+\gamma_2)P_1(i,t)+\lambda P_0(i,t)+\\
\qquad\qquad+\sigma(i+1)P_0(i+1,t)+\lambda P_1(i-1,t),\\
\dfrac{\partial P_2(i,t)}{\partial t} =-(\lambda+\mu_2)P_2(i,t)+\gamma_1P_0(i,t)+\gamma_2P_1(i-1,t)+\lambda P_2(i-1,t).
\end{cases} \tag{1}$$

Let us write down the partial characteristic functions:

$$H_k(u,t)=\sum_{i=0}^{\infty}e^{iuj}P_k(i,t),\quad k=\{0,1,2\},$$

where $j=\sqrt{-1}$.

Multiplying the equations of the system (1) by $e^{iuj}$, we obtain

$$\begin{cases}
\dfrac{\partial H_0(i,t)}{\partial t} =-(\lambda+\gamma_1)H_0(u,t)+j\sigma e^{ju}\dfrac{\partial H_0(u,t)}{\partial u}+\\
\qquad\qquad+\mu_1H_1(u,t)+\mu_2H_2(u,t),\\
\dfrac{\partial H_1(i,t)}{\partial t} =-(\lambda+\mu_1+\gamma_2)H_1(u,t)+\lambda H_0(u,t)-\\
\qquad\qquad-j\sigma\dfrac{\partial H_0(u,t)}{\partial u}+\lambda e^{ju}H_1(u,t),\\
\dfrac{\partial H_2(i,t)}{\partial t} =-(\lambda+\mu_2)H_2(u,t)+\gamma_1H_0(u,t)+\\
\qquad\qquad+\gamma_2e^{ju}H_1(u,t)+\lambda e^{ju}H_2(u,t).
\end{cases} \tag{2}$$

Summing up the equations of the system (2), we write the equation for the characteristic function

$$H(u,t)=H_0(u,t)+H_1(u,t)+H_2(u,t),$$

then we get

$$\frac{\partial H(u,t)}{\partial t}=(e^{ju}-1)\left(H_1(u,t)(\lambda+\gamma_2)+H_2(u,t)\lambda+j\sigma\frac{\partial H_0(u,t)}{\partial u}\right). \tag{3}$$

We will find a characteristic function of the number of customers in orbit under the condition of a long delay. We will investigate in two stages.

## 4. Stage 1. Getting the transfer coefficient

Let us introduce the substitutions in the system (2) and the equation (3)

$$\sigma = \varepsilon, \quad \tau = \varepsilon t, \quad u = \varepsilon\omega, \quad H_k(u,t) = F_k(\omega,\tau,\varepsilon), \quad k = \{0,1,2\}.$$

Then we get the following system:

$$\begin{cases} \varepsilon\dfrac{\partial F_0(\omega,\tau,\varepsilon)}{\partial\tau} = -(\lambda+\gamma_1)F_0(\omega,\tau,\varepsilon) + je^{j\varepsilon\omega}\dfrac{\partial F_0(\omega,\tau,\varepsilon)}{\partial\omega} + \\ \qquad + \mu_1 F_1(\omega,\tau,\varepsilon) + \mu_2 F_2(\omega,\tau,\varepsilon), \\ \varepsilon\dfrac{\partial F_1(\omega,\tau,\varepsilon)}{\partial\tau} = -(\lambda+\mu_1+\gamma_2)F_1(\omega,\tau,\varepsilon) + \lambda F_0(\omega,\tau,\varepsilon) - \\ \qquad - j\dfrac{\partial F_0(\omega,\tau,\varepsilon)}{\partial\omega} + \lambda e^{j\varepsilon\omega}F_1(\omega,\tau,\varepsilon), \\ \varepsilon\dfrac{\partial F_2(\omega,\tau,\varepsilon)}{\partial\tau} = -(\lambda+\mu_2)F_2(\omega,\tau,\varepsilon) + \gamma_1 F_0(\omega,\tau,\varepsilon) + \\ \qquad + \gamma_2 e^{j\varepsilon\omega}F_1(\omega,\tau,\varepsilon) + \lambda e^{j\varepsilon\omega}F_2(\omega,\tau,\varepsilon). \end{cases} \tag{4}$$

The equation (3) will take the form:

$$\varepsilon\frac{\partial F(\omega,\tau,\varepsilon)}{\partial\tau} = (e^{j\varepsilon\omega}-1)\left(F_1(\omega,\tau,\varepsilon)(\lambda+\gamma_2) + F_2(\omega,\tau,\varepsilon)\lambda + j\frac{\partial F_0(\omega,\tau,\varepsilon)}{\partial\omega}\right). \tag{5}$$

In the system (4) and the equation (5), we decompose the exponent into a Taylor series:

$$e^{j\omega\varepsilon} = 1 + j\omega\varepsilon, \quad \lim_{\varepsilon\to 0}\frac{e^{j\omega\varepsilon}}{\varepsilon} = j\omega.$$

Let us perform the transition to the limit at $\varepsilon \to 0$, then we obtain:

$$\begin{cases} -(\lambda+\gamma_1)F_0(\omega,\tau) + j\dfrac{\partial F_0(\omega,\tau)}{\partial\omega} + \mu_1 F_1(\omega,\tau) + \mu_2 F_2(\omega,\tau) = 0, \\ -(\mu_1+\gamma_2)F_1(\omega,\tau) + \lambda F_0(\omega,\tau) - j\dfrac{\partial F_0(\omega,\tau)}{\partial\omega} = 0, \\ -\mu_2 F_2(\omega,\tau) + \gamma_1 F_0(\omega,\tau) + \gamma_2 F_1(\omega,\tau) = 0. \end{cases} \tag{6}$$

$$\frac{\partial F(\omega,\tau)}{\partial\tau} = j\omega\left(F_1(\omega,\tau)(\lambda+\gamma_2) + F_2(\omega,\tau)\lambda + j\frac{\partial F_0(\omega,\tau)}{\partial\omega}\right). \tag{7}$$

We will find a solution to the system (6) and the equation (7) in the form:

$$F_k(\omega,\tau) = R_k e^{j\omega x(\tau)}, \quad k = \{0,1,2\},$$

where $R_k$ has the meaning of the stationary probability that the server is in state $k$, and $x(\tau)$ is a scalar function of the argument $\tau$, which determines at $\varepsilon \to 0$ the average value $\sigma i(\tau/\sigma)$ of the number of customers in orbit normalized by the value at $\varepsilon = \sigma$.

Then the system (6) and the equation (7) will take the form:

$$\begin{cases} -\big(\lambda+\gamma_1+x(\tau)\big)R_0 + \mu_1 R_1 + \mu_2 R_2 = 0, \\ \big(\lambda+x(\tau)\big)R_0 - (\mu_1+\gamma_2)R_1 = 0, \\ \gamma_1 R_0 + \gamma_2 R_1 - \mu_2 R_2 = 0. \end{cases} \tag{8}$$

The probabilities $R_k$ can be found from the system (6) taking into account the normalization condition $R_0 + R_1 + R_2 = 1$.

Since the coefficient of the system of equations (6) depends on $x$, then $R_k$ can also be written as $R_k(x)$.

$$R_0 = \frac{\mu_2(\mu_1 + \gamma_2)}{\big(\gamma_1 + \lambda + \mu_2 + x(\tau)\big)\gamma_2 + \big(\lambda + \mu_1 + x(\tau)\big)\mu_2 + \gamma_1\mu_1},$$
$$R_1 = \frac{\big(\lambda + x(\tau)\big)\mu_2}{\big(\gamma_1 + \lambda + \mu_2 + x(\tau)\big)\gamma_2 + \big(\lambda + \mu_1 + x(\tau)\big)\mu_2 + \gamma_1\mu_1},$$
$$R_2 = \frac{\big(\lambda + \gamma_1 + x(\tau)\big)\gamma_2 + \gamma_1\mu_1}{\big(\gamma_1 + \lambda + \mu_2 + x(\tau)\big)\gamma_2 + \big(\lambda + \mu_1 + x(\tau)\big)\mu_2 + \gamma_1\mu_1}.$$

From the equation (7) we get:

$$x'(\tau) = -x(\tau)R_0 + (\lambda + \gamma_2)R_1 + \lambda R_2.$$

Let us denote the function $a(x) = x'(\tau)$, then

$$a(x) = -x(\tau)R_0 + (\lambda + \gamma_2)R_1 + \lambda R_2,$$

where $a(x)$ is the transfer coefficient.

## 5. Stage 2. Centering and obtaining the diffusion coefficient

Let us introduce the substitutions in the system (2) and the equation (3)

$$H_k(u,t) = H_k^{(2)}(u,t)e^{j\frac{u}{\sigma}x(t)}, \quad k = \{0,1,2\},$$

we get

$$\begin{cases} \dfrac{\partial H_0^{(2)}(u,t)}{\partial t} + H_0^{(2)}(u,t)jux'(\sigma t) = -(\lambda+\gamma_1)H_0^{(2)}(u,t) + \\ \qquad + j\sigma e^{ju}\left(\dfrac{\partial H_0^{(2)}(u,t)}{\partial u} + H_0^{(2)}(u,t)j\dfrac{1}{\sigma}x(\sigma t)\right) + \\ \qquad + \mu_1 H_1^{(2)}(u,t) + \mu_2 H_2^{(2)}(u,t) = 0, \\ \dfrac{\partial H_1^{(2)}(u,t)}{\partial t} + H_1^{(2)}(u,t)jux'(\sigma t) = -(\lambda+\mu_1+\gamma_2)H_1^{(2)}(u,t) + \\ \qquad + \lambda H_0^{(2)}(u,t) - j\sigma\left(\dfrac{\partial H_0^{(2)}(u,t)}{\partial u} + H_0^{(2)}(u,t)j\dfrac{1}{\sigma}x(\sigma t)\right) + \\ \qquad + \lambda e^{ju}H_1^{(2)}(u,t) = 0, \\ \dfrac{\partial H_2^{(2)}(u,t)}{\partial t} + H_2^{(2)}(u,t)jux'(\sigma t) = -(\lambda+\mu_2)H_2^{(2)}(u,t) + \\ \qquad + \gamma_1 H_0^{(2)}(u,t) + \gamma_2 e^{ju}H_1^{(2)}(u,t) + \lambda e^{ju}H_2^{(2)}(u,t) = 0. \end{cases} \tag{9}$$

$$\frac{\partial H^{(2)}(u,t)}{\partial t} + H^{(2)}(u,t)jux'(\sigma t) = (e^{ju}-1)\left(H_1^{(2)}(u,t)(\lambda+\gamma_2) + H_2^{(2)}(u,t)\lambda + j\sigma\left(\frac{\partial H_0^{(2)}(u,t)}{\partial u} + H_0^{(2)}(u,t)j\frac{1}{\sigma}x(\sigma t)\right)\right). \tag{10}$$

Let us introduce the substitutions in the system (9) and the equation (10)

$$\sigma = \varepsilon^2, \quad \tau = \varepsilon^2 t, \quad u = \varepsilon\omega, \quad H_k{}^{(2)}(u,t) = F_k^{(2)}(\omega,\tau,\varepsilon), \quad k = \{0,1,2\}.$$

Then we obtain:

$$\begin{cases} \varepsilon^2 \dfrac{\partial F_0^{(2)}(\omega,\tau,\varepsilon)}{\partial\tau} + F_0^{(2)}(\omega,\tau,\varepsilon) j\varepsilon\omega a(x,\tau) = -(\lambda+\gamma_1)F_0^{(2)}(\omega,\tau,\varepsilon) + \\ \qquad + j\varepsilon e^{j\varepsilon\omega}\dfrac{\partial F_0^{(2)}(\omega,\tau,\varepsilon)}{\partial\omega} - e^{j\varepsilon\omega}F_0^{(2)}(\omega,\tau,\varepsilon)x(\tau) + \\ \qquad + \mu_1 F_1^{(2)}(\omega,\tau,\varepsilon) + \mu_2 F_2^{(2)}(\omega,\tau,\varepsilon), \\ \varepsilon^2 \dfrac{\partial F_1^{(2)}(\omega,\tau,\varepsilon)}{\partial\tau} + F_1^{(2)}(\omega,\tau,\varepsilon) j\varepsilon\omega a(x,\tau) = -(\lambda+\mu_1+\gamma_2)\times \\ \qquad \times F_1^{(2)}(\omega,\tau,\varepsilon) + \lambda F_0^{(2)}(\omega,\tau,\varepsilon) - j\varepsilon\dfrac{\partial F_0^{(2)}(\omega,\tau,\varepsilon)}{\partial\omega} + \\ \qquad + F_0^{(2)}(\omega,\tau,\varepsilon)x(\tau) + \lambda e^{j\varepsilon\omega}F_1^{(2)}(\omega,\tau,\varepsilon), \\ \varepsilon^2 \dfrac{\partial F_2^{(2)}(\omega,\tau,\varepsilon)}{\partial\tau} + F_2^{(2)}(\omega,\tau,\varepsilon) j\varepsilon\omega a(x,\tau) = -(\lambda+\mu_2)F_2^{(2)}(\omega,\tau,\varepsilon) + \\ \qquad + \gamma_1 F_0^{(2)}(\omega,\tau,\varepsilon) + \gamma_2 e^{j\varepsilon\omega}F_1^{(2)}(\omega,\tau,\varepsilon) + \lambda e^{j\varepsilon\omega}F_2^{(2)}(\omega,\tau,\varepsilon). \end{cases} \tag{11}$$

$$\begin{aligned} \varepsilon^2 \frac{\partial F^{(2)}(\omega,\tau,\varepsilon)}{\partial\tau} + F^{(2)}(\omega,\tau,\varepsilon) j\varepsilon\omega a(x,\tau) = (e^{j\varepsilon\omega}-1)\times \\ \times\Big(F_1^{(2)}(\omega,\tau,\varepsilon)(\lambda+\gamma_2) + F_2^{(2)}(\omega,\tau,\varepsilon)\lambda + \\ + j\varepsilon\frac{\partial F_0^{(2)}(\omega,\tau,\varepsilon)}{\partial\omega} - F_0^{(2)}(\omega,\tau,\varepsilon)x(\tau)\Big). \end{aligned} \tag{12}$$

In the system (11), we expand the exponential in a Taylor series and group the terms of order of smallness not higher than $\varepsilon^2$.

$$\begin{cases} F_0^{(2)}(\omega,\tau,\varepsilon) j\varepsilon\omega a(x,\tau) = -(\lambda+\gamma_1)F_0^{(2)}(\omega,\tau,\varepsilon) + \\ \qquad + j\varepsilon(1+j\omega\varepsilon)\dfrac{\partial F_0^{(2)}(\omega,\tau,\varepsilon)}{\partial\omega} - (1+j\omega\varepsilon)F_0^{(2)}(\omega,\tau,\varepsilon)x(\tau) + \\ \qquad + \mu_1 F_1^{(2)}(\omega,\tau,\varepsilon) + \mu_2 F_2^{(2)}(\omega,\tau,\varepsilon) + O(\varepsilon^2), \\ F_1^{(2)}(\omega,\tau,\varepsilon) j\varepsilon\omega a(x,\tau) = -(\lambda+\mu_1+\gamma_2)F_1^{(2)}(\omega,\tau,\varepsilon) + \\ \qquad + \lambda F_0^{(2)}(\omega,\tau,\varepsilon) - j\varepsilon\dfrac{\partial F_0^{(2)}(\omega,\tau,\varepsilon)}{\partial\omega} + F_0^{(2)}(\omega,\tau,\varepsilon)x(\tau) + \\ \qquad + \lambda(1+j\omega\varepsilon)F_1^{(2)}(\omega,\tau,\varepsilon) + O(\varepsilon^2), \\ F_2^{(2)}(\omega,\tau,\varepsilon) j\varepsilon\omega a(x,\tau) = -(\lambda+\mu_2)F_2^{(2)}(\omega,\tau,\varepsilon) + \gamma_1 F_0^{(2)}(\omega,\tau,\varepsilon) + \\ \qquad + \gamma_2(1+j\omega\varepsilon)F_1^{(2)}(\omega,\tau,\varepsilon) + \lambda(1+j\omega\varepsilon)F_2^{(2)}(\omega,\tau,\varepsilon) + O(\varepsilon^2). \end{cases} \tag{13}$$

We will find a solution to the system (13) in the form:

$$F_k^{(2)}(\omega,\tau,\varepsilon) = \Phi(\omega,\tau)(R_k + j\omega\varepsilon f_k) + O(\varepsilon^2), \quad k = \{0,1,2\}, \tag{14}$$

where $R_k = R_k(x,\tau)$.

Substituting expansion (14) into the system (13), we obtain

$$\begin{cases} \Phi(\omega,\tau)j\varepsilon\omega R_0 a(x,\tau) = -(\lambda+\gamma_1)\Phi(\omega,\tau)R_0 - (\lambda+\gamma_1)\Phi(\omega,\tau)j\omega\varepsilon f_0 + \\ \qquad + j\varepsilon\dfrac{\partial\Phi(\omega,\tau)}{\partial\omega}R_0 - \Phi(\omega,\tau)R_0 x(\tau) - j\omega\varepsilon\Phi(\omega,\tau)R_0 x(\tau) - \\ \qquad - \Phi(\omega,\tau)j\omega\varepsilon f_0 x(\tau) + \mu_1\Phi(\omega,\tau)R_1 + \mu_1\Phi(\omega,\tau)j\omega\varepsilon f_1 + \\ \qquad + \mu_2\Phi(\omega,\tau)R_2 + \mu_2\Phi(\omega,\tau)j\omega\varepsilon f_2 + O(\varepsilon^2), \\ \Phi(\omega,\tau)j\varepsilon\omega R_1 a(x,\tau) = -(\mu_1+\gamma_2)\Phi(\omega,\tau)R_1 + \lambda j\omega\varepsilon\Phi(\omega,\tau)R_1 - \\ \qquad - (\mu_1+\gamma_2)\Phi(\omega,\tau)j\omega\varepsilon f_1 + \lambda\Phi(\omega,\tau)R_0 + \lambda\Phi(\omega,\tau)j\omega\varepsilon f_0 - \\ \qquad - j\varepsilon\dfrac{\partial\Phi(\omega,\tau)}{\partial\omega}R_0 + \Phi(\omega,\tau)R_0 x(\tau) + \Phi(\omega,\tau)j\omega\varepsilon f_0 x(\tau) + O(\varepsilon^2), \\ \Phi(\omega,\tau)j\varepsilon\omega R_2 a(x,\tau) = -\mu_2\Phi(\omega,\tau)R_2 + \lambda j\omega\varepsilon\Phi(\omega,\tau)R_2 - \\ \qquad - \mu_2\Phi(\omega,\tau)j\omega\varepsilon f_2 + \gamma_1\Phi(\omega,\tau)R_0 + \gamma_1\Phi(\omega,\tau)j\omega\varepsilon f_0 + \\ \qquad + \gamma_2\Phi(\omega,\tau)R_1 + \gamma_2 j\omega\varepsilon\Phi(\omega,\tau)R_1 + \gamma_2 j\omega\varepsilon f_1 + O(\varepsilon^2). \end{cases}$$

Taking into account the system (8), we get

$$\begin{cases} \Phi(\omega,\tau)j\varepsilon\omega R_0 a(x,\tau) = -(\lambda+\gamma_1)\Phi(\omega,\tau)j\omega\varepsilon f_0 + j\varepsilon\dfrac{\partial\Phi(\omega,\tau)}{\partial\omega}R_0 - \\ \qquad - j\omega\varepsilon\Phi(\omega,\tau)R_0 x(\tau) - \Phi(\omega,\tau)j\omega\varepsilon f_0 x(\tau) + \mu_1\Phi(\omega,\tau)j\omega\varepsilon f_1 + \\ \qquad + \mu_2\Phi(\omega,\tau)j\omega\varepsilon f_2 + O(\varepsilon^2), \\ \Phi(\omega,\tau)j\varepsilon\omega R_1 a(x,\tau) = \lambda j\omega\varepsilon\Phi(\omega,\tau)R_1 - (\mu_1+\gamma_2)\Phi(\omega,\tau)j\omega\varepsilon f_1 + \\ \qquad + \lambda\Phi(\omega,\tau)j\omega\varepsilon f_0 - j\varepsilon\dfrac{\partial\Phi(\omega,\tau)}{\partial\omega}R_0 + \Phi(\omega,\tau)j\omega\varepsilon f_0 x(\tau) + O(\varepsilon^2), \\ \Phi(\omega,\tau)j\varepsilon\omega R_2 a(x,\tau) = \lambda j\omega\varepsilon\Phi(\omega,\tau)R_2 - \mu_2\Phi(\omega,\tau)j\omega\varepsilon f_2 + \\ \qquad + \gamma_1\Phi(\omega,\tau)j\omega\varepsilon f_0 + \gamma_2 j\omega\varepsilon\Phi(\omega,\tau)R_1 + \\ \qquad + \gamma_2 j\omega\varepsilon f_1 + O(\varepsilon^2). \end{cases} \tag{15}$$

Dividing the equations of the system (15) by $j\omega\varepsilon\Phi(\omega,\tau)$ at $\varepsilon\to 0$, we obtain

$$\begin{cases} R_0 a(x,\tau) = -\big(\lambda+\gamma_1+x(\tau)\big)f_0 + \dfrac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}R_0 - R_0 x(\tau) + \\ \qquad + \mu_1 f_1 + \mu_2 f_2, \\ R_1 a(x,\tau) = \lambda R_1 - (\mu_1+\gamma_2)f_1 + \big(\lambda+x(\tau)\big)f_0 - \dfrac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}R_0, \\ R_2 a(x,\tau) = \lambda R_2 - \mu_2 f_2 + \gamma_1 f_0 + \gamma_2 R_1 + \gamma_2 f_1. \end{cases} \tag{16}$$

The inhomogeneous system (16) corresponds to the homogeneous system (8), therefore we will seek a solution to the system (16) in the form:

$$f_k = CR_k + g_k - \varphi_k\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}, \quad k=\{0,1,2\}. \tag{17}$$

Substituting the equation (17) into the system (16), we obtain systems with respect to $\varphi_k$ and $g_k$:

$$\begin{cases} -\big(\lambda+\gamma_1+x(\tau)\big)\varphi_0 + \mu_1\varphi_1 + \mu_2\varphi_2 = R_0, \\ \big(\lambda+x(\tau)\big)\varphi_0 - (\mu_1+\gamma_2)\varphi_1 = -R_0, \\ \gamma_1\varphi_0 + \gamma_2\varphi_1 - \mu_2\varphi_2 = 0. \end{cases} \tag{18}$$

$$\begin{cases} -\big(\lambda+\gamma_1+x(\tau)\big)g_0+\mu_1 g_1+\mu_2 g_2=\big(a(x,\tau)+x(\tau)\big)R_0,\\ \big(\lambda+x(\tau)\big)g_0-(\mu_1+\gamma_2)g_1=\big(a(x,\tau)-\lambda\big)R_1,\\ \gamma_1 g_0+\gamma_2 g_1-\mu_2 g_2=\big(a(x,\tau)-\lambda\big)R_2-\gamma_2 R_1. \end{cases} \tag{19}$$

If we differentiate the equations of the system (8) by $x$, then the resulting equations are identical to the equations of the system (16), from which we can conclude that in the system (18) the following equalities are satisfied:

$$\varphi_k=\varphi_k(x,\tau)=\frac{\partial R_k(x,\tau)}{\partial x},\quad \varphi_0+\varphi_1+\varphi_2=0.$$

Let us consider the system (19), which has an infinite number of solutions, since the determinant of the system matrix is equal to zero, and the rank of the system matrix coincides with the rank of the extended matrix of the system.

To find a solution to the system, we add an additional condition $g_0+g_1+g_2=0$ to the system (19) and obtain:

$$g_0=\frac{(-\mu_1-\gamma_2)\big(a(x)+x(\tau)\big)R_0+R_1(\mu_1-\mu_2)\big(\lambda-a(x)\big)}{\big(\lambda+x(\tau)+\gamma_1+\mu_2\big)\gamma_2+(\gamma_1+\mu_2)\mu_1+\mu_2\big(\lambda+x(\tau)\big)},$$

$$g_1=\frac{\big(\lambda+x(\tau)+\gamma_1+\mu_2\big)\big(\lambda-a(x)\big)R_1-R_0\big(\lambda+x(\tau)\big)\big(a(x)+x(\tau)\big)}{\big(\lambda+x(\tau)+\gamma_1+\mu_2\big)\gamma_2+(\gamma_1+\mu_2)\mu_1+\mu_2\big(\lambda+x(\tau)\big)},$$

$$g_2=\frac{\big(\gamma_2+\lambda+\mu_1+x(\tau)\big)\big(a(x)+x(\tau)\big)R_0+R_1\big(\lambda+x(\tau)+\gamma_1+\mu_1\big)\big(a(x)-\lambda\big)}{\big(\lambda+x(\tau)+\gamma_1+\mu_2\big)\gamma_2+(\gamma_1+\mu_2)\mu_1+\mu_2\big(\lambda+x(\tau)\big)}.$$

Let's return to the equation (12). In this equation we group terms of order of smallness not higher than $\varepsilon^2$.

$$\begin{aligned} \varepsilon^2\frac{\partial F^{(2)}(\omega,\tau,\varepsilon)}{\partial\tau}+F^{(2)}(\omega,\tau,\varepsilon)j\varepsilon\omega a(x,\tau)=\Big(j\omega\varepsilon+\frac{(j\omega\varepsilon)^2}{2}\Big)\times\\ \times\Big(F_1^{(2)}(\omega,\tau,\varepsilon)(\lambda+\gamma_2)+F_2^{(2)}(\omega,\tau,\varepsilon)\lambda+j\varepsilon\frac{\partial F_0^{(2)}(\omega,\tau,\varepsilon)}{\partial\omega}-\\ -F_0^{(2)}(\omega,\tau,\varepsilon)x(\tau)\Big). \end{aligned}$$

$$\begin{aligned} \varepsilon^2\frac{\partial F^{(2)}(\omega,\tau,\varepsilon)}{\partial\tau}+F^{(2)}(\omega,\tau,\varepsilon)j\varepsilon\omega a(x,\tau)=j\omega\varepsilon F_1^{(2)}(\omega,\tau,\varepsilon)(\lambda+\gamma_2)+\\ +j\omega\varepsilon F_2^{(2)}(\omega,\tau,\varepsilon)\lambda+j^2\omega\varepsilon^2\frac{\partial F_0^{(2)}(\omega,\tau,\varepsilon)}{\partial\omega}-j\omega\varepsilon F_0^{(2)}(\omega,\tau,\varepsilon)x(\tau)+\\ +\frac{(j\omega\varepsilon)^2}{2}F_1^{(2)}(\omega,\tau,\varepsilon)(\lambda+\gamma_2)+\frac{(j\omega\varepsilon)^2}{2}F_2^{(2)}(\omega,\tau,\varepsilon)\lambda-\\ -\frac{(j\omega\varepsilon)^2}{2}F_0^{(2)}(\omega,\tau,\varepsilon)x(\tau)+O(\varepsilon^3). \end{aligned} \tag{20}$$

Substituting expansion (14) into the equation (20), we obtain

$$\begin{aligned} &\varepsilon^2\frac{\partial\Phi(\omega,\tau,\varepsilon)}{\partial\tau}+j\varepsilon\omega a(x,\tau)\Phi(\omega,\tau)\big(1+j\omega\varepsilon(f_0+f_1+f_2)\big)=\\ &=j\omega\varepsilon\Big((\lambda+\gamma_2)R_1+\lambda R_2-x(\tau)R_0+j\varepsilon\frac{\partial\Phi(\omega,\tau,\varepsilon)}{\partial\omega}R_0\Big)\Phi(\omega,\tau)+\\ &\quad+\frac{(j\omega\varepsilon)^2}{2}((\lambda+\gamma_2)R_1+\lambda R_2-x(\tau)R_0)\Phi(\omega,\tau)+\\ &\quad+(j\omega\varepsilon)^2((\lambda+\gamma_2)f_1+\lambda f_2-x(\tau)f_0)\Phi(\omega,\tau)+O(\varepsilon^3). \end{aligned} \tag{21}$$

Taking into account that $a(x) = -x(\tau)R_0 + (\lambda + \gamma_2)R_1 + \lambda R_2$, we eliminate the terms of the order of smallness $\varepsilon$ in the equation (21). Then we reduce by $\varepsilon^2$ and perform the transition to the limit at $\varepsilon \to 0$.

$$\begin{aligned}&\frac{\partial\Phi(\omega,\tau)}{\partial\tau} + (j\omega)^2 a(x,\tau)\Phi(\omega,\tau)(f_0 + f_1 + f_2) = \\ &= (j\omega)^2\left(\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega}R_0 + (\lambda+\gamma_2)f_1 + \lambda f_2 - x(\tau)f_0\right)\Phi(\omega,\tau) + \\ &\quad + \frac{(j\omega\varepsilon)^2}{2}a(x,\tau)\Phi(\omega,\tau).\end{aligned} \tag{22}$$

Substituting the equation (17) into the equation (22), we obtain the following equation:

$$\begin{aligned}&\frac{\partial\Phi(\omega,\tau)}{\partial\tau} + (j\omega)^2 a(x,\tau)\Phi(\omega,\tau) = (j\omega)^2\Phi(\omega,\tau)\left((\lambda+\gamma_2)\left(CR_1 + g_1 - \right.\right. \\ &\left. - \varphi_1\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}\right) + \lambda\left(CR_2 + g_2 - \varphi_2\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}\right) - x(\tau)\left(CR_0 + \right. \\ &\left.\left. + g_0 - \varphi_0\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}\right) + \frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}R_0\right) + \frac{(j\omega\varepsilon)^2}{2}a(x,\tau)\Phi(\omega,\tau).\end{aligned} \tag{23}$$

In the system (23), terms containing $C$ are destroyed, then we obtain:

$$\begin{aligned}\frac{\partial\Phi(\omega,\tau)}{\partial\tau} =&(j\omega)^2\Phi(\omega,\tau)\left((\lambda+\gamma_2)\left(g_1 - \varphi_1\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}\right) + \right. \\ &+ \lambda\left(g_2 - \varphi_2\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}\right) - x(\tau)\left(g_0 - \varphi_0\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}\right) + \\ &\left. + \frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}R_0\right) + \frac{(j\omega\varepsilon)^2}{2}a(x,\tau)\Phi(\omega,\tau).\end{aligned} \tag{24}$$

Let us rewrite the equation (24) by collecting identical terms.

$$\begin{aligned}\frac{\partial\Phi(\omega,\tau)}{\partial\tau} =& -(j\omega)^2\Phi(\omega,\tau)\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\omega\Phi(\omega,\tau)}\left((\lambda+\gamma_2)\varphi_1 + \lambda\varphi_2 - \right. \\ &\left. - x(\tau)\varphi_0 - R_0\right) + (j\omega)^2\Phi(\omega,\tau)\left((\lambda+\gamma_2)g_1 + \lambda g_2 - x(\tau)g_0\right) + \\ &+ \frac{(j\omega\varepsilon)^2}{2}a(x,\tau)\Phi(\omega,\tau).\end{aligned} \tag{25}$$

Let us pay attention to the multiplier in the first term of the equation (25), then we get:

$$\begin{aligned}-x(\tau)\varphi_0 + (\lambda+\gamma_2)\varphi_1 + \lambda\varphi_2 - R_0 =& -x(\tau)\frac{\partial R_0(x)}{\partial x} + (\lambda+\gamma_2)\frac{\partial R_1(x)}{\partial x} + \\ &+ \lambda\frac{\partial R_2(x)}{\partial x} - R_0(x) = a'(x).\end{aligned}$$

Then the equation (25) will take the form:

$$\begin{aligned}\frac{\partial\Phi(\omega,\tau)}{\partial\tau} =&\omega\frac{\partial\Phi(\omega,\tau)/\partial\omega}{\Phi(\omega,\tau)}a'(x) + \frac{(j\omega\varepsilon)^2}{2}\Phi(\omega,\tau)(a(x,\tau) + 2((\lambda+\gamma_2)g_1 + \\ &+ \lambda g_2 - x(\tau)g_0)).\end{aligned}$$

Let us denote

$$b(x) = a(x,\tau) + 2\big((\lambda+\gamma_2)g_1(x) + \lambda g_2(x) - x(\tau)g_0(x)\big),$$

we obtain:

$$\frac{\partial \Phi(\omega,\tau)}{\partial \tau} = a'(x)\omega \frac{\partial \Phi(\omega,\tau)/\partial \omega}{\Phi(\omega,\tau)} + b(x)\frac{(j\omega\varepsilon)^2}{2}\Phi(\omega,\tau). \tag{26}$$

In this equation, function $b(x)$ is the diffusion coefficient of the diffusion process for which the transfer coefficient is function $a(x)$.

## 6. Construction of diffuse approximation

Next, applying the inverse Fourier transform to the equation (26), we move on to the equation for the probability density.

Taking into account the following ratios:

$$\begin{cases} \omega \dfrac{\partial \Phi(\omega,\tau)}{\partial \omega} = -\displaystyle\int_{-\infty}^{\infty} e^{j\omega y}\big(yP(y,\tau)\big)'dy, \\ \dfrac{(j\omega\varepsilon)^2}{2}\Phi(\omega,\tau) = \displaystyle\int_{-\infty}^{\infty} e^{j\omega y}\dfrac{\partial^2 P(y,\tau)}{\partial y^2}dy, \end{cases}$$

we obtain an equation that is the Fokker–Planck equation for the probability density of some diffusion process $y(\tau)$ with the transfer coefficient $a'(x)$ and the diffusion coefficient $b(x)$.

Thus, the process $y(\tau)$ is a solution to the stochastic differential equation

$$dy(z) = a'(x)y(\tau)d\tau + \sqrt{b(x)}d\omega(\tau),$$

where $\omega(\tau)$ is the Wiener process.

Let us introduce the diffusion process

$$z(\tau) = x(\tau) + \varepsilon y(\tau),$$

where the function $x(\tau)$ is a solution to the ordinary differential equation

$$dx(\tau) = a(x)d\tau.$$

Then the diffusion process $z(\tau)$ is a solution to the stochastic differential equation:

$$dz(\tau) = (a(x) + \varepsilon a'(x)y(\tau))d\tau + \varepsilon\sqrt{b(x)}d\omega(\tau).$$

Let us write the terms on the right side of the equation

$$a(x) + \varepsilon a'(x)y(\tau) = a(x+\varepsilon y) + O(\varepsilon^2) = a(z) + O(\varepsilon^2),$$

$$\varepsilon\sqrt{b(x)} = \varepsilon\sqrt{b(x+\varepsilon y - \varepsilon y)} = \varepsilon\sqrt{b(z-\varepsilon y)} = \varepsilon\sqrt{b(z)} + O(\varepsilon^2).$$

We will assume that terms of order of smallness greater than $e$ do not make a significant contribution to the solution, which means we can neglect them. Then we get an equation of the form:

$$dz(\tau) = a(z)d\tau + \varepsilon\sqrt{b(z)}d\omega(\tau).$$

Let us denote the probability density of the diffusion process $z(\tau)$:

$$\Pi(z,\tau) = \frac{\partial\big(z(\tau) < z\big)}{\partial z}.$$

Let us write the Fokker–Planck equation for the diffusion process $z(\tau)$:

$$\frac{\partial\Pi(z,\tau)}{\partial\tau} = -\frac{\partial\big(a(z)\Pi(z,\tau)\big)}{\partial z} + \frac{\varepsilon^2}{2}\frac{\partial^2\big(b(z)\Pi(z,\tau)\big)}{\partial^2 z}. \tag{27}$$

In the equation (27) we make the reverse substitution $\sigma = \varepsilon^2$ and move on to the equation for the stationary probability distribution of the diffusion process $z(\tau)$:

$$\begin{aligned} -\big(a(z)\Pi(z)\big)' + \frac{\sigma}{2}\big(b(z)\Pi(z)\big)'' = 0,\\ \big(b(z)\Pi(z)\big)' = \frac{2}{\sigma}a(z)\Pi(z). \end{aligned} \tag{28}$$

To solve the equation (28), we introduce the substitution $G'(z) = b(z)\Pi(z)$, then we get:

$$G'(z) = \frac{2}{\sigma}\frac{a(z)}{b(z)}G(z), \tag{29}$$

where $a(z)$, $b(z)$ are the transfer and diffusion coefficients.

In the equation (29) we make the reverse substitution $\frac{G(z)}{b(z)} = \Pi(z)$, then the stationary probability density of the approximating random process has the form:

$$\Pi(z) = \frac{C}{b(z)}\exp\left(\frac{2}{\sigma}\int\limits_0^z \frac{a(z)}{b(x)}dx\right),$$

where $C$ is a normalizing constant.

Let's construct a diffusion approximation using the formula:

$$PD(i) = \frac{\Pi(i\sigma)}{\sum\limits_{n=0}^{N}\Pi(n\sigma)}.$$

## 7. Results and discussion

We consider a system with parameters: $\lambda = 1$, $\mu_2 = 3$, $\gamma_1 = 0.1$, $\gamma_2 = 0.1$ and different system occupancy parameters $\rho = \frac{\lambda}{\mu_1}$.

Let us determine the accuracy of the approximation using the Kolmogorov distance

$$\begin{aligned} \Delta_1 = \max_{0\le i\le N}\left|\sum_{i=0}^{N} P\_\text{matrix}(i) - \sum_{i=0}^{N} P\_\text{diffusion}(i)\right|,\\ \Delta_2 = \max_{0\le i\le N}\left|\sum_{i=0}^{N} P\_\text{matrix}(i) - \sum_{i=0}^{N} P\_\text{asimpt}(i)\right|, \end{aligned}$$

where $P\_\text{matrix}(i)$ is the distribution obtained by the matrix method, $P\_\text{diffusion}(i)$ is the distribution obtained by the asymptotic-diffusion method and $P\_\text{asimpt}(i)$ is the distribution obtained by the asymptotic method.

As a condition for the applicability of the asymptotic-diffusion method, we take the threshold value of the Kolmogorov distance $\Delta = 0.05$.

Table 1 shows the accuracy between the matrix and asymptotic-diffusion distributions for various parameters and different system loads.

Similarly, Table 2 shows the accuracy between the matrix and asymptotic distributions.

Table 1

**Kolmogorov distance**

| $\Delta_1$ | $\sigma = 2$ | $\sigma = 1$ | $\sigma = 0.5$ | $\sigma = 0.2$ | $\sigma = 0.1$ | $\sigma = 0.05$ |
|---|---|---|---|---|---|---|
| $\rho = 0.6$ | 0.086 | 0.088 | 0.064 | 0.063 | 0.061 | 0.060 |
| $\rho = 0.7$ | 0.062 | **0.049** | **0.047** | **0.045** | **0.043** | **0.041** |
| $\rho = 0.8$ | **0.035** | **0.031** | **0.032** | **0.028** | **0.026** | **0.020** |
| $\rho = 0.9$ | **0.012** | **0.0097** | **0.019** | **0.014** | **0.0085** | **0.0013** |

Table 2

**Kolmogorov distance**

| $\Delta_2$ | $\sigma = 2$ | $\sigma = 1$ | $\sigma = 0.5$ | $\sigma = 0.2$ | $\sigma = 0.1$ | $\sigma = 0.05$ |
|---|---|---|---|---|---|---|
| $\rho = 0.6$ | 0.160 | 0.124 | 0.100 | 0.076 | 0.053 | **0.049** |
| $\rho = 0.7$ | 0.295 | 0.250 | 0.219 | 0.186 | 0.168 | 0.134 |
| $\rho = 0.8$ | 0.439 | 0.391 | 0.348 | 0.304 | 0.223 | 0.267 |
| $\rho = 0.9$ | 0.442 | 0.381 | 0.327 | 0.193 | 0.255 | 0.184 |

According to the data in Table 1 and Table 2, we can conclude that the accuracy of the diffusion approximation increases as the system load factor decreases. The method is applicable when $\rho = 0.8$ for all parameters $\sigma$. The accuracy of the diffusion approximation exceeds the accuracy of the Gaussian approximation.

## 8. Conclusions

In this work, a study of the M/M/1 RQ system with an unreliable server was carried out using the method of asymptotically diffusion analysis. The stationary distribution of server states, the transfer coefficient and the diffusion coefficient are found. A diffusion approximation is constructed. The accuracy of the approximation is determined using the Kolmogorov distance between the distributions constructed by the asymptotic diffusion method and the matrix method. It was proved that the asymptotically diffusion analysis method is more accurate. It is shown that the accuracy of the diffusion approximation exceeds the accuracy of the Gaussian approximation.

**Author Contributions:** Conceptualization, N.Voronina and S.Rozhkova; methodology, N.Voronina and S.Rozhkova; software, N.Voronina; validation, N.Voronina and S.Rozhkova; formal analysis, N.Voronina; investigation, N.Voronina and S.Rozhkova; resources, N.Voronina; data curation, N.Voronina; writing—original draft preparation, N.Voronina; writing—review and editing, N.Voronina and S.Rozhkova. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data sharing is not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Gross, D., Shortle, J. F., Thompson, J. M. & Harris, C. M. *Fundamentals of queueing theory* (John wiley & sons, 2011).
2. G.Gosztony. Repeated call attempts and their effect on trafic engineering. *Budavox Telecommunication Review* **2,** 16–26 (1976).
3. Cohen, J. Basic problems of telephone traffic theory and the influence of repeated calls. *Philips Telecommunication Rev.* **18,** 49 (1957).
4. Hanczewski, S., Stasiak, M., Weissenberg, J. & Zwierzykowski, P. *Queuing model of the access system in the packet network* in *Computer Networks* (2016), 283–293.
5. Falin, G. I. & Templeton, J. G. C. *Retrial Queues* 320 pp. doi:10.1201/9780203740767 (Chapman & Hall, London, 1997).
6. Artalejo, J. R. & Gómez-Corral, A. *Retrial Queueing Systems* doi:10.1007/978-3-540-78725-9 (Springer Berlin Heidelberg, 2008).
7. Dimitriou, I. A queueing model with two classes of retrial customers and paired services. *Annals of Operations Research* **238,** 123–143. doi:10.1007/s10479-015-2059-2 (2016).
8. Wang, J., Cao, J. & Li, Q. Reliability analysis of the retrial queue with server breakdowns and repairs. *Queueing Systems* **38,** 363–380 (2001).
9. Kumar, M. S. & Arumuganathan, R. An MX/G/1 retrial queue with two-phase service subject to active server breakdowns and two types of repair. *International Journal of Operational Research* **8,** 261–291 (2010).
10. Kim, C., Klimenok, V. I. & Orlovsky, D. S. The BMAP/PH/N retrial queue with Markovian flow of breakdowns. *European Journal of Operational Research* **189,** 1057–1072 (2008).
11. Lakaour, L., Aissani, D., Adel-Aissanou, K., Barkaoui, K. & Ziani, S. An unreliable single server retrial queue with collisions and transmission errors. *Communications in Statistics-Theory and Methods* **51,** 1085–1109 (2022).
12. Danilyuk, E. Y., Janos, S., *et al.* Asymptotic analysis of retrial queueing system M/M/1 with impatient customers, collisions and unreliable server. *Journal of Siberian Federal University. Mathematics & Physics* **13,** 218–230. doi:10.17516/1997-1397-2020-13-2-218-230 (2020).
13. Tóth, Á. & Sztrik, J. *Simulation of Finite-Source Retrial Queuing Systems With Collisions, Non-Reliable Server and Impatient Customers in the Orbit.* in *ICAI* (2020), 408–419.
14. Kuki, A., Bérczes, T., Sztrik, J. & Kvach, A. Numerical Analysis of Retrial Queueing Systems with Conflict of Customers and an Unreliable Server. *Journal of Mathematical Sciences* **237,** 673–683. doi:10.1007/s10958-019-04193-1 (2019).
15. Nazarov, A. A., Paul, S. V. & Lizyura, O. D. Two-way communication retrial queue with unreliable server and multiple types of outgoing calls. *Discrete and Continuous Models and Applied Computational Science* **28,** 49–61. doi:10.22363/2658-4670-2020-28-1-49-61 (2020).

16. Dudin, A., Dudina, O., Dudin, S. & Samouylov, K. Analysis of Single-Server Multi-Class Queue with Unreliable Service, Batch Correlated Arrivals, Customers Impatience, and Dynamical Change of Priorities. *Mathematics* **9,** 1257. doi:10.3390/math9111257 (2021).
17. Chakravarthy, S. R., OZKAR, S. & SHRUTI, S. Analysis of M/M/C retrial queue with thresholds, PH distribution of retrial times and unreliable servers. *Journal of applied mathematics & informatics* **39,** 173–196 (2021).
18. Falin, G. An M/G/1 retrial queue with an unreliable server and general repair times. *Performance Evaluation* **67,** 569–582 (2010).
19. Dudin, A., Klimenok, V. & Vishnevsky, V. *Analysis of unreliable single server queueing system with hot back-up server* in *Optimization in the Natural Sciences: 30th Euro Mini-Conference, EmC-ONS 2014, Aveiro, Portugal, February 5-9, 2014. Revised Selected Papers 30* (2015), 149–161.
20. Nazarov, A. A., Paul, S. V., Lizyura, O. D., *et al.* Two-way communication retrial queue with unreliable server and multiple types of outgoing calls. *Discrete and Continuous Models and Applied Computational Science.* doi:10.22363/2658-4670-2020-28-1-49-61 (2020).
21. Voronina, N. M., Fedorova, E. A. & Rozhkova, S. V. Asymptotic analysis of the RQ-system M/M/1 with an unreliable server. *Mathematical and software for information technical and economic systems,* 304–309. doi:10.1007/978-3-031-09331-9_28 (2020).

## Information about the authors

**Nataliya M. Voronina**—Senior Lecturer of Department of Information Technology of School of Information Technology and Robotics Engineering, National Research Tomsk Polytechnic University (e-mail: vnm@tpu.ru, phone: +7 (3822) 701-777-11-54, ORCID: 0000-0001-9044-5211, ResearcherID: AAD-2035-2019, Scopus Author ID: 57802914700)

**Svetlana V. Rozhkova**—Doctor of Physics and Mathematics Sciences, Professor of Department of Mathematics and Mathematical Physics of School of Nuclear Technology Engineering, National Research Tomsk Polytechnic University, Professor of Department of Probability Theory and Mathematical Statistics Institute of Applied Mathematics and Computer Science, National Research Tomsk State University (TSU) (e-mail: rozhkova@tpu.ru, phone: +7 (3822) 701-777-68-98, ORCID: 0000-0002-8888-9291, ResearcherID: F-5512-2017, Scopus Author ID: 6603581666)

УДК 519.872:519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-4-380-394 EDN: DDNAPR

# Асимптотически диффузионный анализ RQ системы с ненадёжным прибором

Н.М. Воронина[1], С.В. Рожкова[1, 2]

[1] Национальный исследовательский Томский политехнический университет, пр. Ленина, д. 30, Томск, 634050, Российская Федерация

[2] Национальный исследовательский Томский государственный университет, пр. Ленина, д. 36, Томск, 634050, Российская Федерация

**Аннотация.** В работе рассматривается однолинейная RQ-система массового обслуживания с ненадёжным прибором. Системы массового обслуживания называются ненадёжными, если их приборы могут время от времени выходить из строя и требовать восстановления (ремонта), только после которого они могут возобновить обслуживание запросов. Исследование проводится методом асимптотически диффузионного анализа в условии большой задержки заявок на орбите. Найдены стационарное распределение состояний прибора, коэффициент переноса и коэффициент диффузии. Построена диффузионная аппроксимация. Доказано, что точность диффузионной аппроксимации превышает точность гауссовской аппроксимации.

**Ключевые слова:** RQ система, асимптотически диффузионный анализ, ненадежный прибор

**Discrete & Continuous Models & Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 395–405**
http://journals.rudn.ru/miph

*Research article*

UDC 519.872:519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-4-395-405 EDN: EQOGQJ

# On the problem of normal modes of a waveguide

Oleg K. Kroytor[1], Mikhail D. Malykh[1,2], Leonid A. Sevastianov[1,2]

[1] RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

[2] Joint Institute for Nuclear Research, 6 Joliot-Curie St, Dubna, 141980, Russian Federation

(received: October 7, 2024; revised: October 20, 2024; accepted: October 22, 2024)

**Abstract.** Various approaches to calculating normal modes of a closed waveguide are considered. A review of the literature was given, a comparison of the two formulations of this problem was made. It is shown that using a self-adjoint formulation of the problem of normal waveguide modes eliminates the occurrence of artifacts associated with the appearance of a small imaginary additive to the eigenvalues. The implementation of this approach for a rectangular waveguide with rectangular inserts in the Sage computer algebra system is presented and tested on hybrid modes of layered waveguides. The tests showed that our program copes well with calculating the points of the dispersion curve corresponding to the hybrid modes of the waveguide.

**Key words and phrases:** polarized electromagnetic radiation, normal modes of a waveguide, spectral problem of waveguide theory, dispersion curve of a waveguide

**For citation:** Kroytor, O. K., Malykh, M. D., Sevastianov, L. A. On the problem of normal modes of a waveguide. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 395–405. doi: 10.22363/2658-4670-2024-32-4-395-405. edn: EQOGQJ (2024).

## 1. Introduction

In classical electrodynamics there are two related spectral problems, the problem of normal modes of a waveguide and the problem of eigenmodes of a resonator [1, 2]. The first of these problems in the vector case turned out to be surprisingly difficult, its solution requiring the use of very subtle theorems from the field of functional analysis.

## 2. Scalar model

Let $S$ be regular domain in $\mathbb{R}^2$, the cylinder $S \times \mathbb{R}$ will be called a waveguide, and the $Oz$ axis of the Cartesian coordinate system used is directed along the axis of the cylinder. A nontrivial solution

$$u = u(x, y)e^{i\omega t - i\gamma z}$$

of the oscillation equation

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = \Delta u$$

in the cylinder $S \times \mathbb{R}$, satisfying the Dirichlet boundary condition

$$u|_{\partial S \times \mathbb{R}} = 0$$

or Neumann boundary condition

$$\frac{\partial u}{\partial n}\Big|_{\partial S \times \mathbb{R}} = 0$$

is called the normal mode of the scalar waveguide, and the corresponding value of the positive parameter $\omega$ is called the natural frequency. The parameter $\gamma/c$ is called the propagation constant. If $\gamma > 0$, then the normal mode runs along the $z$ axis, if $\gamma < 0$, then against it. These modes are called guided modes. If $\gamma$ contains an imaginary part, then the normal modes either increase exponentially or decrease exponentially with increasing $z$. Such modes are called evanescent.

The problem of finding normal modes of a scalar waveguide is reduced to a 2D spectral problem of finding a nontrivial solution $u$ of the equation

$$\Delta_2 u + (k^2 - \gamma^2)u = 0$$

with Dirichlet or Neumann boundary conditions.

Let the eigenvalues of the problem

$$\Delta_2 \phi + \alpha^2 \phi = 0, \quad \phi|_{\partial S} = 0$$

be numbered in ascending order taking into account the multiplicity as $\alpha_1^2, \alpha_2^2, \dots$, and the corresponding eigenfunctions be denoted as $\phi_n$. In this case, functions

$$\phi_n(x, y)e^{i\alpha_n t}$$

describe the natural oscillations of the membrane $S$. For a given domain $S$, the numbers $\alpha_1^2, \alpha_2^2, \dots$ are uniquely determined. For some domains, they can be calculated analytically, for all others they are found using the Galerkin method.

For this reason, the parameters $\omega$ and $\gamma$ of the normal modes of a scalar waveguide are related as

$$k^2 - \gamma^2 = \alpha_n^2.$$

Therefore, for a fixed frequency $\omega$, there are at most a finite number of positive values of the parameter $\gamma$, for which normal modes exist. These modes describe waves traveling along the waveguide and, as already said, are called guided modes. All other normal modes have an imaginary $\gamma$ and for this reason they exponentially increase or decrease along the waveguide axis.

By Steklov's theorem, a monochromatic scalar field in a waveguide can always be represented as a sum

$$\sum \left(a_n e^{i\omega t - i\gamma_n z} + b_n e^{i\omega t + i\gamma_n z}\right) \phi_n(x, y),$$

where $a_n$, $b_n$ are the complex amplitudes. Therefore, the field, say, at large $z$ is a superposition of a finite number of running normal modes, a sum of exponentially decreasing evanescent modes, and a sum of exponentially growing modes. The partial radiation conditions are such that there should be no exponentially growing terms [1].

It should also be noted that normal modes in the framework of the scalar model with Dirichlet conditions exist only for those values of $k$, $\gamma$ that lie on hyperboles

$$k^2 - \gamma^2 = \alpha_n^2, \quad n = 1, 2, \dots .$$

The set of such points $(k, \gamma)$, which correspond to normal modes, is called the dispersion curve of the waveguide. The dispersion curve of a scalar waveguide consists of a countable number of hyperbolas.

The case of Neumann boundary conditions does not present any fundamental difficulties. Let us agree that the eigenvalues of the problem

$$\Delta_2\psi + \beta^2\psi = 0, \quad \frac{\partial\psi}{\partial n}|_{\partial S} = 0$$

are numbered in ascending order and taking into account the multiplicity as $\beta_1^2, \beta_2^2, \dots$, and the corresponding eigenfunctions are denoted as $\psi_n$. Let us add to them the zero eigenvalue $\beta_0 = 0$ and the corresponding eigenfunction $\psi_0 = 1$. The system of functions $\psi_n$ is again complete, and normal modes within the framework of the scalar model with Neumann conditions exist only for those values of $k$, $\gamma$ that lie on the hyperbolas

$$k^2 - \gamma^2 = \beta_n^2, \quad n = 0, 1, 2, \dots .$$

Thus, in the scalar case, classical theorems of mathematical physics are sufficient to construct the theory of waveguides [1].

## 3. Vector model

Let us now turn to the vector model of an electromagnetic waveguide. A nontrivial field of the form

$$\vec{E} = \vec{E}(x, y)e^{i\omega t - i\gamma z}, \quad \vec{H} = \vec{H}(x, y)e^{i\omega t - i\gamma z},$$

satisfying the system of homogeneous Maxwell's equations and the boundary conditions

$$\vec{n} \times \vec{E} = 0, \quad \vec{n} \cdot \vec{H} = 0,$$

is called an eigenmode, and the corresponding value of the positive parameter $\omega$ is called an eigenfrequency. The parameter $\beta = \gamma/c$ is called the propagation constant. To find the eigenfrequencies, it is necessary to solve the eigenvalue problem

$$\overline{\mathrm{rot}}\vec{E} = -ik\mu\vec{H}, \quad \overline{\mathrm{rot}}\vec{H} = ik\epsilon\vec{E} \tag{1}$$

with the boundary conditions

$$\vec{n} \times \vec{E} = 0, \quad \vec{n} \cdot \vec{H} = 0.$$

Here $\overline{\mathrm{rot}}$ is a differential operator in which differentiation with respect to $z$ is replaced with multiplication by $-i\gamma$. As in the scalar case, the points of the $k\gamma$ plane at which this problem has a nontrivial solution form a certain curve called the dispersion curve of the waveguide.

In the case where the waveguide filling is uniform, Tikhonov A.N. and Samarskii A.A. [3] proved a field decomposition theorem, from which it follows that the complete system of waveguide modes can be composed of two types of modes: transverse magnetic (TM, $H_z = 0$) and transverse electric (TE, $E_z = 0$). For a TE mode, from the equation

$$\mathrm{div}\,\epsilon\vec{E} = 0$$

it follows that there is such a function $u$ that

$$E_x = \frac{\partial u}{\partial y}, \quad E_y = -\frac{\partial u}{\partial x}, \quad E_z = 0.$$

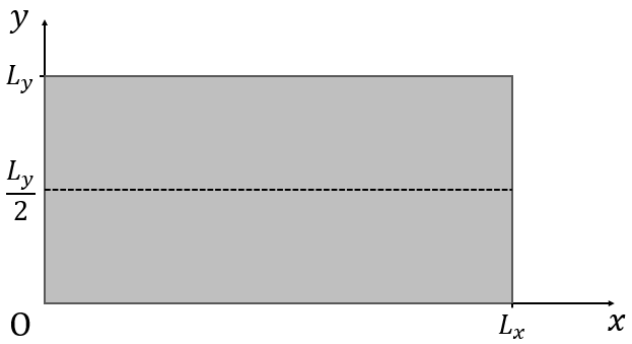


Figure 1. Waveguide filled with layers

This function is called the Borgnis function [1]. By direct substitution of these expressions into Maxwell's equations, it is possible to express all the components of the field through the derivatives of the Borgnis function, and for the Borgnis function itself to obtain a scalar eigenvalue problem with the Dirichlet condition. Similarly, the TM mode can be expressed through the derivatives of the Borgnis function, for which it is possible to obtain a scalar eigenvalue problem with the Neumann condition. It is only necessary to discard the zero eigenvalue, which will correspond to the trivial electromagnetic field.

Thus, the theory of Borgnis functions allows reducing the study of the modes of a waveguide filled with an optically homogeneous substance to the study of the spectrum of the Laplace operator. In this case, the dispersion curve turns out to be the union of a countable number of hyperbolas, which are dispersion curves for a scalar waveguide with the Dirichlet and Neumann conditions.

However, in practice, waveguides with optically inhomogeneous filling are quite common. Such waveguides include waveguides with a core, which are obtained by coating a dielectric cylinder with another dielectric and then with a conducting layer, and multicore waveguides, which are obtained by adding several dielectric cylinders into a bundle covered with a conducting layer on the outside. In this case, it is impossible to decompose the field into TE and TM modes.

## 4. Rectangular waveguide with two layers

As an example, consider a waveguide with rectangular cross-section $L_x \times L_y$ (see Fig. 1), the filling of which is piecewise constant and depends only on $y$. In other words, the waveguide consists of several layers, Fig. 1 shows two such layers of equal thickness $L_y/2$. When one of the layers is air, we say that the waveguide is considered half-filled.

There are two families of normal modes in such a waveguide, the SLE and SLH modes, the former have $E_y = 0$, and the latter have $H_y = 0$ [4]. The theory of these modes is in many ways similar to the theory of TM and TE modes developed by Tikhonov and Samarskii [5].

We will search for SLE modes using the method of separation of variables:

$$\vec{E} = \begin{pmatrix} A_x(y)\cos k_x x \\ 0 \\ A_z(y)\sin k_x x \end{pmatrix} e^{ik_z z - i\omega t} \tag{2}$$

and

$$\vec{H} = \begin{pmatrix} B_x(y)\sin k_x x \\ B_y(y)\cos k_x x \\ B_z(y)\cos k_x x \end{pmatrix} e^{ik_z z - i\omega t}.$$

Here the choice of sines and cosines is determined in such a way as to satisfy the boundary conditions.

First of all, let us consider what Maxwell's equations lead to in a layer, where $\epsilon$ and $\mu$ are constant. From Maxwell's equations we have:

$$\begin{cases} ik\mu B_x + \dfrac{dA_z}{dy} = 0, \\ ik\mu B_y + ik_z A_x - k_x A_z = 0, \\ -ik\mu B_z + \dfrac{dA_x}{dy} = 0, \\ -i\epsilon k A_x - ik_z B_y + \dfrac{dB_z}{dy} = 0, \\ ik_z B_x + k_x B_z = 0, \\ i\epsilon k A_z + k_x By + \dfrac{dB_x}{dy} = 0. \end{cases} \tag{3}$$

Three equations from this system allow us to express $\vec{B}$ in terms of $A_x$ and $A_z$ and their derivatives.

After this substitution, out of 6 equations, 3 non-trivial ones remain:

$$\begin{cases} -i\epsilon k A_x + i(k_z A_x + ik_x A_z)\dfrac{k_z}{k\mu} - i\dfrac{d^2 A_x}{dy^2}\dfrac{1}{k\mu} = 0, \\ -ik_x \dfrac{dA_x}{dy}\dfrac{1}{k\mu} - k_z \dfrac{dAz}{dy}\dfrac{1}{k\mu} = 0, \\ i\epsilon k A_z - (k_z A_x + ik_x A_z)\dfrac{k_x}{k\mu} + i\dfrac{d^2 A_z}{dy^2}\dfrac{1}{k\mu} = 0. \end{cases}$$

The second equation, up to an insignificant constant, allows finding a linear relationship between $A_x$ and $A_z$:

$$A_z = -\frac{ik_x}{k_z} A_x. \tag{4}$$

As a result, two equations that differ only by a constant factor turn out to be nontrivial. Therefore, in the layer, Maxwell's equations are reduced to the equation

$$\frac{d^2 A_x}{dy^2} + (\varepsilon\mu k^2 - k_x^2 - k_z^2)A_x = 0 \tag{5}$$

and equations that allow calculating $A_z$ and $\vec{B}$ from the known $A_x$.

At the waveguide boundary, the modes must satisfy the condition of a wall with ideal conductivity: $\vec{n} \times \vec{E} = 0$.

At the boundary $x = 0, L_x$, these conditions yield:

$$E_y = E_z = 0.$$

For SLE modes, the component $E_y$ is identically zero, so the condition $E_z = 0$ remains valid. We took the sine in (2) for $E_z$ so that this condition is always satisfied at $x = 0$. At $x = L_x$, we obtain the condition $\sin(k_x L_x) = 0$, from which suitable values of $k_x$ are determined as

$$k_x = \frac{\pi n}{L_x}, \quad n \in \mathbb{N}. \tag{6}$$

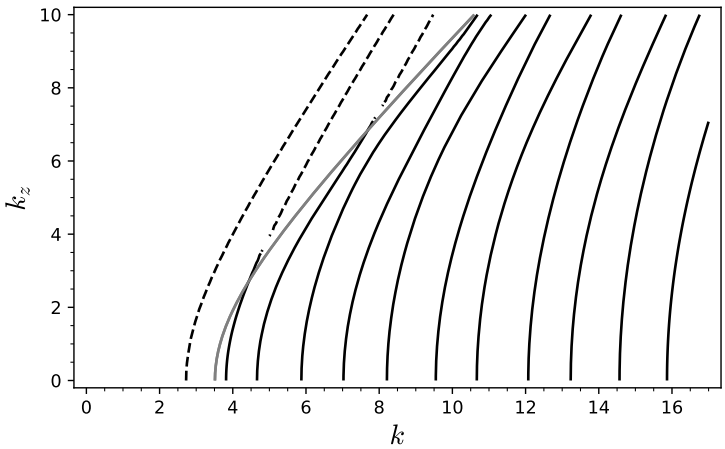


Figure 2. Dispersion curve for a test waveguide with two layers ($\epsilon_0 = 1$, $\epsilon_2 = 1$, $\mu = 1$, $L_x = 1$, $L_y = 2$).

At the boundary $y = 0, L_y$, the conditions of ideal conductivity yield:

$$E_x = E_z = 0.$$

Since $A_x$ and $A_y$ are linearly related in layers, this condition reduces to the Dirichlet condition on $A_x$:

$$A_x(0) = A_x(L_y) = 0.$$

At the boundary of two layers $y = M$, the requirement of continuity of $E_x, E_z$ and $H_x, H_z$ is to be satisfied.

The continuity of $E_x$ indicates the continuity of $A_x$. The coefficient in Eq. (4), relating $A_z$ and $A_x$, does not depend on the filling, so the continuity of $A_x$ implies the continuity of $A_z$, and, consequently, of $E_z$.

Maxwell's equations (3) yield

$$ik\mu B_x = -\frac{dA_z}{dy}, \quad ik\mu B_z = \frac{dA_x}{dy}.$$

From this it is clear that the continuity of $H_x, H_z$ is equivalent to the continuity of $\frac{1}{\mu}A'_x$.

As a test example, we consider a waveguide of rectangular cross-section $L_x \times L_y$ with two layers: for $y < L_y/2$ let $\epsilon = \epsilon_1$, and for $y > L_y/2$ let $\epsilon = \epsilon_0$ (see Fig. 1). We consider $\mu$ to have a constant value.

According to the discussed above, such a waveguide has a family of SLE modes (2), the parameters of which at a given frequency $\omega$ are determined as follows. The number $k_x$ is given by Eq. (6). The number $k_z$ is the eigenvalue of the problem:

$$\begin{cases} \dfrac{d^2A_x}{dy^2} + (\varepsilon\mu k^2 - k_x^2 - k_z^2)A_x = 0, \\ [A_x] = \left[\dfrac{dA_x}{dy}\right] = 0, \quad y = \dfrac{L_y}{2}, \\ A_x(0) = A_x(L_y) = 0. \end{cases} \tag{7}$$

This problem comprises Eq. (5), the boundary conditions and the matching conditions found above. From the eigenfunction $A_x$ we can calculate $A_z$ using Eq. (4), from them we determine $\vec{B}$, and thus determine all the quantities involved in Eq. (2).

For further tests, we are interested in the dependence of the eigenvalues $k_z$ of the problem (7) on $k = \omega/c$.

The solution of (7) for $y < L_y/2$ is

$$A_x = a \sin\left(\sqrt{\varepsilon_1 \mu k^2 - k_x^2 - k_z^2}\, y\right).$$

The solution of (7) for $y > L_y/2$ is

$$A_x = b \sin\left(\sqrt{\varepsilon_0 \mu k^2 - k_x^2 - k_z^2}\,(L_y - y)\right). \tag{8}$$

The choice of sines ensures that the boundary conditions are satisfied. For $y = L_y/2$ we have

$$a \sin\left(\sqrt{\varepsilon_1 \mu k^2 - k_x^2 - k_z^2}\, L_y/2\right) = b \sin\left(\sqrt{\varepsilon_0 \mu k^2 - k_x^2 - k_z^2}\, L_y/2\right)$$

and

$$a\sqrt{\varepsilon_1 \mu k^2 - k_x^2 - k_z^2} \cos\left(\sqrt{\varepsilon_1 \mu k^2 - k_x^2 - k_z^2}\, L_y/2\right) =$$
$$- b\sqrt{\varepsilon_0 \mu k^2 - k_x^2 - k_z^2} \cos\left(\sqrt{\varepsilon_0 \mu k^2 - k_x^2 - k_z^2}\, L_y/2\right).$$

Thus, $k_z$ is the root of the determinant of this homogeneous system of linear equations.

## 5. Results

The calculation result is shown in Fig. 2 as solid lines. Their unexpected discontinuity occurs because in Eq. (8) the sine transforms into the hyperbolic sine. We added the second piece of the program, in which the sine is replaced with the hyperbolic sine. The resulting continuation of the dispersion curves is shown in Fig. 2 with dotted line.

Thus, the conjugation method allows finding a family of normal modes of the test waveguide. These modes are neither TE nor TM modes, so they are often called hybrid modes. The considered example proves that hybrid modes exist. This circumstance makes it a very important test for all kinds of calculations of dispersion curves of waveguides, since it is the hybridization of modes that introduces non-self-conjugation into the known approaches to calculating modes.

## 6. Discussion

Without the decomposition theorem, the normal waveguide mode problem (1) does not decompose into two scalar problems and, thus, does not reduce to any type of problem studied above. It should be noted that the problem (1) contains two parameters, $k = \omega/c$ and $\gamma$, and we must choose one of them as the spectral parameter.

In the early 1990s, in the first works on calculating normal waveguide modes [6], the frequency was used as the spectral parameter. This resulted in a self-adjoint spectral problem with respect to $k$, which could be solved relatively successfully using the software that was available in the early 1990s. The key difficulty at that time was constructing a basis for the Galerkin method that satisfies the condition $\operatorname{div} \vec{H} = 0$.

This approach was soon abandoned in favor of an approach in which $\gamma$ is considered as a spectral parameter, and the frequency $\omega$ is considered given [7–11]. With this approach, the problem (1) is reduced to the study of the spectrum of a non-self-adjoint operator pencil that is quadratic with respect to the spectral parameter $\gamma$. By analogy with the scalar case, it is necessary to prove the completeness of the system of normal modes. However, the conditions under which Keldysh [12–15] proved the completeness of the root vectors of the quadratic operator pencil are not satisfied in all possible notations of the problem of waveguide normal modes. For the first time, the completeness of the system of root vectors of a waveguide with piecewise constant filling was substantiated in the papers by Yu.G. Smirnov [16–19], for an arbitrary filling in [9, 20–22]. This in turn made it possible to substantiate the formulation of partial radiation conditions. The basis property of the system of root vectors of a waveguide could be substantiated only for the axially symmetric case [23, 24].

Even greater difficulties are offered by the numerical calculation of normal modes. The application of the Galerkin method, as well as any other truncation method, leads to the study of a non-self-adjoint matrix pencil. Numerical methods for calculating its spectrum are very whimsical. In a number of works [8, 10, 11], algorithms built into, for example, MatLab were used according to the "black box" principle. In such computer experiments, the dispersion curve turned out to be non-monotonic, real eigenvalues suddenly went into the complex domain, etc. Generally speaking, all these phenomena are inherent in the spectral theory of non-self-adjoint matrices. However, the physical meaning of these phenomena raises many questions.

Over the past 30 years since the publication [6], the situation has changed radically. It seemed then that new methods for approximate calculation of the spectrum of non-self-adjoint matrices would appear in the near future, which would solve the difficulties noted above. However, instead, computer algebra methods came into use, which allow constructing Galerkin method bases that satisfy certain properties. This renewed the interest in the idea of Ref. [6]. The choice of frequency as a spectral parameter has a simple physical background. There is an obvious connection between the modes travelling along the waveguide axis and the standing modes of a cylindrical resonator. Having studied it, we obtained a method for constructing a dispersion curve, which requires solving the spectral problem in a cylindrical resonator, i.e., a classical self-adjoint problem. This approach was implemented as a program for the Sage computer algebra system and presented in Ref. [25].

To test this program, we considered a waveguide in which the insert occupies the lower half (see Fig. 1, $\epsilon_0 = 1$, $\epsilon_1 = 0.1$, $L_x = 1$, $L_y = 2$). It turned out that for the lower modes, the points found in our program lie on the analytical curve with graphical accuracy even with a very small number of basis elements taken into account (three for each direction).

## 7. Conclusion

Using a self-adjoint formulation of the problem of normal waveguide modes eliminates the occurrence of artifacts associated with the appearance of a small imaginary additive to the eigenvalues. We implemented this approach for a rectangular waveguide with rectangular inserts in the Sage computer algebra system. Tests on SLE modes of layered waveguides showed that our program copes well with calculating the points of the dispersion curve corresponding to the hybrid modes of the waveguide.

On the other hand, the approach based on a non-self-adjoint formulation gives important results from a theoretical point of view on the completeness of the system of normal modes and, therefore, allows us to justify the partial conditions of Sveshnikov radiation. At the moment, only a combination of two approaches allows us to bring our knowledge of the vector model of the waveguide closer to the well-studied scalar one.

**Author Contributions:** Conceptualization, L. A. Sevastianov; methodology, M. D. Malykh; software, O. K. Kroytor; validation, O. K. Kroytor; writing—review and editing, O. K. Kroytor; visualization, O. K. Kroytor; supervision, L. A. Sevastianov. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Russian Science Foundation (grant no. 20-11-20257).

**Data Availability Statement:** No new data were created or analysed during this study. Data sharing is not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mogilevskii, I. E. & Sveshnikov, A. G. *Mathematical problems of diffraction theory* Russian. In Russian (MSU, Moscow, 2010).
2. Karliner, M. M. *Microwave electrodynamics: Course of lectures* (NSU, Novosibirsk, 2006).
3. Samarskii, A. A. & Tikhonov, A. N. Russian. *To the theory of excitation of radiowaveguides* in *Selected works of A. A. Samarsky* In Russian. Chap. 1 (Maks Press, Moscow, 2003).
4. Chow, V. T. *Handbook of Applied Hydrology, McGraw-Hill, New York, 1964.* (1964).
5. Tikhonov, A. N. & Samarskii, A. A. *Equations of mathematical physics* Russian (Dover Publications, New York, 1990).
6. Bermfidez, A. & Pedreira, D. G. Mathematical analysis of a finite element method without spurious solutions for computation of dielectric waveguides. *Numer. Math.* **61,** 39–57 (1992).
7. Lezar, E. & Davidson, D. B. *Electromagnetic waveguide analysis* in *Automated solution of differential equations by the finite element method* 629–643 (The FEniCS Project, 2011).
8. Novoselov, N. A., Raevsky, S. B. & Titarenko, A. A. Calculation of symmetrical wave propagation characteristics of a circular waveguide with radially inhomogeneous dielectric filling. Russian. *Proceedings of the Nizhny Novgorod State Technical University named after R.E. Alekseev.* In Russian, 30–38 (2010).
9. Delitsyn, A. L. On the completeness of the system of eigenvectors of electromagnetic waveguides. *Comput. Math. and Math. Phys.* **51,** 1771–1776 (2011).
10. Delitsyn, A. L. & Kruglov, S. I. Mixed finite elements used to analyze the real and complex modes of cylindrical waveguides. Russian. *Moscow University Physics Bulletin* **66,** 546 (2011).
11. Delitsyn, A. L. & Kruglov, S. I. Application of mixed finite element method for calculation of modes of cylindrical waveguides with variable refractive index. *Journal of Radio Electronics.* In Russian, 1–28 (2012).
12. Keldysh, M. V. Russian. *On the completeness of the eigenfunctions of some classes of non-self-adjoint linear operators* in *Selected writings. Mathematics.* In Russian. Chap. 31 (Nauka, Moscow, 1985).
13. Gohberg, I. & Krein, M. *Introduction to the Theory of Linear Nonselfadjoint Operators* (American Mathematical Soc., Providence, Rhode Island, 1969).
14. Markus, A. S. *Introduction to the Spectral Theory of Polynomial Operator Pencils* (American Mathematical Society, Providence, R.I., 1988).
15. Kopachevsky, N. D. *Spectral Theory of Operator Pencils: Special Course of Lectures* (Forma, Simferopol', 2009).
16. Smirnov, Y. G. Completeness of the system of eigen- and associated waves of a partially filled waveguide with an irregular boundary. *Dokl. Math.* **32,** 963–964 (1987).
17. Smirnov, Y. G. The application of the operator pencil method in a problem concerning the natural waves of a partially filled wave guide. *Dokl. Math.* **35,** 430–431 (1990).
18. Smirnov, Y. G. The method of operator pencils in boundary value problems of conjugation for a system of elliptic equations. *Differ. Equ.* **27,** 112–118 (1991).

19. Shestopalov, Y. & Smirnov, Y. Eigenwaves in waveguides with dielectric inclusions: spectrum. *Applicable Analysis* **93,** 408–427. doi:10.1080/00036811.2013.778980 (2014).
20. Bogolyubov, A. N., Delitsyn, A. L. & Sveshnikov, A. G. On the completeness of the set of eigen- and associated functions of a waveguide. *Comput. Math. Math. Phys.* **38,** 1815–1823 (1998).
21. Bogolyubov, A. N., Delitsyn, A. L. & Sveshnikov, A. G. On the problem of excitation of a waveguide filled with an inhomogeneous medium. *Comput. Math. Math. Phys.* **39,** 1794–1813 (1999).
22. Delitsyn, A. L. An approach to the completeness of normal waves in a waveguide with magnetodielectric filling. *Differ. Equ.* **36,** 695–700 (2000).
23. Bogolyubov, A. N., Delitsyn, A. L., Malykh, M. D. & Sveshnikov, A. G. The basis property of root vectors for the radio waveguide. *Moscow University Physics Bulletin* **55,** 22 (2000).
24. Bogolyubov, A. N., Delitsyn, A. L. & Malykh, M. D. On the root vectors of a cylindrical waveguide. *Comput. Math. Math. Phys.* **41,** 121–124 (2001).
25. Kroytor, O. K. & Malykh, M. D. On a dispersion curve of a waveguide filled with inhomogeneous substance. *Discrete and Continuous Models and Applied Computational Science* **30,** 330–341 (2022).

## Information about the authors

**Oleg K. Kroytor**—Candidate of Physical and Mathematical Sciences, Employee of the department of Mathematical Modeling and Artificial Intelligence of RUDN University (e-mail: kroytor_ok@pfur.ru, phone: +7 (495) 952-25-72, ORCID: 0000-0002-5691-7331, ResearcherID: GLS-3788-2022, Scopus Author ID: 57212346588)

**Mikhail D. Malykh**—Doctor of Physical and Mathematical Sciences, Head of the department of Mathematical Modeling and Artificial Intelligence of RUDN University, research fellow of MLIT JINR (Dubna) (e-mail: malykh_md@pfur.ru, phone: +7 (495) 952-25-72, ORCID: 0000-0001-6541-6603, ResearcherID: P-8123-2016, Scopus Author ID: 6602318510)

**Leonid A. Sevastianov**—Doctor of Physical and Mathematical Sciences, Professor of the department of Mathematical Modeling and Artificial Intelligence of RUDN University, research fellow of LTPh JINR (Dubna) (e-mail: sevastianov_la@pfur.ru, phone: +7 (495) 952-25-72, ORCID: 0000-0002-1856-4643, ResearcherID: B-8497-2016, Scopus Author ID: 8783969400)

УДК 519.872:519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-4-395-405 EDN: EQOGQJ

# Задача о нормальных модах волновода

О. К. Кройтор[1], М. Д. Малых[1, 2], Л. А. Севастьянов[1, 2]

[1] Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

[2] Объединённый институт ядерных исследований, ул. Жолио-Кюри, д. 6, Дубна, 141980, Российская Федерация

**Аннотация.** Рассмотрены различные подходы к вычислению нормальных мод закрытого волновода. Дан обзор литературы, проведено сравнение двух формулировок этой задачи. Показано, что использование самосопряжённой постановки задачи о нормальных модах волновода исключает возникновение артефактов, связанных с появлением малой мнимой добавки у собственных значений. Представлена реализация этого подхода для волновода прямоугольного сечения с прямоугольными вставками в системе компьютерной алгебры Sage и протестирована на гибридных модах слоистых волноводов. Тесты показали, что наша программа прекрасно справляется с вычислением точек дисперсионной кривой, отвечающих гибридным модам волновода.

**Ключевые слова:** поляризованное электромагнитное излучение, нормальные моды волновода, спектральная задача теории волновода, дисперсионная кривая волновода

**Discrete & Continuous Models & Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 406–413**
http://journals.rudn.ru/miph

*Research article*

UDC 517.98:517.518.456
PACS 02.30.Nw, 02.30.Gp, 07.05.Tp
DOI: 10.22363/2658-4670-2024-32-4-406-413 EDN: ERLZZM

# On summation of Fourier series in finite form

Mikhail D. Malykh[1, 3], Ksaverii Yu. Malyshev[1, 2]

[1] RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

[2] Lomonosov Moscow State University, 1(2) Leninskie Gory, Moscow, 119991, Russian Federation

[3] Joint Institute for Nuclear Research, 6 Joliot-Curie St, Dubna, 141980, Russian Federation

(received: September 24, 2024; revised: October 10, 2024; accepted: October 15, 2024)

**Abstract.** The problem of summation of Fourier series in finite form is formulated in the weak sense, which allows one to consider this problem uniformly both for classically convergent and for divergent series. For series with polynomial Fourier coefficients $a_n, b_n \in \mathbb{R}[n]$, it is proved that the sum of a Fourier series can be represented as a linear combination of 1, $\delta(x)$, $\cot\frac{x}{2}$ and their derivatives. It is shown that this representation can be found in a finite number of steps. For series with rational Fourier coefficients $a_n, b_n \in \mathbb{R}(n)$, it is shown that the sum of such a series is always a solution of a linear differential equation with constant coefficients whose right-hand side is a linear combination of 1, $\delta(x)$, $\cot\frac{x}{2}$ and their derivatives. Thus, the issue of summing a Fourier series with rational coefficients is reduced to the classical problem of the theory of integration in elementary functions.

**Key words and phrases:** mathematical physics, Fourier series, elementary functions

**For citation:** Malykh, M. D., Malyshev, K. Yu. On summation of Fourier series in finite form. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 406–413. doi: 10.22363/2658-4670-2024-32-4-406-413. edn: ERLZZM (2024).

## 1. Introduction

The problem of summing a functional series in elementary functions naturally arises when solving problems in mathematical physics [1–6]. If desired, even d'Alembert's method of solving the wave equation can be considered as a method of summing a Fourier series [7, 8]. Frequently, results on summation in the final form arose as surprising side effects, for example, when accelerating the convergence of series by A.N. Krylov's method [9–12]. However, the authors of the past avoided considering divergent series, the summation of which, as it seemed then, could yield anything [13, p. 641], [14, Ch. 12, Sect. 4].

With the advent of the theory of generalized functions [15], a reliable basis for considering divergent functional series arose. The surprising fact is that divergent series are usually summed up in a finite form much more easily than convergent ones, and, moreover, the summation of convergent series in a finite form is conveniently reduced to the summation of divergent series. In this paper we illustrate this statement using the example of one-dimensional Fourier series. The possibility of interpreting Krylov's method in terms of generalized functions was mentioned in [16, p. 32].

## 2. Statement of the Problem

**Definition 1.** A periodic function is called piecewise elementary if its period can be divided into a finite number of segments, on each of which an elementary expression in the Liouville sense can be specified for it.

We understand the equality between the sum of a Fourier series and a piecewise elementary function in the weak sense [15], which allows a further uniform consideration of the series summation in elementary functions separate from the issue of its pointwise convergence.

**Definition 2.** The Fourier series

$$u = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx, \tag{1}$$

is called a piecewise elementary function $v$ in the strong sense if the equality

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx = v \tag{2}$$

is satisfied at almost all points of the real axis.

**Definition 3.** The Fourier series (1) is said to be a piecewise elementary function $v$ in the weak sense if the equality

$$\frac{a_0}{2} \int_{x=-\pi}^{\pi} w dx + \sum_{n=1}^{\infty} a_n \int_{x=-\pi}^{\pi} \cos nx w dx + b_n \int_{x=-\pi}^{\pi} \sin nx w dx = \int_{x=-\pi}^{\pi} v w dx \tag{3}$$

is true on any smooth function $w$ with period $2\pi$.

The Fourier coefficients of a smooth function $w$ converge to zero faster than any power of $n$, so the numerical series on the left in Eq. (3) always converges. If this does not lead to misunderstandings, instead of Eq. (3) we will write Eq. (2), implying that it is satisfied in the weak sense.

**Example 1.** For example, by virtue of the classical Dirichlet theorem, the series

$$\sum_{n=1}^{\infty} \frac{\sin nx}{n}$$

converges at all points of the interval $0 < x < 2\pi$ and its sum is equal to $\frac{\pi - x}{2}$. Therefore, this series is not only piecewise elementary, but also piecewise polynomial in the strong sense (Def. 2).

**Example 2.** Although the series

$$\sum_{n=1}^{\infty} \sin nx$$

diverges by Euler's test, the equality

$$\sum_{n=1}^{\infty} \int_{x=-\pi}^{\pi} \sin nx \cdot w dx = \text{v.p.} \int_{x=-\pi}^{\pi} \frac{1}{2} \cot \frac{x}{2} \cdot w dx \tag{4}$$

is true for all smooth $w$ with period $2\pi$. The integrand has a pole at zero, so we specify that the integral is understood in the sense of the Cauchy principal value. Therefore, this series, diverging in the classical sense, is also a piecewise elementary function in the weak sense (Def. 3). In this case, we consider the equality

$$\sum_{n=1}^{\infty} \sin nx = \frac{1}{2} \cot \frac{x}{2}$$

only as a short notation for Eq. (4).

Based on Definitions 1 and 3, we formulate the problem under consideration.

**Problem 1**. The coefficients of the Fourier series $a_n, b_n$ are given as rational functions of number $n$:

$$a_n, b_n \in \mathbb{R}(n).$$

It is required to find out whether this series is a piecewise elementary function and, if the answer is affirmative, indicate this function.

## 3. Summation of series with polynomial coefficients

Problem 1 is solved very simply for the polynomial case when $a_n$ and $b_n \in \mathbb{R}[n]$. However, this interesting case escaped the attention of authors of the 19th century, since in this case the general term of the Fourier series does not tend to zero, and therefore the series diverges. This difficulty is removed in Definition 3.

Indeed, let

$$a_n = \sum_{m=0}^{M} \alpha_m n^m, \quad b_n = \sum_{m=0}^{M} \beta_m n^m,$$

then the Fourier series under consideration can be rewritten as

$$u = \frac{a_0}{2} + \sum_{m=0}^{M} \alpha_m \sum_{n=1}^{\infty} n^m \cos nx + \sum_{m=0}^{M} \beta_m \sum_{n=1}^{\infty} n^m \sin nx.$$

The series that arise here are derivatives of the two main series

$$s(x) = \sum_{n=1}^{\infty} \sin nx$$

and

$$c(x) = \sum_{n=1}^{\infty} \cos nx.$$

For example,

$$\sum_{n=1}^{\infty} n^{2m} \cos nx = (-1)^m D^{2m} c(x).$$

We understand series in the sense of Definition 3, therefore

$$s(x) = \sum_{n=1}^{\infty} \sin nx = \frac{1}{2} \cot \frac{x}{2}$$

and

$$c(x) = \sum_{n=1}^{\infty} \cos nx = -\frac{1}{2} + \pi\delta(x)$$

for $-\pi < x < \pi$.

**Theorem 1.** *If $a_n, b_n \in \mathbb{R}[n]$, then the sum of the Fourier series* (1) *can be represented as a linear combination of 1, $\delta(x)$, $\cot\frac{x}{2}$ and their derivatives, this representation can be found in a finite number of steps.*

## 4. Summation of series with rational coefficients

Let us now turn to the solution of Problem 1 in the case when $a_n, b_n \in \mathbb{R}(n)$. Differentiation of $a_n \cos nx$ and $b_n \sin x$ reduces to multiplication by $\pm n$ and permutation of sine and cosine. Therefore, there always exists a linear differential operator $L$ such that

$$L[u] = \sum_{n=1}^{\infty} A_n \cos nx + B_n \sin nx, \quad A_n, B_n \in \mathbb{R}[n]. \tag{5}$$

We will say that the operator $L$ annihilates the denominator of the Fourier coefficients of the original series, and $A_n$ and $B_n$ are the Fourier coefficients obtained after the annihilation. The divergent series in the right-hand side of Eq. (5) has polynomial coefficients and is summed as described in the previous Section.

**Example 3.** Consider the Fourier series

$$u = \sum_{n=1}^{\infty} \frac{\sin nx}{1+n^2}.$$

We have

$$(-D^2+1)u = \sum_{n=1}^{\infty} \sin nx = \frac{1}{2}\cot\frac{x}{2}.$$

By Theorem 1, Problem 1 is reduced to the following problem.

**Problem 2.** A linear differential operator $L$ and a linear combination $f$ of functions 1, $\delta(x)$, $\cot\frac{x}{2}$ and their derivatives are given. It is required to find out whether the equation

$$L[u] = f$$

has a solution in piecewise elementary functions.

Since the coefficients of the operator $L$ are constant, the general solution of the equation $L[u] = f$ can be written in quadratures using the method of variation of constants. Quadratures containing the $\delta$-function and its derivative are always taken.

**Theorem 2.** *If the given series converges and after annihilation the coefficient $A_n$ is an even function of n, and the coefficient $B_n$ is an odd function of n, then Problem 1 is solvable.*

Numerous examples illustrating this theorem were considered in classical studies of accelerating the summation of Fourier series [9, 10]. In the general case, the solution $L[u] = f$ will contain quadratures of the form

$$\int x^p e^{\lambda x} D^q \cot\frac{x}{2} dx.$$

The conditions found in Liouville theory [17] under which integrals of this type are taken in elementary functions provide sufficient conditions for the solvability of Problem 1. Thus, Problem 1 is reduced to the classical problem of computer algebra [18].

**Example 4.** Returning to Example 3, we see that $u$ is a solution of the linear differential equation

$$u'' - u = -\frac{1}{2}\cot\frac{x}{2}$$

whose general solution is given by the quadrature

$$u = \frac{e^{-x}}{4}\int e^{x}\cot\frac{x}{2}dx - \frac{e^{x}}{4}\int e^{-x}\cot\frac{x}{2}dx.$$

Thus, the solution of Problem 1 for the series from Example 3 is reduced to the study of the elementariness of this expression.

## 5. Results

The sum of the Fourier series

$$u = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n\cos nx + b_n\sin nx,$$

whose coefficients are rational functions of $n$, is always a solution of some linear differential equation

$$L[u] = f,$$

whose right-hand side is the sum of a divergent Fourier series with polynomial coefficients. This series can always be represented as a linear combination of 1, $\delta(x)$, $\cot\frac{x}{2}$ and their derivatives, so the original series can be represented as a quadrature of piecewise elementary functions. The conditions under which these quadratures are taken in elementary functions provide sufficient conditions for the summation of the Fourier series in piecewise elementary functions.

## 6. Discussion

In this paper, we propose a simple approach to summation of a certain class of trigonometric series. Its distinctive feature is the term-by-term differentiation of Fourier series, which inevitably leads to the appearance of divergent series. We believe that working with them can serve as the basis for symbolic algorithms for summation of eigenfunction series, and significantly supplement the generally accepted methods for summation, see [19, 20]. From the point of view of computer algebra, the approach under consideration allows us to establish a connection between the problem of summation of a certain class of series and the classical problem of integration in elementary functions. This is achieved by adding the Dirac $\delta$-function and other distributions to the set of elementary functions.

## 7. Conclusions

The transition from convergent to divergent series using an annihilation operator allows us to divide the problem of summing a convergent Fourier series into two simpler ones: summing a series with polynomial coefficients, which is solved explicitly, and integrating LDEs with constant coefficients. The development of an algorithm for constructing an annihilation operator for a given Fourier series with rational coefficients and its implementation in computer algebra systems will allow a wide class of Fourier series to be summed in a finite form.

Thus, the transition to summation of Fourier series in the weak sense allows reducing the problem of summation of series in a finite form (Problem 1) to calculating integrals of elementary functions in this form, i.e., a classic problem of computer algebra.

**Author Contributions:** Conceptualization, K.M; methodology, Ksaverii Yu. Malyshev and Mikhail D. Malykh; software, Ksaverii Yu. Malyshev and Mikhail D. Malykh; validation, Ksaverii Yu. Malyshev; formal analysis, Mikhail D. Malykh; investigation, Ksaverii Yu. Malyshev; resources, Mikhail D. Malykh; data curation, Ksaverii Yu. Malyshev; writing—original draft preparation, Ksaverii Yu. Malyshev; writing—review and editing, Mikhail D. Malykh and Ksaverii Yu. Malyshev; visualization, Ksaverii Yu. Malyshev; supervision, Mikhail D. Malykh; project administration, Mikhail D. Malykh; funding acquisition, Mikhail D. Malykh and Ksaverii Yu. Malyshev. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Russian Science Foundation (grant no. 20-11-20257).

**Data Availability Statement:** No new data were created or analysed during this study. Data sharing is not applicable.

**Acknowledgments:** The authors are grateful to Prof. L.A. Sevastianov (RUDN), PhD student M.V. Alexeev (HSE), Assoc. Prof. A.V. Bad'in (MSU), PhD O.K. Kroytor (RUDN) for valuable discussions.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Courant, R. & Hilbert, D. *Methods of mathematical physics* (Interscience Publishers, New York, 1953).
2. Titchmarsh, E. C. *Eigenfunction expansions associated with second-order differential equations* (Clarendon press, Oxford, 1962).
3. Tikhonov, A. N. & A., S. A. *Equations of Mathematical Physics* (Dover Publications, inc., New York, 2013).
4. Naydyuk F.O.; Pryadiev, V. & Sitnik, S. Laguerre Polynomials in the Forward and Backward Wave Profile Description for the Wave Equation on an Interval with the Robin Condition or the Attached Mass Condition. *Math. Notes* **115,** 789–799 (2024).
5. Rykhlov, V. S. Generalized solution of the initial-boundary-value problem for the wave equation with a mixed derivative and a general potential. Russian. *Itogi Nauki i Tekhniki. Sovrem. Mat. Pril. Temat. Obz.* **232,** 99–121 (2024).
6. Khromov, A. P. Divergent series and generalized mixed problem for wave equation. Russian. *Izv. Saratov Univ. Math. Mech. Inform.* **24,** 351–358 (2024).
7. Dolya, P. G. Solution to the homogeneous boundary value problems of free vibrations of a finite string. *Zh. Mat. Fiz. Anal. Geom.* **4,** 237–251 (2008).
8. Lomov, I. Generalized d'Alembert formula for the telegraph equation. *Itogi Nauki Tekh. Sovrem. Mat. Pril. Tematich. Obzory* **199,** 66–79 (2021).
9. Krylov, A. N. *On some differential equations of mathematical physics* (GITTL, Moscow-Leningrad, 1950).
10. Khromov, A. P. & Burlutskaya, M. S. Classical solution by the Fourier method of mixed problems with minimum requirements on the initial data. *Izv. Saratov Univ. Math. Mech. Inform.* **14,** 171–198 (2014).
11. Nersessian, A., Poghosyan, A. & Barkhudaryan, R. Convergence acceleration for Fourier series. *Izvestiya Natsionalnoi Akademii Nauk Armenii. Matematika* **41,** 39–51 (2006).
12. Nersessian, A. & Poghosyan, A. Accelerating the convergence of trigonometric series. *Central European Journal of Mathematics* **3,** 435–448 (2006).

13. Kozlov, V. Euler and mathematical methods in mechanics (on the 300th anniversary of the birth of Leonhard Euler). *Russian Math. Surveys* **62,** 639–661 (2007).
14. Reed, M. C. & Simon, B. *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic Press, New York, San Francisko, London, 1978).
15. Schwartz, L. *Mathematics for the physical sciences* (Hermann, Paris, 1966).
16. Gelfand, I. M. & Shilov, G. E. *Generalized Functions* (Academic Press, New-York, London, 1964).
17. Bronstein, M. *Symbolic Integration I: Transcendental Functions* (Springer, Berlin, Heidelberg, 1997).
18. Moses, J. Symbolic Integration: The Stormy Decade. *Communications of the ACM* **14,** 548–560 (1971).
19. Malyshev, K. & Malykh, M. D. Implementation of A.N. Krylov Series Convergence Acceleration in the CAS Sage. *International Conference Polynomial Computer Algebra'2023,* 78–79 (2023).
20. Malyshev, K. Divergent Fourier Series and Summation in Finite Terms Using the Krylov Method in CAS. *International Conference Polynomial Computer Algebra'2024,* 100–102 (2024).

## Information about the authors

**Mikhail D. Malykh**—Doctor of Physical and Mathematical Sciences, Head of the department of Mathematical Modeling and Artificial Intelligence of RUDN University, research fellow of MLIT JINR (Dubna) (e-mail: malykh_md@pfur.ru, phone: +7 (495) 952-25-72, ORCID: 0000-0001-6541-6603, ResearcherID: P-8123-2016, Scopus Author ID: 6602318510)

**Ksaverii Yu. Malyshev**—PhD student of the chair of Mathematical Modeling and Artificial Intelligence of RUDN University, engineer of Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University (e-mail: kmalyshev08102@mail.ru, phone: +7 (495) 952-25-72, ORCID: 0000-0001-8823-9136, Scopus Author ID: 57221615001)

УДК 517.98:517.518.456
PACS 02.30.Nw, 02.30.Gp, 07.05.Tp
DOI: 10.22363/2658-4670-2024-32-4-406-413 EDN: ERLZZM

# О суммировании рядов Фурье в конечном виде

М. Д. Малых[1, 3], К. Ю. Малышев[1, 2]

[1] Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

[2] Московский государственный университет имени М. В. Ломоносова, ул. Ленинские горы, д. 1, стр. 2, Москва, 119991, Российская Федерация

[3] Объединённый институт ядерных исследований, ул. Жолио-Кюри, д. 6, Дубна, 141980, Российская Федерация

**Аннотация.** Задача о суммировании рядов Фурье в конечном виде сформулирована в слабом смысле, что позволяет единообразно рассматривать эту задачу как для сходящихся в классическом смысле рядов, так и для расходящихся. Для рядов с полиномиальными коэффициентами Фурье $a_n$, $b_n \in \mathbb{R}[n]$ доказано, что сумма ряда Фурье может быть представлена как линейная комбинация $1$, $\delta(x)$, $\cot\frac{x}{2}$ и их производных. Показано, что это представление может быть найдено за конечное число действий. Для рядов с рациональными коэффициентами Фурье $a_n$, $b_n \in \mathbb{R}(n)$ показано, что сумма такого ряда всегда является решением линейного дифференциального уравнения с постоянными коэффициентами, правая часть которого является линейной комбинацией $1$, $\delta(x)$, $\cot\frac{x}{2}$ и их производных. Тем самым вопрос о суммировании рядов Фурье с рациональными коэффициентами сведен к классическому вопросу теории интегрирования в элементарных функциях.

**Ключевые слова:** математическая физика, ряды Фурье, элементарные функции

**Discrete & Continuous Models**
**& Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 414–424**
http://journals.rudn.ru/miph

*Research article*

UDC 519.6:004.94
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-4-414-424 EDN: DHGEBY

# Solving a two-point second-order LODE problem by constructing a complete system of solutions using a modified Chebyshev collocation method

Konstantin P. Lovetskiy[1], Mikhail D. Malykh[1,2], Leonid A. Sevastianov[1,2], Stepan V. Sergeev[1]

[1] RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

[2] Joint Institute for Nuclear Research, 6 Joliot-Curie St, Dubna, 141980, Russian Federation

(received: May 29, 2024; revised: June 25, 2024; accepted: June 30, 2024)

**Abstract.** Earlier we developed a stable fast numerical algorithm for solving ordinary differential equations of the first order. The method based on the Chebyshev collocation allows solving both initial value problems and problems with a fixed condition at an arbitrary point of the interval with equal success. The algorithm for solving the boundary value problem practically implements a single-pass analogue of the shooting method traditionally used in such cases. In this paper, we extend the developed algorithm to the class of linear ODEs of the second order. Active use of the method of integrating factors and the d'Alembert method allows us to reduce the method for solving second-order equations to a sequence of solutions of a pair of first-order equations. The general solution of the initial or boundary value problem for an inhomogeneous equation of the second order is represented as a sum of basic solutions with unknown constant coefficients. This approach ensures numerical stability, clarity, and simplicity of the algorithm.

**Key words and phrases:** linear ordinary differential equation of the second order, stable method, Chebyshev collocation method, d'Alembert method, integrating factor

**For citation:** Lovetskiy, K. P., Malykh, M. D., Sevastianov, L. A., Sergeev, S. V. Solving a two-point second-order LODE problem by constructing a complete system of solutions using a modified Chebyshev collocation method. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 414–424. doi: 10.22363/2658-4670-2024-32-4-414-424. edn: DHGEBY (2024).

## 1. Introduction

The paper studies a method for solving linear ordinary differential equations (ODEs) of the second order using integrating factors [1–3]. The method of integrating factors in combination with the Chebyshev collocation method [4] was previously applied by the authors to solve first-order ODEs (of general form) [5]. Moreover, the Chebyshev collocation method was successfully applied by the authors to solve second-order linear ODEs (LODEs) using both differentiation matrices [6] and integration matrices [7]. K.P. Lovetskiy et al. developed and applied a modified Chebyshev collocation method, which turned out to be not only more reliable, but also significantly more efficient compared to previous versions of the collocation method and other Runge–Kutta-type methods (see [5–9]) or shooting method [10].

At the first stage of the two-stage modified method proposed by the authors, when expanding the approximate solution in Chebyshev polynomials (of the first or second kind), the corresponding special Gauss–Chebyshev–Lobatto grids are used, on which the search for a part of the coefficients of the approximate general solution of the ODE is reduced to solving non-degenerate and well-conditioned (with diagonal matrices) System of Linear Algebraic Equations (SLAE). At the second stage, the solution is refined by using correctly formulated initial (or boundary) conditions. In this case, the SLAE with a positive definite diagonal matrix is solved first, and then the low-dimensional (one- or two-dimensional) SLAE is solved with respect to the first coefficients of the expansion of the solution in Chebyshev polynomials. The method allows solving with equal efficiency both initial problems and problems with conditions at arbitrary points, previously solved, e.g., by the shooting method, which thus loses its relevance.

Thus, we propose a constructive algorithm for approximate numerical solution of a wide class of LODEs. At the same time, the stage of the algorithm, consisting of solving the SLAE with a diagonal matrix, actually does not require computational costs, because it is reduced to a set of a small number of the simplest computational procedures. And only at the final stage, comprising the calculation of the first pair of coefficients of the expansion of the final particular solution, it is necessary to solve two-dimensional linear algebraic systems of equations determined by the initial or boundary conditions.

The method has proven itself to be perfect in solving one-point problems for first-order ODEs (see [5, 8, 9]). The application of the modified Chebyshev collocation method to solving second-order ODEs has also demonstrated high efficiency. We solve two-point problems for second-order linear ODEs using [11] the two-stage Chebyshev collocation method. The first stage is devoted to finding an approximate solution to the ODE in the form of a Chebyshev polynomial [12] with undetermined first coefficients. At the second stage, the first coefficients (if they exist) are found by solving a $2 \times 2$ SLAE [5–7, 13]. The first stage can be implemented in several not entirely equivalent ways [14]. Ref. [6] presents the Chebyshev collocation method for obtaining a solution to a second order LODE using the Chebyshev differentiation matrix [15]. The paper [7] implements the Chebyshev collocation method for obtaining a solution to a second order LODE using the Chebyshev antidifferentiation matrix. The authors noted that constructing a general (complete) solution from the individual partial solutions of the LODE obtained in this way seems to be a computationally complex task. At the same time, using an intermediate method that makes use of integrating factors to reduce the LODE to the form of a total derivative allows one to obtain general (complete) solutions of the second-order LODE more efficiently.

In the present article we seek approximate solutions of linear second-order ODEs of a rather general form

$$a(x)y''_{xx} + b(x)y'_x + c(x)y + f(x) = 0, \tag{1}$$

by the Chebyshev collocation method

## 2. Methods and algorithms

Let us consider step-by-step the methods for calculating contributions to the complete solution of a second order LODE. In each specific case, the solution of the original problem is divided into two stages. At the first stage, we find out the conditions that the coefficients of the second-order equations under study must satisfy, allowing us to reduce the search for the first of a pair of linearly independent solutions of a second-order linear equation to the solution of a first-order equation.

It turns out that such conditions can be determined at least for

– linear ordinary differential equations with constant coefficients;
– exact linear ordinary differential equations;
– linear equations reducible to a total differential form by means of integrating factors.

In the case of a homogeneous equation with constant non-zero coefficients $a$, $b$, $c$ linearly independent solutions of such an ODE can be found directly using the characteristic equation [16, 17]. In the worst case, i.e., when the discriminant $b^2 - 4ac$ is equal to zero, at least one of the solutions is easily determined.

In this case, the corresponding homogeneous equation takes the form

$$ay''(x) + by'(x) + \frac{b^2}{4a}y(x) = 0,$$

from which it follows that the characteristic equation allows finding only one solution

$$y_1(x) = e^{\frac{-b}{2a}x}.$$

If the coefficient functions depend continuously on the argument, a theoretical study of the conditions for reducing the second order LODE to the form of a full derivative of the first-order LODE is given below, in Section 3. The conditions that the coefficients of the inhomogeneous equation must satisfy for the possible construction of the potential are investigated. When sufficient conditions are met, a particular solution of the homogeneous first order LODE is constructed, which becomes the first necessary basic solution of the main nonhomogeneous second order LODE.

After obtaining the first linearly independent solution $y_1(t)$ of the second order LODE, at the next step, using several known algorithms [16–19], one can find the second linearly independent solution $y_2(t)$ and, consequently, the general solution. The most general and convenient method for finding the second solution numerically is the order reduction method [19, 20] (d'Alembert reduction).

Let one solution $y_1(t)$ of the linear homogeneous equation of the second order (1) be known and it is required to find the second linearly independent solution $y_2(t)$, thereby constructing a fundamental system of solutions of the inhomogeneous equation [16, 19]. For the brevity of presenting the method, we introduce the notation $p(x) = \dfrac{b(x)}{a(x)}$, $q(x) = \dfrac{c(x)}{a(x)}$, $g(x) = \dfrac{-f(x)}{a(x)}$.

Equation (1) takes the form

$$y'' + p(x)y' + q(x)y = g(x).$$

When the solution $y_1(t)$ of the homogeneous equation

$$y'' + p(x)y' + q(x)y = 0, \tag{2}$$

is found, we find another linear independent solution of the nonhomogeneous equation in the form of a product of the first solution $y_1(x)$ and an unknown function $v(x)$:

$$y(x) = y_1(x)v(x).$$

The search for the solution in the form of a product of the known solution $y_1(x)$ of the homogeneous equation (2) and a non-constant function $v(x)$ is explained by the fact that such a product is guaranteed to be a function linearly independent of $y_1(x)y$ and thus restricts the search for $y(x)$ to a one-dimensional subspace of the space of solutions of our ODE that is not covered by $y_1(x)$.

Actually, such an approach allows finding a general solution to an inhomogeneous equation. Namely, substituting $y_1(x)v(x)$ into $y'' + p(x)y' + q(x)y = g(x)$, and taking into account that $y_1(t)$ is a solution to the homogeneous equation, we obtain a nonhomogeneous equation with respect to the unknown derivative of the desired function $v'(x)$:

$$y_1v'' + \left(2y_1^{'} + p(x)y_1\right)v' = g(x).$$

As a result, for calculating the factor $v(x)$ in the second solution (assuming, that $y_1(x) \neq 0$) we obtain a differential equation of the first order with respect to $v'(x)$

$$v'' + \left(2\frac{y_1'(x)}{y_1(x)} + p(x)\right)v' = \frac{g(x)}{y_1(x)}. \tag{3}$$

Applying the technique of solving nonhomogeneous ordinary differential equations of the first order, based on integrating factors and approved in Refs. [5, 21], we get the desired solution

$$v'(x) = V(x)\left[C_1 + \int_{x_0}^{x} \frac{g(t)}{y_1(t)V(t)}dt\right], \tag{4}$$

where, considering that the solution passes through a certain point $x_0$, the following notation is introduced:

$$V(x) = \left[\frac{y_1'(x_0)}{y_1(x)}\right]^2 \exp\left[\int_{x_0}^{x} p(t)dt\right].$$

By integrating the ODE (4), we calculate the desired function $v(x)$ and become able to determine the solution of the nonhomogeneous equation by substituting into $y(x) = y_1(x)v(x)$. Hence, the general solution of the nonhomogeneous ODE (1) takes the form

$$y(x) = y_1(x)\left[C_2 + C_1\int_{x_0}^{x} V(t)dt + \int_{x_0}^{x} V(t)dt\int_{x_0}^{x} \frac{g(z)}{y_1(z)V(z)}dz\right]. \tag{5}$$

Finally, we obtained the complete (two-parametric family) solution of the inhomogeneous LODE. If it is necessary to solve a Cauchy problem of a boundary value problem with Eq. (1), we apply the second stage of the modified Chebyshev collocation method to calculate the constants $C_1$ and $C_2$.

The technique of finding the first coefficients of expansion of the desired solutions in Chebyshev polynomials is described in enough detail in our papers [6, 7, 13] for all kinds of "boundary" conditions: the Dirichlet, Neuman, and Robin ones.

## 3. The search for the first solution by reducing a linear ODE to the total derivative form

We consider the nonhomogeneous linear ODE of the second order with coefficients depending on the independent variable:

$$a(x)y''_{xx} + b(x)y'_x + c(x)y + f(x) = 0. \tag{6}$$

This equation is exact, if there exists such function $u(x, y, y')$, that

$$a(x)y''_{xx} + b(x)y'_x + c(x)y + f(x) = \frac{du}{dx}. \tag{7}$$

We want to reduce the search for a solution of the linear second-order ODE to the search for a solution of a linear first-order ODE and, therefore, restrict ourselves to a particular case when

$$u = A(x)y' + B(x)y + F(x). \tag{8}$$

By substituting expression (8) for $u$ into Eq. (7), we obtain equality in the form

$$a(x)y''_{xx} + b(x)y'_x + c(x)y + f(x) = A(x)y'' + \big(A'(x) + B(x)\big)y' + B'(x)y + F'.$$

It will be valid for any smooth $y(x)$ then and only then, when the coefficients of the expressions in the left-hand side and in the right-hand side of the equality coincide:

$$\begin{aligned} &A(x) = a(x);\\ &A'(x) + B(x) = b(x);\\ &B'(x) = c(x);\\ &F'(x) = f(x). \end{aligned}$$

This system of four equations allows unambiguous determination of $A(x)$, $B(x)$, $(x)$ from the coefficients of linear ODE (6):

$$\begin{aligned} &A(x) = a(x);\\ &B(x) = b(x) - a'(x);\\ &F(x) = \int f(x)dx, \end{aligned}$$

provided that one more condition is fulfilled,

$$c(x) = \big(b(x) - a'(x)\big)',$$

which, therefore, is a necessary and sufficient condition for the possibility to represent the linear ODE (6) in the form (7) with linear potential (8). Hence, the following theorem is valid.

**Theorem 3.** *The linear ODE* (6) *is exact and has a linear potential*

$$a(x)y'' + b(x)y' + c(x)y + f(x) = \frac{d}{dx}\big(A(x)y' + B(x)y + F(x)\big),$$

*when and only when the coefficients of the linear ODE* (6) *satisfy the condition*

$$c(x) = \big(b(x) - a'(x)\big)'$$

*and the potential has the form*

$$u(x, y, y') \equiv a(x)y' + \big(b(x) - a'(x)\big)y + \int f(x)dx = \text{const}. \tag{9}$$

*Corollary 1.* The linear homogeneous ODE

$$a(x)y''_{xx} + b(x)y'_x + c(x)y = 0 \tag{10}$$

is exact and has linear potential

$$a(x)y'' + b(x)y' + c(x)y = \frac{d}{dx}\big(A(x)y' + B(x)y\big),$$

then and only then, when the coefficients of the linear ODE (6) satisfy the condition

$$c(x) = \big(b(x) - a'(x)\big)'. \tag{11}$$

If for a certain second-order equation of the form (6) the condition (11) is fulfilled, then the search for one of its solutions can be reduced to a search for a solution of the linear first-order ODE with an arbitrary constant const

$$Ay' + By + F = \text{const}.$$

Any solution $y_{part}(x)$ of the linear first-order ODE (8) at any value of the constant const is at the same time a solution to the initial equation (10).

Knowing one solution of the linear homogeneous equation of the second order (10), one can find its other linear independent solution using the d'Alembert method.

We have implemented checking of Eq. (8) fulfilment and searching for integral (9) in Sage as function `lsolve`.

*Example 1.* Consider a LODE

$$y'' + xy' + y + \cos x = 0.$$

The application of

```
sage: lsolve(diff(y, x, 2)+x*diff(y, x)+y+cos(x))
```

returns

$$x + \sin x + y'.$$

Thus, the integration of the initial second-order equation is analytically reduced to the integration of the first order LODE

$$x + \sin x + y' = \text{const}.$$

Now let us assume that Eq. (6) is not exact. In this case it is possible to try searching for an integrating factor $\mu(x)$ such that the equation

$$\mu(x)a(x)y''_{xx} + \mu(x)b(x)y'_x + \mu(x)c(x)y + \mu(x)f(x) = 0 \tag{12}$$

would be exact.

**Theorem 4.** *After introducing the factor $\mu(x)$ LODE* (6) *becomes exact and possesses a linear potential*

$$\mu(x) \cdot (a(x)y'' + b(x)y' + c(x)y + f(x)) = \frac{d}{dx}\left(A(x)y' + B(x)y + F(x)\right)$$

*then and only then, when the coefficients in the LODE* (10) *satisfy the condition*

$$\mu(x)c(x) = \Big(\mu(x)b(x) - \big(\mu(x)a(x)\big)'\Big)'. \tag{13}$$

*In this case the potential has the form*

$$u(x, y, y') \equiv \mu(x)a(x)y' + \Big(\mu(x)b(x) - \big(\mu(x)a(x)\big)'\Big)y + \int \mu(x)f(x)dx = \text{const}. \tag{14}$$

*Any potential solution at any value of the constant is a solution to Eq.* (12).

Having one solution of the linear inhomogeneous second-order equation (12) it is possible to find its other linearly independent solution, using the d'Alembert algorithm.

Equation (13) is a homogeneous linear ODE of the second order with respect to $\mu(x)$. If the initial linear ODE is also homogeneous, then it is possible to formulate a very simple method to find the factor.

*Corollary 2.* If the linear ODE (12) is homogeneous ($f(x) = 0$) and its coefficients satisfy the relation

$$b'(x)a(x) - a'(x)b(x) - c(x)a(x) = 0 \tag{15}$$

then the linear ODE (12) has an integrating factor

$$\mu(x) = \frac{1}{a(x)}. \tag{16}$$

In this case, the potential for the initial linear second-order ODE is given by the linear ODE of the first order

$$u(x, y, y') \equiv y'(x) + \frac{b(x)}{a(x)} y(x) = \text{const}. \tag{17}$$

**Proof.** By Theorem 2, in order to reduce a second-order LODE to a first-order LODE it is sufficient to find the factor $\mu(x)$ from Eq. (13). Substitution of expression (16) into it leads to relation (15). To determine the coefficients of potential (14) at $f(x) = 0$ we have:

$$A = \mu a = 1;$$
$$B = \mu b - (\mu a)' = \frac{b}{a};$$
$$F = \mu c - (\mu b)' + (\mu a)'' = \frac{c}{a} - \left(\frac{b}{a}\right)' = 0.$$

Any solution of potential (17) at any value of the constant is a solution to Eq. (12).

Having one solution of the linear homogeneous solution of the second order (12), it is possible to find its another linearly independent solution by using the d'Alembert method.

We implemented the checking of the search for the factor in Sage within the function `lsolve` mentioned above. This function checks the fulfilment of condition (15). In the case of success, it divides the LODE by $a(x)$ and finds the first-order LODE by the methods described in Corollary 1. In the case of failure, the system tries to integrate Eq. (13).

*Example 2.* Consider a LODE

$$(x^2 + 1)(y' + xy)' = 0.$$

The application of

```
sage: lsolve((x^2+3)*diff(diff(y, x)+x*y, x))
```

returns

$$y' + xy.$$

Thus, the integration of the initial second-order equation is analytically reduced to the integration of the first order LODE

$$y' + xy = \text{const}.$$

*Example 3.* Consider the LODE

$$y'' + y + \sin x = 0.$$

Our function

```
sage: lsolve(diff(y, x, 2)+y+sin(x))
```

returns a family of factors of this equation:

$$K_1 \sin x + K_2 \cos x.$$

It is possible to take any element of this family: the query

```
sage: lsolve((diff(y, x, 2)+y+sin(x))*sin(x))
```

returns the LODE of the first order

$$\sin x y' + \frac{x}{2} - \cos x - \frac{\sin 2x}{4} = \text{const}.$$

## 4. Results

Previously, the authors proposed a method for finding a solution to a non-homogeneous linear ordinary differential equation of the second order using a modified Chebyshev collocation method using spectral (Chebyshev) matrices of differentiation and anti-differentiation [6, 7]. In this paper, a method for finding a solution to a second-order LODE is implemented by reducing it to the form of a total derivative either directly or using an integrating factor.

The modified Chebyshev collocation method allows one to obtain a complete system of linearly independent solutions to a linear ordinary differential equation of the second order using the d'Alembert method based on one known existing solution and to construct a general solution to a two-point problem for the corresponding second-order LODE in the case where it exists. In this case, the problem of the existence of a solution to a two-point problem for the corresponding second-order LODE is reduced to the problem of the existence of a solution to a two-dimensional system of linear algebraic equations for the first two coefficients in the expansion of the desired solution to the original problem in Chebyshev polynomials using the collocation method on the Chebyshev–Gauss–Lobatto grid.

## 5. Discussion

The D'Alembert method allows us to derive Eq. (3), the solution of which using integrating factors gives us the factor $v(x)$ of the general solution $y_1(x)v(x)$ of the inhomogeneous equation. As in the previous case, the numerical solution of Eq. (3) with respect to $v'(x)$ is carried out approximately using the Chebyshev collocation method [5, 7]. Integrating expression (4), we obtain $v(x)$ in the form of an interpolation polynomial. Substituting the obtained expression into the product $y_1(x)v(x)$, we obtain the general solution of the inhomogeneous ODE of the second order in the form (5).

The coefficients $C_1$, $C_2$ in the general solution formula are further determined based on the initial or boundary conditions defining the initial or boundary value problem for a second-order differential inhomogeneous equation. In the case of the Cauchy problem, the coefficients are always uniquely determined [16, 17]. In the case where a boundary value problem is considered, the system of resulting equations may have an infinite number of solutions, have no solutions, or have a unique solution. This is determined by the coincidence or difference of the ranks of the proper and extended matrices of the SLAE depending on the "boundary conditions". Thus, if a two-point boundary value problem has a solution, we obtain its approximate solution using the proposed approach—reducing the LODE to the total derivative form.

## 6. Conclusion

The paper considers an approach to solving linear inhomogeneous second-order ODEs based on the d'Alembert method of order reduction. The method allows, given one solution $y_1(x)$ of the complementary homogeneous equation, calculating both the general solution of the homogeneous equation and the general solution of the inhomogeneous equation. The method for obtaining the first solution of the homogeneous linear differential equation remains a difficult problem within this approach.

We have formulated the conditions for reducing the second-order LODE to the form of a total derivative of the solution using the Chebyshev collocation method. In cases where reduction to the form of a total derivative is not immediately possible, we assume, in the future, the use of a numerical method for solving a second-order equation using the method of integrating factors based on the Chebyshev collocation [7] to obtain the first solution of the accompanying first solution of the homogeneous equation.

The paper proposes an algorithm for obtaining the first basic solution of a complementary homogeneous ODE in cases of an equation with constant coefficients, an exact linear ordinary differential equation, or an equation that can be reduced to the form of a total differential using integrating factors. At the first step of the algorithm, the fulfillment of the conditions of belonging to exact equations or the possibility of finding such an integrating factor with which it is possible to reduce the equation to an exact one is checked. When the conditions set out in the corollaries to Theorems 1 and 2 are met, it is possible to construct a potential for a homogeneous equation—a first-order ODE. The solution to the potential equation can be found numerically [5, 21] using the efficient and stable Chebyshev collocation method. It is this solution to the homogeneous equation that is used further in the D'Alembert algorithm as the first known solution $y_1(x)$ of a second-order ODE.

**Author Contributions:** Conceptualization, software: Konstantin P. Lovetskiy; formal analysis: Mikhail D. Malykh; theoretical proof of the consistency of Chebyshev collocation method in the task used: Leonid A. Sevastianov; software: Stepan V. Sergeev. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the RUDN University Scientific Projects Grant System, project No 021934-0-000 (Konstantin P. Lovetskiy, Mikhail D. Malykh, Leonid A. Sevastianov). This research was supported by the RUDN University Strategic Academic Leadership Program (Stepan V. Sergeev).

**Data Availability Statement:** Data sharing is not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Tenenbaum, M. & Pollard, H. *Ordinary Differential Equations: An Elementary Textbook for Students of Mathematics, Engineering, and the Sciences* (Dover, Mineola, New York, 1986).
2. Yeomans, J. M. *Complex Numbers and Differential Equations. Lecture Notes for the Oxford Physics course* 2014.
3. Binney, J. J. *Complex Numbers and Ordinary Differential Equations. Lecture Notes for the Oxford Physics course* 2002.
4. Boyd, J. P. *Chebyshev and Fourier Spectral Methods* 2nd. 665 pp. (Dover, Mineola, New York, 2000).
5. Sevastianov, L. A., Lovetskiy, K. P. & Kulyabov, D. S. *Multistage collocation pseudo-spectral method for the solution of the first order linear ODE* Russian. in *2022 VIII International Conference on Information Technology and Nanotechnology (ITNT)* (2022), 1–6. doi:10.1109/ITNT55410.2022.9848731.
6. Lovetskiy, K. P., Kulyabov, D. S., Sevastianov, L. A. & Sergeev, S. V. Multi-stage numerical method of collocations for solving second-order ODEs. Russian. *Tomsk State University Journal of Control and Computer Science* **2023,** 45–52. doi:10.17223/19988605/63/6 (2023).
7. Lovetskiy, K. P., Kulyabov, D. S., Sevastianov, L. A. & Sergeev, S. V. Chebyshev collocation method for solving second order ODEs using integration matrices. *Discrete and Continuous Models and Applied Computational Science* **31,** 150–163. doi:10.22363/2658-4670-2023-31-2-150-163 (2023).
8. Lovetskiy, K. P., Sevastianov, L. A., Kulyabov, D. S. & Nikolaev, N. E. Regularized computation of oscillatory integrals with stationary points. *Journal of Computational Science* **26,** 22–27. doi:10.1016/j.jocs.2018.03.001 (2018).
9. Lovetskiy, K. P., Sevastianov, L. A., Hnatič, M. & Kulyabov, D. S. Numerical Integration of Highly Oscillatory Functions with and without Stationary Points. *Mathematics* **12,** 307. doi:10.3390/math12020307 (2024).

10. Zill, D. G. & Cullen, M. R. *Differential Equations with Boundary-Value Problems. Cengage Learning* 2008.
11. Canuto, C., Hussaini, M. Y., Quarteroni, A. & Zang, T. A. *Spectral Methods: Fundamentals in Single Domains* (Springer Verlag, 2006).
12. Mason, J. C. & Handscomb, D. C. *Chebyshev polynomials* doi:10.1201/9781420036114 (Chapman and Hall/CRC Press, 2002).
13. Lovetskiy, K. P., Sergeev, S. V., Kulyabov, D. S. & Sevastianov, L. A. Application of the Chebyshev collocation method tosolve boundary value problems of heat conduction. *Discrete and Continuous Models and Applied Computational Science* **32,** 74–85. doi:10.22363/2658-4670-2024-32-1-74-85 (2024).
14. Fornberg, B. *A Practical Guide to Pseudospectral Methods* (Cambridge University Press, New York, 1996).
15. Amiraslani, A., Corless, R. M. & Gunasingam, M. Differentiation matrices for univariate polynomials. *Numerical Algorithms* **83,** 1–31. doi:10.1007/s11075-019-00668-z (2023).
16. Lebl, J. *Notes on Diffy Qs: Differential Equations for Engineers* 2024.
17. Trench, W. F. *Elementary Differential Equations with Boundary Value Problems* 2013.
18. Polyanin, A. D. & V.F. Zaitsev, V. F. *Handbook of Ordinary Differential Equations: Exact Solutions, Methods, and Problems* (CRC Press, Boca Raton–London, 2018).
19. Schekochihin, A. A. *Lectures on ordinary differential equations. Lectures for the Oxford 1st-year undergraduate physics course, paper CP3/4* 2018.
20. Tikhonov, A. N., Vasil'eva, A. B. & Sveshnikov, A. G. *Differential Equations* (Springer, Berlin, 1985).
21. Sevastianov, L. A., Lovetskiy, K. P. & Kulyabov, D. S. A new approach to the formation of systems of linear algebraic equations for solving ordinary differential equations by the collocation method. Russian. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics* **23,** 36–47. doi:10.18500/1816-9791-2023-23-1-36-47 (2023).

## Information about the authors

**Konstantin P. Lovetskiy**—Candidate of Sciences in Physics and Mathematics, Associate Professor of Department of Computational Mathematics and Artificial Intelligence of RUDN University (e-mail: lovetskiy-kp@rudn.ru, phone: +7 (495) 952-25-72, ORCID: 0000-0002-3645-1060)

**Mikhail D. Malykh**—Doctor of Physical and Mathematical Sciences, Head of the Department of Computational Mathematics and Artificial Intelligence of RUDN University (e-mail: malykh-md@rudn.ru, phone: +7 (495) 955-09-27, ORCID: 0000-0001-6541-6603, ResearcherID: P-8123-2016, Scopus Author ID: 6602318510)

**Leonid A. Sevastianov**—Professor, Doctor of Sciences in Physics and Mathematics, Professor at the Department of Computational Mathematics and Artificial Intelligence of RUDN University, Leading Researcher of Bogoliubov Laboratory of Theoretical Physics, Joint Institute for Nuclear Research (e-mail: sevastianov-la@rudn.ru, phone: +7 (495) 952-25-72, ORCID: 0000-0002-1856-4643)

**Stepan V. Sergeev**—PhD student of Department of Computational Mathematics and Artificial Intelligence of RUDN University (e-mail: 1142220124@rudn.ru, ORCID: 0009-0004-1159-4745)

УДК 519.6:004.94
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-4-414-424 EDN: DHGEBY

# Решение двухточечной задачи ЛОДУ второго порядка построением полной системы решений модифицированным методом Чебышевской коллокации

К. П. Ловецкий[1], М. Д. Малых[1,2], Л. А. Севастьянов[1,2], С. В. Сергеев[1]

[1] Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

[2] Объединённый институт ядерных исследований, ул. Жолио-Кюри, д. 6, Дубна, 141980, Российская Федерация

**Аннотация.** В предыдущих работах мы разработали устойчивый быстрый численный алгоритм для решения обыкновенных дифференциальных уравнений первого порядка. Метод, основанный на чебышевской коллокации, позволяет одинаково успешно решать как начальные задачи, так и с фиксированным условием в произвольной точке интервала. Алгоритм решения краевой задачи практически реализует однопроходный аналог традиционно применяющегося в таких случаях метода стрельбы (Shooting method). В настоящей работе мы расширяем разработанный алгоритм на класс линейных ОДУ второго порядка. Активное использование метода интегрирующих множителей и метода Даламбера позволяет свести метод решения уравнений второго порядка к последовательности решений пары уравнений первого порядка. Общее решение начальной или краевой задачи для неоднородного уравнения 2-го порядка представляется в виде суммы базисных решений с неизвестными постоянными коэффициентами. Такой подход позволяет обеспечить численную устойчивость, наглядность и простоту алгоритма.

**Ключевые слова:** линейное обыкновенное дифференциальное уравнение второго порядка, устойчивый метод, метод чебышевской коллокации, метод Даламбера, интегрирующий множитель

**Discrete & Continuous Models & Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 425–444**
http://journals.rudn.ru/miph

*Research article*

UDC 519.872:519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-4-425-444 EDN: DIOZWP

# Development and adaptation of higher-order iterative methods in $R^n$ with specific rules

T. Zhanlav[1,2], Kh. Otgondorj[2]

[1] Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences, Ulaanbator, 13330, Mongolia

[2] Mongolian University of Science and Technology, Ulaanbator, 14191, Mongolia

(received: September 21, 2024; revised: October 5, 2024; accepted: October 6, 2024)

**Abstract.** In this article, we propose fourth- and fifth-order two-step iterative methods for solving the systems of nonlinear equations in $R^n$ with the operations of multiplication and division of vectors. Some of the proposed optimal fourth-order methods are considered as an extension of well-known methods that designed only for solving the nonlinear equations. We also developed $p\,(5 \leqslant p \leqslant 8)$—order three-point iterative methods for solving the systems of nonlinear equations, that contain some known iterations as particular cases. The computational efficiency of the new methods has been calculated and compared. The outcomes of numerical experiments are given to support the theoretical results concerning convergence order and computational efficiency. Comparative analysis demonstrates the superiority of the developed numerical techniques.

**Key words and phrases:** nonlinear systems, newton-type methods, order of convergence, computational efficiency, three-step iteration

**For citation:** Zhanlav, T., Otgondorj, K. Development and adaptation of higher-order iterative methods in $R^n$ with specific rules. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 425–444. doi: 10.22363/2658-4670-2024-32-4-425-444. edn: DIOZWP (2024).

## 1. Introduction

The problem to find a real solution of nonlinear system

$$F(x) = 0, \quad x = (x_1, x_2, \dots, x_n)^T \in R^n \tag{1}$$

has many applications in sciences and engineering [1–19]. In general, the root (zero) of equation (1) cannot be computed exactly. Most of the numerical methods used for solving this problem are iterative ones. Recently, many high-order iterative methods are presented in literature, see [1–9, 14, 17, 19] and references therein. Some methods [7, 8] of multiplication and division of two vectors, understood component-wise, are used. Let $a = (a_1, a_2, \dots, a_n)^T \in R^n$ and $b = (b_1, b_2, \dots, b_n)^T \in R^n$. Then

$$a \cdot b = (a_1 b_1, a_2 b_2, \dots, a_n b_n)^T \in R^n, \tag{2}$$

$$\frac{a}{b} = \left(\frac{a_1}{b_1}, \frac{a_2}{b_2}, \dots, \frac{a_n}{b_n}\right)^T \in R^n. \tag{3}$$

The direct consequence of (2) is

$$a \cdot b = b \cdot a. \tag{4}$$

If $a = b$ then (2) and (3) can be written as

$$a^2 = a \cdot a = \left(a_1^2, a_2^2, \dots, a_n^2\right)^T \in R^n,$$
$$\mathbf{1} = \left(1, 1, \dots, 1\right)^T \in R^n.$$

The purpose of this paper is to construct higher-order iterative methods in $R^n$ with multiplication and division rules (2) and (3). For obtaining the numerical solutions of the equation (1) we often use the two-step and three-step iterative methods as shown below:

$$\begin{aligned} &y_k = x_k - F'(x_k)^{-1}F(x_k), \\ &x_{k+1} = y_k - \bar{T}_k F'(x_k)^{-1}F(y_k) \text{ or } x_{k+1} = x_k - T_k F'(x_k)^{-1}F(x_k), \end{aligned} \tag{5}$$

and

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= y_k - \bar{T}_k F'(x_k)^{-1}F(y_k), \\ x_{k+1} &= z_k - \bar{\Psi}_k F'(x_k)^{-1}F(z_k). \end{aligned} \tag{6}$$

Here $\bar{T}_k$ and $\bar{\Psi}_k$ are iteration parameters to be determined properly. The convergence order of iterations (5) and (6) we denote by $p$ and $\rho = p + q$ respectively. We use $C''$- convergence order [2] based on nonlinear residual:

$$\frac{\|F(x_{k+1})\|}{\|F(x_k)\|^\sigma} = \text{const}$$

or

$$F(x_{k+1}) = O(h^\sigma), \quad h = F(x_k).$$

In our previous papers [11, 14] we find the sufficient convergence conditions in term of parameters $\bar{T}_k$ and $\bar{\Psi}_k$:

$$\bar{T}_k = I + O(h), \tag{7a}$$
$$\bar{T}_k = I + 2\Theta_k + O(h^2), \tag{7b}$$
$$\bar{T}_k = I + 2\Theta_k + 3d_k + 5\Theta_k^2 + O(h^3), \tag{7c}$$

and

$$\bar{\Psi}_k = I + O(h), \tag{8a}$$
$$\bar{\Psi}_k = I + 2\Theta_k + O(h^2), \tag{8b}$$
$$\bar{\Psi}_k = I + 2\Theta_k + 3d_k + 6\Theta_k^2 + O(h^3), \tag{8c}$$

where

$$\begin{aligned} \Theta_k &= \frac{1}{2}F'(x_k)^{-1}F''(x_k)\xi_k, \\ d_k &= -\frac{1}{6}F'(x_k)^{-1}F'''(x_k)\xi_k^2, \\ \xi_k &= F'(x_k)^{-1}F(x_k). \end{aligned} \tag{9}$$

Table 1
Summarizing the results of [11]

| $p$ | $\bar{T}_k$ | $q$ | $\bar{\Psi}_k$ |
|---|---|---|---|
| 3 | (7a) | 1 | (8a) |
| 4 | (7b) | 2 | (8b) |
| 5 | (7c) | 3 | (8c) |

Table 2
Summarizing the results of [14]

| $\rho = p + q$ | $\bar{T}_k$ | $\bar{\Psi}_k$ |
|---|---|---|
| 5 | (7a) | (8b) |
| | (7b) | (8a) |
| 6 | (7b) | (8b) |
| | (7c) | (8a) |
| | (7a) | (8c) |
| 7 | (7b) | (8c) |
| | (7c) | (8b) |
| 8 | (7c) | (8c) |

Summarizing the results of [11, 14], we present in Tables 1, 2 the convergence orders $p$ and $\rho$ of iterations (5) and (6) respectively.

It is worth noting that the convergence of iterations (5) and (6) was proved under the following condition

$$\frac{1}{2}F'(x_k)^{-1}F''(x_k)\Theta_k F'(x_k)^{-1}F(x_k) = \Theta_k^2 + O(h^3), \tag{10}$$

which holds true due to permutation properties of $q$-derivatives ($q \geq 1$) [20]. This paper is organized in five sections. Section 2 is devoted to constructing fourth-order two-step iterations and family of parametric two-step iterations. Extensions of some well known scalar methods with fourth order of convergence to solve systems of nonlinear equations are also discussed in this section. In Section 3, new two-parametric family of sixth, seventh, and eighth-order three-point iterative methods are constructed. Computational efficiency of the developed iterations is discussed in Section 4. In Section 5, we describe the outcomes of numerical experiments to confirm the theoretical analysis and made a comparison of some methods. Finally, we close with conclusions.

## 2. Construction of new iterative methods

First, we consider the following two-step iterations:

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ x_{k+1} &= x_k - \Omega_k, \end{aligned} \tag{11}$$

where

$$\Omega_k = \big(aF'(x_k) + bF'(\eta_k) + cF'(y_k)\big)^{-1}\big(\alpha F(y_k) + \beta F(x_k) + \gamma F(w_k)\big), \tag{12}$$

$w_k = x_k + \xi_k$, $\eta_k = x_k - \frac{2}{3}\xi_k$ and $a + b + c \neq 0$, $a$, $b$, $c$, $\alpha$, $\beta$, $\gamma$ are real constants. The convergence of our iteration (11) is established.

**Theorem 5.** *The convergence order of iteration* (11) *is equal to four, iff*

$$\begin{aligned} &b = \frac{3}{2}(1-2\gamma)(1-4\gamma), \quad a = 1-2\gamma-b, \quad c = 2\gamma, \\ &\beta = 1-2\gamma, \quad \alpha = 1-5\gamma-\frac{4}{3}b. \end{aligned} \tag{13}$$

**Proof.** The Taylor expansions of $F'(y_k)$ and $F'(\eta_k)$ at point $x_k$ give

$$F'(y_k) = F'(x_k)(I - 2\Theta_k - 3d_k) + O(h^3),$$
$$F'(\eta_k) = F'(x_k)(I - \frac{4}{3}(\Theta_k + d_k) + O(h^3).$$

Hence, we have

$$aF'(x_k) + bF'(\eta_k) + cF'(y_k) = F'(x_k)\Big((a+b+c)I - \Big(\frac{4}{3}b(\Theta_k + d_k) + c(2\Theta_k + 3d_k)\Big)\Big) + O(h^3).$$

From this we get

$$\big(aF'(x_k) + bF'(\eta_k) + cF'(y_k)\big)^{-1} = \frac{1}{a+b+c}\Big(I + \frac{1}{a+b+c}\Big(\Big(2c + \frac{4}{3}b\Big)\Theta_k + \\ + \Big(\frac{4}{3}b + 3c\Big)d_k\Big) + \frac{1}{(a+b+c)^2}\Big(\frac{16}{9}b^2 + 4c + \frac{16}{3}bc\Big)\Theta_k^2\Big)F'(x_k)^{-1} + O(h^3).$$

Similarly, using Taylor expansions of $F(y_k)$ and $F(w_k)$, we get

$$F'(x_k)^{-1}\big(\alpha F(y_k) + \beta F(x_k) + \gamma F(w_k)\big) = \Big((\beta + 2\gamma)I + (\alpha + \gamma)\Theta_k + (\alpha - \gamma)d_k\Big)\xi_k + O(h^4).$$

Hence, $\Omega_k$ given by (12) can be rewritten as

$$\Omega_k = \frac{1}{a+b+c}\Big((\beta + 2\gamma)I + \Big(\frac{\beta + 2\gamma}{a+b+c}\big(2c + \frac{4}{3}b\big) + \alpha + \gamma\Big)\Theta_k + \Big(\alpha - \gamma + \frac{\beta + 2\gamma}{a+b+c}\big(3c + \frac{4}{3}b\big)\Big)d_k + \\ + \Big(\frac{\beta + 2\gamma}{(a+b+c)^2}\big(\frac{16}{9}b^2 + 4c + \frac{16}{3}bc\big) + \frac{\alpha + \gamma}{a+b+c}\big(2c + \frac{4}{3}b\big)\Big)\Theta_k^2\Big)\xi_k + O(h^3).$$

We find the unknown coefficients in (11) such that [11, 14]

$$\Omega_k = (I + \Theta_k + d_k + 2\Theta_k^2)\xi_k.$$

This condition gives us

$$a + b + c = 1, \quad \beta + 2\gamma = 1, \quad 2c + \frac{4}{3}b + \alpha + \gamma = 1,$$
$$\alpha - \gamma + 3c + \frac{4}{3}b = 1, \quad \frac{16}{9}b^2 + 4c + \frac{16}{3}bc + (\alpha + \gamma)\big(2c + \frac{4}{3}b\big) = 2. \tag{14}$$

The solution of system (14) is (13). □

Thus, the iteration (11) becomes as:

$$y_k = x_k - F'(x_k)^{-1}F(x_k),$$
$$x_{k+1} = x_k - \Omega_k(\gamma), \tag{15}$$

where

$$\Omega_k(\gamma) = \Big((1 - 2\gamma - b)F'(x_k) + bF'(\eta_k) + 2\gamma F'(y_k)\Big)^{-1}\Big((1 - 5\gamma - \frac{4}{3}b)F(y_k) \\ + (1 - 2\gamma)F(x_k) + \gamma F(w_k)\Big), \quad b = \frac{3}{2}(1 - 2\gamma)(1 - 4\gamma),$$

$$\bar{T}_k = \frac{\Omega_k(\gamma)}{\Theta_k\xi_k}. \tag{16}$$

This family includes some well known methods as particular cases. We consider the particular cases of family (15):

1. Let $\gamma = \frac{1}{2}$. Then by (13) we get $a = b = \beta = 0$, $c = 1$, $\alpha = -\frac{3}{2}$ and (15) leads to

$$y_k = x_k - F'(x_k)^{-1}F(x_k),$$
$$x_{k+1} = x_k - F'(y_k)^{-1}\Big(-\frac{3}{2}F(y_k) + \frac{1}{2}F(w_k)\Big),$$

   which is the first iteration obtained by Su in [9].
2. Let $\gamma = 0$. Then by (13) we get $\beta = 1$, $c = 0$, $b = \frac{3}{2}$, $a = -\frac{1}{2}$, $\alpha = -1$ and (15) leads to

$$y_k = x_k - F'(x_k)^{-1}F(x_k),$$
$$x_{k+1} = x_k - \Big(\frac{1}{2}F'(x_k) - \frac{3}{2}F'(y_k)\Big)^{-1}\Big(F(y_k) - F(x_k)\Big),$$

   which is the second iteration obtained by Su in [9].
3. Let $\gamma = \frac{1}{4}$. Then by (13) we get $a = c = \beta = \frac{1}{2}$, $b = 0$, $\alpha = -\frac{1}{4}$ and (15) leads to

$$y_k = x_k - F'(x_k)^{-1}F(x_k),$$
$$x_{k+1} = x_k - \Big(F'(x_k) + F'(y_k)\Big)^{-1}\Big(F(x_k) + \frac{1}{2}F(w_k) - \frac{1}{2}F(y_k)\Big).$$

The following lemma plays key role in constructing high-order iterations.

**Lemma 1.** *The $\Theta_k$ given by* (9) *is equal to*

$$\Theta_k = \frac{F(y_k)}{F(x_k)} + O(h^2). \tag{17}$$

**Proof.** The Taylor expansion of $F(y_k)$ at point $x_k$ gives

$$F(y_k) = \frac{1}{2}F''(x_k)\xi_k^2 - \frac{1}{6}F'''(x_k)\xi_k^3 + O(h^4).$$

Then

$$F(x_k)^{-1}F(y_k) = (\Theta_k + d_k)\xi_k = \Theta_k\xi_k + O(h^3),$$

and thereby using (10) we obtain

$$\begin{aligned} F'(x_k)^{-1}F''(x_k)F'(x_k)^{-1}F(y_k) &= F'(x_k)^{-1}F''(x_k)\Theta_k\xi_k + O(h^3) \\ &= 2\Theta_k^2 + O(h^3). \end{aligned} \tag{18}$$

On the other hand, the left-hand side of (18) can be described as:

$$F'(x_k)^{-1}F''(x_k)F'(x_k)^{-1}F(x_k)\frac{F(y_k)}{F(x_k)} = 2\Theta_k\frac{F(y_k)}{F(x_k)}. \tag{19}$$

From (18) and (19) we reach (17). □

Note that the same definition (17) for $\Theta_k$ was used in the scalar equation case [11]. In $\mathbb{R}^n$ with operations (2) and (3) the iteration parameters $\bar{T}_k$ and $\bar{\Psi}_k$ turn out to be determined as vectors, that permit essentially simplification of implementation algorithms as compared with other existing methods with same order of convergence.

**Theorem 6.** *The two-step iteration* (5) *has a third, fourth and fifth-order convergence if and only if the parameter* $\bar{T}_k$ *satisfies*

$$\bar{T}_k = \mathbf{1} + O(h), \tag{20a}$$

$$\bar{T}_k = \mathbf{1} + 2\Theta_k + O(h^2), \tag{20b}$$

$$\bar{T}_k F'(x_k)^{-1} F(y_k) = \{\mathbf{1} + \Theta_k^2\} F'(y_k)^{-1} F(y_k), \tag{20c}$$

*where* $\Theta_k = \dfrac{F(y_k)}{F(x_k)}$, $\mathbf{1} = (1, 1, \dots, 1)^T \in \mathbb{R}^n$.

**Proof.** In [21] was shown that the fourth order convergence condition [12] is equivalent to:

$$\bar{T}_k = \mathbf{1} + \Theta_k + 2\Theta_k^2 + O(h^3). \tag{21}$$

Using (17), (21) and

$$\bar{T}_k = \frac{T_k - I}{\Theta_k},$$

we obtain

$$\bar{T}_k = \mathbf{1} + 2\Theta_k + O(h^2).$$

Analogously, using (20c) we obtain $F(x_{k+1}) = O(h^5)$. □

Using the expansion

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots, \quad \|x\| = \max_i |x_i| < 1, \quad x \in \mathbb{R}^n,$$

it is easy to show that $\bar{T}_k$ given by formula

$$\bar{T}_k = \frac{\mathbf{1} + a\Theta_k + b\Theta_k^2}{\mathbf{1} + (a-2)\Theta_k + d\Theta_k^2}, \quad a, b, d \in \mathbb{R}, \quad \Theta_k = \frac{F(y_k)}{F(x_k)}, \tag{22}$$

satisfies the conditions (20a) and (20b). In this case, the two-step iteration with (22) can be written as:

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1} F(x_k), \\ x_{k+1} &= y_k - \frac{F(x_k)^2 + aF(x_k)F(y_k) + bF(y_k)^2}{F(x_k)^2 + (a-2)F(x_k)F(y_k) + dF(y_k)^2} F'(x_k)^{-1} F(y_k), \quad a, b, d \in \mathbb{R}. \end{aligned} \tag{23}$$

Thus, we have obtained another family of fourth-order iterations (23). Now we consider some particular cases of this family.

1. Let $b = d = 0$. Then (23) leads to

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1} F(x_k), \\ x_{k+1} &= y_k - \frac{\mathbf{1} + a\Theta_k}{\mathbf{1} + (a-2)\Theta_k} F'(x_k)^{-1} F(y_k), \quad a \in \mathbb{R}. \end{aligned}$$

   This is a generalization of King's method for the system (1).
2. Let $a = b = 0$, $d = 1$. Then (23) leads to

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1} F(x_k), \\ x_{k+1} &= y_k - \frac{\mathbf{1}}{(\mathbf{1} - \Theta_k)^2} F'(x_k)^{-1} F(y_k), \quad a \in \mathbb{R}, \end{aligned}$$

   which is a generalization of King and Traub's method for the system (1).

3. Let $a = 1$, $b = -1$, $d = 0$. Then (23) leads to

$$y_k = x_k - F'(x_k)^{-1}F(x_k),$$
$$x_{k+1} = y_k - \frac{\mathbf{1} + \Theta_k - \Theta_k^2}{\mathbf{1} - \Theta_k}F'(x_k)^{-1}F(y_k),$$

which is a generalization of Maheshwari's method for the system (1).

Note that similar extension of King's, King and Traub's and Chun's methods are suggested in [20]. According to the definition given in [20], the family of iteration (23) is the optimal fourth-order one. Similarly, one can construct the generalization of all known fourth-order methods for system (1). Analogously, the following theorem is proved:

**Theorem 7.** *The convergence order of two-step iterations* (5) *equal to five, if* $\bar{T}_k$ *satisfies*

$$\bar{T}_k = (dI + \alpha t_k + \beta t_k^2)^{-1}(aI + bt_k + ct_k^2), \tag{24}$$

*where*

$$a = \frac{(3\alpha - 26\beta + 13c)}{5}, \quad b = \frac{(\alpha + 18\beta - 14c)}{5}$$
$$d = \frac{(4c - 13\beta - \alpha)}{5}, \quad \alpha + \beta + c \neq 0, \quad \alpha, \beta, c \in R. \tag{25}$$

**Proof.** Using the formula (18) and

$$t_k = F'(x_k)^{-1}F'(y_k) = I - 2\Theta_k - 3d_k + O(h^3), \quad s_k = F'(y_k)^{-1}F'(x_k),$$

one can easily shown that (24) satisfies (7c) under (25). □

We consider some special case of (24), (25).

1. Let $c = \beta = 0$. Then by (25) we get $a = \frac{3\alpha}{5}$, $b = -\frac{\alpha}{5}$, $d = \frac{3\alpha}{5}$. Substituting these values into (24) we obtain
$$\bar{T}_k = (5t_k - I)^{-1}(3I + t_k),$$
which is obtained by Wang in [6]. Note that his result is a generalization of method Ham and Chun (HC5) constructed for the scalar equation case [5].
2. Let $\alpha = \beta = 0$. Then by (25) we get $a = \frac{13c}{5}$, $b = -\frac{14c}{5}$, $d = \frac{4c}{5}$. Substituting these values into (24) we obtain
$$\bar{T}_k = \frac{13}{4}I - \frac{7}{2}t_k + \frac{5}{4}t_k^2,$$
which coincides with result [14, 18].

The formula (24) can be rewritten also in term of $s_k$ as:

$$\bar{T}_k = (\beta I + \alpha s_k + ds_k^2)^{-1}(cI + bs_k + as_k^2),$$

which includes choices of Cordero et. al. [1] as particular cases.

Now we consider the following two-step iterations

$$y_k = x_k - \bar{a}F'(x_k)^{-1}F(x_k), \quad \bar{a} \in R, \quad \bar{a} \neq 0,$$
$$z_k = y_k - \bar{T}_k F'(x_k)^{-1}F(y_k). \tag{26}$$

For iteration (26), the following result holds:

**Theorem 8.** *The convergence order of the family of iterations* (26) *equal to four (five, when* $\bar{a} = 1$ *) if* $\bar{T}_k$ *is given by*

$$\bar{T}_k = \mathbf{1} + (\bar{a}+1)\bar{\Theta}_k + (\bar{a}^2 + 2\bar{a} + 2)\bar{\Theta}_k^2 + (\bar{a}^2 + \bar{a} + 1)d_k, \tag{27}$$

*where*

$$\bar{\Theta}_k = \frac{1}{\bar{a}^2}\left(\frac{F(y_k) + F(w_k)}{2F(x_k)} - \mathbf{1}\right), \tag{28}$$

$$d_k = \frac{1}{\bar{a}^2}\left(\frac{F(y_k) - F(w_k)}{2\bar{a}F(x_k)} + \mathbf{1}\right), \tag{29}$$

$$w_k = x_k + \bar{a}F'(x_k)^{-1}F(x_k).$$

**Proof.** As in proof of lemma (1), it is easy to show that

$$\begin{aligned} \frac{F(y_k)}{F(x_k)} &= (1-\bar{a})\mathbf{1} + \bar{a}^2\bar{\Theta}_k + \bar{a}^3 d_k + O(h^3), \\ \frac{F(w_k)}{F(x_k)} &= (1+\bar{a})\mathbf{1} + \bar{a}^2\bar{\Theta}_k - \bar{a}^3 d_k + O(h^3). \end{aligned} \tag{30}$$

From (30) we find $\bar{\Theta}_k$ and $d_k$ and obtain (28), (29) with accuracy $O(h^3)$. So fourth order (fifth order when $\bar{a} = 1$) convergence condition [14] is satisfied for (27) with (28), (29). □

The adaptation of formula (27) in $R^n$ with operations of multiplication and division of vectors extremely easily realized by (28), (29).

Note that in (26) only one inverse of $F'(x_k)$ is required, whereas in the iteration (5) with $\bar{T}_k$ defined by (20c) two inverses of $F'(x_k)$ and $F'(y_k)$ are required.

## 3. The three-step methods

Now consider three-step iterations

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= \phi_k(x_k, y_k), \\ x_{k+1} &= z_k - \bar{\Psi}_k F'(x_k)^{-1}F(z_k). \end{aligned} \tag{31}$$

Here $z_k = \phi_k(x_k, y_k)$ is the iteration function of order $p \geq 2$.

**Theorem 9.** *The methods* (31) *have order of convergence* $p+1$, $p+2$, $p+3$ *if and only if the parameter* $\bar{\Psi}_k$ *satisfies*

$$\bar{\Psi}_k = \mathbf{1} + O(h), \tag{32a}$$

$$\bar{\Psi}_k = \mathbf{1} + 2\Theta_k + O(h^2), \tag{32b}$$

$$\bar{\Psi}_k F'(x_k)^{-1}F(z_k) = \{\mathbf{1} + 2\Theta_k^2\}F'(y_k)^{-1}F(z_k), \tag{32c}$$

respectively. The proof of this theorem is the same as Theorem 7 thus we omit it here. By virtue of Theorem 2 in [14] the iterations (31) has order of convergence $p+3$ if and only if $\bar{\Psi}_k$ satisfies (see also Table 1).

$$\bar{\Psi}_k = I + 2\Theta_k + 3d_k + 6\Theta_k^2 + O(h^3). \tag{33}$$

Using Taylor expansions of $F'(y_k)$ and $F'(y_k)^{-1}$ one can easily show that (33) equivalent to:

$$\bar{\Psi}_k = F'(y_k)^{-1}F'(x_k) + 2\Theta_k^2 + O(h^3), \quad \Theta_k = \frac{F(y_k)}{F(x_k)}.$$

Then by (4) we get

$$\begin{aligned}\bar{\Psi}_k F'(x_k)^{-1}F(z_k) &= F'(y_k)^{-1}F(z_k) + 2F'(x_k)^{-1}F(z_k)\Theta_k^2 + O(h^3)\\ &= F'(y_k)^{-1}(\mathbf{1} + 2\Theta_k^2)F(z_k) + O(h^3),\end{aligned}$$

in which we used $F'(x_k)^{-1} = F'(y_k)^{-1} + O(h)$.

Thus, we obtain $p+3$ order three-step iterations

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k),\\ z_k &= \phi_p(x_k, y_k),\\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(z_k).\end{aligned} \tag{34}$$

If we use $\bar{T}_k$ given by (24), (25) in (34) we obtain a family of eighth-order three-step iterations:

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k),\\ z_k &= y_k - (dI + \alpha t_k + \beta t_k^2)^{-1}(aI + bt_k + ct_k^2)F'(x_k)^{-1}F(y_k),\\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(z_k),\end{aligned} \tag{35}$$

where $a$, $b$ and $d$ are given by (25). Besides of (31), we can consider the family of three-step iterations:

$$\begin{aligned} y_k &= x_k - \bar{a}F'(x_k)^{-1}F(x_k), \quad \bar{a} \in R, \quad \bar{a} \neq 0,\\ z_k &= y_k - \bar{T}_k F'(x_k)^{-1}F(y_k),\\ x_{k+1} &= z_k - \bar{\Psi}_k F'(x_k)^{-1}F(z_k).\end{aligned} \tag{36}$$

For method (36) holds true:

**Theorem 10.** *The convergence order of the family of iterations* (36) *equal to seven (eight, when* $\bar{a} = 1$ *) if* $\bar{T}_k$ *is given by* (27) *and*

$$\bar{\Psi}_k = \mathbf{1} + 2\bar{\Theta}_k + 6\bar{\Theta}_k^2 + 3d_k. \tag{37}$$

*where* $\Theta_k$ *and* $d_k$ *are given by* (28) *and* (29) *respectively.*

**Proof.** By Theorem 8 we prove that $F(z_k) = O(h^p)$, $p = 4$ in case $\bar{a} \neq 1$ and $p = 5$ in case $\bar{a} = 1$. The $p+3$ order of convergence condition (33) of (36) is realized as (37). So the convergence order of iterations equals to seven (eight, when $\bar{a} = 1$). □

The combination of (15), (23), (20), (24), (27)–(29) and (32) (or Theorem 5–8 with Theorem 9, 10) gives us a wide set of iterative methods with convergence order $\rho$ (see Table 3). Below we list only the most effective methods of them.

– The fifth-order methods:

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k),\\ z_k &= y_k - F'(x_k)^{-1}F(y_k),\\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\frac{F(y_k)}{F(x_k)}\Big)F'(x_k)^{-1}F(z_k);\end{aligned} \tag{38}$$

Table 3

Iterative methods with convergence order $\rho$

| $\rho$ | $\bar{T}_k$ | $\bar{\Psi}_k$ | $\rho$ | $\bar{T}_k$ | $\bar{\Psi}_k$ | $\rho$ | $\bar{T}_k$ | $\bar{\Psi}_k$ | $\rho$ | $\bar{T}_k$ | $\bar{\Psi}_k$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | (20a) | (32b) | 6 | (20a) | (32c) | 7 | (16) | (32c) | 8 | (20c) | (32c) |
| | (20b) | (32a) | | (20b) | (32b) | | (20b) | (32c) | | (24) | (32c) |
| | (16) | (32a) | | (22) | (32b) | | (22) | (32c) | | (27) $a = 1$ | (32c) |
| | (22) | (32a) | | (27) $a \neq 1$ | (32b) | | (27) $a \neq 1$ | (32c) | | (7c) | (8c) |
| | (27) $a = 0$ | (32a) | | (16) | (32b) | | (20c) | (32b) | | | |
| | | | | (7c) | (32a) | | (24) | (32b) | | | |
| | | | | (24) | (32a) | | (27) $a = 1$ | (32b) | | | |

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= y_k - \Big(\mathbf{1} + 2\frac{F(y_k)}{F(x_k)}\Big)F'(x_k)^{-1}F(y_k), \\ x_{k+1} &= z_k - F'(x_k)^{-1}F(z_k), \end{aligned} \tag{39}$$

and

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= x_k - \Big(\mathbf{1} + \frac{F(y_k)}{F(x_k)}\Big)F'(x_k)^{-1}F(x_k), \\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\frac{F(y_k)}{F(x_k)}\Big)F'(x_k)^{-1}F(z_k). \end{aligned} \tag{40}$$

– The sixth order methods:

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= x_k - \Omega_k(\gamma), \\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\frac{F(y_k)}{F(x_k)}\Big)F'(x_k)^{-1}F(z_k). \end{aligned} \tag{41}$$

Note that the iteration (2.19) in [8] is a particular case of (41) when $\gamma = 0$.

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= y_k - \frac{F(x_k)^2 + 2F(x_k)F(y_k) + cF(y_k)^2}{F(x_k)^2 + (a-2)F(x_k)F(y_k) + dF(y_k)^2}F'(x_k)^{-1}F(y_k), \quad a, c, d \in \mathbb{R}, \\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\frac{F(y_k)}{F(x_k)}\Big)F'(x_k)^{-1}F(z_k). \end{aligned} \tag{42}$$

Most easy case of (42) is obtained when $a = 2$, $c = d = 0$. Another sixth-order iteration obtained from (34) is

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= y_k - F'(x_k)^{-1}F(y_k), \\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(z_k), \end{aligned} \tag{43}$$

because of $\bar{\tau}_k = \mathbf{1}$ in (43) and $F(z_k) = O(h^3)$ (see Table 2).

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= y_k - \Big(\mathbf{1} + \Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(y_k), \\ x_{k+1} &= z_k - F'(x_k)^{-1}F(z_k). \end{aligned} \tag{44}$$

– The seventh order methods:

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= x_k - \Omega_k(\gamma), \\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(z_k). \end{aligned} \tag{45}$$

Note that when $\gamma = 1/2$, the iteration (45) converted to iteration (2.18) given in [8].

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= y_k - \frac{F(x_k)^2 + 2F(x_k)F(y_k) + cF(y_k)^2}{F(x_k)^2 + (a-2)F(x_k)F(y_k) + dF(y_k)^2}F'(x_k)^{-1}F(y_k), \quad a, c, d \in \mathbb{R}, \\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(z_k), \end{aligned} \tag{46}$$

and

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= y_k - \Big(\mathbf{1} + \Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(y_k), \\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\frac{F(y_k)}{F(x_k)}\Big)F'(x_k)^{-1}F(z_k). \end{aligned} \tag{47}$$

– The eighth order methods:

$$\begin{aligned} y_k &= x_k - F'(x_k)^{-1}F(x_k), \\ z_k &= y_k - \Big(\mathbf{1} + \Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(y_k), \\ x_{k+1} &= z_k - \Big(\mathbf{1} + 2\Big(\frac{F(y_k)}{F(x_k)}\Big)^2\Big)F'(y_k)^{-1}F(z_k), \end{aligned} \tag{48}$$

and method (36) with $\bar{a} = 1$.

Note that in (36) only one inverse of $F'(x_k)$ is required, whereas in other three-step iterations (44), (47), (48), (35) with seventh and eighth-order of convergence two inverses of $F'(x_k)$ and $F'(y_k)$ are required.

**Remark 1.** Using the generating function method [13] one can easily show that the following replacements

$$\mathbf{1} + 2\Theta_k \Rightarrow \frac{1 + a_1\Theta_k + b_1\Theta_k^2}{1 + (a_1 - 2)\Theta_k + c_1\Theta_k^2}, \quad a_1,\, b_1,\, c_1 \in R,$$
$$\mathbf{1} + \Theta_k^2 \Rightarrow \frac{1 + a_2\Theta_k + b_2\Theta_k^2}{1 + a_2\Theta_k + (b_2 - 1)\Theta_k^2}, \quad a_2,\, b_2 \in R,$$
$$\mathbf{1} + 2\Theta_k^2 \Rightarrow \frac{1 + a_3\Theta_k + b_3\Theta_k^2}{1 + a_3\Theta_k + (b_3 - 2)\Theta_k^2}, \quad a_3,\, b_3 \in R,$$

in the above mentioned methods are also possible and in this case the convergence order maintained. In this way, we obtain multi-parametric families of iterations.

Note that in [3] the three-step iterations (3), (4), (5) were considered. The original idea is to have the weight functions $Q_1$ and $Q_2$ chosen in such a way that the method will be of order higher than 4. But this was not successful as the numerical experiments will show. If we use the operations (2) and (3) in [3] then $Q_1$ and $Q_2$ satisfy the conditions (20) and (32). So their method indeed has a sixth order convergence. It should be also pointed out that in [10] another definition of division of vectors was introduced and the extensions of some iterations in multidimensional case were considered by means of matrix $X$ such that $Xa = b$ i.e. $a \overset{X}{\rightarrow} b$. But find such matrix $X$ is not easy task.

## 4. Computational efficiency

In practice, a method is considered computationally efficient if it has higher convergence order and low computational cost. The computational efficiency index of iterative technique is calculated by [15]

$$E = p^{\frac{1}{C}},$$

where $p$ is order of convergence and $C = d + op$ stands for the total computational cost per iteration, $d$ is the number of function evaluations and $op$ is the number of products and quotients per iteration. We discuss the computational efficiency of the proposed methods and made comparisons between these and existing methods of similar nature. We denote by $C_i$ and $E_i$ the total cost and efficiency index for $i$-th method. The $E_i$ of the presented iterative methods is given in Table 4.

From Table 4, we see that the iterative method (36) has higher efficiency index. Specific property of our proposed methods is that they are much easier to implement as compared to other methods. In fact, each step of our methods requires to solve only one linear system. So passing from $x_k$ to $x_{k+1}$ is realized by solving three linear systems all together. While all other methods of order $p = 5, 6, 7, 8$ require to solve at least seven or eight linear systems [14, 16]. It makes the algorithms computationally more efficient. Thus, our methods are more simple and guarantee high computational efficiency as compared to other same order existing iterative techniques.

## 5. Results and discussion

The numerical experiments are carried out to confirm the theoretical results obtained in the previous sections. To get this aim, we consider several test problems, some of them are from real-life problems, e.g., Lane-Emden type equation and 2D Bratu problem.

Table 4

Comparison of computational efficiency

| № | methods | $p$ | $C_i$ | $E_i$ |
|---|---|---|---|---|
| 1 | (38) | 5 | $C_1 = \frac{1}{3}n^3 + 3n^2 + \frac{17}{3}n$ | $5^{1/C_1}$ |
| 2 | (39) | 5 | $C_2 = \frac{1}{3}n^3 + 4n^2 + \frac{17}{3}n$ | $5^{1/C_2}$ |
| 3 | (40) | 5 | $C_3 = \frac{1}{3}n^3 + 4n^2 + \frac{14}{3}n$ | $5^{1/C_3}$ |
| 4 | (41), $\gamma = 0$ | 6 | $C_4 = \frac{1}{3}n^3 + 5n^2 + \frac{16}{3}n$ | $6^{1/C_4}$ |
| 5 | (42), a=2 $c = d = 0$ | 6 | $C_5 = \frac{1}{3}n^3 + 4n^2 + \frac{17}{3}n$ | $6^{1/C_5}$ |
| 6 | (43) | 6 | $C_6 = \frac{2}{3}n^3 + 5n^2 + \frac{13}{3}n$ | $6^{1/C_6}$ |
| 7 | (44) | 6 | $C_{11} = \frac{1}{3}n^3 + 4n^2 + \frac{14}{3}n$ | $6^{1/C_{11}}$ |
| 8 | (45) | 7 | $C_7 = \frac{2}{3}n^3 + 5n^2 + \frac{16}{3}n$ | $7^{1/C_7}$ |
| 9 | (46) | 7 | $C_8 = \frac{1}{3}n^3 + 5n^2 + \frac{16}{3}n$ | $7^{1/C_8}$ |
| 10 | (47) | 7 | $C_9 = \frac{1}{3}n^3 + 5n^2 + \frac{16}{3}n$ | $7^{1/C_9}$ |
| 11 | (48) | 8 | $C_{10} = \frac{2}{3}n^3 + 5n^2 + \frac{13}{3}n$ | $8^{1/C_{10}}$ |
| 12 | (36) | 8 | $C_{12} = \frac{1}{3}n^3 + 4n^2 + \frac{17}{3}n$ | $8^{1/C_{12}}$ |

The experiments were made with an Intel Core processor i5-4590, with a CPU of 3.30 GHz and 4096 MB of RAM memory. For comparison, we consider the proposed $\rho$-order methods ($\rho = 5, 6, 7, 8$) and methods proposed in [17], [19], [18] and [14], namely, $T_1$, PM7, NLM8, and ZMO8, respectively. We also consider the sixth and seventh order methods (2.18) and (2.19) in [8]. In Tables 5–8, we give the error between two consecutive iterations $\|x_k - x_{k-1}\|$, computational order of convergence $\rho_c$ (see [17, 19]) is given by

$$\rho_c = \frac{\ln(\|x_{k+1} - x_k\|/\|x_k - x_{k-1}\|)}{\ln(\|x_k - x_{k-1}\|/\|x_{k-1} - x_{k-2}\|)}.$$

In addition, we include the elapsed CPU time (in seconds) in these tables. For each case, the following stopping criterion is used to terminate the iterations:

$$\|x_k - x_{k-1}\|_2 \le 10^{-150}.$$

**Example 1.** As a first example, we have taken the following small system (see [8]):

$$\begin{aligned}
x_{(1)}x_{(2)} + x_{(3)}(x_{(2)} + x_{(4)}) &= 2,\\
x_{(1)}x_{(3)} + x_{(2)}(x_{(1)} + x_{(4)}) &= 1,\\
x_{(1)}x_{(4)} + x_{(3)}(x_{(1)} + x_{(2)}) &= 3,\\
x_{(3)}x_{(2)} + x_{(1)}(x_{(2)} + x_{(4)}) &= 1.
\end{aligned}$$

The initial vector is $x_0 = \{-4, -3, -6, -6\}^T$ for the solution $x^* = \{-1.04, 0.26, -1.64, -1.64\}^T$.

Table 5

Comparison of methods for Example 1

| Methods | $k$ | $\|x_k - x_{k-1}\|$ | $\rho_c$ | e-time |
|---|---|---|---|---|
| (38) | 4 | 0.9003e-321 | 4.99 | 0.9486 |
| (39) | 4 | 0.1245e-342 | 4.99 | 0.9656 |
| (40) | 4 | 0.1983e-306 | 4.99 | 1.2112 |
| SSK5 [22] | 4 | 0.3751e-433 | 4.99 | 1.2507 |
| $T_1$ [17] | 4 | 0.4856e-591 | 4.99 | 1.2586 |
| (41), $\gamma = 1/2$ | 4 | 0.4756e-801 | 5.99 | 1.6553 |
| (41), $\gamma = 0$ [8] | 4 | 0.1132e-714 | 5.99 | 1.3688 |
| (42), a=2 $c = d = 0$ | 4 | 0.7142e-602 | 5.99 | 1.1804 |
| (43) | 4 | 0.7623e-634 | 5.99 | 1.3362 |
| (44) | 4 | 0.4561e-711 | 5.99 | 1.2558 |
| (45), $\gamma = 0$ | 4 | 0.2456e-1002 | 6.99 | 1.4639 |
| (45), $\gamma = 1/2$ [8] | 4 | 0.3452e-1220 | 6.99 | 1.1154 |
| (46), a=2 $c = d = 0$ | 4 | 0.1138e-1009 | 6.99 | 1.0325 |
| (47) | 4 | 0.1988e-1179 | 6.99 | 1.7041 |
| PM7 [19] | 4 | 0.7145e-1051 | 6.99 | 1.9620 |
| (48) | 4 | 0.9001e-2007 | 7.99 | 1.5592 |
| (36) | 4 | 0.8124e-4832 | 7.99 | 1.1801 |
| NLM8 [18] | 4 | 0.4356e-2089 | 7.99 | 1.7416 |
| ZMO8 [14] | 4 | 0.9140e-1991 | 7.99 | 1.1874 |

**Example 2.** Consider a system with 20 equations (see [6, 19]):

$$x_{(i)} - \cos\left(2x_{(i)} - \sum_{j=1}^{4} x_{(j)}\right) = 0,$$
$$i = 1, 2, \dots, 20.$$

The solution of this system is $x^* = \{0.5149, 0.5149, \dots, 0.5149\}^T$. We choose the initial approximation $x_0 = \{1, 1, \dots, 1\}^T$ for obtaining the solution $x^*$.

**Example 3.** We consider the singular boundary value problem (SBVP) [23]:

$$u''(x) + \frac{2}{x}u'(x) + \sin u(x) = 0, \quad u(0) = 1, \quad u'(0) = 0.$$

Table 6

Comparison of methods for Example 2

| Methods | $k$ | $\|x_k - x_{k-1}\|$ | $\rho_c$ | e-time |
|---|---|---|---|---|
| (38) | 4 | 0.6894e-589 | 4.99 | 15.838 |
| (39) | 4 | 0.2611e-306 | 4.99 | 15.837 |
| (40) | 4 | 0.6894e-589 | 4.99 | 15.791 |
| SSK5 [22] | 4 | 0.5781e-568 | 4.99 | 23.215 |
| $T_1$ [17] | 4 | 0.4856e-591 | 4.99 | 24.622 |
| (41), $\gamma = 1/2$ | 4 | 0.1374e-1121 | 5.99 | 12.036 |
| (41), $\gamma = 0$ [8] | 4 | 0.7735e-1156 | 5.99 | 12.823 |
| (42), a=2 $c = d = 0$ | 4 | 0.8521e-1107 | 5.99 | 11.022 |
| (43) | 4 | 0.9195e-1225 | 5.99 | 15.005 |
| (44) | 4 | 0.9353e-1292 | 5.99 | 12.828 |
| (45), $\gamma = 0$ | 4 | 0.3459e-2153 | 6.99 | 16.432 |
| (45), $\gamma = 1/2$ [8] | 4 | 0.5977e-2102 | 6.99 | 18.955 |
| (46), a=2 $c = d = 0$ | 4 | 0.9521e-2080 | 6.99 | 11.885 |
| (47) | 4 | 0.1142e-2181 | 6.99 | 11.891 |
| PM7 [19] | 4 | 0.4326e-2298 | 6.99 | 37.862 |
| (48) | 4 | 0.1717e-3730 | 7.99 | 8.9289 |
| (36) | 4 | 0.7355e-4562 | 8.00 | 8.1871 |
| NLM8 [18] | 4 | 0.7145e-4451 | 7.99 | 26.246 |
| ZMO8 [14] | 4 | 0.9784e-4962 | 7.99 | 26.384 |

After applying finite difference formulas the problem is reduced to a system of $n-1$ nonlinear equations with $n-1$ unknowns:

$$\frac{u_{k-1} - 2u_k + u_{k+1}}{h^2} + \frac{1}{x_k}\left(\frac{u_{k-1} - u_{k+1}}{h}\right) + \sin u_k = 0, \quad k = 1, 2, 3, \dots n-1.$$

We set $n = 101$ and take the initial guess $u_0 = (0.2, 0.2, \dots, 0.2)^T$.

**Example 4.** We consider the 2D Bratu problem [19]:

$$\begin{aligned} &u_{xx} + u_{yy} + \lambda e^u = 0, \\ &\Omega : \{(x, y) \in 0 \le x \le 1,\ 0 \le y \le 1\}, \\ &u = 0 \qquad \text{on } \partial\Omega, \end{aligned}$$

where $u = u(x, y)$.

Table 7

Comparison of methods for Example 3

| Methods | $k$ | $\|x_k - x_{k-1}\|$ | $\rho_c$ | e-time |
|---|---|---|---|---|
| (38) | 4 | 0.6894e-419 | 4.99 | 33.349 |
| (39) | 4 | 0.2611e-400 | 4.99 | 33.353 |
| (40) | 4 | 0.6894e-409 | 4.99 | 32.975 |
| SSK5 [22] | 4 | 0.7561e-457 | 4.99 | 42.121 |
| $T_1$ [17] | 4 | 0.8245e-492 | 4.99 | 41.714 |
| (41), $\gamma = 1/2$ | 4 | 0.2456e-1007 | 5.99 | 64.782 |
| (41), $\gamma = 0$ [8] | 4 | 0.4781e-1089 | 5.99 | 64.209 |
| (42), a=2 $c = d = 0$ | 4 | 0.8521e-1107 | 5.99 | 29.247 |
| (43) | 4 | 0.2145e-1059 | 5.99 | 43.619 |
| (44) | 4 | 0.9353e-1020 | 5.99 | 61.839 |
| (45) | 4 | 0.3785e-1992 | 6.99 | 62.167 |
| (45), $\gamma = 1/2$ [8] | 4 | 0.5977e-1988 | 6.99 | 61.453 |
| (46), a=2 $c = d = 0$ | 4 | 0.9521e-1997 | 6.99 | 59.098 |
| (47) | 4 | 0.1142e-1999 | 6.99 | 59.917 |
| PM7 [19] | 4 | 0.1756e-2001 | 6.99 | 101.86 |
| (48) | 4 | 0.1717e-2650 | 7.99 | 63.181 |
| (36) | 4 | 0.2751e-5821 | 8.00 | 58.567 |
| NLM8 [18] | 4 | 0.1457e-2775 | 7.99 | 120.09 |
| ZMO8 [14] | 4 | 0.6789e-2811 | 7.99 | 121.55 |

Applying the finite-difference formulas the problem is reduced to the nonlinear systems:

$$u_{i,j+1} - 4u_{i,j} + u_{i+1,j} + u_{i,j-1} + u_{i-1,j} + h^2\lambda e^{u_{i,j}} = 0,$$

where $u_{i,j}$ is $u$ at $(x_i, y_j)$ and $1 \leq i, j \leq N$. For obtaining a large nonlinear system of size $100 \times 100$, we take $N = 11$ and $\lambda = 0.1$. The initial vector is $u_0 = 0.1\big(\sin(\pi x_1)\sin(\pi y_1),\ \sin(\pi x_2)\sin(\pi y_2), \dots, \sin(\pi x_{10})\sin(\pi y_{10})\big)^T$ for the nonlinear system.

As can be observed from the Tables 5–8, the performance of the proposed methods is better than that of existing methods in terms of accuracy and CPU time. The comparison for considered problems shows that our method (36) is the fastest as compared to the other methods. The main reason is that for method (36) the inverse of $F'$ is used only once in per iteration.

Table 8

Comparison of methods for Example 4

| Methods | $k$ | $\|x_k - x_{k-1}\|$ | $\rho_c$ | e-time |
|---|---|---|---|---|
| (38) | 4 | 0.4215e-655 | 4.99 | 50.452 |
| (39) | 4 | 0.7653e-696 | 4.99 | 50.955 |
| (40) | 4 | 0.4579e-678 | 4.99 | 49.857 |
| SSK5 [22] | 4 | 0.1658e-622 | 4.99 | 60.921 |
| $T_1$ [17] | 4 | 0.4901e-602 | 4.99 | 62.103 |
| (41), $\gamma = 1/2$ | 4 | 0.2145e-812 | 5.99 | 83.325 |
| (41), $\gamma = 0$ [8] | 4 | 0.3457e-805 | 5.99 | 111.35 |
| (42), a=2 $c = d = 0$ | 4 | 0.9114e-833 | 5.99 | 57.112 |
| (43) | 4 | 0.1342e-799 | 5.99 | 82.717 |
| (44) | 4 | 0.2456e-843 | 5.99 | 83.840 |
| (45), $\gamma = 0$ | 4 | 0.4589e-1411 | 6.99 | 82.105 |
| (45), $\gamma = 1/2$ [8] | 4 | 0.9756e-1543 | 6.99 | 81.535 |
| (46), a=2 $c = d = 0$ | 4 | 0.7946e-1611 | 6.99 | 81.397 |
| (47) | 4 | 0.6789e-1589 | 6.99 | 81.912 |
| PM7 [19] | 4 | 0.1456e-1566 | 6.99 | 123.82 |
| (48) | 4 | 0.8599e-2316 | 7.99 | 85.183 |
| (36) | 4 | 0.6981e-2712 | 7.99 | 81.127 |
| NLM8 [18] | 4 | 0.7895e-2178 | 7.99 | 143.08 |
| ZCO8 [14] | 4 | 0.4879e-2002 | 7.99 | 143.66 |

## Conclusions

The main contributions of this paper are:

- We propose new fourth and fifth order two-step methods for solving the system of nonlinear equations in $R^n$ with the operations of multiplication and division of vectors.
- We extend the well-known two-point optimal fourth-order methods that designed for solving nonlinear equations to the systems of nonlinear equations. These are unexpected and elegant results.
- We also proposed $p$ $(5 \leq p \leq 8)$- order three-step iterative methods for solving the systems of nonlinear equations. These families of methods include some known methods as particular cases. Moreover, if we use generating function methods in these methods we obtain multi-parametric families of iterative methods.

– The proposed methods (exception of (35)) are simple and require solving three linear systems, whereas the existing methods of the same order convergence require to solve at least seven or eight linear systems. Moreover, they based on a multiplicity of vector by vector, instead of matrix-vector multiplication that inherent in other methods. Both these two factors make our algorithms computationally efficient and in principle new approach to construct higher order iterations.
– To illustrate the high efficiency and accuracy of the proposed methods, the numerical experiments are carried out on both academic and real world problems. Finally, based on numerical results, one can conclude that our methods are the most efficient and fastest than the existing ones of similar nature.

**Author Contributions:** Conceptualization, T. Zhanlav; methodology, T. Zhanlav; software, Kh. Otgondorj; formal analysis, T. Zhanlav, and Kh. Otgondorj; validation, Kh. Otgondorj; writing—original draft preparation, T. Zhanlav, and Kh. Otgondorj; writing—review and editing, Kh. Otgondorj. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was supported partially by the Foundation of Science and Technology of Mongolia under Grant Number MAS_2022/09.

**Data Availability Statement:** No new data were created or analysed during this study. Data sharing is not applicable.

**Acknowledgments:** The authors wish to thank the editor and the anonymous referees for their valuable suggestions and comments on the first version of this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Behl, R., Cordero, A., Motsa, S. S. & Torregrosa, J. R. Construction of fourth-order optimal families of iterative methods and their dynamics. *Applied Mathematics and Computation* **271.** doi:10.1016/j.amc.2015.08.113 (2015).
2. Cătinas, E. A survey on the high convergence orders and computational convergence orders of sequences. *Applied Mathematics and Computation* **343.** doi:10.1016/j.amc.2018.08.006 (2019).
3. Changbum, C. & Neta, B. Developing high order methods for the solution of systems of nonlinear equations. *Applied Mathematics and Computation* **344.** doi:10.1016/j.amc.2018.09.032 (2019).
4. Changbum, C. & Neta, B. An efficient derivative-Free method for the solution of systems of equations. *Numerical Functional Analysis and Optimization* **42.** doi:10.1080/01630563.2021.1931313 (2021).
5. Ham, Y. & Changbum, C. A fifth-order iterative method for solving nonlinear equations. *Applied Mathematics and Computation* **194.** doi:10.1016/j.amc.2007.04.005 (2007).
6. Wang, X. Fixed-point iterative method with eighth-order constructed by undetermined parameter technique for solving nonlinear systems. **13.** doi:10.3390/sym13050863 (2021).
7. Cuyt, A. A. & van der Cruyssen, P. Abstract Padé-approximants for the solution of a system of nonlinear equations. *Computers & Mathematics with Applications* **9.** doi:10.1016/0898-1221(83)90119-0 (1983).
8. Dehghan, M. & Shirilord, A. Three-step iterative methods for numerical solution of systems of nonlinear equations. *Engineering with Computers* **38.** doi:10.1007/s00366-020-01072-1 (2020).
9. Su, Q. A unified model for solving a system of nonlinear equations. *Applied Mathematics and Computation* **290.** doi:10.1016/j.amc.2016.05.047 (2016).
10. Ghanbari, B. & Changbum, C. A constructive method for solving the equation $Xa = b$ in $R^n$: A generalization of division in $R^n$. *Applied Mathematics and Computation* **364.** doi:10.1016/j.amc.2019.124673 (2020).

11. Zhanlav, T., Chuluunbaatar, O. & Ulziibayar, V. The necessary and sufficient conditions for two and three-point iterative method of Newton's type iterations. *Computational Mathematics and Mathematical Physics* **57.** doi:10.1134/S0965542517070120 (2017).
12. Zhanlav, T., Chun, C., Otgondorj, K. & Ulziibayar, V. High-order iterations for systems of nonlinear equations. *International Journal of Computer Mathematics* **97.** doi:10.1080/00207160.2019.1652739 (2020).
13. Zhanlav, T., Chuluunbaatar, O. & Ulziibayar, V. Generating function method for constructing new iterations. *Applied Mathematics and Computation* **315.** doi:10.1016/j.amc.2017.07.078 (2017).
14. Zhanlav, T., Mijiddorj, R. & Otgondorj, K. A family of Newton-type methods with seventh and eighth-order of convergence for solving systems of nonlinear equations. *Hacettepe Journal of Mathematics and Statistics* **54.** doi:10.15672/hujms.1061471 (2023).
15. Ostrowski, A. M. *Solution of equations and systems of equations* 202 pp. (Academic Press, New York, 1966).
16. Ahmad, F., Soleymani, F., Haghani, F. K. & Serra-Capizzano, S. Higher order derivative-free iterative methods with and without memory for systems of nonlinear equations. *Applied Mathematics and Computation* **314.** doi:10.1016/j.amc.2017.07.012 (2017).
17. Singh, H. & Sharma, J. R. Simple and efficient fifth order solvers for systems of nonlinear problems. *Mathematical Modelling and Analysis* **28.** doi:10.3846/mma.2023.16244 (2023).
18. Sharma, J. & Arora, H. Improved Newton-like methods for solving systems of nonlinear equations. *SeMA Journal* **74.** doi:10.1007/s40324-016-0085-x (2017).
19. Behl, R. & Arora, H. A novel scheme having seventh-order convergence for nonlinear systems. *Journal of Computational and Applied Mathematics* **404.** doi:10.1016/j.cam.2020.113301 (2020).
20. Cordero, A., Rojas-Hiciano, R. V., Torregrosa, J. R. & Vassileva, M. P. A highly efficient class of optimal fourth-order methods for solving nonlinear systems. *Numerical Algorithms* **95.** doi:10.1007/s11075-023-01631-9 (2024).
21. Zhanlav, T., Chun, C. & Otgondorj, K. Construction and dynamics of efficient high-Order methods for nonlinear systems. *International Journal of Computational Methods* **19.** doi:10.1142/S0219876222500207 (2022).
22. Singh, H., Sharma, J. R. & Kumar, S. A simple yet efficient two-step fifth-order weighted-Newton method for nonlinear models. *Numerical Algorithms* **93.** doi:10.1007/s11075-022-01412-w (2023).
23. Parand, K., Dehghan, M., Rezaei, A. R. & Ghaderi, S. M. An approximation algorithm for the solution of the nonlinear Lane–Emden type equations arising in astrophysics using Hermite functions collocation method. *Computer Physics Communications* **181.** doi:10.1016/j.cpc.2010.02.018 (2010).

## Information about the authors

**Tugal Zhanlav**—Academician, Professor, Doctor of Sciences in Physics and Mathematics, (e-mail: tzhanlav@yahoo.com, ORCID: 0000-0003-0743-5587, Scopus Author ID: 24484328800)

**Khuder Otgondorj**—Associate Professor of Department of Mathematics at School of Applied Sciences, Mongolian University of Science and Technology (e-mail: otgondorj@gmail.com, ORCID: 0000-0003-1635-7971, Scopus Author ID: 57209734799)

УДК 519.872:519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-4-425-444 EDN: DIOZWP

# Разработка и адаптация итерационных методов высшего порядка в $R^n$ с конкретными правилами

Т. Жанлав[1, 2], Х. Отгондорж[2]

[1] Институт математики и цифровой технологии, Монгольская академия наук, Улан-Батор, 13330, Монголия

[2] Монгольский государственный университет науки и технологии, Улан-Батор, 14191, Монголия

**Аннотация.** В данной работе мы предлагаем двухшаговые итерационные методы четвёртого и пятого порядков для решения систем нелинейных уравнений в $R^n$ с использованием операций векторного умножения и деления. Некоторые из предложенных оптимальных методов четвёртого порядка рассматриваются как расширение известных методов, разработанных исключительно для решения нелинейных уравнений. Мы также разработали трёхточечные итерационные методы $p$-порядка ($5 \leq p \leq 8$) для решения систем нелинейных уравнений, которые включают некоторые известные итерации как частные случаи. Проведён расчёт и сравнение вычислительной эффективности новых методов. Представлены результаты численных экспериментов для подтверждения теоретических выводов относительно порядка сходимости и вычислительной эффективности. Сравнительный анализ демонстрирует превосходство разработанных численных методов.

**Ключевые слова:** нелинейные системы, методы типа Ньютона, порядок сходимости, вычислительная эффективность, трёхшаговая итерация

**Discrete & Continuous Models & Applied Computational Science**
ISSN 2658-7149 (Online), 2658-4670 (Print)

**2024, 32 (4) 445–451**
http://journals.rudn.ru/miph

*Research article*

UDC 538.91
DOI: 10.22363/2658-4670-2024-32-4-445-451
EDN: DFMUXX

# Superconductivity and special symmetry of twisted tri-layer graphene in chiral model

Yuri P. Rybakov, Medina Umar

RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

(received: March 1, 2024; revised: March 20, 2024; accepted: May 10, 2024)

**Abstract.** Superconducting properties of twisted tri-layer graphene (TTG) are studied within the scope of the chiral model based on using the unitary matrix $U \in SU(2)$ as an order parameter. To check the superconductor behavior of this system, the interaction with the external magnetic field $B_0$ oriented along the graphene sheets is switched on and the internal magnetic intensity in the center is calculated as the function of the twisting angle. Vanishing of this function, due to the Meissner effect, being the important feature of the superconductivity, the corresponding dependence of the magic twisting angle on $B_0$ is calculated. The unusual effect of re-entrant superconductivity for large values of $B_0$ is discussed.

**Key words and phrases:** tri-layer graphene, chiral model, superconductivity

**For citation:** Rybakov, Y. P., Umar, M. Superconductivity and special symmetry of twisted tri-layer graphene in chiral model. *Discrete and Continuous Models and Applied Computational Science* **32** (4), 445–451. doi: 10.22363/2658-4670-2024-32-4-445-451. edn: DFMUXX (2024).

## 1. Introduction

It should be noticed that since the discovery of mono-atomic carbon layers called graphenes [1, 2] this material attracted high attention of researchers due to its extraordinary properties concerning magnetism, stiffness and considerable electric and thermal conductivity [3, 4]. The important connection was revealed with other graphene-based materials: Fullerenes [5] and carbon nanotubes [6]. A very simple explanation of these unusual properties of graphene was suggested in [7], where the idea of massless Dirac-like excitations of honeycomb carbon lattice was discussed, the latter one being considered as a superposition of two triangular sublattices. The further development of this idea was realized in [8, 9].

The unprecedented raise of interest has emerged to graphene-based materials and especially to moiré super-lattice patterns, this fact being motivated by their unconventional characteristics. In particular, specific magic-angle systems constructed by stacking two or three graphene layers twisted relative to each other have shown superconducting behavior [10–18]. However, these systems exhibit superconducting properties also for the very strong external magnetic fields (up to 10 T) [19], and therefore the standard superconductivity model by J. Bardeen, L. Cooper, J. Schrieffer and N. Bogoliubov [20] appears to be non suitable for the explanation of this fact. Thus, the superconductivity in TTG is likely to be driven by a mechanism that results in non-spin-singlet Cooper pairs. Nevertheless, it can be shown that the phenomenological approach based on the Landau theory of phase transitions [21] and on the corresponding chiral model of graphene suggested earlier [8] seems to be well suitable for the description of TTG.

## 2. Lagrangian density for the chiral model of graphene

In accordance with the hexagonal structure of the graphene mono-atomic carbon lattice, the three valence electrons of the atom form strong covalent bonds with the neighbours, but the forth electron belongs to the so-called hybridized state and appears to be "free". Thus, let us combine scalar $a_0$ and 3-vector $\mathbf{a}$ fields corresponding to the $s$-orbital and the $p$-orbital states of the "free" electron, respectively, into the unitary matrix $U \in SU(2)$ serving as the order parameter in our model:

$$U = a_0\,\tau_0 + \imath\,\mathbf{a}\cdot\boldsymbol{\tau}. \tag{1}$$

Here $\tau_0$ is the unit $2\times 2$-matrix and $\boldsymbol{\tau}$ stands for the three Pauli matrices, with the subsidiary $SU(2)$-condition being imposed: $a_0^2 + \mathbf{a}^2 = 1$ . To describe a single graphene sheet, one can use the Lagrangian density of the sigma-model form:

$$\mathcal{L} = -\frac{1}{4} I\,\mathrm{Sp}(l_\mu l^\mu) - \frac{1}{2}\lambda^2\mathbf{a}^2, \tag{2}$$

involving the so-called left chiral current $l_\mu = U^+\partial_\mu U$ and the coordinates $x^i$, $i = 1, 2, 3$ and the time $x^0 = ct$ derivatives. Comparing the Lagrangian density (2) with that of the Landau–Lifshitz theory [22] corresponding to the quasi-classical long-wave approximation to the Heisenberg ferromagnetic model, one can interpret the parameter $I$ in (2) as the exchange energy between carbon atoms (per spacing). The equations of motion corresponding to (2) admit the kink-like or the domain-wall solution [8]:

$$U = \exp(\imath\hat{n}\Theta), \quad \hat{n} = \mathbf{n}\cdot\boldsymbol{\tau}, \quad \Theta = 2\arctan\exp(-z/\ell_0); \tag{3}$$

describing the electrons distribution in an ideal graphene plane oriented along the unit vector $\mathbf{n}$ and orthogonal to the $z$-axis. The configuration 3 contains the characteristic length $\ell_0 = I^{1/2}/\lambda$, which can be identified with the diameter of the carbon atom $\ell_0 = 0.26$ nm.

It is worth while to underline that the interaction with an external electromagnetic field can be included via extending the derivatives in accordance with the gauge invariance principle:

$$\partial_\mu \Rightarrow D_\mu - \imath e_0 A_\mu[\tau_3,\, U],$$

where $e_0, \tau_3, A_\mu$ denote the electromagnetic coupling constant, the charge operator and the 4-potential, respectively. In particular case of the interaction with the uniform magnetic field oriented along the $y$-axis the Lagrangian density reads:

$$\mathcal{L} = -\frac{1}{4} I\,\mathrm{Sp}(l_\mu l^\mu) - \frac{1}{2}\lambda^2\mathbf{a}^2 - \frac{\mathbf{B}^2}{8\pi}, \tag{4}$$

where

$$L_\mu = U^+ D_\mu U, \quad \mathbf{B} = (0,\, B,\, 0), \quad B = A'(z), \quad A(z) = A_1, \quad B(\pm\infty) = B_0 = const.$$

The unitary matrix $U$ for the TTG configuration has the form:

$$U = U_1 U_0 U_2, \quad U_1 = \exp\bigl(\imath\hat{n}_j\Theta_j\bigr), \quad \Theta_j = \Theta_j(z), \quad \hat{n}_j = \mathbf{n}_j\cdot\boldsymbol{\tau}; \tag{5}$$

$$\mathbf{n}_1 = \bigl(\cos(\alpha/2),\, \sin(\alpha/2),\, 0\bigr), \quad \mathbf{n}_2 = \bigl(\cos(\alpha/2),\, -\sin(\alpha/2),\, 0\bigr), \quad \mathbf{n}_0 = (1,\, 0,\, 0), \tag{6}$$

with the vector $\mathbf{a}$ being defined as follows:

$$\mathbf{a} = -(\imath/2)tr(\boldsymbol{\tau}U).$$

Here $j = 0,\ 1,\ 2$ is the number of the correspondent sheet.

In accordance with (4), (5) and (6) the Lagrangian density takes the form:

$$\mathcal{L} = -I\Big[S + \cos(\alpha/2)\Theta_0'(\Theta_1' + \Theta_2') + \Theta_1'\Theta_2'\,(\sin^2\Theta_0 + \cos\alpha\cos^2\Theta_0)\Big] \\ - Ie_0^2A^2(1 - P - Q + R) - (\lambda)^2/2(X + Y + Z) - A'^2/(8\pi);$$

where the following denotations are used:

$$S = (\Theta_1'^2 + \Theta_2'^2 + \Theta_0'^2)/2,$$
$$P = 2\sin^2(\alpha/2)\sin^2\Theta_0\sin 2\Theta_1\sin 2\Theta_2,$$
$$Q = \cos 2\Theta_0\cos 2(\Theta_1 - \Theta_2),$$
$$R = \cos(\alpha/2)\sin 2\Theta_0\sin 2(\Theta_1 + \Theta_2),$$
$$X = \cos^2\Theta_0\Big[\sin^2(\Theta_1 + \Theta_2) - \sin^2(\alpha/2)\sin 2\Theta_1\sin 2\Theta_2 + \sin^2\alpha\sin^2\Theta_1\sin^2\Theta_2\Big],$$
$$Y = \sin^2\Theta_0\Big[\sin^2(\alpha/2) + \cos^2(\alpha/2)\cos^2(\Theta_1 + \Theta_2)\Big],$$
$$Z = \cos(\alpha/2)\sin 2\Theta_0\sin(\Theta_1 + \Theta_2)[\cos\Theta_1\cos\Theta_2 - \cos\alpha\sin\Theta_1\sin\Theta_2].$$

The boundary conditions read:

$$\Theta_j(-\infty) = \pi, \qquad \Theta_j(+\infty) = 0, \tag{7}$$

and central phases are chosen equal: $\Theta_0(0) = \Theta_1(-2l) = \Theta_2(2l) = \pi/2$, where $2l$ stands for the distance between the sheets.

## 3. Asymptotic structure of solutions to the equations of motion

At large $z \to \pm\infty$ one can put $\tan\Theta_j = u_j \to 0$ with the discrete symmetry being $u_1 = u_2 = u$. The asymptotic Lagrangian density

$$\mathcal{L} = -(I/2)[u_0' + 2\cos(\alpha/2)u']^2 - [u_0 + 2\cos(\alpha/2)u]^2\,(2Ie_0^2A^2 + \lambda^2/2),$$

where $A \approx B_0z$, admits the symmetry $u_0 \leftrightarrow 2u\cos(\alpha/2)$, with the solution being derived through the substitution $u^{-1} = \sinh w$, $u_0 = u\cos(\alpha/2)$. The asymptotic estimation reads:

$$u = 2\exp\big(-e_0B_0z^2\big). \tag{8}$$

As a result, for the vector potential $A = B_0z + a(z)$, where $a'(\infty) = 0$, one finds the equation:

$$\frac{a''}{4\pi} = 128\,Ie_0^2B_0z(1 + \cos\alpha)\exp\big(-2e_0B_0z^2\big)$$

with the evident solution:

$$A' = B_0 - 128\pi Ie_0(1 + \cos\alpha)\exp\big(-2e_0B_0z^2\big),$$
$$A = B_0z - 128\pi Ie_0(1 + \cos\alpha)\int_0^z \exp\big(-2e_0B_0z^2\big)\,dz, \tag{9}$$

where the anti-symmetric property of the vector potential found later was taken into account.

Now let us investigate the behavior of our system at small $z$, where one can put due to (7) the Lagrangian density for the vector potential taking the form:

$$\mathcal{L} = -2I\,e_0^2A^2 - A'^2/(8\pi).$$

The corresponding equation of motion reads:

$$A'' - 16\pi I\,e_0^2A = 0$$

and admits the evident solution:

$$A = C\sinh(k\,z), \tag{10}$$

where $C$ is an arbitrary constant and $k^2 = 16\pi I\,e_0^2$. Taking the derivative, it is not difficult to find the magnetic intensity $B = A' = k\,C\cosh(k\,z)$.

Now, to fix the value of the constant $C$ in (10), let us perform the smooth matching of the expressions (3), (9) and (10) at some intermediate point $z = \bar{l}$. However, to simplify this operation, let us introduce some denotations:

$$y = k\,C/B_0, \qquad x^2 = 2\,e_0B_0\bar{l}^2, \qquad \Lambda = 16(1+\cos\alpha)\exp\bigl(-x^2\bigr);$$

$$\Gamma = 8\pi I\,e_0/B_0, \quad \xi = \frac{\sinh\Bigl(x\sqrt{2\Gamma}\Bigr)}{x\sqrt{2\Gamma}}, \quad \eta = \cosh\Bigl(x\sqrt{2\Gamma}\Bigr).$$

Also the special representation for the error function is used [23]:

$$\frac{1}{x}\int\limits_0^x \exp\bigl(-s^2\bigr)\,d\,s = \frac{\pi^{1/2}}{2x}\,\mathrm{erf}(x) = (1+g)\exp\bigl(-x^2\bigr); \quad g = \sum_{n=1}^{i\infty}\frac{2^n\,x^{2n}}{(2n+1)!!}.$$

As a result, one obtains the following system of equations

$$\xi\,y = 1 - \Gamma\Lambda(1+g), \tag{11}$$

$$\eta\,y = 1 - \Gamma\Lambda.$$

Now it is worth while to stress that, in accordance with the Meissner effect [20], our system reveals superconducting properties if the relative magnetic field $y$ vanishes in the central domain. Let us first recall some information about graphene properties [24]. For numerical illustration of the twist effect one can use the following parameters of the chiral model: the spacing $a = 0.287$ nm, the exchange energy between atoms $E_0 = 2.9$ eV with the value $I = E_0/a = 1.619$ nN, the coupling constant $e_0 = e/(\hbar c)$, with $-e$ being the electron charge, the value $I\,e_0 = 0.246$ T being known as the effective (internal) "magnetic" intensity in graphene, the distance between the sheets $2l = 0.34$ nm. Taking into account that for standard graphene experiments

$$x^2 \ll 1, \quad g \ll 1, \quad \Gamma = (8\pi)246/B_0(m\,T) \gg 1, \quad \alpha \approx \pi - \zeta, \quad \zeta \ll 1, \quad \Lambda \approx 8\zeta^2;$$

one concludes that small values of $y$ can be provided by so-called "magic" values of twisting angle:

$$\zeta(rad) \approx (8\Gamma)^{-1/2}.$$

It should be noted that the other possible magic twisting angle can be obtained through the reflection $\alpha \Rightarrow \pi - \alpha$, which leaves the moiré super-lattice invariant.

Let us now discuss, in view of (11), the case of strong magnetic fields, when the quantity $\Gamma(1+g)$ retains large values. This fact implies the so-called re-entrant superconductivity. Experimental verification of this effect can be found in [19], the peculiar symmetry properties of TTG system being underlined earlier in connection with the boundary conditions (7).

## 4. Results and Discussions

In our paper the Landau phase transitions method is applied to the twisted tri-layer graphene model, the order parameter being the unitary matrix, depending on the twisting angle $\alpha$. The superconductivity property of the TTG model is proven for the special "magic" twisting angle, the cases of small and large external magnetic fields being considered.

## 5. Conclusions

The superconducting properties of the TTG configuration were studied within the framework of the chiral graphene model suggested in [8]. The product-ansatz being used for the description of the TTG system, the Lagrangian density and the asymptotic solutions to the equations of motion at small and large distances were found. Using the anti-symmetric behavior of the vector potential and matching these solutions at some intermediate point, a pair of algebraic equations for the magnetic field in the central domain and the twisting angle were obtained. Finally, in view of the Meissner effect, the correlation between the magic angle and the external magnetic intensity was established. The important effect of the re-entrant superconductivity was mentioned for the case of strong magnetic fields.

**Author Contributions:** Conceptualization, Yu.P. Rybakov; methodology, Yu.P. Rybakov; validation, Yu.P. Rybakov; writing—original draft preparation, Yu.P. Rybakov; writing—review and editing, Yu.P. Rybakov and M. Umar; supervision, Yu.P. Rybakov; project administration, Yu.P. Rybakov. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data sharing is not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306,** 666–669 (2004).
2. Geim, A. K. Graphine: status and prospects. *Science* **324,** 1530–1534 (2009).
3. Lee, C. *et al.* Measurement of elastic properties and intrinsic strength of monolayer graphene. *Science* **321,** 385–388 (2008).
4. Bolotin, K. I. *et al.* Ultrahigh electron mobility in suspended graphene. *Solid State Comm.* **146,** 351–355 (2008).
5. Lu, X. & Chen, Z. Curved $\pi$-conjugation aromaticity, and the related chemistry of small Fullerenes ($C_60$) and single-walled carbon nanotubes. *Chem. Rev.* **105,** 3643–3696 (2005).
6. Terrones, M. Science and Technology of the twenty-first century: synthesis, properties, and applications of carbon nanotubes. *Ann. Rev. Mat. Res.* **33,** 419–501 (2003).
7. Semenoff, G. W. Condensed-matter simulation of a three-dimensional anomaly. *Phys. Rev. Lett.* **53,** 2449–2452 (1984).
8. Rybakov, Y. P. On chiral model of graphene. *Solid State Phenomena* **190,** 59–62 (2012).
9. Rybakov, Y. P. Spin excitations in chiral model of graphene. *Solid State Phenomena* **233–234,** 16–19 (2015).
10. Cao, Y. *et al.* Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556,** 43–50 (2018).

11. Park, J. M. *et al.* Tunable strongly coupled superconductivity in magic-angle twisted tri-layer graphene. *Nature* **590,** 249–255 (2021).
12. Hao, Z. *et al.* Elastic field-tunable superconductivity in alternating-twist magic-angle tri-layer graphene. *Science* **371,** 1133–1138 (2021).
13. Yankowicz, M. *et al.* Tuning superconductivity in twisted bilayer graphene. *Science* **363,** 1059–1064 (2019).
14. Lu, X. *et al.* Superconductors, orbital magnets and correlated states in magic-angle bilayer graphene. *Nature* **574,** 653–657 (2019).
15. Zondiner, U. *et al.* Cascade of phase transitions and Dirac revivals in magic-angle graphene. *Nature* **582,** 203–208 (2020).
16. Wong, D. *et al.* Cascade of electronic transitions in magic-angle twisted bi-layer gtaphene. *Nature* **582,** 198–202 (2020).
17. Xu, C. & Balents, L. Topological superconductivity in twisted multi-layer graphene. *Phys. Rev. Lett.* **121,** 087001 (2018).
18. Scheurer, M. S. & Samajdar, R. Pairing in graphene-based moiré superlattices. *Phys. Rev. Lett.* **2,** 033062 (2020).
19. Cao, Y. *et al.* Pauli limit violation and re-entrant superconductivity in moiré graphene. *Nature* **595,** 526–531 (2021).
20. Kittel, C. *Quantum theory of solids* (John Wiley & Sons, New York, London, 1963).
21. Landau, L. D. & Lifshitz, E. M. *Statistical Physics, volume 5, Course of theoretical physics* (Butterwert–Heinemann, Oxford, Boston, Johannesburgh, Melburn, New Delhi, Singapore, 1990).
22. Ter Haar, D. *Collected papers of L. D. Landau* (Pergamon Press, Oxford, London, Edinburgh, New York, Paris, Frankfurt, 1965).
23. Abramowitz, M. & Stegun, I. A. *Handbook of mathematical functions* (National Bureau of Standards, Washington, 1964).
24. Sheka, E. *Spin chemical physics of graphene* (Pan Stanford Publishing, Singapore, 2018).

## Information about the authors

**Yuri P. Rybakov**—Professor, Doctor of Sciences in Physics and Mathematics, Professor at the Institute of Physical Research and Technologies of RUDN University (e-mail: rybakov-yup@rudn.ru, phone: +7 (916) 262-55-36, ORCID: 0000-0002-7744-9725, ResearcherID: S-4813-2018, Scopus Author ID: 16454766600)

**Medina Umar**—Exchange student of the Institute of Physical Research and Technologies of RUDN University (e-mail: mails4medina@gmail.com)

УДК 538.91
DOI: 10.22363/2658-4670-2024-32-4-445-451 EDN: DFMUXX

# Сверхпроводимость и особая симметрия скрученного трехслойного графена в киральной модели

Ю. П. Рыбаков, М. Умар

Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

**Аннотация.** Сверхпроводящие свойства скрученного трёхслойного графена изучаются в рамках киральной модели, основанной на использовании унитарной матрицы $U \in SU(2)$ в качестве параметра порядка. Для проверки сверхпроводящего поведения этой системы включается взаимодействие с внешним магнитным полем $B_0$, ориентированным вдоль листа графена, и вычисляется внутренняя магнитная напряжённость в центре как функция угла закручивания. Обращение этой функции в нуль, вследствие эффекта Мейсснера, являющегося важной особенностью сверхпроводимости, вычисляется соответствующая зависимость магического угла закручивания от $B_0$. Обсуждается необычный эффект возвратной сверхпроводимости при больших значениях $B_0$.

**Ключевые слова:** трёхслойный графен, киральная модель, сверхпроводимость