ISSN 2500-316X (Online)

RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ ЖУРНАЛ

Information systems.
Computer sciences.
Issues of information security

Multiple robots (robotic centers) and systems.
Remote sensing and non-destructive testing

Modern radio engineering and telecommunication systems

Micro- and nanoelectronics.
Condensed matter physics

Analytical instrument engineering and technology

Mathematical modeling

Economics of knowledge-intensive and high-tech enterprises and industries.
Management in organizational systems

Product quality management. Standardization

Philosophical foundations of technology and society

12+
www.rtj-mirea.ru

МИРЭА
Российский технологический университет

11(4) 2023

ISSN 2500-316X (Online)

RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ ЖУРНАЛ

- Information systems. Computer sciences. Issues of information security
- Multiple robots (robotic centers) and systems. Remote sensing and non-destructive testing
- Modern radio engineering and telecommunication systems
- Micro- and nanoelectronics. Condensed matter physics
- Analytical instrument engineering and technology
- Mathematical modeling
- Economics of knowledge-intensive and high-tech enterprises and industries. Management in organizational systems
- Product quality management. Standardization
- Philosophical foundations of technology and society

- Информационные системы. Информатика. Проблемы информационной безопасности
- Роботизированные комплексы и системы. Технологии дистанционного зондирования и неразрушающего контроля
- Современные радиотехнические и телекоммуникационные системы
- Микро- и наноэлектроника. Физика конденсированного состояния
- Аналитическое приборостроение и технологии
- Математическое моделирование
- Экономика наукоемких и высокотехнологичных предприятий и производств. Управление в организационных системах
- Управление качеством продукции. Стандартизация
- Мировоззренческие основы технологии и общества

Russian Technological Journal
2023, Vol. 11, No. 4

Russian Technological Journal
2023, том 11, № 4

https://www.rtj-mirea.ru

https://doi.org/10.32362/2500-316X

ISSN 2500-316X (Online)

Russian Technological Journal
2023, Vol. 11, No. 4

Publication date July 31, 2023.

The peer-reviewed scientific and technical journal highlights the issues of complex development of radio engineering, telecommunication and information systems, electronics and informatics, as well as the results of fundamental and applied interdisciplinary researches, technological and economical developments aimed at the development and improvement of the modern technological base.

Periodicity: bimonthly.
The journal was founded in December 2013. The titles were «Herald of MSTU MIREA» until 2016 (ISSN 2313-5026) and «Rossiiskii tekhnologicheskii zhurnal» from January 2016 until July 2021 (ISSN 2500-316X).

Founder and Publisher:
Federal State Budget
Educational Institution of Higher Education
«MIREA – Russian Technological University»
78, Vernadskogo pr., Moscow, 119454 Russia.

The journal is included into the List of peer-reviewed science press of the State Commission for Academic Degrees and Titles of Russian Federation. The Journal is included in Russian State Library (RSL), Russian Science Citation Index, eLibrary, Socionet, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Ulrich's International Periodicals Directory.

The registration number ПИ № ФС 77 - 81733 was issued in August 19, 2021 by the Federal Service for Supervision of Communications, Information Technology, and Mass Media of Russia.

The subscription index of *Pressa Rossii*: **79641**.

Russian Technological Journal
2023, том 11, № 4

Дата опубликования 31 июля 2023 г.

Научно-технический рецензируемый журнал освещает вопросы комплексного развития радиотехнических, телекоммуникационных и информационных систем, электроники и информатики, а также результаты фундаментальных и прикладных междисциплинарных исследований, технологических и организационно-экономических разработок, направленных на развитие и совершенствование современной технологической базы.

Периодичность: один раз в два месяца.
Журнал основан в декабре 2013 года. До 2016 г. издавался под названием «Вестник МГТУ МИРЭА» (ISSN 2313-5026), а с января 2016 г. по июль 2021 г. под названием «Российский технологический журнал» (ISSN 2500-316X).

Учредитель и издатель:
федеральное государственное бюджетное образовательное учреждение высшего образования
«МИРЭА – Российский технологический университет»
119454, РФ, г. Москва, пр-т Вернадского, д. 78.

Журнал входит в Перечень ведущих рецензируемых научных журналов ВАК РФ, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук и доктора наук, индексируется в РГБ, РИНЦ, eLibrary, Соционет, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Ulrich's International Periodicals Directory.

Регистрационный номер и дата принятия решения о регистрации СМИ ПИ № ФС 77 - 81733 от 19.08.2021 г. СМИ зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор).

Индекс по объединенному каталогу «Пресса России» **79641**.

https://www.rtj-mirea.ru

Editorial Board

Редакционная коллегия

Contents

Содержание

Information systems. Computer sciences. Issues of information security
Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.89
https://doi.org/10.32362/2500-316X-2023-11-4-7-15

RESEARCH ARTICLE

Password strength verification based on machine learning algorithms and LSTM recurrent neural networks

Vladimir V. Belikov [1, @], Ivan A. Prokuronov [2]

[1] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[2] *SFB Laboratory, Moscow, 127083 Russia*
[@] *Corresponding author, e-mail: belikov_v@mirea.ru*

Abstract

Objectives. One of the most commonly used authentication methods in computer systems, password authentication is susceptible to various attacks including brute-force and dictionary attacks. This susceptibility requires not only the strict protection of user credentials, but also the definition of criteria for increasing a password's strength to minimize the possibility of its exploitation by an attacker. Thus, an important task is the development of a verifier for checking passwords for strength and prohibiting the user from setting passwords that are susceptible to cracking. The use of machine learning methods to construct a verifier involves algorithms for formulating requirements for password complexity based on lists of known passwords available for each strength category.

Methods. The proposed supervised machine learning algorithms comprise support vector machines, random forest, boosting, and long short-term memory (LSTM) recurrent neural network types. Embedding and term frequency–inverse document frequency (TF-IDF) methods are used for data preprocessing, while cross-validation is used for selecting hyperparameters.

Results. Password strength recommendations and requirements from international and Russian standards are described. The existing methods of password strength verification in various operating systems are analyzed. The experimental results based on existing datasets comprising passwords having an associated level of strength are presented.

Conclusions. A LSTM recurrent neural network is highlighted as one of the most promising areas for building a password strength verifier.

Keywords: cybersecurity, password strength, supervised machine learning, recurrent neural network, LSTM

• **Submitted:** 11.12.2022 • **Revised:** 01.02.2023 • **Accepted:** 02.05.2023

For citation: Belikov V.V., Prokuronov I.A. Password strength verification based on machine learning algorithms and LSTM recurrent neural networks. *Russ. Technol. J*. 2023;11(4):7–15. https://doi.org/10.32362/2500-316X-2023-11-4-7-15

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Построение верификатора стойкости пароля с использованием классических методов машинного обучения и рекуррентной LSTM нейронной сети

В.В. Беликов [1, @],
И.А. Прокуронов [2]

[1] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[2] *СФБ Лаборатория, Москва, 127083 Россия*
[@] *Автор для переписки, e-mail: belikov_v@mirea.ru*

Резюме

Цель. Аутентификация с использованием паролей является одним из наиболее распространенных способов проверки подлинности в компьютерных системах. Существующие атаки на пароли, включающие в себя, в т.ч. атаки перебора и атаки по словарю, требуют не только защиты учетных данных пользователя на этапе эксплуатации паролей, но и определения требований к паролю, позволяющих повысить стойкость пароля к атакам, минимизируя возможность их реализации злоумышленником. Важной задачей при этом становится разработка верификатора, осуществляющего проверку пароля на стойкость и позволяющего исключить задание пользователем паролей, подверженных взлому. Построение верификатора с использованием методов машинного обучения позволяет алгоритмам самим формулировать требования к сложности пароля в произвольно комплексной форме, отталкиваясь только от инцидентов, имеющихся для каждой категории стойкости списков известных паролей.

Методы. Предложены алгоритмы машинного обучения с учителем: метод опорных векторов, случайный лес, бустинг, рекуррентная LSTM (long short-term memory) нейронная сеть. В эксперименте для предобработки данных применены метод простой индексации символов с последующей обработкой embedding-слоем и метод TF-IDF (term frequency-inverse document frequency). Для выбора гиперпараметров алгоритмов была использована кросс-валидация.

Результаты. Проведен анализ рекомендаций и требований к паролям в международных и отечественных стандартах и возможности их реализации в виде верификатора стойкости пароля в различных операционных системах. Приведены результаты эксперимента на существующем наборе помеченных по уровню стойкости паролей. Проведена их оценка с использованием macro f1-меры.

Выводы. Использование рекуррентной LSTM нейронной сети выделено как одно из наиболее перспективных направлений для построения верификатора стойкости пароля.

Ключевые слова: компьютерная безопасность, стойкость пароля, машинное обучение с учителем, рекуррентная нейронная сеть, LSTM

• **Поступила:** 11.12.2022 • **Доработана:** 01.02.2023 • **Принята к опубликованию:** 02.05.2023

Для цитирования: Беликов В.В., Прокуронов И.А. Построение верификатора стойкости пароля с использованием классических методов машинного обучения и рекуррентной LSTM нейронной сети. *Russ. Technol. J.* 2023;11(4):7–15. https://doi.org/10.32362/2500-316X-2023-11-4-7-15

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Password authentication represents one of the most common authentication methods used in computer systems [1]. Existing password threats, including brute-force-, dictionary-, and rainbow table attacks, require not only the protection of user credentials during password exploitation, but also the definition of password requirements when setting a password. Such requirements should increase password strength to withstand specified attacks, minimizing the possibility of their being compromised by an attacker [2]. In this connection, an important task involves the development of a verifier checking the password for strength to prohibit users from setting passwords that are vulnerable to cracking [3].

ANALYSIS OF EXISTING APPROACHES TO BUILDING A PASSWORD STRENGTH VERIFIER

The most commonly used Russian standards for setting a strong password are set out [4] in the following documents:

1) National Institute of Standards and Technology (NIST) Special Publication "Digital Identity Guidelines. Authentication and Lifecycle Management"[1];
2) Methodological document of the Federal Service for Technical and Export Control (FSTEC) "Information Protection Measures in State Information Systems"[2].

These guidelines contain recommendations for users when creating passwords, as well as requirements and recommendations for verifiers (websites, software, etc.) that contain a system for checking and processing passwords.

The first document refers to passwords as "memorized secrets." Among the general provisions are the following:

1) memorized secrets (passwords) must be at least 8 characters long if selected by the user;
2) memorized secrets (passwords) randomly generated by the computer or verifier must be at least 6 characters long and may consist entirely of digits;
3) if the computer or verifier prohibits the selected memorized secret (password) on the grounds that it is contained in the previously accepted blacklist of compromised values, the user should select another memorized secret (password).

In the FSTEC document, the recommended minimum password length of 6 characters achieves the requirements of the simplest level (level 4) of personal data security [5].

When processing requests to establish and change memorized secrets, verifiers should compare assumed secrets with a list of frequently used, expected or compromised passwords. For example, the list may include:

1) passwords obtained from the database of cracked passwords;
2) vocabulary words;
3) repetitive or sequential characters (for example, "qqqqqq," "qwerty12345");
4) context-dependent words such as service name, user name, and their derivatives (e.g., "mireastudent," "ivanivanov").

A convenient way to check a password for strength is to use special services such as those offered on the Kaspersky website[3]. This service provides information on password strength, occurrence of the password in leaked databases, as well as how long it would take to crack a particular password using a brute-force attack (Fig. 1).

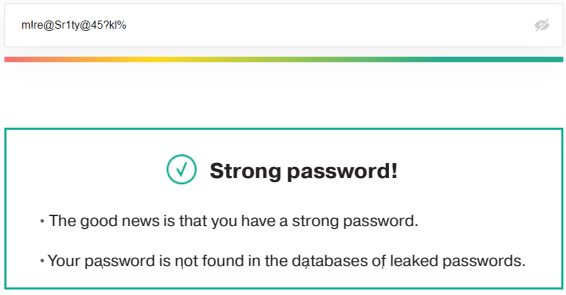


Fig. 1. Demonstration of password verification on the Kaspersky website

The complexity of user-selected passwords can be characterized using the concept of information theory, whose founder is widely considered to be the American mathematician, Claude Shannon. Although entropy, representing a measure of the information capacity of the system, can be easily computed for data with deterministic distribution functions, evaluation of entropy for user-chosen passwords represents a complicated task, which defeats attempts to derive an exact result on this basis. For this reason, for example, the NIST publication uses an approach based primarily on password length.

[1] https://pages.nist.gov/800-63-3/sp800-63b.html. Accessed February 01, 2023.

[2] https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-11-fevralya-2014-g (in Russ.). Accessed February 01, 2023.

[3] https://password.kaspersky.com/ru/ (in Russ.). Accessed February 01, 2023.

In order to provide password strength testing in different operating systems, special modules are used to evaluate passwords against some criteria that adhere to generally accepted recommendations, as well as offering the possibility of setting such criteria by the system administrator. In general, this is referred to in terms of password policy, comprising a set of rules improving the security of user accounts by encouraging them to use stronger passwords.

Linux operating systems have long used the pluggable authentication module (PAM) *pam_cracklib* designed for verifying passwords by dictionary words [6]. In recent versions of Linux, this module has been replaced by the *pam_pwquality* module based on the *pam_cracklib* module, which is fully backward-compatible with its predecessors. This module offers a simpler password policy approach for verifying that users have conformed with the administrator's password requirements.

The following list of requirements represents an example of a password policy set using the specified module:

1) minimum password length is 10 characters;
2) the new password must contain 6 new characters not present in the old password;
3) the password must contain lowercase letters, uppercase letters, special characters, and numbers;
4) the password must not contain more than two characters in a row;
5) the password must have no more than six consecutive characters from the same class;
6) the presence of GESOC check;
7) prohibiting the words "mirea," "security," "admin," "password," and "cyber."

The configuration file */etc/security/pwquality.conf* reflecting the described security policy is shown in Fig. 2.

Fig. 2. Configuration file *pwquality.conf*

The *pwscore* utility included in the *libpwquality* package can be used for checking the correctness of the specified password policy. An example of testing the password that contains a forbidden word and one that meets all requirements using the *pwscore* utility, which gives a score from 0 to 100, is shown in Fig. 3.

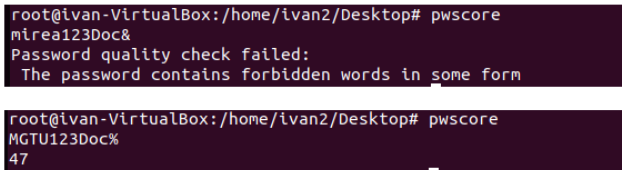


Fig. 3. Password testing

In Windows operating systems (OS), the password policy setting is set in the "Local Security Policy" section of the *gpedit.msc* utility [7] (Fig. 4).

Offering only eight configurable options, Windows OS allows the password policy to be set and changed according to strictness criteria, including and/or combining the set values of these options. One of the most important of these is the "Password must meet complexity requirements" option. Enabling this option specifies that all passwords should meet the following requirements:

1) the password must not contain a user account name or parts of a full user name longer than two consecutive characters;
2) the minimum password length is 6 characters;
3) the password must contain characters from at least three classes: Latin uppercase letters, Latin lowercase letters, numbers, and special characters. In Windows 11, the fourth class requiring any Unicode character is added.

Thus, the analysis of existing approaches to password strength verifier construction shows that the most common solution consists in the allocation of rules focusing primarily on password length and its presence in existing password dictionaries. This solution has a number of disadvantages including the difficulty of implementing more complex requirements reflecting rigorous password-strength properties. In contrast, machine-learning approaches to building a verifier uses algorithms to formulate these requirements themselves in an arbitrarily complex form relying only on incidents that comprise lists of known passwords available for each strength category. One of the properties of this approach is its versatility, since creating and customizing a password policy (number of special characters, number of digits, etc.) for each specific case is not required. The machine simulates a mixture of many password verification algorithms. The trained machine can be used as a separate module (for example, PAM module in Linux) for ensuring that password strength is accurately determined. For example, if a developer is creating some new social network, it is sufficient to teach the machine using a dataset from other popular social

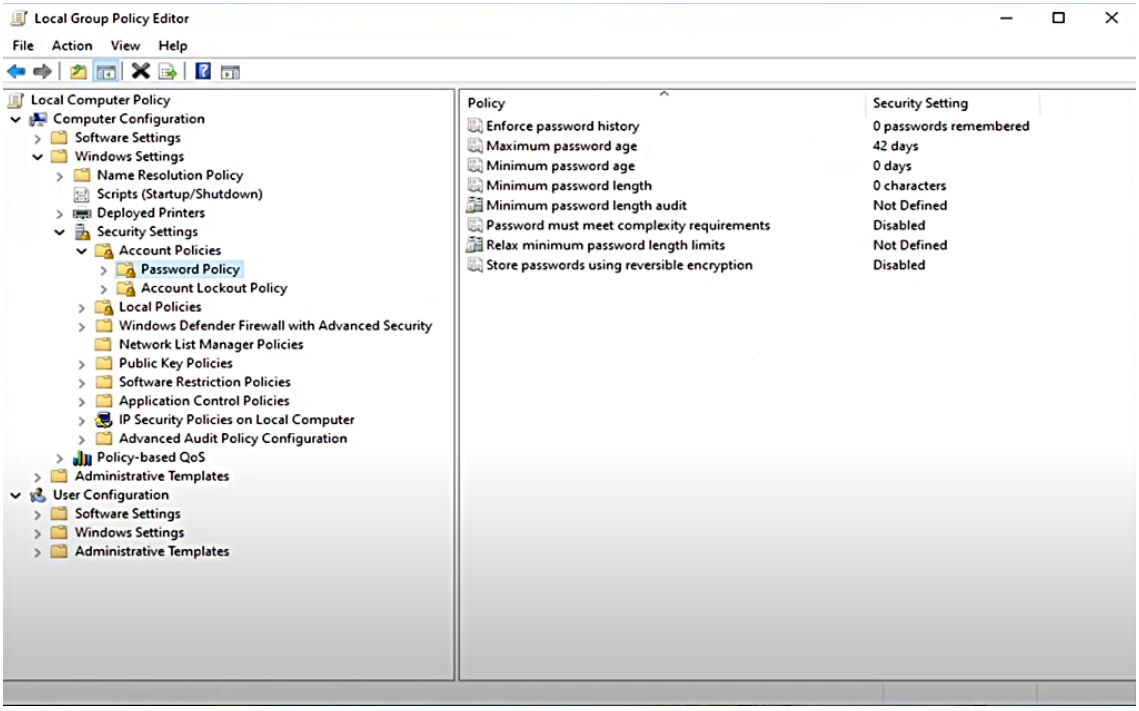


Fig. 4. Password policy settings window in Windows OS

networks, where the password policy is considered to be effective: TikTok[4], Twitter[5] (banned in Russia), Facebook[6] (banned in Russia), etc. The machine is then built in as a verifier evaluating the password and providing the corresponding information to the user when registering. In this case, if the entered password is given the highest stability score, the user can be sure that his/her password is really stable without being bound to specific requirements and recommendations.

METHODOLOGY AND EXPERIMENTAL RESULTS

The task of password strength verification can be presented as a classification problem where the data object is a password while the class is its strength level.

The dataset of one of the largest password leaks, comprising the hosting site 000webhost [8], is used as the dataset for the experiment. Using the *PARS* tool [9] containing many counters evaluating password strength, the password database and its evaluation are based on three different algorithms implemented by Twitter, Microsoft[7], and Battle.net[8], which were created by an independent developer. The dataset itself contains only the passwords rated equally by all three measures. There are three classes of password scores in the dataset (0 is low, 1 is medium, and 2 is high). The dataset contains unbalanced data: 496 649 passwords rated "1," 89 662 passwords rated "0," and 83 113 strong passwords.

The examples of passwords for each strength level are shown in Table 1. The distributions of password length for the whole dataset, as well as for individual strength levels, are shown in Figs. 5 and 6.

Table 1. Examples of passwords from training dataset

Password strength	Password examples
0	wewes19 asdas95
1	oyeleye1 80188063JA
2	JFRTgxTQyNQTh9ZD d7a6AoTMxMw0dLVy

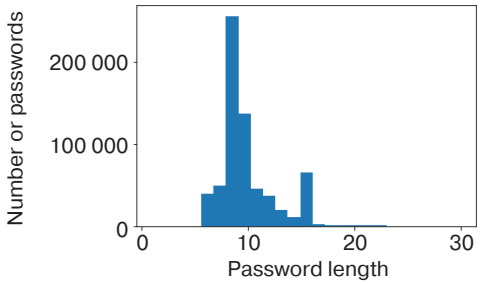


Fig. 5. Histogram of the password length in the dataset

[4] https://www.tiktok.com/. Accessed February 01, 2023.
[5] https://twitter.com/. Accessed February 01, 2023.
[6] http://facebook.com/. Accessed February 01, 2023.
[7] https://www.microsoft.com/. Accessed February 01, 2023.
[8] https://www.battle.net. Accessed February 01, 2023.

Table 2. Learning results evaluated on the training dataset by cross–validation

Algorithm	Found hyperparameter	Macro f1-measure
Support vector machine	Regularization coefficient: 1	0.806
Random forest	Maximum tree depth: 32	0.903
Boosting	Maximum tree depth: 32	0.946

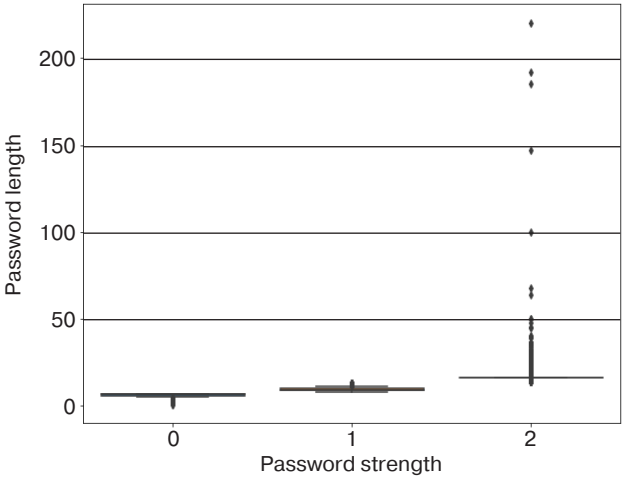


Fig. 6. Distribution of password length depending on strength in the dataset

For the experiment, the set is split into a training set (80%) and a test set (20%) after random mixing.

The term frequency–inverse document frequency (TF-IDF) method [10] is used for representing the text password as a vector of numeric values in order to consider not only the presence, but also the weight of each individual character, as follows:

$$TF = \frac{\text{number of character occurencies in password}}{\text{total number of characters in password}},$$

$$IDF = \log_2 \frac{\text{total number of passwords}}{\text{number of passwords with character}},$$

$$TFIDF = TF \cdot IDF,$$

After applying the TF-IDF method to the set of passwords, a set is obtained where each password has the same length as the size of the dictionary, while each element vector has TF-IDF weight of the character whose serial number in the dictionary corresponds to the serial number of the vector element.

The equalization of the number of instances in each class with an undersampling technique and k-fold cross–validation [11] with $k = 5$ is used for learning classical algorithms and for selecting hyperparameters, respectively. The results of learning the algorithms on the training dataset evaluated using the macro f1-measure [12] are presented in Table 2.

Table 3. Learning results evaluated on the test set

Value	Boosting	LSTM recurrent network
Precision	0.972	0.9995
Accuracy	0.971	0.9994
Completeness	0.971	0.9990
Macro f1-measure	0.971	0.9992

At the same time, the use of classical machine learning methods is complicated by the need to transform passwords using TF-IDF algorithm; due to the necessity of recalculating weights when the password set is updated, this is a costly procedure. In addition, these algorithms represent passwords as an unordered set of characters (a simplification used in the "bag of words"[9] approach), which does not consider the mutual arrangement of characters. This can be partially fixed by using bigram- and trigram models, which in turn makes the learning process much more difficult. In contrast, recurrent neural networks work directly with character sequences of arbitrary length [13]. Thus, for a recurrent neural network, the "PASSWORD" and "AWSSODPR" passwords are determined by two different vectors, whereas when using TF-IDF they are determined by one. Therefore, for evaluating password strength in the experiment, the neural network implemented using the PyTorch library [14] is also used as a method alternative to classical machine learning algorithms. The network consists of the following layers:

1) embedding layer, used for converting a password consisting of characters into a vector of numeric values;
2) long short-term memory (LSTM) layer, representing a special kind of recurrent neural network architecture, capable of learning long-term dependencies (which is important when working with long-length passwords [15]);
3) linear layer used for converting the internal state of LSTM layer into category scores.

The results of comparing the performance of the boosting algorithm and recurrent LSTM network on the test set are shown in Table 3.

The source code of the experiment can be found on the website[10].

[9] A simplified representation of the text as a bag (multiset) of its words without any regard to grammar or word order but with retention of information about their number.

[10] https://github.com/james116blue/password_strength_verifier. Accessed February 01, 2023.

It should be noted that the approach used in the experiment in no way reduces the hacker's capabilities in the case of phishing-, keylogger-, and man-in-the-middle attacks. However, according to the 2020 Data Breach Investigation Report [16], 89% of all hackings involve some type of credential abuse (brute force attacks and their subtypes, as well as attacks aimed at reusing credentials). This fact suggests that the approach to password strength using machine learning minimizes the risks for most attack vectors, thus also explaining a significant amount of current research in this area [17–21].

Thus, it may be concluded that the recurrent LSTM network outperforms conventional machine learning methods, bringing the password strength classification quality score closer to one.

CONCLUSIONS

The present work, which proposes an approach to building a password strength verifier using machine learning methods, compares several algorithms on a set of password data tagged by strength level. This approach is shown to have a number of advantages over classical password strength verification methods operating without machine learning techniques. Thus, using machine learning methods for verifier construction allows the formulation of password strength requirements in an arbitrary complex form, which relies on incidents only. In addition, the proposed approach allows better resistance to attacks comprising a subset of brute force attacks, as well as rainbow table attacks. In the first case, this is achieved due to the fact that a complex password whose strength is evaluated using the machine learning algorithm makes brute force attack impossible due to the huge time required for the attacker to accomplish this task. In the second case, attacking a complex password would only be possible by applying a large rainbow table, which would require, in turn, the use of a significant amount of attacker resources, thus making the risks of such an attack almost trivial.

Among the considered machine learning methods, recurrent neural networks demonstrate a particular efficiency. Here, the representation learning of the text password as a vector of numerical values, consisting in finding the embedding layer weights, occurs simultaneously with training of the entire network to maximize classification accuracy. This allows the neural network itself to choose such a password vectorization that is effective exactly for the task at hand. In addition, neural networks work with a sequence of characters, not just their presence in the password, thus allowing avoiding the simplification used in the "bag of words" approach. This allows the recurrent neural network to be recommended as one of the most promising research areas for building a password strength verifier.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Conklin A., Dietrich G., Walz D. Password-based authentication: a system perspective. In: *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*. 2004; IEEE. https://doi.org/10.1109/HICSS.2004.1265412
2. Dell'Amico M., Michiardi P., Roudier Y. Password strength: An empirical analysis. In: *2010 Proceedings IEEE INFOCOM.* 2010; IEEE. https://doi.org/10.1109/INFCOM.2010.5461951
3. Chakrabarti S., Singhal M. Password-based authentication: Preventing dictionary attacks. *Computer*. 2007;40(6):68–74. https://doi.org/10.1109/MC.2007.216
4. Shay R., Komanduri S., Kelley P.G., Leon P.G., Mazurek M.L., Bauer L., Christin N., Cranor L.F. Encountering stronger password requirements: user attitudes and behaviors. In: *Proceedings of the Sixth Symposium on Usable Privacy and Security.* 2010; Article 2. https://doi.org/10.1145/1837110.1837113
5. Selifanov V.V. Evaluation of the efficiency of the information protection system of state information systems from unauthorized access. In: *Integration of Science, Society, production and Industry: Collection of Articles of the International Scientific and Practical Conference*. 2016. P. 109–113 (in Russ.).
6. Ferreira J.F., Johnson S.A., Mendes A., Brooke P.J. Certified password quality: a case study using Coq and Linux pluggable authentication modules. In: *Integrated Formal Methods. IFM 2017. Lecture Notes in Computer Science.* V. 10510. Springer International Publishing; 2017. P. 407–421. https://doi.org/10.1007/978-3-319-66845-1_27
7. Alshare K.A., Lane P.L., Lane M.R. Information security policy compliance: a higher education case study. *Information & Computer Security*. 2018;26(1):91–108. https://doi.org/10.1108/ICS-09-2016-0073
8. AlSabah M., Oligeri G., Riley R. Your culture is in your password: An analysis of a demographically-diverse password dataset. *Computers & Security*. 2018;77:427–441. https://doi.org/10.1016/j.cose.2018.03.014
9. Ji S., Yang S., Wang T., Liu C., Lee W.H., Beyah R. Pars: A uniform and open-source password analysis and research system. In: *ACSAC' 15*: *Proceedings of the 31st Annual Computer Security Applications Conference.* 2015. P. 321–330. https://doi.org/10.1145/2818000.2818018
10. Aizawa A. An information-theoretic perspective of TF–IDF measures. *Information Processing & Management*. 2003;39(1):45–65. https://doi.org/10.1016/S0306-4573(02)00021-3

11. Bishop C.M. *Pattern Recognition and Machine Learning*. New York: Springer; 2006. 738 p.
12. Lever J., Krzywinski M., Altman N. Classification evaluation: It is important to understand both what a classification metric expresses and what it hides. *Nat. Methods*. 2016;13(8):603–604. https://doi.org/10.1038/nmeth.3945
13. Medsker L.R., Jain L.C. (Eds.). *Recurrent Neural Networks*. *Design and Applications*. CRC Press; 2001. P. 64–67.
14. Imambi S., Prakash K.B., Kanagachidambaresan G.R. PyTorch. In: Prakash K.B., Kanagachidambaresan G.R. (Eds.). *Programming with TensorFlow*. EAI/Springer Innovations in Communication and Computing (book series). Springer; 2021. P. 87–104. https://doi.org/10.1007/978-3-030-57077-4_10
15. Yu Y., Si X., Hu C., Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput*. 2019;31(7):1235–1270. https://doi.org/10.1162/neco_a_01199
16. Jartelius M. The 2020 Data Breach Investigations Report–a CSO's perspective. *Network Security*. 2020;2020(7):9–12. https://doi.org/10.1016/S1353-4858(20)30079-9
17. Sarkar S., Nandan M. Password Strength Analysis and its Classification by Applying Machine Learning Based Techniques. In: *2022 Second International Conference on Computer Science, Engineering and Applications* (*ICCSEA*). IEEE, 2022. P. 1–5. https://doi.org/10.1109/ICCSEA54677.2022.9936117
18. Sakya S.S., Mauparna M.N. Building a Multi-class Password Strength Generator and Classifier Model by Augmenting Supervised Machine Learning Techniques. Preprint. 2022. https://doi.org/10.21203/rs.3.rs-1820885/v1
19. Murmu S., Kasyap H., Tripathy S. PassMon: A Technique for Password Generation and Strength Estimation. *J. Network Syst. Manage*. 2022;30(1):13. https://doi.org/10.1007/s10922-021-09620-w
20. Tran L., Nguyen T., Seo C., Kim H., Choi D. *A Survey on Password Guessing*. *arXiv preprint arXiv*:2212.08796. 2022. https://doi.org/10.48550/arXiv.2212.08796
21. Xiao Y., Zeng J. Dynamically generate password policy via Zipf distribution. *IEEE Transactions on Information Forensics and Security*. 2022;17:835–848. https://doi.org/10.1109/TIFS.2022.3152357

About the authors

Vladimir V. Belikov, Cand. Sci. (Military), Assistant Professor, Department of Information Security, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: belikov_v@mirea.ru. Scopus Author ID 57983605100, https://orcid.org/0000-0003-1423-1072

Ivan A. Prokuronov, Cryptographic Analysis Specialist, SFB Laboratory (56/2, Mishina ul., Moscow, 127083 Russia). E-mail: miltumultik@gmail.com. https://orcid.org/0000-0003-0999-311X

Об авторах

Беликов Владимир Вячеславович, к.воен.н., доцент, доцент кафедры информационной безопасности № 252 Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: belikov_v@mirea.ru. Scopus Author ID 57983605100, https://orcid.org/0000-0003-1423-1072

Прокуронов Иван Андреевич, специалист инженерно-криптографического анализа, ООО «СФБ Лаборатория» (127083, Россия, Москва, ул. Мишина, д. 56, стр. 2). E-mail: miltumultik@gmail.com. https://orcid.org/0000-0003-0999-311X

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Information systems. Computer sciences. Issues of information security
Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.89
https://doi.org/10.32362/2500-316X-2023-11-4-16-25

RESEARCH ARTICLE

Approach to knowledge management and the development of a multi-agent knowledge representation and processing system

Evgeniy I. Zaytsev [@], **Elena V. Nurmatova**

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] *Corresponding author, e-mail: zajcev@mirea.ru*

Abstract

Objectives. A multi-agent knowledge representation and processing system (MKRPS) comprises a distributed artificial intelligence system designed to solve problems that are difficult or impossible to solve using monolithic systems. Solving complex problems in an MKRPS is accomplished by communities of intelligent software agents that use cognitive data structures, logical inference, and machine learning. Intelligent software agents are able to act rationally under conditions of incompleteness and ambiguity of incoming information. The aim of the present work is to identify models and methods, as well as software modules and tools, for use in developing a highly efficient MKRPS.

Methods. Agent-based modeling methods were used to formally describe and programmatically simulate the rational behavior of intelligent agents, expert evaluation methods, the mathematical apparatus of automata theory, Markov chains, fuzzy logic, neural networks, and reinforcement learning.

Results. An MKRPS structure diagram, a multi-agent solver, and microservices access control diagram were developed. Methods for distribution of intelligent software agents on the MKRPS nodes are proposed along with algorithms for optimizing the logical structure of the distributed knowledge base (DKB) to improve the performance of the MKRPS in terms of volume, cost and time criteria.

Conclusions. The proposed approach to the development and use of intelligent software agents combines knowledge-based reasoning mechanisms with neural network models. The developed MKRPS structure and DKB control diagram includes described methods for optimizing the DKB, determining the availability of microservices used by the agents, ensuring the reliability assurance and coordinated functioning of the computing nodes of the system, as well as instrumental software tools to simplify the design and implementation of the MKRPS. The results demonstrate the effectiveness of the presented approach to knowledge management and the development of a high-performance problem-oriented MKRPS.

Keywords: multi-agent system, intelligent software agents, multi-agent intelligent solver, knowledge representation and processing system, reinforcement learning

• **Submitted:** 24.10.2022 • **Revised:** 27.01.2023 • **Accepted:** 02.05.2023

For citation: Zaytsev E.I., Nurmatova E.V. Approach to knowledge management and the development of a multi-agent knowledge representation and processing system. *Russ. Technol. J.* 2023;11(4):16–25. https://doi.org/10.32362/2500-316X-2023-11-4-16-25

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

О подходе к управлению знаниями и разработке мультиагентной системы представления и обработки знаний

Е.И. Зайцев [@], **Е.В. Нурматова**

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: zajcev@mirea.ru*

Резюме

Цели. Мультиагентная система представления и обработки знаний (МСПОЗ) – это распределенная система искусственного интеллекта, предназначенная для решения проблем, которые трудно или невозможно решить с помощью монолитной интеллектуальной системы. Решение сложных проблем в МСПОЗ осуществляется интеллектуальными программными агентами, которые инкапсулируют в программных классах когнитивные структуры данных, методы логического вывода и машинного обучения. Интеллектуальные программные агенты МСПОЗ способны рационально действовать в условиях неполноты и нечеткости поступающей информации. Целями работы являются исследование и разработка моделей, методов, программных модулей и инструментальных программных средств, которые позволяют создать высокоэффективную МСПОЗ.

Методы. В работе использовались методы агентного моделирования, позволяющие формально описывать и программно имитировать рациональное поведение интеллектуальных агентов, методы экспертных оценок, математический аппарат теории автоматов, марковские цепи, нечеткая логика, нейронные сети, алгоритмы машинного обучения с подкреплением.

Результаты. Разработаны структурная схема МСПОЗ, мультиагентный решатель, схема управления доступом к микросервисам. Предложены методы распределения интеллектуальных программных агентов по узлам МСПОЗ, а также алгоритмы оптимизации логической структуры распределенной базы знаний (РБЗ), позволяющие повысить эффективность объемных, стоимостных и временны́х характеристик МСПОЗ.

Выводы. Предложен подход к разработке и использованию интеллектуальных программных агентов, который объединяет механизмы рассуждений на основе знаний с нейросетевыми моделями. Разработаны структура МСПОЗ, схема управления РБЗ, методы оптимизации РБЗ, определения доступности используемых агентами микросервисов, обеспечения надежности и скоординированного функционирования вычислительных узлов системы, а также инструментальные программные средства, позволяющие упростить процесс проектирования и реализации МСПОЗ. Полученные результаты демонстрируют эффективность представленного подхода к управлению знаниями и разработке высокопроизводительной проблемно-ориентированной МСПОЗ.

Ключевые слова: мультиагентная система, интеллектуальные программные агенты, мультиагентный интеллектуальный решатель, система представления и обработки знаний, обучение с подкреплением

• **Поступила:** 24.10.2022 • **Доработана:** 27.01.2023 • **Принята к опубликованию:** 02.05.2023

Для цитирования: Зайцев Е.И., Нурматова Е.В. О подходе к управлению знаниями и разработке мультиагентной системы представления и обработки знаний. *Russ. Technol. J.* 2023;11(4):16–25. https://doi.org/10.32362/2500-316X-2023-11-4-16-25

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

In a multi-agent knowledge representation and processing system (MKRPS), complex, ill-defined problems are solved by intelligent software agents, which are able to act rationally under conditions of uncertainty, including the incompleteness and ambiguity of incoming information [1–5]. Making decisions and carrying out rational actions, software agents use a knowledge base and event-driven microservices, which are designed as separate interacting processes with lightweight inter-process communications. Agents interact with microservices through events, which can be simple notifications or complex state-supported structures.

The solution of complex problems in the MKRPS is performed by decomposing the problems into subtasks, which are jointly solved by reactive and cognitive application software agents. Both horizontal decomposition, which creates a multi-connected system having a flat structure, and vertical decomposition, which creates a hierarchical system with several levels, are used.

Due to the implementation of Reinforcement Learning (RL) machine learning methods in the MKRPS, the behavior of applied software agents becomes more rational when solving problems repeatedly. The Actor-Critic algorithm [6–10] is used to train the applied software agents in the MKRPS.

In order to improve performance, special application programming interfaces (API) and system software modules associated with system software agents are implemented in the MKRPS. System software agents plan and manage the computational resources of the MKRPS, as well as providing mobility for the application software agents. Application agents can roam nodes of the MKRPS that provide the necessary environment for them. In contrast to containers implemented on the basis of namespaces (e.g., by the Docker platform[1]), specialized LibOS (Library Operating System) modules are used to support the technology of mobile agents in the MKRPS.

The performance of an MKRPS is largely determined by the selected approach for structuring, storing and processing knowledge [11, 12]. A high-performance problem-oriented multi-agent solver has been developed, the logical structure of whose DKB has been optimized to support minimal total processing time of queries and transactions.

MKRPS STRUCTURE

The structural diagram of the MKRPS is shown in Fig. 1. There are two types of applied (intelligent) software agents at each computational node of the MKRPS: reactive and cognitive [13]. Application software agents interact with each other, as well as with system software agents that are part of the external shared user-level library LibOS, which is oriented on the exokernel architecture of the OS. System software agents are used for planning and managing computing resources, as well as load balancing and system monitoring.

Special software methods and Cognitive Data Structures (CDS) associated with cognitive software agents are used to represent agent-based abstractions (goals, desires, intentions, beliefs of agents) and implement logical inferences.

Figure 2 shows an example of a state-transition diagram of a cognitive software agent comprising one of the MKRPS nodes.

As follows from the diagram, a cognitive software agent can be in one of five states, two of which are composite, i.e., they have nested states. A change of the state of a software agent occurs as a result of an event. It is possible to switch to a new state without committing an event, which is carried out immediately after performing actions (or activities) associated with the previous state.

From the “Initialization” state, the cognitive agent switches to the “Choice” composite state. In this state, a strategy is generated and the necessary knowledge source is selected, taking into account the links established during the initialization stage and the informative messages received from other software agents. Then, the cognitive software agent enters the “Coordination” composite state, in which new knowledge sources are activated and the actions of reactive software agents are coordinated.

If the reactive software agents do not find a consistent solution (the “No Solution” state), the cognitive software agent returns to the nested partial solution state of the problem. If a solution is found (the “Solution” state), the data obtained at this stage are used to form new queries to the knowledge base.

To work with the knowledge base, four types of methods are implemented to form queries and process the results of these queries:

- analysis (ANS method) for implementing a logical analysis of events;
- association (ASS method) used to get responses to queries about links between objects and events;
- comparison of events or objects (CMP method);
- object specification (VAL method).

Both explicit and fuzzy queries to the knowledge base can be used to specify objects. When implementing fuzzy queries, different types of affiliation functions can be used; these are chosen by the knowledge engineer on the basis of the results of computational experiments.

Cognitive software agents coordinate the work of reactive software agents associated with local knowledge sources. An example of an interaction diagram of reactive software agents of one of the MKRPS nodes is shown in Fig. 3.

[1] https://www.docker.com/. Accessed March 20, 2023.

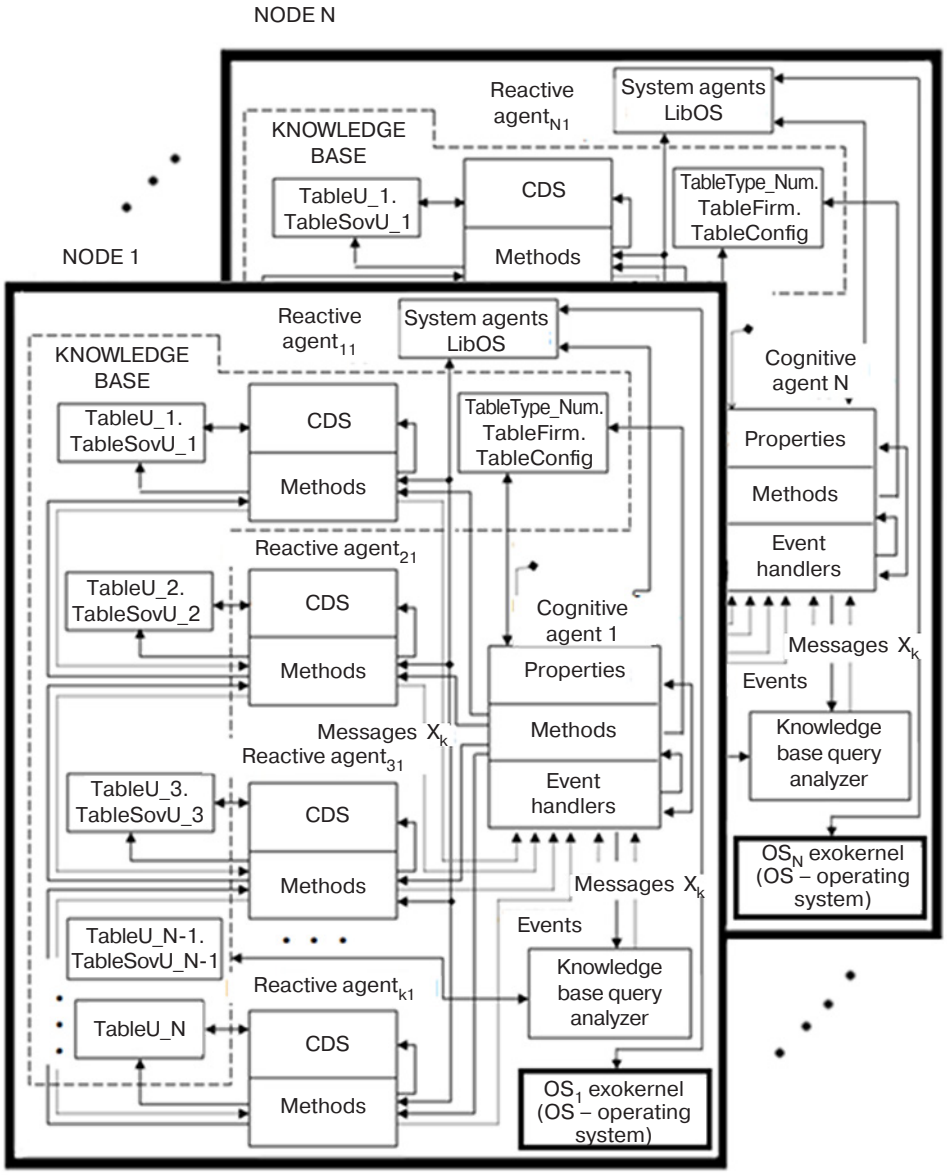


Fig. 1. MKRPS structure

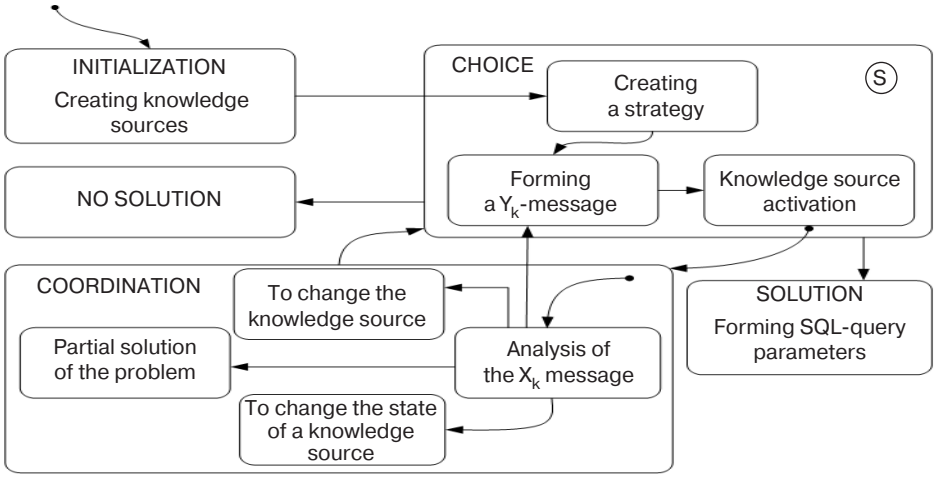


Fig. 2. State-transition diagram of a cognitive software

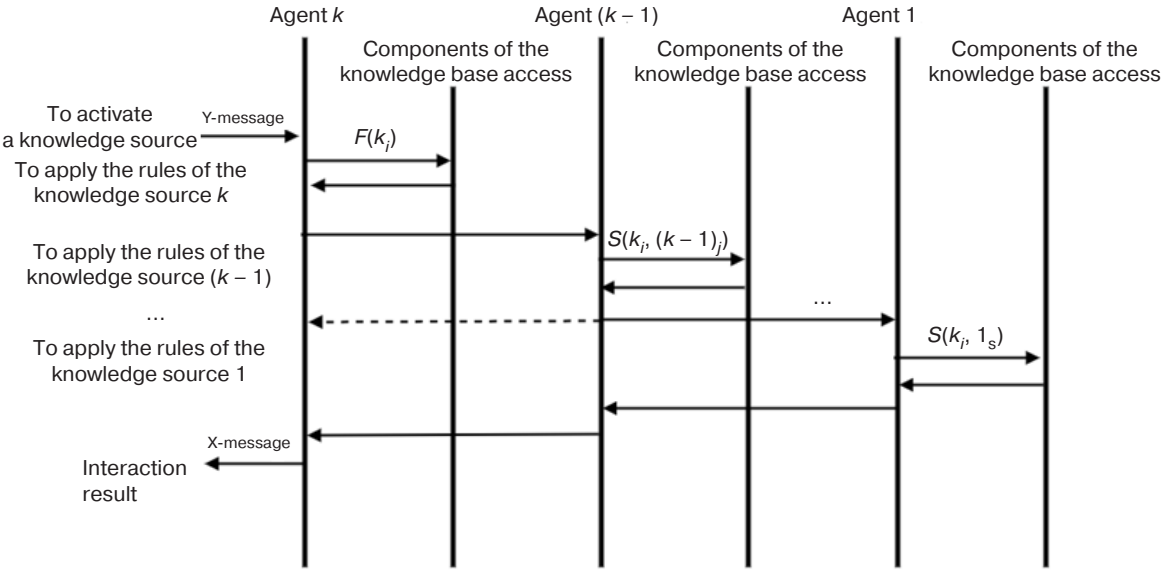


Fig. 3. Diagram of reactive software agent interaction

In the interaction diagram, time moves from top to bottom. Knowledge base access software components and agents are represented by vertical lines. Messages between agents (components) are marked with horizontal arrows. Upon receipt of a message, each agent (source of knowledge) invokes the corresponding program method (a member function of class $F()$ or $S()$) and returns the result. The figure on the left shows the comments. In this case, each reactive software agent with number k interacts only with its nearest neighbor having the number $(k - 1)$. The software agents in this group sequentially perform their assigned tasks within a single process (without switching the context). The priorities of the reactive software agents comprising the associated MKRPS node are set according to the agent's sequence number. The first reactive agent uses high-priority frames associated with TableU_1 and TableSovU_1. The software agent with number k has the lowest priority and is associated with TableU_N.

If the MKRPS node is a multiprocessor system, the agents of this node can act simultaneously. To organize parallel computations, the application software agents of each MKRPS node are distributed into groups using compatibility and inclusion matrices.

The compatibility matrix **S** has the following form:

$$\mathbf{S} = \begin{bmatrix} 0 & s_{12} & s_{13} & \dots & s_{1M} \\ s_{21} & 0 & & \dots & s_{2M} \\ s_{31} & s_{32} & 0 & \dots & s_{3M} \\ \dots & \dots & \dots & \dots & \dots \\ s_{M1} & s_{M2} & s_{M3} & \dots & 0 \end{bmatrix} \begin{matrix} S_1 \\ S_2 \\ S_3 \\ \dots \\ S_M \end{matrix},$$

where $s_{ij} = 1$ if agents A_i and A_j should work in parallel, otherwise $s_{ij} = 0$; S_i is the ith row of the coherence matrix **S**; M is the number of agents.

The inclusion matrix **R** is used to distribute the software agents of the node into groups:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1M} \\ r_{21} & r_{22} & \dots & r_{2M} \\ \dots & \dots & \dots & \dots \\ r_{N1} & r_{N2} & \dots & r_{NM} \end{bmatrix} \begin{matrix} R_1 \\ R_2 \\ \dots \\ R_N \end{matrix},$$

where N is the number of groups; $r_{ij} = 1$ if agent A_i is included in group Y_j. Agent A_i is included in group Y_j if $S_i \cap R_j = \varnothing$, i.e., matrix rows do not intersect.

For optimal partitioning of the set of agents into subsets when using compatibility and inclusion matrices, it is necessary to consider the structure of the MKRPS node, the functional features of software agents, and their requirements to computing resources.

In Multi-Agent Reinforcement Learning (MARL), the environment depends on all software agents. Unlike centralized learning, in which software agents have full control over the computational process (the relevant policies being distributed by a central agent), in the decentralized model used in MKRPS, independent agents can share experiences and policies. In the decentralized model, execution and learning are implemented locally, allowing the application software agents to adapt to the local perception of the environment (Fig. 4).

A decentralized model of multi-agent reinforcement learning allows the use of standard RL algorithms. Applied software agents of MKRPS are trained through a series of rewards and punishments based on the Actor-Critic algorithm, in which a strategy generates actions, while a value function critiques those actions.

Since there is an actor and a critic for each software agent, agents may have different strategies (policies) and

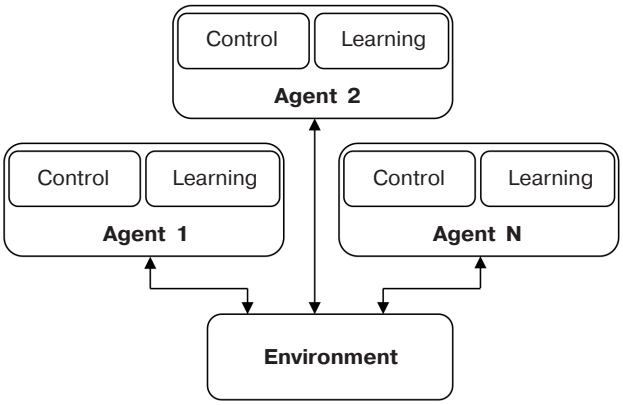


Fig. 4. Decentralized model of multi-agent reinforcement learning

rewards. The agents of MKRPS cooperate to optimize the overall long-term goal. Function approximation in MKRPS is implemented based on neural networks that model both policy and value functions.

For effective implementation of the multi-agent reinforcement learning, problems of multidimensional and multimodal targets, scalability, instability, and optimality must be solved [7].

In the process of solving subtasks, software agents use microservices, which are duplicated on different nodes of the MKRPS to improve reliability and performance. The system software agents distribute the computational load and manage the microservices based on the data provided by the monitor agents (Fig. 5).

SYNTHESIS OF DISTRIBUTED COGNITIVE DATA STRUCTURES

Due to the large dimensionality of the created cognitive data structures, operations with data structures in MKRPS, whether representing data replenishment or retrieval, are performed by keys. To do this, the entire logical CDS structure must be broken down into a number of clusters that have the smallest interconnection under various constraints. In this case, we will accept restrictions on the dimensionality of clusters and restrictions on the degree of semantic proximity of logical records included in the clusters, taking into account the type of storage systems used.

Let us introduce a binary parameter Z_{kj}^{i} to characterize the use by the *k*th query of the *i*th data group

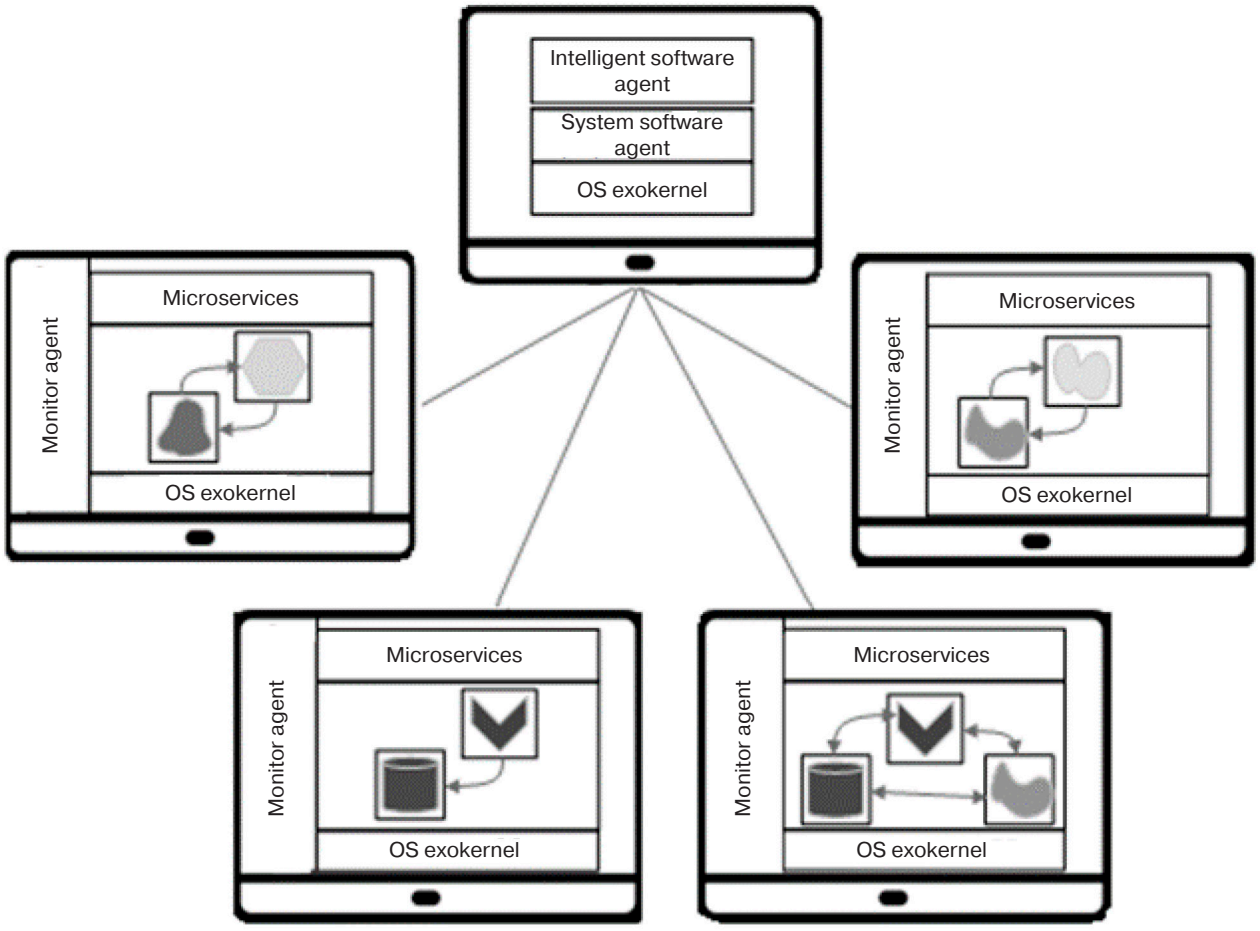


Fig. 5. Microservices and system monitor agents

related to the *j*th logical record. The calculation of this parameter is based on the binary variable a_{ik}, which is equal to one if the *i*th data group is included in the *k*th query, and zero if it is not. This parameter, which is given by the link logic that data groups are included in logical records, is amplified by the variable x_{ij}, which, by analogy, is equal to one if the *i*th data group is included in the *j*th logical record; $x_{ij} = 0$ if it is not.

We have the following calculation of the Z_{kj}^{i} parameter:

$$Z_{kj} = \begin{cases} 1, \text{ at } \sum_{i=1}^{I} a_{ik} x_{ij} \geq 1, \\ 0, \text{ at } \sum_{i=1}^{I} a_{ik} x_{ij} = 0. \end{cases}$$

The complete problem of synthesis of the distributed data structure for MKRPS will be solved by taking into account the criterion of the minimum total time of execution of user requests under such constraints as the uniqueness of the data sets in the record, the length of the logical record, the total number of types of logical records in the structure, the data search time structured by duration of requests, as well as the uniqueness of the input nodes in the structure and their total number.

In the present work, we give an approximate algorithm for solving the problem of synthesis of the optimal distributed data structure by the criterion of minimum total query execution time. Thus, in order to determine the distribution of groups by the criterion of the minimum total traffic, we first use an approximate algorithm for the distribution of data clusters between the server and clients of the local network, then reduce the canonical graph of the data structure to an uncoupled graph with the calculation of the weight of each data group.

The data group weight includes the weights of the group itself, as well as the weights of the arcs, taking into account the requirements of MKRPS users:

$$V_i = V_i^{\text{gr}} + V_{ii'}^{\text{cg}},$$

where V_i^{gr} is the total weight of the data group; $V_{ii'}^{\text{cg}}$ is the weight of the arcs of the connected graph of the canonical data structure; *i'* is the index of the group adjacent to the *i*th data group.

$$V_i^{\text{gr}} = \sum_{k=1}^{k_0} \sum_{p=1}^{p_0} \gamma_{kp}^{\text{q}} \delta_{kp}^{\text{q}} \vartheta_{pi},$$

$$V_{ii'}^{\text{cg}} = \sum_{k=1}^{k_0} \sum_{p=1}^{p_0} \gamma_{kp}^{\text{q}} \delta_{kp}^{\text{q}} \vartheta_{pi} \sum_{i' \neq i}^{I} \vartheta_{pi'} a_{ii'}^{\text{G}},$$

where γ_{kp}^{q} is the frequency of query usage by user; δ_{kp}^{q} are elements of user query matrix; $\vartheta_{pi'}$ is the matrix of data groups in query processing; $a_{ii'}^{\text{G}}$ is the matrix of semantic adjacency of the *i*th data group with the data group that has index *i'*.

For a particular *i*th group, the weight will be:

$$V_i = \sum_{k=1}^{k_0} \sum_{p=1}^{p_0} \gamma_{kp}^{\text{q}} \delta_{kp}^{\text{q}} \vartheta_{pi} \left(1 + \sum_{i' \neq i}^{I} \vartheta_{pi'} a_{ii'}^{\text{G}} \right).$$

Then the graph of the computer network is converted into an unconnected graph with calculation of node weight:

$$V_r = t_r + \sum_{r' \neq r}^{R_0} t_{rr'},$$

where t_r is the total average duration of data processing in the *r*th node, consisting of the time of decomposition of the request into subrequests, route selection, and connection establishment; $t_{rr'}$ is the average duration of data transmission between nodes, determined on the basis of the matrix of logical distances between servers of the computer network nodes.

Then a matrix $\mathbf{V} = \|v_{ir}\|$ is formed, the elements of which are defined as the Cartesian product of the weight of each node by the weight of each data group:

$$v_{ir} = V_i \times V_i \text{ for } i = \overline{1, I}, \; r = \overline{1, R_0}.$$

Next, we solve the problem

$$\min_{\{x_{ir}\}} \sum_{i=1}^{I} \sum_{r=1}^{R_0} v_{ir} x_{ir}$$

with the following restrictions:

- by the number of data groups that can be localized on one node:

$$\sum_{i=1}^{I} x_{ir} \leq N_r, \; r = \overline{1, r_0};$$

- on the permissible redundancy of groups by network nodes:

$$\sum_{r=1}^{r_0} x_{ir} \leq M_i, \; i = \overline{1, I};$$

- the capacity of the available external memory of the data storage system:

$$\sum_{i=1}^{I} x_{ir}\rho_i\pi_i \le \eta_r^{\text{DSD}},$$

where ρ_i are values of data group lengths; π_i is the amount of instances in groups; η_r^{DSD} is the amount of available external data storage device (DSD) memory on the node; $x_{ir} = 1$ if the ith data group is included in the rth network node, $x_{ir} = 0$ if it is not.

The solution of the linear integer programming problem is used not only to determine the optimal localization of data groups by network nodes, but also to define the optimal structure of data groups placed on the network nodes.

At the next stage we solve the problem of optimal distribution of node data groups in each node of the network by types of logical records according to the criterion of minimum total time of local data processing. Here, the number of synthesis tasks is determined by the number of network nodes. The initial data are comprised of the subgraphs of the canonical data structure graph, as well as the temporal and volumetric characteristics of the subgraphs of their canonical structure, a set of user requests, and network nodes [14, 15]. The synthesis problem is solved by approximate algorithms with restrictions on the number of groups in a record, on the uniqueness of the inclusion of groups in a record, on the cost of storing information, and on the total time of the request service. As a result, the logical structures of the database are determined for each node of the network along with the formed distribution matrices of the set of data groups by types of logical records and subsequent distribution matrices of the set of record groups by network nodes.

The results of solving the problem of synthesis of distributed cognitive data structures have important practical application in the design of the optimal DKB structure and the possibility of forming specifications for queries and adjustments of distributed data.

CONCLUSIONS

The paper presents a methodological approach to the development of MKRPS. The functional and structural organization of the multi-agent solver are described in terms of models of applied and system software agents, as well as methods of designing and managing DKB. Methods for distributing software agents to the nodes of MKRPS, determining the availability of microservices, as well as providing reliable and coordinated work of computing nodes, are considered. Examples of interaction-, state- and transition diagrams for cognitive and reactive software agents are given. Algorithms for the optimization of logical DKB structure are described. The optimization of DKB improves the efficiency of MKRPS in terms of time-, volume- and cost characteristics to make it more productive, flexible and functional. The obtained results confirm the effectiveness of the presented approach to the development of MKRPS.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Zaytsev E.I., Khalabiya R.F., Stepanova I.V., Bunina L.V. Multi-Agent System of Knowledge Representation and Processing. In: Kovalev S., Tarassov V., Snasel V., Sukhanov A. (Eds.). *Proceedings of the Fourth International Scientific Conference "Intelligent Information Technologies for Industry"* (*IITI'19*). *Advances in Intelligent Systems and Computing*. Springer; 2020. V. I. P. 131–141. https://doi.org/10.1007/978-3-030-50097-9_14
2. Baranauskas R., Janaviciute A., Jasinevicius R., Jukavicius V. On Multi-Agent Systems Intellectics. *Inf. Technol. Control*. 2015;44(1):112–121. https://doi.org/10.5755/j01.itc.44.1.8768
3. Darweesh S., Shehata H. Performance Evaluation of a Multi-Agent System using Fuzzy Model. *2018 First International Workshop on Deep and Representation Learning* (*IWDRL*). 2018. P. 7–12. https://doi.org/10.1109/IWDRL.2018.8358208

СПИСОК ЛИТЕРАТУРЫ

1. Zaytsev E.I., Khalabiya R.F., Stepanova I.V., Bunina L.V. Multi-Agent System of Knowledge Representation and Processing. In: Kovalev S., Tarassov V., Snasel V., Sukhanov A. (Eds.). *Proceedings of the Fourth International Scientific Conference "Intelligent Information Technologies for Industry"* (*IITI'19*). *Advances in Intelligent Systems and Computing*. Springer; 2020. V. I. P. 131–141. https://doi.org/10.1007/978-3-030-50097-9_14
2. Baranauskas R., Janaviciute A., Jasinevicius R., Jukavicius V. On Multi-Agent Systems Intellectics. *Inf. Technol. Control*. 2015;44(1):112–121. https://doi.org/10.5755/j01.itc.44.1.8768
3. Darweesh S., Shehata H. Performance Evaluation of a Multi-Agent System using Fuzzy Model. *2018 First International Workshop on Deep and Representation Learning* (*IWDRL*). 2018. P. 7–12. https://doi.org/10.1109/IWDRL.2018.8358208

4. Russel S., Norvig P. *Iskusstvennyi intellekt: sovremennyi podkhod.* T. 2. *Znaniya i rassuzhdeniya v usloviyakh neopredelennosti* (*Artificial Intelligence: A Modern Approach.* V. 2. *Knowledge and Reasoning under Uncertainty*): transl. from Engl. St. Petersburg: Dialektika; 2021. 480 p. (in Russ.).
5. Russel S., Norvig P. *Iskusstvennyi intellekt: sovremennyi podkhod.* T. 3. *Obuchenie, vospriyatie i deistvie* (*Artificial Intelligence: A Modern Approach.* V. 3. *Learning, Perception, and Action*): transl. from Engl. St. Petersburg: Dialektika; 2022. 640 p. (in Russ.).
6. Sutton R.S., Barto E.J. *Obuchenie s podkrepleniem* (*Reinforcement Learning*): transl. from Engl. Moscow: DMK Press; 2020. 552 p. (in Russ.).
7. Graesser L., Keng W.L. *Glubokoe obuchenie s podkrepleniem: teoriya i praktika na yazyke Python* (*Reinforcement Learning: Theory and Practice in Python*): transl. from Engl. St. Petersburg: Piter; 2022. 416 p. (in Russ.).
[Graesser L., Keng W.L. *Foundations of Deep Reinforcement Learning: Theory and Practice in Python*. Addison-Wesley Professional; 2020. 416 p.]
8. Pumperla M., Ferguson K. *Glubokoe obuchenie i igra v go* (*Deep Learning and the Game of Go*): transl. from Engl. Moscow: DMK Press; 2020. 372 p. (in Russ.).
9. Red'ko V.G. *Evolyutsiya, neironnye seti, intellekt: Modeli i kontseptsii evolyutsionnoi kibernetiki* (*Evolution, Neural Networks, Intelligence: Models and Concepts of Evolutionary Cybernetics*). Moscow: LIBROKOM; 2013. 224 p. (in Russ.).
10. Winder P. *Reinforcement Learning. Industrial Applications of Intelligent Agents*. O'Reilly Media, Inc.; 2021. 382 p.
11. Houhamdi Z., Athamena B., Abuzaineddin R., Muhairat M. A Multi-Agent System for Course Timetable Generation. *TEM Journal*. 2019;8(1):211–221. https://doi.org/10.18421/TEM81-30
12. Aly S., Badoor H. Performance Evaluation of a Multi-Agent System using Fuzzy Model. *2018 First International Workshop on Deep and Representation Learning* (*IWDRL*). 2018. P. 175–189. https://doi.org/10.1109/IWDRL.2018.8358208
13. Zaytsev E.I. Method of date representation and processing in the distributed intelligence information systems. *Avtomatizatsiya i sovremennye tekhnologii = Automation. Modern Technologies*. 2008;1:29–34 (in Russ.).
14. Batouma N., Sourrouille J.-L. Dynamic adaption of resource aware distributed applications. *Int. J. Grid Distrib. Comput*. 2011;4(2):25–42. Available from URL: http://article.nadiapub.com/IJGDC/vol4_no2/3.pdf
15. Nurmatova E.V., Gusev V.V., Kotliar V.V. Analysis of the features of the optimal logical structure of distributed databases. In: *GRID Workshop Proceedings - GRID 2018 Selected Papers of the 8th International Conference "Distributed Computing and Grid-technologies in Science and Education"*. Dubna; 2018. V. 2267. P. 579–584. Available from URL: https://ceur-ws.org/Vol-2267/579-584-paper-111.pdf

4. Рассел С., Норвиг П. *Искусственный интеллект: современный подход*. Т. 2. *Знания и рассуждения в условиях неопределенности*: пер. с англ. СПб.: Диалектика; 2021. 480 с.
5. Рассел С., Норвиг П. *Искусственный интеллект: современный подход*. Т. 3. *Обучение, восприятие и действие*: пер. с англ. СПб.: Диалектика; 2022. 640 с.
6. Саттон Р.С., Барто Э.Дж. *Обучение с подкреплением*: пер. с англ. М.: ДМК Пресс; 2020. 552 с.
7. Грессер Л., Кенг В.Л. *Глубокое обучение с подкреплением: теория и практика на языке Python:* пер. с англ. СПб.: Питер; 2022. 416 с.
[Graesser L., Keng W.L. *Foundations of Deep Reinforcement Learning: Theory and Practice in Python*. Addison-Wesley Professional; 2020. 416 p.]
8. Памперла М. Фергюсон К. *Глубокое обучение и игра в го:* пер. с англ. М.: ДМК Пресс; 2020. 372 с.
[Pumperla M., Ferguson K. *Deep Learning and the Game of Go.* Manning; 2019. ISBN 978-1-6172-9532-4. 384 p.]
9. Редько В.Г. *Эволюция, нейронные сети, интеллект: Модели и концепции эволюционной кибернетики*. М.: ЛИБРОКОМ; 2013. 224 с.
10. Winder P. *Reinforcement Learning. Industrial Applications of Intelligent Agents*. O'Reilly Media, Inc.; 2021. 382 p.
11. Houhamdi Z., Athamena B., Abuzaineddin R., Muhairat M. A Multi-Agent System for Course Timetable Generation. *TEM Journal*. 2019;8(1):211–221. https://doi.org/10.18421/TEM81-30
12. Aly S., Badoor H. Performance Evaluation of a Multi-Agent System using Fuzzy Model. *2018 First International Workshop on Deep and Representation Learning* (*IWDRL*). 2018. P. 175–189. https://doi.org/10.1109/IWDRL.2018.8358208
13. Зайцев Е.И. Методология представления и обработки знаний в распределенных интеллектуальных информационных системах. *Автоматизация и современные технологии*. 2008;1:29–34.
14. Batouma N., Sourrouille J.-L. Dynamic adaption of resource aware distributed applications. *Int. J. Grid Distrib. Comput*. 2011;4(2):25–42. URL: http://article.nadiapub.com/IJGDC/vol4_no2/3.pdf
15. Nurmatova E.V., Gusev V.V., Kotliar V.V. Analysis of the features of the optimal logical structure of distributed databases. In: *GRID Workshop Proceedings – GRID 2018 Selected Papers of the 8th International Conference "Distributed Computing and Grid-technologies in Science and Education."* Dubna: 2018. V. 2267. P. 579–584. URL: https://ceur-ws.org/Vol-2267/579-584-paper-111.pdf

About the authors

Evgeniy I. Zaytsev, Cand. Sci. (Eng.), Associate Professor, Department of Hardware Software and Mathematical Support of Computing System, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (20, Stromynka ul., Moscow, 107996 Russia). E-mail: zajcev@mirea.ru. Scopus Author ID 57218190023, ResearcherID ABA-4823-2020, RSCI SPIN-code 9662-7658, https://orcid.org/0000-0002-1979-5611

Elena V. Nurmatova, Cand. Sci. (Eng.), Associate Professor, Department of Hardware Software and Mathematical Support of Computing System, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (20, Stromynka ul., Moscow, 107996 Russia). E-mail: nurmatova@mirea.ru. Scopus Author ID 57205460003, ResearcherID GQI-3212-2022, RSCI SPIN-code 7036-3661, https://orcid.org/0000-0001-8511-0978

Об авторах

Зайцев Евгений Игоревич, к.т.н., доцент, кафедра «Аппаратное, программное и математическое обеспечение вычислительных систем» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (107996, Россия, Москва, ул. Стромынка, д. 20). E-mail: zajcev@mirea.ru. Scopus Author ID 57218190023, ResearcherID ABA-4823-2020, SPIN-код РИНЦ 9662-7658, https://orcid.org/0000-0002-1979-5611

Нурматова Елена Вячеславовна, к.т.н., доцент, кафедра «Аппаратное, программное и математическое обеспечение вычислительных систем» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (107996, Россия, Москва, ул. Стромынка, д. 20). E-mail: nurmatova@mirea.ru. Scopus Author ID 57205460003, ResearcherID GQI-3212-2022, SPIN-код РИНЦ 7036-3661, https://orcid.org/0000-0001-8511-0978

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Multiple robots (robotic centers) and systems. Remote sensing and non-destructive testing

**Роботизированные комплексы и системы.
Технологии дистанционного зондирования неразрушающего контроля**

UDC 004.89; 004.93
https://doi.org/10.32362/2500-316X-2023-11-4-26-35

RESEARCH ARTICLE

Algorithms for the visual analysis of an environment by an autonomous mobile robot for area cleanup

Maksim E. Beliakov [@], Sekou Abdel Kader Diane

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: beliakow.m@gmail.com

Abstract

Objectives. At present, increasing rates of pollution of vast areas by various types of household waste are becoming an increasingly serious problem. In this connection, the creation of a robotic complex capable of performing autonomous litter collection functions becomes an urgent need. One of the key components of such a complex comprises a vision system for detecting and interacting with target objects. The purpose of this work is to develop the underlying algorithmics for the vision system of robots executing area cleaning functions.

Methods. Within the framework of the proposed structure of the system for visual analysis of the external environment, algorithms for detecting and classifying objects of various appearance have been developed using convolutional neural networks. The neural network detector was set up by gradient descent on the open dataset of TACO training samples. To determine the geometric parameters of a surface in the field of view of the robot and estimate the coordinates of objects on the ground, a homography matrix was formed to take into account information about the characteristics and location of the video camera.

Results. The developed software and algorithms for a mobile robot equipped with a monocular video camera are capable of implementing the functions of neural network detection and classification of litter objects in the frame, as well as projection of found objects on a terrain map for their subsequent collection.

Conclusions. Experimental studies have shown that the developed system of visual analysis of the external environment of an autonomous mobile robot has sufficient efficiency to solve the tasks of detecting litter in the field of view of an autonomous mobile robot.

Keywords: neural detection, computer vision, homography, mobile robots, territory cleaning

• **Submitted:** 10.03.2022 • **Revised:** 28.02.2023 • **Accepted:** 02.05.2023

For citation: Beliakov M.E., Diane S.A.K. Algorithms for the visual analysis of an environment by an autonomous mobile robot for area cleanup. *Russ. Technol. J*. 2023;11(4):26–35. https://doi.org/10.32362/2500-316X-2023-11-4-26-35

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Алгоритмы визуального анализа внешней среды автономным мобильным роботом в задаче уборки территории

М.Э. Беляков [@], **С.А.К. Диане**

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: beliakow.m@gmail.com*

Резюме

Цели. В настоящее время опасной глобальной тенденцией становятся нарастающие темпы загрязнения огромных по площади территорий различными типами бытовых отходов. В связи с этим актуальной потребностью является создание робототехнических комплексов, способных в автономном режиме осуществлять сбор такого мусора. Одной из ключевых составляющих подобных комплексов должна стать система технического зрения для детекции и взаимодействия с целевыми объектами. Цель работы – разработка алгоритмического обеспечения системы технического зрения робототехнических комплексов в задаче уборки территории.

Методы. В рамках предложенной структуры системы визуального анализа внешней среды были оптимизированы под задачу распознавания мусора алгоритмы детекции и классификации объектов различного внешнего вида с применением технологии сверточных нейронных сетей. Настройка нейросетевого детектора производилась методом градиентного спуска на открытой базе обучающих примеров TACO. Для определения геометрических параметров плоского пространства в поле зрения робота и оценки координат объектов на местности использована матрица гомографии, формируемая с учетом информации о характеристиках и расположении видеокамеры в пространстве.

Результаты. Разработанное программно-алгоритмическое обеспечение системы технического зрения для мобильного робота, оснащаемого монокулярной видеокамерой, реализует функции нейросетевой детекции и классификации объектов в кадре, а также проекции найденных объектов на карту местности для их последующего сбора.

Выводы. Проведенные экспериментальные исследования показали, что разработанная система визуального анализа внешней среды автономного мобильного робота обладает достаточной эффективностью для решения поставленных задач, в т.ч. для обнаружения мусора в поле зрения автономного мобильного робота.

Ключевые слова: нейросетевая детекция, техническое зрение, гомография, мобильные роботы, уборка территории

• **Поступила:** 10.03.2022 • **Доработана:** 28.02.2023 • **Принята к опубликованию:** 02.05.2023

Для цитирования: Беляков М.Э., Диане С.А.К. Алгоритмы визуального анализа внешней среды автономным мобильным роботом в задаче уборки территории. *Russ. Technol. J.* 2023;11(4):26–35. https://doi.org/10.32362/2500-316X-2023-11-4-26-35

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

In most countries of the world, including Russia, waste produced in the course of daily human activity is generally disposed of in one of two ways: incinerated in waste incinerators or buried in landfills. Both methods of disposal have a negative impact on the environment. The incineration of waste is accompanied by the release of toxic gases and dust into the atmosphere, contributing to global warming and the pollution of water bodies, forests and cities far from the location of their release into the atmosphere. While incinerators used in post-industrial countries dispose of the toxic and polluting part of the waste in such a way as to reduce the impact of air pollution, the majority of waste continues to be buried in landfill sites, which also involves negative impacts in terms of emissions, contamination of ground water, etc.

These factors determine the importance of sorting waste into categories in order to reduce the amount of hazardous pollution as a result of waste disposal. The wastes produced by human activities can be divided into hazardous and safe [1]. Safe wastes include food residues, cardboard and paper, cellophane, and other organic wastes. Since such wastes when decomposed do not poison the soil or water sources, their negative impact on the environment is minimal; moreover, such materials can be easily recycled into new products. Hazardous wastes such as electric batteries, paint and varnish products, polyethylene, etc., can poison the soil and water bodies within a radius of several kilometers. Thus, waste sorting helps to reduce the amount of pollution by decreasing the amount of hazardous waste being sent to landfill sites instead of proper disposal sites, as well as increasing the quantity of recycled wastes and reducing associated waste removal costs.

An equally noteworthy problem involves the unintentional or intentional littering of public places such as streets, parks, picnic areas and beaches. According to Report Park Litter2020[1], the most common objects of litter in public spaces are cigarette butts, food wrappers, and plastic bottles. Given the diversity and vastness of the areas exposed to pollution and the general trend towards automation, a need arises to create automated robotic systems capable of autonomously picking up litter.

This present work considers a visual environmental analysis system for an autonomous mobile robot to search for and recognize different categories of waste, and localize the litter in a given area for subsequent collection. For such purposes, it is sufficient for the robot to have a single on-board camera.

[1] https://www.legambienteverona.it/wp-content/uploads/2021/01/Report-Park-Litter_English-final.pdf. Accessed January 25, 2022.

ANALYSIS OF DEVELOPMENTS IN THE FIELD OF ROBOTIC TERRITORY CLEANING

A number of R&D and production organizations around the world are working on the issue of automating litter pickup. Thus, in [2], a robotic system (RS) for the recognition of litter objects and their further collection, equipped with a gripper, a camera, and a visual sensing system, is described. This RS comprises an easy-to-implement hardware and software complex, forming the basis for creating an autonomous mobile robot for performing area cleaning tasks.

Since the visual analysis system of this RS uses the MobileNet pre-trained machine learning model [3], the system is only able to recognize bottles as litter. The work uses a simplified system of pointing and estimating the distance to the object. After detecting an object on the camera frame, the difference between the center of the frame and the center of the object's dimensional frame obtained from the MobileNet detector output is used to rotate the robot and point it at the object.

A PID (Proportional-Integral-Differential) controller can be used to control the motion of the robot [4]. The distance to the object is calculated using the parameters of the camera location on the base of the robot, taking tilt angle, height from the floor, and camera opening angle into account. The disadvantage of this approach is that, when recognizing multiple objects, there is no possibility of building an optimal route for their collection. Moreover, the system offers no litter sorting function.

In [5], a robotic system for litter processing is also described. However, in this case, the described system is not a mobile robot, but a conveyor system equipped with a mechanical manipulator, as well as a camera for visual classification and subsequent segregation assembly of waste arriving on the conveyor belt. The visual analysis system is capable of classifying waste into four classes: paper, metal, glass, plastic. Instead of using neural network algorithms, this system is based on the classical image processing algorithms, consisting of the following steps: recognition of object boundaries on the basis of the corresponding Canny algorithm [6]; threshold detection to separate objects from the background and remove noise; Gaussian filtering to blur object details; photo conversion into black-and-white format in order to use the brightness channel for edge detection; object contour recognition on the binary image. Further processing steps are the use of Hu Moments [7] and Fourier [8] descriptors to describe the shape of objects, along with the Hyperplane K-Nearest Neighbors method [9] for classification by object descriptors.

MODELS AND ALGORITHMS FOR RECOGNIZING THE TYPICAL OBJECTS IN THE ROBOT'S FIELD OF VIEW

Recognition and localization of objects is typically performed by the processing of raster images by methods of computer vision. However, there are different approaches and methods of computer vision. Classical methods include algorithms for finding key points, selecting object boundaries, and geometric transformations. Such algorithms are well suited for simpler and more deterministic tasks in terms of external factors, e.g., lighting and distance to the subject, as well as slight variations in the shape of objects. Alternatively, recognition algorithms based on deep learning are more robust to false positives when the illumination, size or foreshortening of the object changes. Such algorithms, which are usually built using convolutional neural networks, allow the creation of more advanced and fault-tolerant computer vision systems.

The convolutional network architecture is so named due to the convolution operation, whose essence is that each image fragment is multiplied element-by-element by the convolution matrix (kernel), with the result being summed and written in the similar position of the output image. The best results in the generalizability of the network and its computational efficiency are achieved when the convolutional layers and sub-sampling layers are alternated. Since this leads to a consistent reduction in the dimensionality of the input data, it makes the network robust to minor transformations of the analyzed image.

Among the computer vision problems solved by convolutional neural network methods, we can single out the problem of detecting objects in an image [10] as most suitable for the problem discussed in the present work. The detection problem consists in the need to determine the class and coordinates of target objects in order to permit processing of the image algorithm.

Since the neural network can be represented as a multidimensional function, the learning process consists in the optimization of the numerous internal parameters by the method of gradient descent [11].

To train a neural network model for the task of litter detection in the robot's field of view an open dataset TACO[2] [12] was chosen, comprising 1500 images with 4784 labeled objects sorted into 28 classes, which are globally divided into paper, glass, plastic, and metal (Fig. 1).

[2] TACO is a growing image dataset of waste in the wild. It contains images of litter taken under diverse environments: woods, roads and beaches.

As the architecture of the neural network detector, the YOLOv4 algorithm was selected [13]. This architecture offers a compromise between accuracy of object detection and speed (up to several hundreds frames per second when running on a graphics gas pedal). As a result of training, the model was optimized to 0.13 mAP (mean average precision) by 4 classes.

Fig. 1. Examples of marked images in the TACO training examples database

The complex and comprehensive mAP metric [14] takes into account both classification errors (FP, FN) and errors in localization of objects in the frame. Therefore, the obtained value of accuracy according to mAP metric is sufficient for practical application. The result of the detector operation is shown in Fig. 2.

Fig. 2. Result of the YOLOv4 detector operation

MODELS AND ALGORITHMS FOR LOCALIZATION OF TYPICAL OBJECTS ON THE TERRAIN MAP

Since it is assumed that the coordinates of the robot on a given territory are known, their coordinates can be calculated relative to the robot in order to determine the coordinates of litter objects (Fig. 3). This can be done by using the homography matrix. However, this transformation is valid only for objects on a flat surface. Thus, the proposed model of coordinate determination is valid only for objects lying on a flat surface.

Perspective distortion can be eliminated by converting the previously determined coordinates of litter objects on the frame from the mobile robot's onboard camera into coordinates in the top view of the space in front of the robot on the basis of the tilt angle, height above the surface and focal length of the camera.

The coordinates on the shot plane and spatial plane are defined by the following relation:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \mathbf{H}_0 \times \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}. \quad (1)$$

According to the geometrical explanation in Fig. 4, the homography matrix $\mathbf{H}_0$ can be described as follows [15]:

$$\mathbf{H}_0 = \begin{bmatrix} f & x_0 \cos\alpha & x_0 h \cos\alpha \\ 0 & f\sin\alpha + y_0 \cos\alpha & y_0 h \cos\alpha - fh\sin\alpha \\ 0 & \cos\alpha & h\cos\alpha \end{bmatrix}. \quad (2)$$

where f is the focal length of camera; h is the camera height; α is the camera tilt angle; (x_0, y_0) are coordinates of intersection point of image axes.

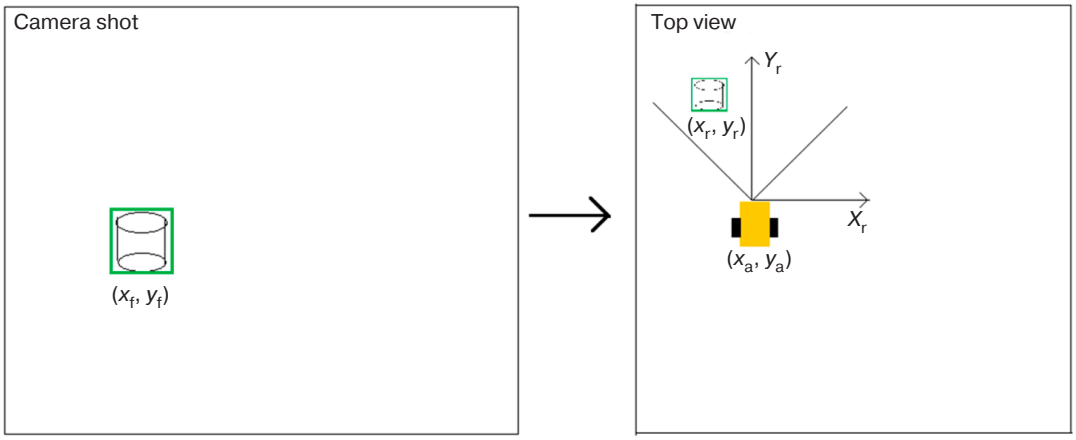


Fig. 3. Calculating the coordinates of the object relative to the robot: (x_r, y_r) – object coordinates in the coordinate system relative to the robot; (X_r, Y_r) – coordinate system relative to the robot; (x_a, y_a) – robot coordinates in the global coordinate system; (x_f, y_f) – coordinates of the object on the shot

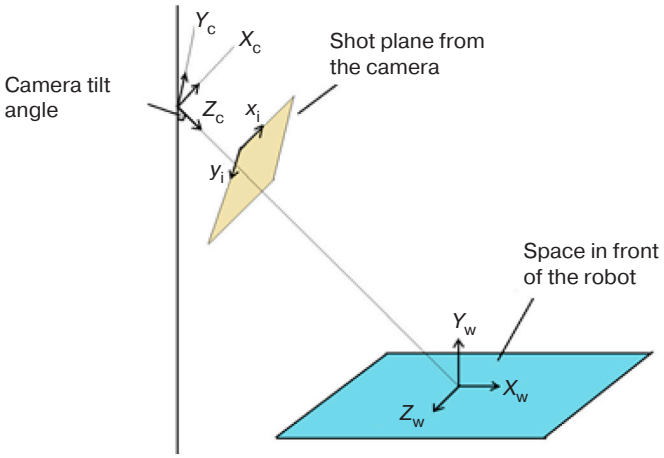


Fig. 4. Geometric representation of the homography matrix: (X_c, Y_c, Z_c) – coordinate system relative to camera; (X_w, Y_w, Z_w) – coordinate system in the plane of space in front of the robot; (x_i, y_i) – coordinates of points of rectangle framing object

Provided the camera angle is different from 0° and 90°, there is an inverse to this transformation, which can be used to obtain the top view from an image with a distorted perspective and vice versa.

Figure 5 shows a distorted perspective image taken at a tilt angle of α = 45°, at height $h = 0.7$ m, as well as its corresponding transformation to the top view.

The frame from the robot's camera is analyzed by the neural network algorithm YOLOv4, which determines the pixel coordinates of the object on the frame as four coordinates of points $p_i = (x_i, y_i)$ of the framing rectangle (Fig. 2); thus, the center of the rectangle is taken as a single-valued coordinate of this object $p_c = \left\{\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right\}$.

This coordinate undergoes the transformation (3) to determe the relative pixel coordinates of the object within the visible space in front of the robot:

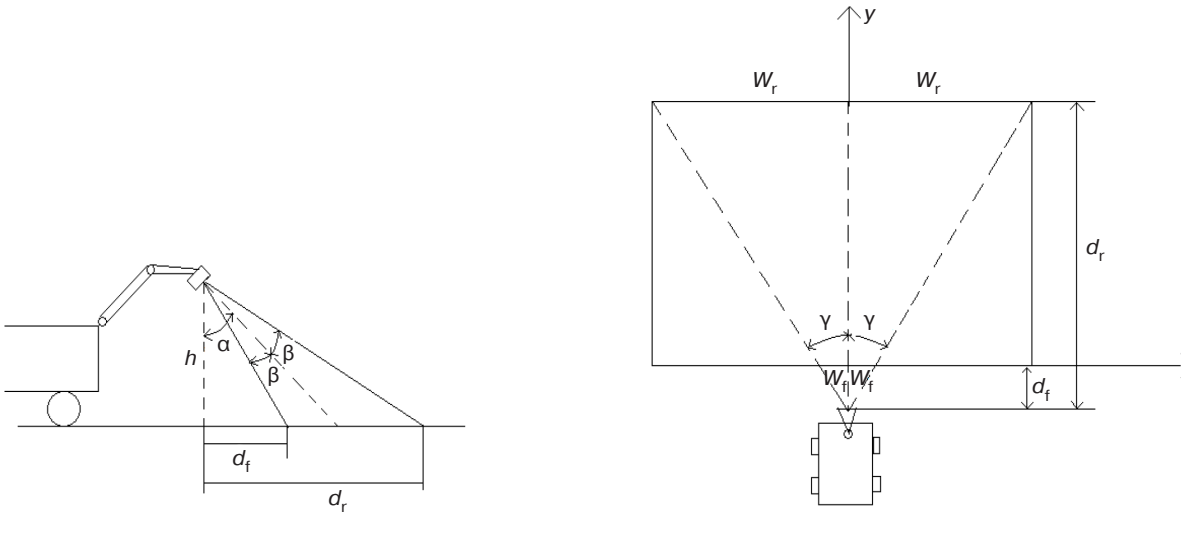
$$\begin{bmatrix} x_m \\ y_m \\ 1 \end{bmatrix} = \mathbf{H}_0^{-1} \times \begin{bmatrix} x_f \\ y_f \\ 1 \end{bmatrix}, \quad (3)$$

where x_m, y_m are coordinates of the object on the space in front of the robot.

To convert the pixel coordinates of objects in the visible space in front of the robot into meter coordinates relative to the robot, it is necessary to calculate the dimensions of the visible space. According to the explanations in Fig. 6, the geometric parameters of the space in front of the robot are unambiguously specified

(a) (b)

Fig. 5. Top view: (a) image taken with the camera; (b) image after correction of perspective distortion



(a) (b)

Fig. 6. Geometry of the observed space in front of the robot: (a) camera field of view in the vertical projection; (b) camera field of view in the horizontal projection

by the horizontal opening angle, vertical opening angle, height and camera tilt angle.

The distance from the camera to the near edge of the visible space d_f, the distance from the camera to the far edge of the visible space d_r, and the length of the visible space Y are determined by the following relations:

$$d_f = h \cdot \mathrm{tg}(\alpha - \beta), \quad (4)$$

$$d_r = h \cdot \mathrm{tg}(\alpha + \beta), \quad (5)$$

$$Y = d_r - d_f. \quad (6)$$

The half-width of the near edge of the visible space w_f, the half-width of the far edge of the visible space w_r, and the width of the visible space X are determined by the following relations:-

$$w_f = \mathrm{tg}\gamma\sqrt{d_f^2 + h^2}, \quad (7)$$

$$w_r = \mathrm{tg}\gamma\sqrt{d_r^2 + h^2}, \quad (8)$$

$$X = 2w_r. \quad (9)$$

According to the aforesaid, the coordinates of objects relative to the robot are defined by the following ratios:-

$$x_0 = \frac{X \cdot x_{pxl}}{w_{pxl}} - w_r, \quad (10)$$

$$y_0 = Y\left(1 - \frac{X \cdot x_{pxl}}{h_{pxl}}\right) + d_f, \quad (11)$$

where x_{pxl}, y_{pxl} are the screen coordinates of the object in the visible space; w_{pxl}, h_{pxl} are the screen width and height of the visible space, respectively.

SOFTWARE STRUCTURE

The structure of the software and algorithmic support is determined by the previously established problems and tasks (Fig. 7). The software includes a subsystem of the user interface, libraries of image acquisition and preprocessing, as well as modules of target object recognition and calculation of its coordinates relative to the mobilebot.

The object recognition module includes not only the procedures directly responsible for the detection and classification process, but also a file with the preconfigured neural network architecture and its preconfigured weighting coefficients.

The Python programming language provides a convenient means for integrating different computer vision technologies and reducing debugging time.

EXPERIMENTAL RESEARCH

The conducted experiments confirmed the performance of the software and algorithmic software. Thus, Fig. 8 demonstrates procedures for recognizing and localizing several litter objects in the robot's field of view, removing perspective distortion, and calculating the coordinates of objects relative to the robot according to a camera of height $h = 0.5$ m, tilt angle $\alpha = 45°$, vertical camera opening angle $\beta = 23.75°$, and horizontal camera opening angle $\gamma = 30.41°$.

In the following experimental study (Fig. 9), the camera is in a position described by height $h = 0.5$ m, tilt angle $\alpha = 45°$, operating against a different background and identifying objects of different categories

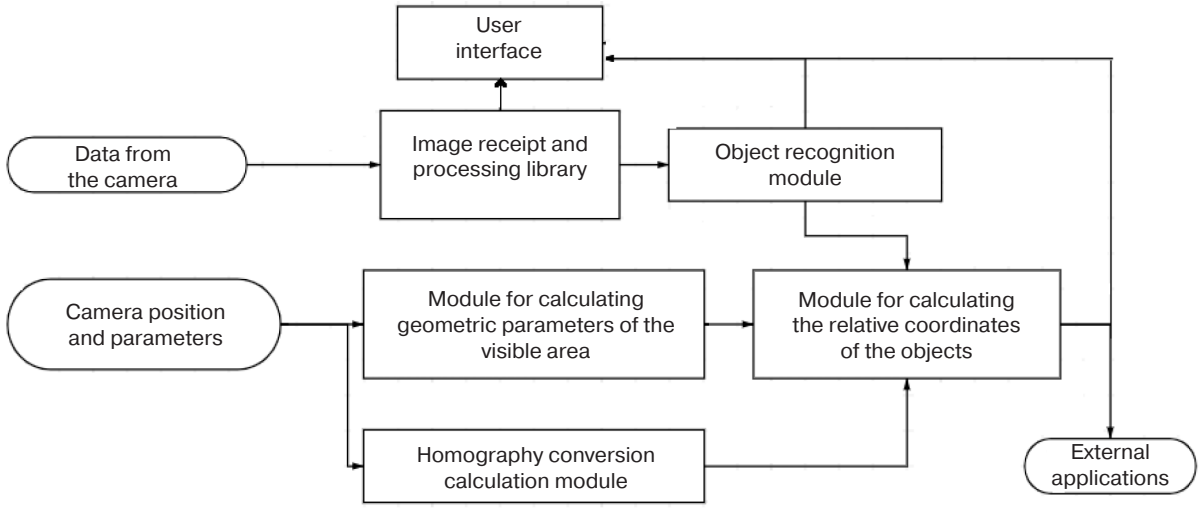
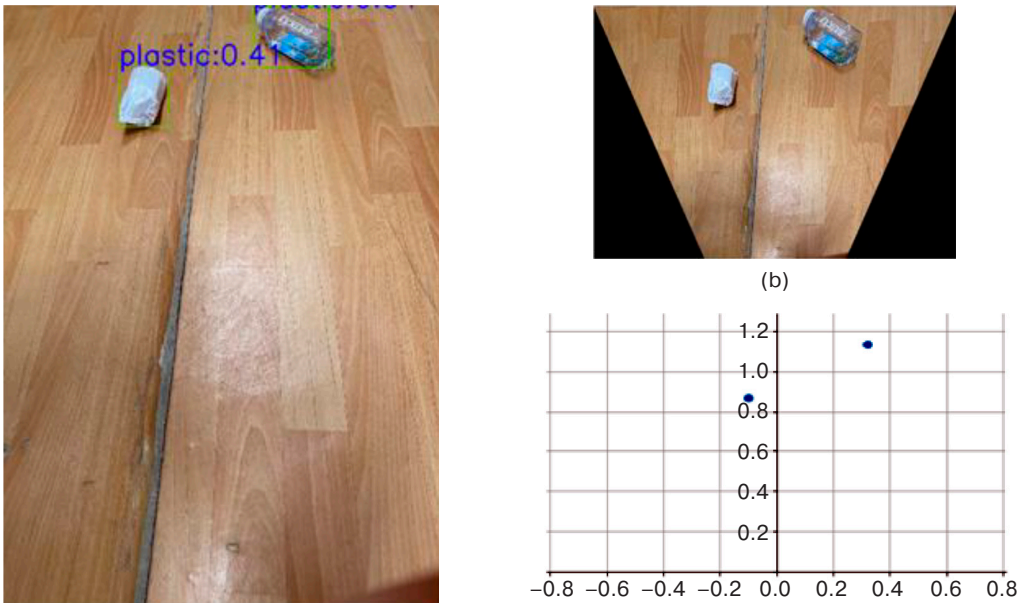
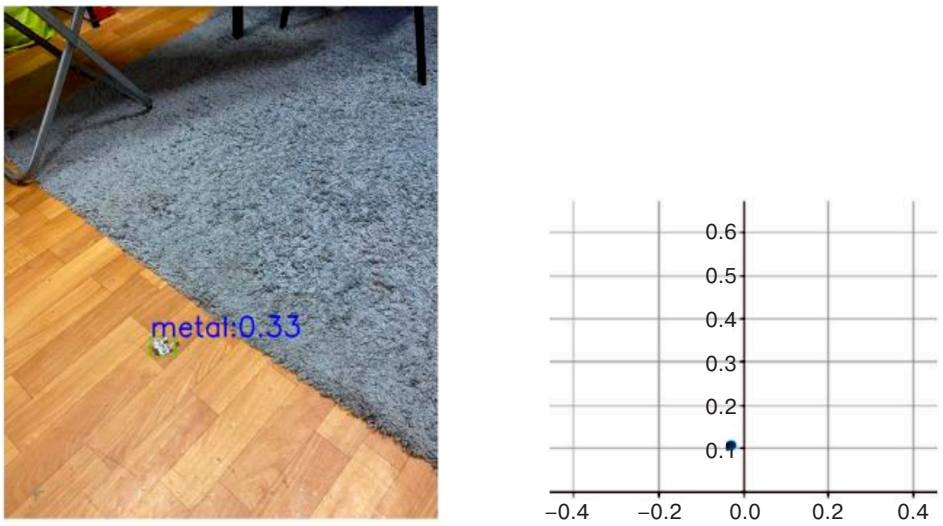


Fig. 7. Structure of software and algorithms for visual analysis of the mobile robot environment



(a) (b) (c)

Fig. 8. Object recognition: (a) camera view; (b) top view;
(c) map of objects in the robot's field of view with coordinates in the metric system



(a) (b)

Fig. 9. Object recognition: (a) camera view;
(b) map of objects in the robot's field of view with coordinates in the metric system

of household waste. This case demonstrates that the software-algorithmic software is invariant both to the background, the number and categories of objects, as well as their spatial positions relative to the camera.

CONCLUSIONS

Unlike existing studies and approaches in this field, which do not combine the mobility of the platform and intelligent separate waste collection, the present study demonstrates the relevance and feasibility of automating these separate functions. The possibility of separate waste collection consists in the flexible intelligent collection with autonomous search for objects in a given area. In the framework of the study, the following results were achieved.

1. Proposed structure (Fig. 7) of software and algorithms for visual analysis of the mobile robot environment in the task of area cleaning with integrated waste sorting function.
2. Optimized algorithms for recognizing objects of different appearance and different waste categories on robot camera images using neural network algorithms.

3. Developed model for determining the geometric parameters of planar space in the field of view of the robot.
4. Developed model for determining the coordinates of objects in the field of view relative to the robot according to the position of the camera in space.

The experimental studies confirmed the performance and universality of the developed algorithms. Correct data were obtained for the recognition of various objects and further determination of their relative coordinates. In addition, the system demonstrated its performance when detecting objects of different categories and different spatial camera positions, taking into account uncertainties in the background of the underlying surface and the number of searched objects.

Further prospects for the development of the research presented consist in the development of algorithms for path planning on the ground and motion control of the mobile robot in the process of collecting and transporting the detected objects of litter.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Chernyaeva T.K. Actual problems of the impact of production and consumption waste on environment and public health (review of literature). *Gigiena i sanitariya = Hygiene and Sanitation*. 2013;3:32–35 (in Russ.).
2. Bansal S., Patel S., Shah I., Patel A., Makwana J., Thakker R. *AGDC: Automatic Garbage Detection and Collection*. arXiv preprint arXiv:1908.05849. 2019. https://doi.org/10.48550/arXiv.1908.05849
3. Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. arXiv preprint arXiv:1704.04861. 2017. https://doi.org/10.48550/arXiv.1704.04861
4. Karpov V.E. *PID-upravlenie v nestrogom izlozhenii* (*PID-Control in a Loose Presentation*). Moscow; 2012. 34 p. (in Russ.). Available from URL: http://radioservice.at.ua/_fr/0/Karpov_mobline1.pdf
5. Salmador A., Cid J.P., Novelle I.R. Intelligent Garbage Classifier. *Int. J. Interact. Multimedia Artif. Intell*. 2008;1(1):31–36.
6. Xu Z., Ji X., Wang M., Sun X. Edge detection algorithm of medical image based on Canny operator. *J. Phys.: Conf. Ser*. 2021;1955(1):012080. https://doi.org/10.1088/1742-6596/1955/1/012080
7. Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Inform. Theory*. 1962;8(2): 179–187. https://doi.org/10.1109/TIT.1962.1057692
8. El-ghazal A., Basir O., Belkasim S. A novel curvature-based shape Fourier Descriptor. *2008 15th IEEE International Conference on Image Processing* (*ICIP*). 2008. P. 953–956. https://doi.org/10.1109/ICIP.2008.4711914
9. Pascal V., Yoshua B. K-Local Hyperplane and Convex Distance Nearest Neighbor Algorithms. In: Dietterich T.G., Becker S., Ghahramani Z. (Eds.). *Advances in Neural Information Processing Systems*. The MIT Press; 2001. V. 14. P. 985–992.
10. Deng J., Xuan X., Wang W., Li Z., Yao H., Wang Z. A review of research on object detection based on deep learning. *J. Phys.: Conf. Ser.* 2020;1684(1):012028. https://doi.org/10.1088/1742-6596/1684/1/012028
11. Bredikhin A.I. Training algorithms for convolutional neural networks. *Vestnik Yugorskogo gosudarstvennogo universiteta = Yugra State University Bulletin*. 2019;1(52):41–54 (in Russ.). https://doi.org/10.17816/byusu20190141-54
12. Proença P.F., Simões P. *TACO: Trash Annotations in Context for Litter Detection*. arXiv preprint arXiv: 2003.06975. 2020. https://doi.org/10.48550/arXiv.2003.06975
13. Bochkovskiy A., Wang C.Y., Liao H.Y.M. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. arXiv preprint arXiv: 2004.10934. 2020. https://doi.org/10.48550/arXiv.2004.10934
14. Liu L., Özsu M.T. (Eds.). *Encyclopedia of Database Systems*. New York, NY, USA: Springer; 2009. V. 6. https://doi.org/10.1007/978-0-387-39940-9
15. Yasuda T. (Ed.) *Multi-Robot Systems, Trends and Development*. London, United Kingdom: IntechOpen; 2011. 598 p. https://doi.org/10.5772/544

About the authors

Maksim E. Beliakov, Bachelor, Department of Control Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). beliakow.m@gmail.com. https://orcid.org/0000-0001-7193-049X

Sekou Abdel Kader Diane, Cand. Sci. (Eng.), Associate Professor, Department of Control Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: sekoudiane1990@gmail.com. ResearcherID T-5560-2017, Scopus Author ID 57188548666, RSCI SPIN-code 8691-0290, https://orcid.org/0000-0002-8690-6422

Об авторах

Беляков Максим Эдуардович, бакалавр, кафедра проблем управления Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: beliakow.m@gmail.com. https://orcid.org/0000-0001-7193-049X

Диане Секу Абдель Кадер, к.т.н., доцент, кафедра проблем управления Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: sekoudiane1990@gmail.com. ResearcherID T-5560-2017, Scopus Author ID 57188548666, SPIN-код РИНЦ 8691-0290, https://orcid.org/0000-0002-8690-6422

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Modern radio engineering and telecommunication systems

Современные радиотехнические и телекоммуникационные системы

UDC 621.314:681.586.7:004.942
https://doi.org/10.32362/2500-316X-2023-11-4-36-48

RESEARCH ARTICLE

Analysis of the DC/DC Zeta topology converter ripples by applying its limiting continuous mathematical model

Vladimir K. Bityukov,
Alexey I. Lavrenov @,
Daniil A. Malitskiy

MIREA – Russian Technological University, Moscow, 119454 Russia
@ Corresponding author, e-mail: lavrenov@mirea.ru

Abstract

Objectives. A DC/DC Zeta topology converter represents a unipolar electronic device for converting an input positive voltage into a stabilized output voltage of the same polarity, which can be set at voltages both below and above the input voltage. The aim of this work is to analyze Zeta converter circuitry, which requires the following tasks to be solved: using Kirchhoff's Circuit Laws, obtain systems of equations describing converter operation in the phase of energy accumulation and in the phase of energy transfer; using a method proposed by A.I. Korshunov, combine the resulting systems of equations into a marginal continuous mathematical model; using expressions describing constant components of currents and voltages in Zeta converter, analyze their ripples and obtain equations for their calculation; compare the current and voltage values obtained from the continuous limiting mathematical model with the Zeta simulation results.

Methods. The tasks are solved using Kirchhoff's rules and the method for obtaining the limiting continuous mathematical model proposed by A.I. Korshunov. The results are analyzed using a circuit modelling in NI *Multisim*.

Results. It is shown that the phase coordinates of the mathematical model tend to real values of converter currents and voltages at a switching frequency of the power switch of more than 200 kHz. A strong correspondence was established between the calculated ripple values and their values obtained in the simulation (when changing the duty factor).

Conclusions. Mathematical models comprise the basis of unified calculation methods for any radio electronic circuit. The developed limiting continuous mathematical model allows a range of changes in current flowing through the choke windings and voltages on capacitor plates to be evaluated, including their maximum and minimum values for various converter parameters, such as power switch switching frequency, duty factor, element ratings, etc. Obtaining this information in turn enables the rational selection of the electronic component base of the converter.

Keywords: DC/DC converter, step-up and step-down converter, equivalent circuit, Zeta topology, converter, limiting continuous mathematical model, Kirchhoff's rules, ripple spreading

• **Submitted:** 28.12.2022 • **Revised:** 10.02.2023 • **Accepted:** 05.05.2023

For citation: Bityukov V.K., Lavrenov A.I., Malitskiy D.A. Analysis of the DC/DC Zeta topology converter ripples by applying its limiting continuous mathematical model. *Russ. Technol. J.* 2023;11(4):36–48. https://doi.org/10.32362/2500-316X-2023-11-4-36-48

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Анализ пульсаций DC/DC-преобразователя, построенного по Zeta-топологии, с использованием его предельной непрерывной математической модели

В.К. Битюков,
А.И. Лавренов [@],
Д.А. Малицкий

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] Автор для переписки, e-mail: lavrenov@mirea.ru

Резюме

Цели. DC/DC-преобразователь, построенный по Zeta-топологии, является униполярным электронным устройством, которое обеспечивает преобразование входного положительного напряжения в стабилизированное выходное напряжение той же полярности с возможностью его регулирования как ниже входного напряжения, так и выше. Цель работы – проанализировать схемотехнику Zeta-преобразователя. Для этого необходимо решить следующие задачи: при помощи правил Кирхгофа получить системы уравнений, описывавшие работу преобразователя в режимах накопления и передачи энергии; по методике, предложенной А.И. Коршуновым, объединить системы уравнений в предельную непрерывную математическую модель преобразователя; при помощи выражений, описывающих постоянные составляющие токов и напряжений в Zeta-преобразователе, провести анализ их пульсаций и получить уравнения для их расчета; провести сравнение полученных при помощи предельной непрерывной математической модели значений токов и напряжений с результатами моделирования Zeta-преобразователя.

Методы. Задача решена при помощи правил Кирхгофа и методики получения предельной непрерывной математической модели, предложенной А.И. Коршуновым. Результаты проанализированы с использованием схемотехнического моделирования в среде *Multisim*.

Результаты. Показано, что фазовые координаты математической модели стремятся к значениям реальных токов и напряжений преобразователя при частоте коммутации силового ключа более 200 кГц. Установлено высокое соответствие расчетных значений пульсаций и их значений, полученных при моделировании (при изменении коэффициента заполнения).

Выводы. Математические модели являются основой унифицированных методик расчета любых радиоэлектронных схем. Полученная предельная непрерывная математическая модель Zeta-преобразователя позволяет оценить диапазон изменения токов, протекающих через обмотки дросселей, и напряжений на обкладках конденсаторов, их максимальные и минимальные значения при различных параметрах преобразователя, таких как частота коммутации силового ключа, коэффициент заполнения, номиналы элементов и т.д. Эта модель позволяет выполнить рациональный подбор электронной компонентной базы преобразователя.

Ключевые слова: DC/DC-преобразователь, понижающе-повышающий преобразователь, эквивалентная схема, топология Zeta, преобразователь, предельная модель, непрерывная модель, математическая модель, правила Кирхгофа, размах пульсаций

• **Поступила:** 28.12.2022 • **Доработана:** 10.02.2023 • **Принята к опубликованию:** 05.05.2023

Для цитирования: Битюков В.К., Лавренов А.И., Малицкий Д.А. Анализ пульсаций DC/DC-преобразователя, построенного по Zeta-топологии, с использованием его предельной непрерывной математической модели. *Russ. Technol. J.* 2023;11(4):36–48. https://doi.org/10.32362/2500-316X-2023-11-4-36-48

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

DC/DC converters are widely used in autonomous battery-powered devices such as unmanned aerial vehicles (UAVs), pyrometers, pacemakers, automotive electronics, robots, etc. [1–8]. Since the presence of a DC/DC converter in the power supply device determines both its mass-size parameters and efficiency, as well as the energy efficiency of the entire radio-electronic means (REM), it also affects the maximum period of autonomous operation without recharging. The complexity of manufacturing such power supply devices is also due to the tendency of autonomous devices to require stabilized low-voltage potentials. All this predetermines the need to develop specialized methods, algorithms, and design tools for DC/DC converters.

DC/DC converters implementing basic topologies of step-up, step-down, and polar inverting types [9, 10] into complex DC/DC converters—i.e., step-up and step-down (buck–boost) converters—require the development of design methods and a sufficient detail of supporting research [2, 4, 11–13].

The development of DC/DC converters, like that of other REMs, is based on appropriate mathematical models, which are the basis for the unified methodical approach to developing, designing and researching devices. Limiting continuous mathematical models of basic DC/DC converters are proposed in [9, 14–17]. However, so far, a converter for buck–boost DC/DC converters based on the Ćuk topology exists only in the form of a mathematical model [18, 19]. Meanwhile, there is no mathematical model of unipolar DC/DC converters based on Zeta topology or single-ended primary-inductor converter (SEPIC) topology. Here it should be noted that, while SEPIC, Zeta, and Ćuk converters are identical in terms of the electronic component base, they differ significantly from the schematic viewpoint.

MATHEMATICAL MODEL

A phase plane comprising a set of points of its possible states for each continuous mathematical model can be constructed to represent processes of a real device; here, a point on the phase plane represents the current state of the model and a change in the state when it moves. The trace from the representing point movement is referred to as a phase trajectory, while the point itself represents a phase coordinate. Continuity of the system means that the state of the system, i.e., the values of phase coordinates, can be established at any time.

The mathematical model of the device provides a means of obtaining the relationship between phase coordinates, which correspond to real currents and voltages of the DC/DC converter. Phase coordinates and real values of currents and voltages of the mathematical model coincide when the switching period T of the electronic switch tends to zero. Such mathematical models of key devices are commonly referred to as limiting models [14, 15].

The circuit of the unipolar Zeta converter (Fig. 1) first proposed in [12] comprises two chokes, L1 and L2, two capacitors, C1 and C2, an electronic switch VT1 typically implemented by field-effect transistor, as well as a control unit determining the transistor mode. Chokes L1 and L2 perform the function of energy storage and transfer via electromagnetic induction, while capacitor C1 is present in the circuit to separate the converter input from its output. The separating capacitor C1 is also sometimes referred to as a "flying capacitor" since it performs not only the function of separation but also the function of energy storage and transfer between sections of the converter [11, 12]. The remaining radio components perform traditional functions.

Two operation phases may be distinguished for the DC/DC converter, as well as for the majority of key devices. The first phase determines the energy storage mode, while the second determines the energy transfer mode. Thus, the mathematical model requires equivalent circuits to be constructed for each phase of the converter. Based on equivalent diagrams, systems of equations describing two operation phases are written.

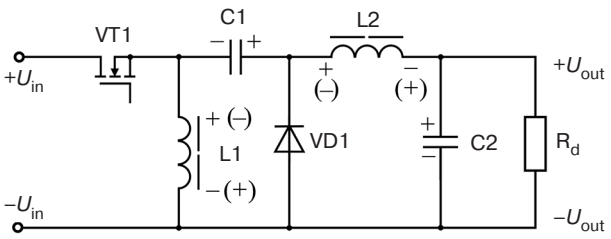


Fig. 1. Schematic diagram of the buck–boost converter based on Zeta topology. U_{in} and U_{out} are input and output voltages; R_d is the duty resistor. (In the diagrams that follow, the designations adopted in the GOST 2.710-81[1] standard are used.)

In [20], only the mathematical model of the converter with the same chokes L1 and L2 is presented. Therefore, it makes sense to construct a mathematical model of the buck–boost DC/DC converter based on Zeta topology, but in a general form.

The system of equations describing both operation phases are made according to Kirchhoff's rules. The DC/DC converter in the energy storage mode (first phase) is described by five equations of algebraic sums of currents and voltages, as follows:

[1] GOST 2.710-81. *Unified system for design documentation. Alpha-numerical designations in electrical diagrams*. Moscow: Standartinform; 2008 (in Russ.).

$$i_{in} - i_{L1} - i_{L2} = 0, \quad (1)$$

$$\begin{cases} \dfrac{du_{C2}}{dt} = \dfrac{1}{C_2} i_{L2} - \dfrac{1}{R_d C_2} u_{C2}, \\ \dfrac{du_{C1}}{dt} = \dfrac{1}{C_1} i_{L2}, \\ \dfrac{di_{L1}}{dt} = \dfrac{1}{L_1} U_{in} - \dfrac{r_1}{L_1} i_{L1}, \\ \dfrac{di_{L2}}{dt} = -\dfrac{1}{L_2} u_{C1} - \dfrac{1}{L_2} u_{C2} - \dfrac{r_2}{L_2} i_{L2} - \dfrac{1}{L_2} U_{in}, \end{cases} \quad (2)$$

where L_1 is the inductance of a choke L1; L_2 is the inductance of a choke L2; C_1 is the capacitance of a capacitor C1; C_2 is the capacitance of a capacitor C2; R_d is the resistance of a duty resistor R_d; i_{L1} and i_{L2} are instantaneous currents flowing through the winding of chokes L1 and L2; r_1 and r_2 are active resistances of the L1 and L2 choke windings; and u_{C1} and u_{C2} are instantaneous voltages at the coatings of capacitors C1 and C2.

Equation (1) containing input current i_{in}, according to Kirchhoff's rules, is necessary for describing the first phase of the inverter operation in detail. However, considering the equations of system (2) not depending on the input current, Eq. (1) can be excluded from the system.

The system of equations determining the inverter operation in the energy transfer mode (second phase) may be written as follows:

$$\begin{cases} \dfrac{du_{C2}}{dt} = \dfrac{1}{C_2} i_{L2} - \dfrac{1}{R_d C_2} u_{C2}, \\ \dfrac{du_{C1}}{dt} = \dfrac{1}{C_1} i_{L1}, \\ \dfrac{di_{L1}}{dt} = -\dfrac{r_1}{L_1} i_{L1} - \dfrac{1}{L_1} u_{C1}, \\ \dfrac{di_{L2}}{dt} = -\dfrac{r_2}{L_2} i_{L2} - \dfrac{1}{L_2} u_{C2}. \end{cases} \quad (3)$$

For building the mathematical model, it is necessary to combine systems of Eqs. (2)–(3) according to the method [14, 15]. The mathematical model may be written as the following matrix system of equations:

$$\mathbf{X} = \begin{bmatrix} i_{L1} \\ i_{L2} \\ u_{C1} \\ u_{C2} \end{bmatrix}, \quad (4)$$

$$\mathbf{A} = \begin{bmatrix} -\dfrac{r_1}{L_1} & 0 & -(1-D)\dfrac{1}{L_1} & 0 \\ 0 & -\dfrac{r_2}{L_2} & -D\dfrac{1}{L_2} & -\dfrac{1}{L_2} \\ (1-D)\dfrac{1}{C_1} & D\dfrac{1}{C_1} & 0 & 0 \\ 0 & \dfrac{1}{C_2} & 0 & -\dfrac{1}{RC_2} \end{bmatrix}, \quad (5)$$

$$\mathbf{B} = \begin{bmatrix} D\dfrac{1}{L_1} \\ -D\dfrac{1}{L_2} \\ 0 \\ 0 \end{bmatrix}, \quad (6)$$

where **X** is the matrix of the system phase coordinates; **A** is the coefficient matrix of the phase coordinates; **B** is the coefficient matrix of an external source, such as the input voltage; **D** is the fill factor of the pulse-width modulated signal controlling by power switch VT1.

Thus, the system of Eqs. (4)–(6) is the limiting continuous mathematical model of DC/DC converter based on Zeta topology.

Instantaneous currents and voltages contain constant and variable (called ripple) components. Analysis of the mathematical model shows that it would be advisable to determine separately the constant components of currents I_{L1}, I_{L2}, voltages U_{C1}, U_{C2}, and their ripples.

Solving the system of Eqs. (4)–(6) of the mathematical model for the constant components of currents I_{L1}, I_{L2} and voltages U_{C1}, U_{C2}, the following equations may be written:

$$I_{L1} = \left| \frac{U_{in} D^2}{R_d D^2 - (r_1 + 2R_d) D + R_d} \right|, \quad (7)$$

$$I_{L2} = \left| \frac{U_{in} D(D-1)}{R_d D^2 - (r_1 + 2R_d) D + R_d} \right|, \quad (8)$$

$$U_{C1} = \left| \frac{R_d U_{in} D(D-1)}{R_d D^2 - (r_1 + 2R_d) D + R_d} \right|, \quad (9)$$

$$U_{C2} = \left| -\frac{U_{in} D((r_2 + r_1 + R_d) D - (r_2 + R_1))}{R_d D^2 - (r_1 + 2R_d) D + R_d} \right|. \quad (10)$$

Using Eqs. (1), (7), and (8), the following equation for the constant component of the input current I_{in} may be written:

$$I_{in} = \left(\left| \frac{U_{in} D^2}{R_d D^2 - (r_1 + 2R_d)D + R_d} \right| + \left| \frac{U_{in} D(D-1)}{R_d D^2 - (r_1 + 2R_d)D + R_d} \right| \right) D. \quad (11)$$

Equations (7)–(11) are the basis for the preliminary calculation of the converter and hence for the selection of its electronic component base.

CURRENT AND VOLTAGE RIPPLES

Instantaneous currents i_{L1} and i_{L2} flowing through the windings of chokes L1 and L2 as well as instantaneous voltages u_{C1} and u_{C2} on capacitors C1 and C2 contain constant and variable components

$$\begin{aligned} i_{L1} &= I_{L1} + \delta i_{L1}, \quad u_{C1} = U_{C1} + \delta u_{C1}, \\ i_{L2} &= I_{L2} + \delta i_{L2}, \quad u_{C2} = U_{C2} + \delta u_{C2}, \end{aligned} \quad (12)$$

where δi_{L1} is the variable component of current i_{L1}; δi_{L2} is the variable component of current i_{L2}; δu_{C1} is the variable component of voltage u_{C1}; δu_{C2} is the variable component of voltage u_{C2}.

Substituting (12) into the systems of Eqs. (2) and (3), which describe both phases of the converter operation, the following may be written:

$$\begin{cases} \dfrac{d\left(I_{L1} + \delta i_{L1}\right)}{dt} = \dfrac{1}{L_1} U_{in} - \dfrac{r_1}{L_1}\left(I_{L1} + \delta i_{L1}\right), & (13.1) \\ \dfrac{d\left(I_{L2} + \delta i_{L2}\right)}{dt} = -\dfrac{1}{L_2}\left(U_{C1} + \delta u_{C1}\right) - \dfrac{1}{L_2}\left(U_{C2} + \delta u_{C2}\right) - \dfrac{r_2}{L_2}\left(I_{L2} + \delta i_{L2}\right) - \dfrac{1}{L_2} U_{in}, & (13.2) \\ \dfrac{d\left(U_{C1} + \delta u_{C1}\right)}{dt} = \dfrac{1}{C_1}\left(I_{L2} + \delta i_{L2}\right), & (13.3) \\ \dfrac{d\left(U_{C2} + \delta u_{C2}\right)}{dt} = \dfrac{1}{C_2}\left(I_{L2} + \delta i_{L2}\right) - \dfrac{1}{R_d C_2}\left(U_{C2} + \delta u_{C2}\right). & (13.4) \end{cases} \quad (13)$$

$$\begin{cases} \dfrac{d\left(I_{L1} + \delta i_{L1}\right)}{dt} = -\dfrac{r_1}{L_1}\left(I_{L1} + \delta i_{L1}\right) - \dfrac{1}{L_1}\left(U_{C1} + \delta u_{C1}\right), & (14.1) \\ \dfrac{d\left(I_{L2} + \delta i_{L2}\right)}{dt} = -\dfrac{r_2}{L_2}\left(I_{L2} + \delta i_{L2}\right) - \dfrac{1}{L_2}\left(U_{C2} + \delta u_{C2}\right), & (14.2) \\ \dfrac{d\left(U_{C1} + \delta u_{C1}\right)}{dt} = \dfrac{1}{C_1}\left(I_{L1} + \delta i_{L1}\right), & (14.3) \\ \dfrac{d\left(U_{C2} + \delta u_{C2}\right)}{dt} = \dfrac{1}{C_2}\left(I_{L2} + \delta i_{L2}\right) - \dfrac{1}{R_d C_2}\left(U_{C2} + \delta u_{C2}\right). & (14.4) \end{cases} \quad (14)$$

Given that $\dfrac{U_{C2}}{R_d} \approx I_{L2}$, Eqs. (13.4) and (14.4) may be simplified, as follows:

$$\frac{d\left(U_{C2} + \delta u_{C2}\right)}{dt} \approx \frac{1}{C_2}\delta i_{L2} - \frac{1}{R_d C_2}\delta u_{C2}.$$

The constant components of currents and voltages are usually much larger than the ripple, so in the right-hand sides of Eqs. (13.1)–(13.3) and (14.1)–(14.3), the ripple can be neglected compared to the corresponding constant components:

$$\begin{cases} \dfrac{d\delta i_{L1}}{dt} = \dfrac{1}{L_1}U_{in} - \dfrac{r_1}{L_1}I_{L1}, \quad (15.1) \\ \dfrac{d\delta i_{L2}}{dt} = -\dfrac{1}{L_2}U_{C1} - \dfrac{1}{L_2}U_{C2} - \dfrac{r_2}{L_2}I_{L2} - \dfrac{1}{L_2}U_{in}, \quad (15.2) \\ \dfrac{d\delta u_{C1}}{dt} = \dfrac{1}{C_1}I_{L2}, \quad (15.3) \\ \dfrac{d\delta u_{C2}}{dt} = \dfrac{1}{C_2}\delta i_{L2} - \dfrac{1}{R_d C_2}\delta u_{C2}. \quad (15.4) \end{cases} \tag{15}$$

$$\begin{cases} \dfrac{d\delta i_{L1}}{dt} = -\dfrac{r_1}{L_1}I_{L1} - \dfrac{1}{L_1}U_{C1}, \quad (16.1) \\ \dfrac{d\delta i_{L2}}{dt} = -\dfrac{r_2}{L_2}I_{L2} - \dfrac{1}{L_2}U_{C2}, \quad (16.2) \\ \dfrac{d\delta u_{C1}}{dt} = \dfrac{1}{C_1}I_{L1}, \quad (16.3) \\ \dfrac{d\delta u_{C2}}{dt} = \dfrac{1}{C_2}\delta i_{L2} - \dfrac{1}{R_d C_2}\delta u_{C2}. \quad (16.4) \end{cases} \tag{16}$$

Using Eqs. (15.1)–(15.3), the following equations for variable component (ripple) δi_{L1}, δi_{L2}, and δu_{C1} of the first phase may be written:

$$\delta i_{L1} = \int\left(\frac{1}{L_1}U_{in} - \frac{r_1}{L_1}I_{L1}\right)dt = \left(\frac{1}{L_1}U_{in} - \frac{r_1}{L_1}I_{L1}\right)t, \tag{17}$$

$$\delta i_{L2} = \int\left(-\frac{1}{L_2}U_{C1} - \frac{1}{L_2}U_{C2} - \frac{r_2}{L_2}I_{L2} - \frac{1}{L_2}U_{in}\right)dt = \left(-\frac{1}{L_2}U_{C1} - \frac{1}{L_2}U_{C2} - \frac{r_2}{L_2}I_{L2} - \frac{1}{L_2}U_{in}\right)t, \tag{18}$$

$$\delta u_{C1} = \int\left(\frac{1}{C_1}I_{L2}\right)dt = \left(\frac{1}{C_1}I_{L2}\right)t. \tag{19}$$

Using Eqs. (16.1)–(16.3), the following equations for variable components δi_{L1}, δi_{L2}, and δu_{C1} of the second phase may be written:

$$\delta i_{L1} = \int\left(-\frac{r_1}{L_1}I_{L1} - \frac{1}{L_1}U_{C1}\right)dt = \left(-\frac{r_1}{L_1}I_{L1} - \frac{1}{L_1}U_{C1}\right)t, \tag{20}$$

$$\delta i_{L2} = \int\left(-\frac{r_2}{L_2}I_{L2} - \frac{1}{L_2}U_{C2}\right)dt = \left(-\frac{r_2}{L_2}I_{L2} - \frac{1}{L_2}U_{C2}\right)t, \tag{21}$$

$$\delta u_{C1} = \int\left(\frac{1}{C_1}I_{L1}\right)dt = \left(\frac{1}{C_1}I_{L1}\right)t. \tag{22}$$

Equations (15.4) and (16.4) have the same form, so they may be written as follows:

$$\frac{d\delta u_{C2}}{dt} = y' = px - qy, \tag{23}$$

where $q = \dfrac{1}{R_d C_2}$, coefficient p is expressed by the equation

$$p_1 = \frac{1}{C_2}\left(-\frac{1}{L_2}U_{C1} - \frac{1}{L_2}U_{C2} - \frac{r_2}{L_2}I_{L2} - \frac{1}{L_2}U_{in}\right)$$

for the first phase and $p_2 = \dfrac{1}{C_2}\left(-\dfrac{r_2}{L_2}I_{L2} - \dfrac{1}{L_2}U_{C2}\right)$ for the second phase.

Solving differential Eq. (23), the following equations may be written:

$$y = \left(\frac{p(qx-1)e^{qx}}{q^2} + K\right)e^{-qx} \text{ or}$$

$$\delta u_{C2} = -\frac{p}{q^2} + \frac{pt}{q} + Ke^{-qt}. \tag{24}$$

For simplifying calculations, the boundary condition $y(0) = 0$ may be taken. Equations of constants for the first and second phases are written as follows:

$$K_1 = \frac{p_1}{q^2} \text{ and } K_2 = \frac{p_2}{q^2}. \tag{25}$$

Using Eqs. (17)–(19), the following equations for determining the ripple spreading Δi_{L1}, Δi_{L2}, and Δu_{C1} of the first phase may be written:

$$\Delta i_{L1} = \delta i_{L1}(TD) - \delta i_{L1}(0) = \left(\frac{1}{L_1}U_{in} - \frac{r_1}{L_1}I_{L1}\right)TD = -\frac{U_{in}DT\left(\left(r_1 - R_d\right)D^2 + \left(r_1 + 2R_d\right)D - R_d\right)}{L_1\left(R_d D^2 - \left(r_1 + 2R_d\right)D + R_d\right)}, \tag{26}$$

$$\Delta i_{L2} = \delta i_{L2}(TD) - \delta i_{L2}(0) = \left(-\frac{1}{L_2}U_{C1} - \frac{1}{L_2}U_{C2} - \frac{r_2}{L_2}I_{L2} - \frac{1}{L_2}U_{in}\right)TD = \frac{U_{in}DT\left(\left(r_1 - R_d\right)D^2 + \left(r_1 + 2R_d\right)D - R_d\right)}{L_2\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}, \quad (27)$$

$$\Delta u_{C1} = \delta u_{C1}(TD) - \delta u_{C1}(0) = \left(\frac{1}{C_1}I_{L2}\right)TD = \frac{U_{in}D^2T\left(D-1\right)}{C_1\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}. \quad (28)$$

Using Eqs. (20)–(22), the equations for determining the ripple spreading Δi_{L1}, Δi_{L2}, and Δu_{C1} of the second phase may be written:

$$\Delta i_{L1} = \delta i_{L1}(TD) - \delta i_{L1}(T) = \left(-\frac{r_1}{L_1}I_{L1} - \frac{1}{L_1}U_{C1}\right)T(1-D) = -\frac{U_{in}DT(1-D)\left(\left(r_1 + R_d\right)D - R_d\right)}{L_1\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}, \quad (29)$$

$$\Delta i_{L2} = \delta i_{L2}(TD) - \delta i_{L2}(T) = \left(-\frac{r_2}{L_2}I_{L2} - \frac{1}{L_2}U_{C2}\right)T(1-D) = \frac{U_{in}DT(1-D)\left(\left(r_1 + R_d\right)D - R_d\right)}{L_2\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}, \quad (30)$$

$$\Delta u_{C1} = \delta u_{C1}(TD) - \delta u_{C1}(T) = \left(\frac{1}{C_1}I_{L1}\right)T\left(D-1\right) = \frac{U_{in}D^2T\left(D-1\right)}{C_1\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}. \quad (31)$$

Considering that the equations for determining the current ripple spreading Δi_{L1} and Δi_{L2} both are different in two phases and are equivalent, it would be advisable to determine the average ripple spreading Δi_{L1av} and Δi_{L2av} of currents flowing through the choke windings:

$$\Delta i_{L1av} = \frac{\left|-\frac{U_{in}DT\left(\left(r_1 - R_d\right)D^2 + \left(r_1 + 2R_d\right)D - R_d\right)}{L_1\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}\right|}{2} + \frac{\left|-\frac{U_{in}DT(1-D)\left(\left(r_1 + R_d\right)D - R_d\right)}{L_1\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}\right|}{2}, \quad (32)$$

$$\Delta i_{L2av} = \frac{\left|\frac{U_{in}DT\left(\left(r_1 - R_d\right)D^2 + \left(r_1 + 2R_d\right)D - R_d\right)}{L_2\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}\right|}{2} + \frac{\left|\frac{U_{in}DT(1-D)\left(\left(r_1 + R_d\right)D - R_d\right)}{L_2\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}\right|}{2}. \quad (33)$$

Using Eq. (24), the following equation for determining the ripple spreading Δu_{C2} may be written:

$$\Delta u_{C2} = \left|\delta u_{C21}\left(\frac{TD}{2}\right)\right| + \left|\delta u_{C22}\left(T\frac{1-D}{2}\right)\right| = -\frac{p_1}{q^2} + \frac{p_1\frac{TD}{2}}{q^2} + K_1e^{-q\frac{TD}{2}} - \frac{p_2}{q^2} + \frac{p_2T\frac{1-D}{2}}{q^2} + K_2e^{-qT\frac{1-D}{2}} = \frac{1}{2}\left(\left|\frac{R_dU_{in}D\left(\left(r_1 + R_d\right)D - R_d\right)}{L_2\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}\times\left(2R_dC_2e^{\frac{T(D-1)}{2R_dC_2}} + \left((1-D)T - 2R_dC_2\right)\right)\right| + \left|\frac{R_dU_{in}D\left(\left(r_1 + R_d\right)D - R_d\right)}{L_2\left(R_dD^2 - \left(r_1 + 2R_d\right)D + R_d\right)}\times\left(2R_dC_2e^{\frac{-TD}{2R_dC_2}} + \left(DT - 2R_dC_2\right)\right)\right|\right), \quad (34)$$

where δu_{C21} is the function of the voltage ripple variable component on the capacitor C2 obtained by solving the differential equation in the first phase; δuC_{22} is the function of the voltage ripple variable component on the capacitor C2 obtained by solving the differential equation in the second phase.

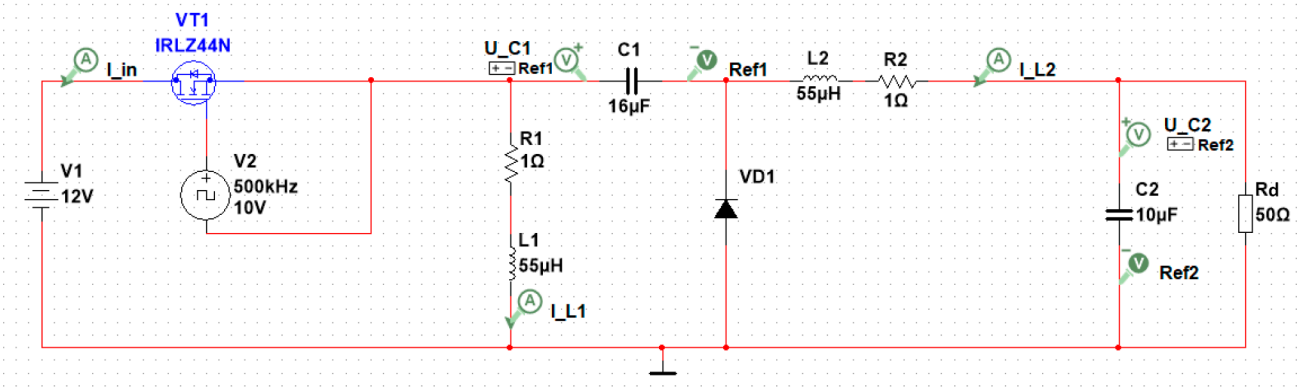


Fig. 2. Simulation scheme of the DC/DC converter

Using Eqs. (1), (32), and (33), we can write the following equation for determining the ripple spreading of input current Δi_{in}:

$$\Delta i_{in} = \frac{\Delta i_{L1av} + \Delta i_{L2av}}{2} + \frac{1}{D} I_{in} =$$
$$= \frac{1}{4}\left(\left| -\frac{U_{in} DT\left((r_l - R_d)D^2 + (r_l + 2R_d)D - R_d\right)}{L_1\left(R_d D^2 - (r_l + 2R_d)D + R_d\right)} \right| + \right.$$
$$+ \left| -\frac{U_{in} DT(1-D)\left((r_l + R_d)D - R_d\right)}{L_1\left(R_d D^2 - (r_l + 2R_d)D + R_d\right)} \right| +$$
$$+ \left| \frac{U_{in} DT\left((r_l - R_d)D^2 + (r_l + 2R_d)D - R_d\right)}{L_2\left(R_d D^2 - (r_l + 2R_d)D + R_d\right)} \right| + \quad (35)$$
$$+ \left. \left| \frac{U_{in} DT(1-D)\left((r_l + R_d)D - R_d\right)}{L_2\left(R_d D^2 - (r_l + 2R_d)D + R_d\right)} \right| \right) +$$
$$+ \left(\left| \frac{U_{in} \cdot D^2}{R_d \cdot D^2 - (r_l + 2 \cdot R_d) \cdot D + R_d} \right| + \right.$$
$$+ \left. \left| \frac{U_{in} \cdot D \cdot (D-1)}{R_d \cdot D^2 - (r_l + 2 \cdot R_d) \cdot D + R_d} \right| \right).$$

Thus, Eqs. (28) or (31) and (32)–(35) allow calculating the ripple spreading of currents i_{L1}, i_{L2} flowing through the windings of chokes L1 and L2, input current i_{in} and voltages u_{C1}, u_{C2} at the coatings of capacitors C1 and C2, if the nominal values of selected electronic components and the converter operating mode (duty factor D and period T) are known.

SIMULATION IN THE *MULTISIM* ENVIRONMENT

For checking the validity of the obtained expressions for determining the ripple currents and voltages of the DC/DC converter based on Zeta topology, circuit simulation in the *Multisim*[2] environment is used. The derivation of analytical formulas and the results of simulating constant components of currents and voltages of the considered converter are presented in [20–22].

The simulation is preceded by the study of MOSFET power switches, as recommended in [23, 24], in static and dynamic mode. On this basis, the IRLZ44N transistor, whose model characteristics correspond to the Datasheet[3] data, was selected.

In the simulation scheme depicted in Fig. 2, the switch VT1 commutes the current of the input power supply V1 with the frequency of the clock pulses set by generator V2. Components from the *Multisim* database are selected as elements. The chokes are represented by equivalent circuits. The active resistance of the choke having an inductance of 55 µH does not exceed 1 Ohm.

To study the impact of the duty factor D on ripples, the circuit is simulated in the transient analysis mode. Here, the current and voltage ripples are recorded in the steady-state mode 5–12 ms after the simulation has started. The study results of the impact of duty factor D as the main parameter determining the converter operation mode are shown in Figs. 3–5. The impact of switching frequency f on the ripple spreading value is shown in Figs. 6–8.

The studies of the impact of duty factor D on the ripple spreading value show a good coincidence of the results of the mathematical model and simulation. However, at duty factors D less than 0.3 and greater than 0.7, a significant difference is observed. This discrepancy is caused by imperfection of mathematical model and impact of parasitic parameters of radio components on converter operation. The coincidence of calculated ripple spreading Δi_{in}, Δi_{Lav}, and Δu_C along with values

[2] https://www.ni.com/ru-ru.html. Accessed March 27, 2023 (in Russ.).

[3] International rectifier, IRFZ44N HEXFET Power MOSFET, Data Sheet. https://static.chipdip.ru/lib/158/DOC000158617.pdf. Accessed March 27, 2023.

Δi_{inm}, Δi_{Lm}, and Δu_{Cm} obtained during simulation is observed at duty factor D equal to 0.5. Here, Δi_{inm} is input current ripple spreading obtained by simulation; Δi_{Lm} is the ripple spreading of current flowing through winding of chokes L1 and L2 obtained by simulation; Δu_{Cm} is the voltage ripple spreading on capacitors C1 and C2 obtained by simulation. The difference between calculated values and simulation results for input current ripple is 6 mA at Δi_{in} = 718 mA. The difference between calculated values and simulation results for current ripple i_{L1}, i_{L2} is 7 mA at Δi_{Lav} = 218 mA. The difference between calculated values and simulation results is 0.3 mV at Δu_{C1} = 15 mV and 0.05 mV at Δu_{C2} = 5.45 mV for voltage ripple spreading Δu_{C1} and Δu_{C2}, respectively.

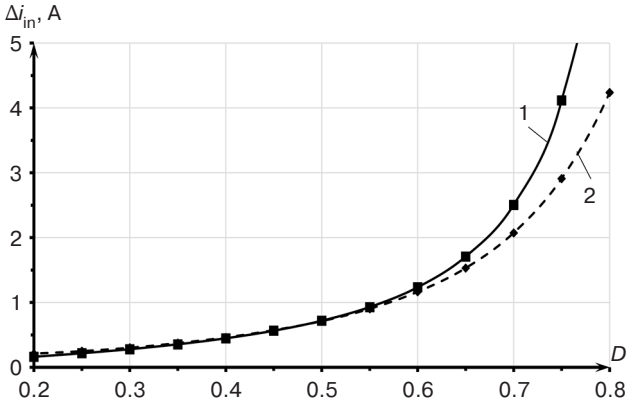


Fig. 3. Impact of the duty factor on the input current ripple: 1 is the calculated value Δi_{in}; 2 is the simulation result Δi_{inm}

From Fig. 3 it can be seen that the calculated value of input current ripple corresponds to the simulation results across almost the whole variation range of duty factor D. However, the calculation and simulation results differ at duty factors $D < 0.25$ and $D > 0.75$. This is especially evident at a duty factor D greater than 0.75. For example, at duty factor D = 0.8, the difference of calculated and simulated ripple is ~4 A at Δi_{in} = 4.3 A while at duty factor D = 0.2, this difference is ~48 mA at Δi_{in} = 163 mA.

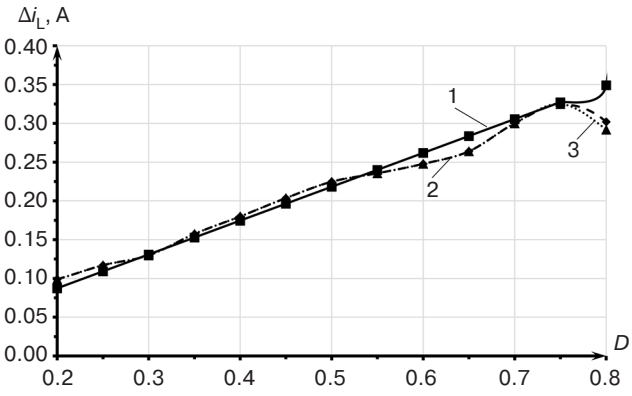


Fig. 4. Impact of the duty factor on ripple currents flowing through windings of chokes L1 and L2: 1 is the calculated value Δi_L; 2 is the simulation result Δi_{L1m}; 3 is the simulation result Δi_{L2m}

Figure 4 shows that the ripple currents flowing through the choke windings practically coincide. The calculation and simulation results, as well as in case of input current, differ systematically at duty factor $D < 0.25$ and $D > 0.75$. At duty factor D = 0.8, the difference of calculated and simulated ripple values is ~47 mA at Δi_L = 349 mA while at duty factor D = 0.2, this difference is ~13 mA at Δi_L = 87 mA.

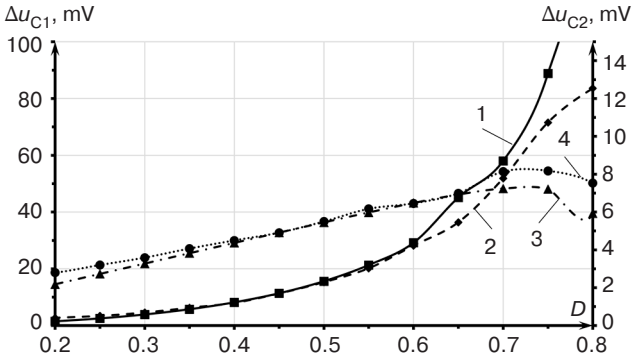


Fig. 5. Impact of the duty factor on the voltage ripple at capacitors C1 and C2: 1 is the calculated value Δu_{C1}; 2 is the simulation result Δu_{C1m}; 3 is the calculated value Δu_{C2}; 4 is the simulation result Δu_{C2m}

It follows from Fig. 5 that at duty factor of 0.5, the calculated ripples and simulation results coincide. However, the voltage ripple on capacitor C2 differs by 1.59 mV at 6 mV if D = 0.8, and by 0.6 mV at Δu_{C2} = 2.2 mV if D = 0.2.

The difference between calculated voltage value Δu_{C1} on capacitor C1 and simulation result Δu_{C1m} reaches 76 mV at Δu_{C1} = 160 mV if D ~ 0.8; however, the difference between the calculation and simulation results would be insignificant at D ~ 0.2.

When the switching frequency of the power transistor is increased, the ripple currents and voltages are significantly reduced. As shown in Fig. 6, the calculated ripple values Δi_{in} and simulation results Δi_{inm} coincide at switching frequencies from 50 to 800 kHz. The maximum difference of the calculated ripples and simulation in the operating frequency band is observed at the frequency of 50 kHz (in the enlarged scale, the difference is shown in the inset of Fig. 7) and is 0.35 A at Δi_{in} = 4.2 A.

It is shown in Fig. 7 that calculated ripple Δi_L and simulated ripple Δi_{Lm}, as well as the input current, coincide at switching frequencies from 50 to 800 kHz. The maximum difference between the calculated ripple and simulated ripple in the operating frequency band is observed at 50 kHz (in the enlarged scale, the difference is shown in the inset of Fig. 7) and is 0.26 A for Δi_{L1} and 0.2 A for Δi_{L2} at Δi_L = 2.18 A.

The limit of the continuous mathematical model is clearly visible in Fig. 8, in contrast to Fig. 7. For

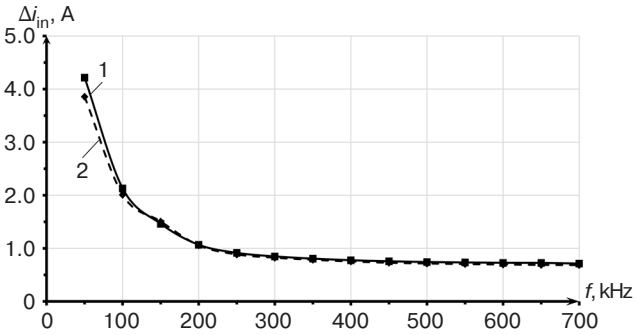


Fig. 6. Impact of switching frequency on the input current ripple at the duty factor equal to 0.5: 1 is the calculated value Δi_{in}; 2 is the simulation result Δi_{inm}

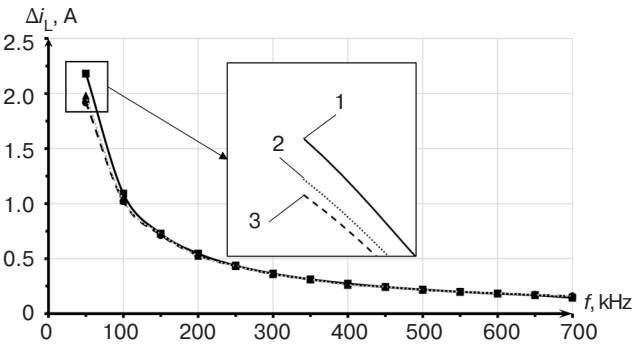


Fig. 7. Impact of switching frequency on ripple currents flowing through the winding of chokes L1 and L2 at the duty factor equal to 0.5: 1 is the calculated value Δi_L; 2 is the simulation result Δi_{L2m}; 3 is the simulation result Δi_{L1m}

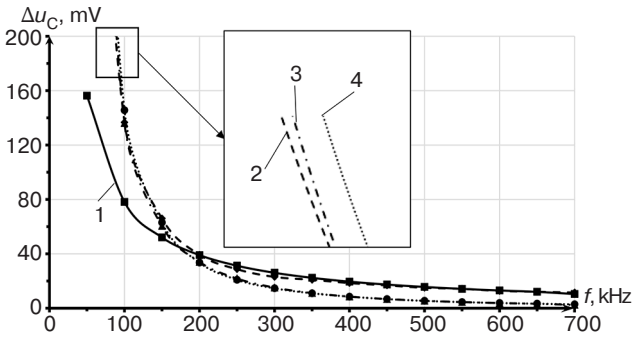


Fig. 8. Impact of switching frequency on the voltage on capacitors C1, C2 at the duty factor equal to 0.5: 1 is the calculated value Δu_{C1}; 2 is the simulation result Δu_{C1m}; 3 is the calculated value Δu_{C2}; 4 is the simulation result Δu_{C2m}

example, at the frequency of 100 kHz, the calculated value Δu_{C1} is 78 mV while simulation Δu_{C1} is 136 mV. The values of phase coordinates of the limiting continuous mathematical model tend to the values of currents and voltages of the converter at frequencies higher than 200 kHz. It can be seen from Fig. 8 that the calculated ripple and the simulated ripple coincide at frequencies higher than 200 kHz. The maximum difference is observed at ~230 kHz and is equal to ~3 mV for Δu_{C1} at Δu_{C1} = 31 mV. For ripples Δu_{C2}, the difference is ~0.9 mV at Δu_{C2} = 21.8 mV.

CONCLUSIONS

The present work presents equivalent circuits Zeta topology-based DC/DC converters in the modes of energy storage and transfer. Using Kirchhoff's rules, systems of equations describing each phase of the device are set up. In order to formulate the mathematical model in matrix form, the systems of equations are converted into a phase coordinate matrix, a coefficient matrix of phase coordinates, and a coefficient matrix of the external source. To permit the complete analysis of the limiting continuous mathematical model, equations for constant components are written.

Using the representation of currents and voltages as a sum of constant and variable components, systems of equations describing the converter in the modes of energy storage and transfer are written. Derived expressions of constant components of currents and voltages are integrated with systems of equations describing each phase and expressions for determining the ripple currents flowing through the winding of chokes and voltages on capacitors.

The results of ripple calculations when using the limiting continuous mathematical model and converter simulation are compared. The dependences of ripples on duty factor D and switching frequency of power switch f are obtained. At duty factor D = 0.5, the ripples obtained by the mathematical model coincide with those obtained by simulation. The difference is ~7 mA at Δi_L = 220 mA for ripple currents Δi_{L1}, Δi_{L2}. The difference between the voltage ripples is ~0.3 mV at Δu_{C1} = 15 mV for the Δu_{C1} ripple and ~0.05 mV at Δu_{C2} = 5.5 mV for the Δu_{C2} ripple. The maximum deviations of calculated values from simulation results at duty factor D = 0.8 are 47 mA for Δi_{L1} and Δi_{L2}, 76.4 mV for Δu_{C1}, and 1.59 mV for Δu_{C2}.

The presented limiting continuous mathematical model allows variation range of currents flowing through the choke windings and voltages at the coatings of the capacitor to be estimated along with their maximum and minimum values at different converter parameters such as switching frequency of the power switch, duty factor, nominal values of elements, etc. The obtained dependences can be used to rationally select electronic components for constructing Zeta topology converters.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Bityukov V.K., Ivanov A.A., Mironov A.V., Mikhnevich N.G., Perfiliev V.S., Petrov V.A. Test bench for studying characteristics integrated circuit chips of secondary regulated charge pump power supply. *Russ. Technol. J.* 2016;4(3):37–52 (in Russ.). https://doi.org/10.32362/2500-316X-2016-4-3-37-52
2. Zakharov L.F. Pulsed voltage regulator with wide input voltage change limits. *Elektrosvyaz = The Elektrosvyaz Magazine*. 2018;11:81−82 (in Russ.).
3. Garelina S.A., Latyshenko K.P., Frunze A.V., Gorbunov R.A. Practical implementation of pyrometers for measuring flame temperature and objects through a flame. *Nauchnye i obrazovatel'nye problemy grazhdanskoi zashchity = Scientific and Educational Tasks of Civil Defence*. 2020;2(45):110–115 (in Russ.).
4. Odinokov A.O., Kremzukov Yu.A. Selecting DC converter topology. SEPIC or Zeta. *Prakticheskaya silovaya elektronika = Practical Power Electronics*. 2022;4(88):44–47 (in Russ.).
5. Vasyukov I.V., Pavlenko A.V., Batishchev D.V. Review and analysis of topologies of converters of power supply systems on hydrogen fuel cells for unmanned aerial vehicles of kilowatt power class. *Izvestiya vysshikh uchebnykh zavedenii. Elektromekhanika = Russian Electromechanics*. 2022;65(2):19–26 (in Russ.). https://doi.org/10.17213/0136-3360-2022-2-19-26
6. Choudhary V. Transient resistant primary DC/DC converter solutions for automotive electronics power systems. *Silovaya elektronika = Power Electronics*. 2017;3(66):30–34 (in Russ.). Available from URL: https://power-e.ru/wp-content/uploads/6630.pdf
7. Kazantsev D.P., Shcherbak V.F., Zakamaldin D.A., Kazantsev Yu.E., Pronin A.V., Molodykh S.V. *Electric Cardiac Pacemaker*: RF Pat. 2531695. Publ. 27.10.2014 (in Russ.).
8. Kastrov M.Yu., Makarov V.V. Zeta converter based on SIPEX SP16125/6/7 controllers design baseline. *Prakticheskaya silovaya elektronika = Practical Power Electronics*. 2010;2(38):15–18 (in Russ.).
9. Korshunov A.I. Methodology for constructing continuous models of pulsed DC voltage converters. *Komponenty i tekhnologii = Components & Technologies*. 2006;8(61):124−130 (in Russ.). Available from URL: https://kit-e.ru/powerel/metodika-postroeniya-nepreryvnyh-modelej-impulsnyh-preobrazovatelej-napryazheniya-postoyannogo-toka/
10. Bityukov V.K., Simachkov D.S., Babenko V.P. *Istochniki vtorichnogo elektropitaniya* (*Secondary Power Sources*). Moscow: Infra-Inzheneriya; 2020. 376 p. (in Russ.).
11. Erickson R.W., Maksimović D. *Fundamentals of Power Electronics*. NY: Springer; 2020. 1084 p. https://doi.org/10.1007/978-3-030-43881-4
12. Jozwik J.J., Kazimierczuk M.K. Dual Sepic PWM Switching-Mode DC/DC Power Converter. *IEEE Transactions on Industrial Electronics*. 1989;36(1):64–70. https://doi.org/10.1109/41.20346

СПИСОК ЛИТЕРАТУРЫ

1. Битюков В.К., Иванов А.А., Миронов А.В., Михневич Н.Г., Перфильев В.С., Петров В.А. Стенд для исследования характеристик микросхем источников вторичного электропитания с накачкой заряда. *Russ. Technol. J.* 2016;4(3):37–52. https://doi.org/10.32362/2500-316X-2016-4-3-37-52
2. Захаров Л.Ф. Импульсный стабилизатор напряжения с широкими пределами изменения входного напряжения. *Электросвязь*. 2018;11:81−82.
3. Гарелина С.А., Латышенко К.П., Фрунзе А.В., Горбунов Р.А. Практическая реализация пирометров для измерения температуры пламени и объектов сквозь пламя. *Научные и образовательные проблемы гражданской защиты*. 2020;2(45):110–115.
4. Одиноков А.О., Кремзуков Ю.А. Выбор топологии преобразователя постоянного напряжения. SEPIC или Zeta. *Практическая силовая электроника*. 2022;4(88):44–47.
5. Васюков И.В., Павленко А.В., Батищев Д.В. Обзор и анализ топологий преобразователей систем электропитания на водородных топливных элементах для беспилотных летательных аппаратов киловаттного класса мощности. *Известия ВУЗов. Электромеханика*. 2022;65(2):19–26. https://doi.org/10.17213/0136-3360-2022-2-19-26
6. Чоудхари В. Решения устойчивого к переходным процессам первичного DC/DC-преобразователя системы питания автомобильной электроники. *Силовая электроника*. 2017;3(66):30–34. URL: https://power-e.ru/wp-content/uploads/6630.pdf
7. Казанцев Д.П., Щербак В.Ф., Закамалдин Д.А., Казанцев Ю.Е., Пронин А.В., Молодых С.В. *Оптико-электронный аппарат*: пат. 2531695 РФ. Заявка № 2012156979/14; заявл. 24.12.2012; опубл. 27.10.2014. Бюл. № 30.
8. Кастров М.Ю., Макаров В.В. Основы разработки преобразователя Zeta на базе контроллеров SIPEX SP16125/6/7. *Практическая силовая электроника*. 2010;2(38):15–18.
9. Коршунов А.И. Методика построения непрерывных моделей импульсных преобразователей напряжения постоянного тока. *Компоненты и технологии*. 2006;8(61):124−130. URL: https://kit-e.ru/powerel/metodika-postroeniya-nepreryvnyh-modelej-impulsnyh-preobrazovatelej-napryazheniya-postoyannogo-toka/
10. Битюков В.К., Симачков Д.С., Бабенко В.П. *Источники вторичного электропитания*. М.: Инфра-Инженерия; 2020. 376 с.
11. Erickson R.W., Maksimović D. *Fundamentals of Power Electronics*. NY: Springer; 2020. 1084 p. https://doi.org/10.1007/978-3-030-43881-4
12. Jozwik J.J., Kazimierczuk M.K. Dual Sepic PWM Switching-Mode DC/DC Power Converter. *IEEE Transactions on Industrial Electronics*. 1989;36(1):64–70. https://doi.org/10.1109/41.20346
13. Пономарев Ю.Г., Присмотров Н.И., Шураков И.А. Вентильный электропривод с высоким коэффициентом мощности. *Интеллектуальная электротехника*. 2022;1(17):27–41. https://doi.org/10.46960/2658-6754_2022_1_27

13. Ponomarev Yu.G., Prismotrov N.I., Shurakov I.A. BLDC and PMBLAC motor drives with high power factor. *Intellektual'naya elektrotekhnika = Smart Electrical Engineering*. 2022;1(17):27–41 (in Russ.). https://doi.org/10.46960/2658-6754_2022_1_27
14. Korshunov A.I. Limiting continuous model of a system with high-frequency structure variation. *Izvestiya vysshikh uchebnykh zavedenii. Priborostroenie = J. Instrument Eng.* 2009;52(9):42−48 (in Russ.). Available from URL: https://pribor.ifmo.ru/file/article/4902.pdf
15. Korshunov A.I. Limiting continuous model of a system with periodic high-frequency structure variation. *Silovaya elektronika = Power Electronics*. 2021;5(92):48−51 (in Russ.).
16. Korshunov A.I. Improving the quality of output voltage stabilisation of a pulsed DC converter. *Izvestiya vysshikh uchebnykh zavedenii. Priborostroenie = J. Instrument Eng.* 2013;56(3):48−57 (in Russ.). Available from URL: https://pribor.ifmo.ru/file/article/6176.pdf
17. Korshunov A.I. Two approaches to stability analysis of DC-DC converters with variable structure of power stage. *Prakticheskaya silovaya elektronika = Practical Power Electronics*. 2017;2(66):12−19 (in Russ.).
18. Korshunov A.I. Pulsed DC voltage converter based on Chuk scheme. *Silovaya elektronika = Power Electronics*. 2017;4(67):60−66 (in Russ.).
19. Korshunov A.I. Specifics of DC voltage stabilization by Chuk converter. *Prakticheskaya silovaya elektronika = Practical Power Electronics*. 2017;4(68):2−9 (in Russ.).
20. Bityukov V.K., Lavrenov A.I. Mathematical model of a DC/DC voltage converter based on Zeta topology. In: *Fundamental, Prospecting, Applied Research and Innovative Projects*: *Proceedings of the National Scientific and Practical Conference.* Moscow: RTU MIREA; 2022. P. 209−215 (in Russ.).
21. Bityukov V.K., Lavrenov A.I., Malitskiy D.A. Mathematical model of DC/DC converter based on Zeta topology (Part 1). *Proektirovanie i tekhnologiya elektronnykh sredstv = Design and Technology of Electronic Means*. 2022;4:53−57 (in Russ.).
22. Bityukov V.K., Lavrenov A.I., Malitskiy D.A. Mathematical model of DC/DC converter based on Zeta topology (Part 2). *Proektirovanie i tekhnologiya elektronnykh sredstv = Design and Technology of Electronic Means*. 2023;1:48−53 (in Russ.).
23. Babenko V.P., Bityukov V.K. Simulation of Switching of High-Power FETs Using the Electronics Workbench Software. *J. Commun. Technol. Electron.* 2019;64(2): 176–181. https://doi.org/10.1134/S1064226919020025 [Original Russian Text: Babenko V.P., Bityukov V.K. Simulation of Switching of High-Power FETs Using the Electronics Workbench Software. *Radiotekhnika i elektronika = J. Commun. Technol. Electron.* 2019;64(2):199−205 (in Russ.). https://doi.org/10.1134/S0033849419020025]
24. Babenko V.P., Bityukov V.K., Kuznetsov V.V., Simachkov D.S. Simulation of static and dynamic losses in MOSFET keys. *Russ. Technol. J.* 2018;6(1):20−39 (in Russ.). https://doi.org/10.32362/2500-316X-2018-6-1-20-39
14. Коршунов А.И. Предельная непрерывная модель системы с высокочастотным периодическим изменением структуры. *Известия ВУЗов. Приборостроение*. 2009;52(9):42−48. URL: https://pribor.ifmo.ru/file/article/4902.pdf
15. Коршунов А.И. Предельная непрерывная модель системы с периодическим высокочастотным изменением структуры. *Силовая электроника*. 2021;5(92):48−51.
16. Коршунов А.И. Повышение качества стабилизации выходного напряжения импульсного преобразователя постоянного тока. *Известия ВУЗов. Приборостроение*. 2013;56(3):48−57. URL: https://pribor.ifmo.ru/file/article/6176.pdf
17. Коршунов А.И. Два подхода к анализу устойчивости стабилизаторов напряжения постоянного тока с переменной структурой силовой части. *Практическая силовая электроника*. 2017;66:12−19.
18. Коршунов А.И. Импульсный преобразователь напряжения постоянного тока по схеме Чука. *Силовая электроника*. 2017;4(67):60−66.
19. Коршунов А.И. Особенности стабилизации напряжения постоянного тока с помощью преобразователя Чука. *Практическая силовая электроника*. 2017;4(68):2−9.
20. Битюков В.К., Лавренов А.И. Математическая модель DC/DC преобразователя напряжения, построенного по Zeta топологии. *Фундаментальные, поисковые, прикладные исследования и инновационные проекты: сборник трудов Национальной научно-практической конференции*; под ред. С.У. Увайсова. М.: РТУ МИРЭА; 2022. С. 209−215.
21. Битюков В.К., Лавренов А.И., Малицкий Д.А. Математическая модель DC/DC преобразователя, построенного по Zeta топологии (часть 1). *Проектирование и технология электронных средств*. 2022;4:53−57.
22. Битюков В.К., Лавренов А.И., Малицкий Д.А. Математическая модель DC/DC преобразователя, построенного по Zeta топологии (часть 2). *Проектирование и технология электронных средств*. 2023;1:48−53.
23. Бабенко В.П., Битюков В.К. Имитационное моделирование процессов переключения силовых полевых транзисторов в программе Electronics Workbench. *Радиотехника и электроника*. 2019;64(2):199−205. https://doi.org/10.1134/S0033849419020025
24. Бабенко В.П., Битюков В.К., Кузнецов В.В., Симачков Д.С. Моделирование статических и динамических потерь в MOSFET ключах. *Russ. Technol. J.* 2018;6(1):20−39. https://doi.org/10.32362/2500-316X-2018-6-1-20-39

About the authors

Vladimir K. Bityukov, Dr. Sci. (Eng.), Professor, Department of Radio Wave Processes and Technology, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: bitukov@mirea.ru. ResearcherID Y-8325-2018, Scopus Author ID 6603797260, RSCI SPIN-code 3834-5360, https://orcid.org/0000-0001-6448-8509

Alexey I. Lavrenov, Assistant, Department of Radio Wave Processes and Technology, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: lavrenov@mirea.ru. SPIN-код РИНЦ 6048-5027, https://orcid.org/0000-0001-5722-541X

Daniil A. Malitskiy, Assistant, Department of Radio Wave Processes and Technology, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: malickij@mirea.ru. SPIN-код РИНЦ 4912-3018, https://orcid.org/0000-0003-4558-9085

Об авторах

Битюков Владимир Ксенофонтович, д.т.н., профессор, кафедра радиоволновых процессов и технологий Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: bitukov@mirea.ru. ResearcherID Y-8325-2018, Scopus Author ID 6603797260, SPIN-код РИНЦ 3834-5360, https://orcid.org/0000-0001-6448-8509

Лавренов Алексей Игоревич, ассистент, кафедра радиоволновых процессов и технологий Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: lavrenov@mirea.ru. SPIN-код РИНЦ 6048-5027, https://orcid.org/0000-0001-5722-541X

Малицкий Даниил Александрович, ассистент, кафедра радиоволновых процессов и технологий Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: malickij@mirea.ru. SPIN-код РИНЦ 4912-3018, https://orcid.org/0000-0003-4558-9085

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Modern radio engineering and telecommunication systems

Современные радиотехнические и телекоммуникационные системы

UDC 621.391
https://doi.org/10.32362/2500-316X-2023-11-4-49-58

RESEARCH ARTICLE

Multi-task neural network for solving the problem of recognizing the type of QAM and PSK modulation under parametric a priori uncertainty

Aleksei A. Paramonov [@], Van Minh Nguyen, Minh Tuong Nguyen

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: paramonov@mirea.ru

Abstract

Objectives. Automatic modulation recognition of unknown signals is an important task for various fields of technology such as radio control, radio monitoring, and identification of interference and sources of radio emission. The paper aims to develop a method for recognizing the types of signal modulation under conditions of parametric a priori uncertainty, including the uncertainty of carrier frequency- and initial signal phase values. An additional task consists in estimating the offset values of the carrier frequency or signal phase at the initial stage of the recognition process.

Methods. A multi-task learning with artificial neural network and the theory of cumulants of random variables are used.

Results. For signals with a carrier frequency and initial phase shift, cumulant approaches for QAM-8, APSK-16, QAM-64, and PSK-8 modulations are calculated. A multi-task learning with artificial neural network using cumulant features and a data standardization algorithm is presented. The results of the experiment show that using multi-task learning with an artificial neural network provides high accuracy of recognizing QAM-8 and APSK-16, QAM-64 and PSK-8 modulations with small mismatches of the carrier frequency or initial phase. The accuracy of determining the offset values from the carrier frequency or the initial phase for QAM-8, APSK-16, QAM-64, and PSK-8 modulation is high.

Conclusions. The multi-task learning with neural network using high-order signal cumulants makes it possible not only to recognize modulation types with high accuracy under conditions of a priori uncertainty of signal parameters, but also to determine the offset values of carrier frequency or initial signal phase from expected values.

Keywords: recognition, neural network, carrier frequency, initial phase, cumulant feature

• **Submitted:** 24.03.2023 • **Revised:** 11.04.2023 • **Accepted:** 02.05.2023

For citation: Paramonov A.A., Nguyen V.M., Nguyen M.T. Multi-task neural network for solving the problem of recognizing the type of QAM and PSK modulation under parametric a priori uncertainty. *Russ. Technol. J*. 2023;11(4):49–58. https://doi.org/10.32362/2500-316X-2023-11-4-49-58

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Многозадачная нейронная сеть в задаче распознавания вида QAM- и PSK-модуляции в условиях параметрической априорной неопределенности

А.А. Парамонов [@], В.М. Нгуен, М.Т. Нгуен

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] Автор для переписки, e-mail: paramonov@mirea.ru

Резюме

Цели. Автоматическое распознавание видов модуляции неизвестных сигналов является важной задачей для различных областей техники: радиоконтроля и радиомониторинга, идентификации помех и источников радиоизлучения. Основная цель работы – разработка метода распознавания видов модуляции сигналов в условиях параметрической априорной неопределенности, в т.ч. неопределенности значений несущей частоты и начальной фазы сигнала. Дополнительной задачей является оценка значений отстроек от несущей частоты или фазы сигнала на начальном этапе процесса распознавания.

Методы. Использована многозадачная искусственная нейронная сеть, теория кумулянтов случайных величин.

Результаты. Для сигналов со сдвигом несущей частоты и начальной фазы вычислены кумулянты для модуляции QAM-8, APSK-16, QAM-64 и PSK-8. Представлена использующая кумулянтные признаки и алгоритм стандартизации данных многозадачная нейронная сеть. Результаты эксперимента показали, что использование многозадачной нейронной сети обеспечивает высокую точность распознавания модуляции QAM-8 и APSK-16, QAM-64 и PSK-8 в случае небольших отстроек несущей частоты или начальной фазы. Точность определения значений отстройки несущей частоты или начальной фазы сигнала для модуляции QAM-8, APSK-16, QAM-64 и PSK-8 оказывается высокой.

Выводы. Многозадачная нейронная сеть, использующая кумулянты сигналов высокого порядка, позволяет не только распознавать с высокой точностью виды модуляции в условиях априорной неопределенности параметров сигналов, но определять при этом значения отстроек несущей частоты или начальной фазы сигнала от ожидаемых значений.

Ключевые слова: распознавание, нейронная сеть, несущая частота, начальная фаза, кумулянтный признак

• **Поступила:** 24.03.2023 • **Доработана:** 11.04.2023 • **Принята к опубликованию:** 02.05.2023

Для цитирования: Парамонов А.А., Нгуен В.М., Нгуен М.Т. Многозадачная нейронная сеть в задаче распознавания вида QAM- и PSK-модуляции в условиях параметрической априорной неопределенности. *Russ. Technol. J.* 2023;11(4):49–58. https://doi.org/10.32362/2500-316X-2023-11-4-49-58

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

One of the promising development areas of modern telecommunication systems involves the introduction of intelligent technologies offering the possibility to recognize and analyze the information transmitted through the communication channel. Such intelligent systems can be used to process data in real time, make predictions, and take decisions based on the obtained results. Here, an associated problem consists in recognizing the various types of digital modulation used in data transmission [1–8]. Knowledge of the parameters of received signals informs identification of the transmitting device and restoration of the transmitted information, as well as making it possible to introduce interference into location and communication radio channels. However, parametric uncertainties that can significantly affect the accuracy of recognition often occur under real conditions of information transmission.

At present, the most effective method of automatic recognition of signal modulation types involves the use of multi-task neural networks. The algorithm for recognizing modulation types using a multi-task neural network under conditions of a priori certainty of values of carrier frequency and initial phase of signals is presented in [1, 4]. In [1], for recognizing ten modulations (GMSK, 8-QAM, 16-QAM, 64-QAM, 16-APSK, 32-APSK, BPSK, QPSK, 8-PSK, and 2-FSK)[1], up to 9th order cumulants are used as information features. The results of computer simulation show that a multi-task neural network using high-order cumulants can not only be used to recognize modulation types but also to determine the value of signal-to-noise ratio (SNR) of the received signal with high accuracy. At SNR = 0 dB, the detection accuracy for GMSK, QAM-8, APSK-16, APSK-32, BPSK, and QPSK modulation is 0.98; however, for QAM-16, QAM-64, PSK-8, and FSK-2 modulation, the detection accuracy is lower. Hereinafter, recognition accuracy refers to the probability of correctly identifying a particular type of modulation signal among all the considered types of modulation. The present work aims to analyze the recognition of four modulation types—QAM-8 and APSK-16, QAM-64, and PSK-8—under conditions of parametric a priori uncertainty, as well as to determine the offset of carrier frequency Δω and initial phase $\Delta\varphi_0$ of the received signal from the expected values at SNR = 3 dB which, according to simulation results, is sufficient for the recognition of signals with an acceptable accuracy. It is assumed that the received signal is subjected to preprocessing providing the transfer to zero frequency, filtering, and signal sampling from the low-pass filter output[2]. The signal obtained as a result of preprocessing is described by the following expression:

$$r_k(t) = A(t)\{\cos[\Delta\omega t + \varphi(t) + \Delta\varphi_0] + i\sin[\Delta\omega t + \varphi(t) + \Delta\varphi_0]\} = I_k(t) + iQ_k(t),$$

where $A(t)$ and $\varphi(t)$ are the envelope and phase of the signal; Δω is the carrier frequency offset; $\Delta\varphi_0$ is the initial phase offset; $I_k(t)$ and $Q_k(t)$ are the in-phase and quadrature components of the signal, respectively.

The resulting complex signal $r_k(t)$ and complex-conjugate signal $\overline{r}_k(t) = I_k(t) - iQ_k(t)$ comprise the initial data for calculating moments and cumulants. Formulas for calculating moments and high-order cumulants are described in detail in [1, 9]. Table 1 presents examples of values of cumulants up to 9th order for QAM-64 and PSK-8 modulation with different offset values of carrier frequency Δω and initial phase $\Delta\varphi_0$. According to the analysis of the resulting cumulant values, it can be asserted that the information content of a particular cumulant about the type of signal modulation depends significantly on Δω and $\Delta\varphi_0$ offsets. For example, in the absence of offsets, cumulant $C_{2,0}$, being the first in the table, has the same negative sign for both distinguishable types of QAM-64 and PSK-8 modulation; at Δω = 900 Hz, the cumulant signs are different; while, at Δφ0 = 0.04 rad, signs of cumulants are positive.

ARTIFICIAL NEURAL NETWORKS

Neural network methods of modulation type recognition are based on selecting certain information features that can help in determining the type of modulation and on constructing a knowledge base based on the analysis of these features. Each neural network of modulation type recognition has its own set of used information features and parameters including the type of data processing and activation function. These differences can affect the efficiency of recognition and the overall accuracy of the system. The process of neural network learning is shown in Fig. 1.

[1] GMSK is a Gaussian minimum shift keying.
8-QAM, 16-QAM, and 64-QAM are quadrature amplitude modulations with 8, 16, or 64 levels.
16-APSK and 32-APSK are amplitude and phase-shift keying with 16 or 32 levels.
BPSK is a binary phase-shift keying.
QPSK is a quaternary phase-shift keying.
8-PSK is an 8-phase-shift keying.
2-FSK is a binary frequency-shift keying.

[2] Caravan O.V. *Distinguishing constellations of signals with quadrature amplitude modulation under parametric a priori uncertainty*. Cand. Sci. Thesis (Phys.-Math.). Voronezh; 2010. 120 p. (in Russ.).

Table 1. Values of cumulants of different orders for QAM-64 and PSK-8 modulation

Cumulant \ Modulation	$\Delta\omega = 0$ and $\Delta\varphi_0 = 0$		$\Delta\omega = 900$ Hz		$\Delta\varphi_0 = 0.04$ rad	
	QAM-64	PSK-8	QAM-64	PSK-8	QAM-64	PSK-8
$C_{2,0}$	−0.00913	−0.0063	0.01209	−0.01225	1.08181	1.12291
$C_{3,0}$	−0.03605	0.04699	−0.04641	0.04051	−0.01704	0.00593
$C_{2,1}$	−0.00800	0.00439	0.01381	0.01362	−0.01714	0.00621
$C_{4,0}$	−0.56435	−0.01765	3.12438	3.20105	−0.36449	−0.27393
$C_{2,2}$	−0.61285	−0.97795	0.92006	0.85290	−0.36504	−0.27938
$C_{5,0}$	−0.20728	0.092786	0.04405	0.05804	−0.20601	0.20011
$C_{3,2}$	−0.00930	−0.00627	−0.05664	0.01384	−0.21641	0.19121
$C_{6,0}$	0.05008	−0.06516	0.97633	1.33265	1.17776	1.29652
$C_{3,3}$	1.62832	3.84222	−0.99868	−0.566	1.14810	1.33444
$C_{7,0}$	−1.56151	−0.47311	−1.01121	0.51321	1.33945	−0.51879
$C_{6,1}$	0.59933	−0.50570	−1.20013	−1.02734	1.40554	−0.47957
$C_{4,3}$	−0.12922	0.01030	−0.33111	−0.76300	3.20184	−1.62158
$C_{8,0}$	−12.8750	−1.42191	−191.842	−201.990	−4.24775	−11.0519
$C_{6,2}$	−11.6686	0.01107	−28.2597	−22.6214	−4.12208	−11.3511
$C_{4,4}$	1356.218	1413.430	1412.013	1359.459	1041.279	1197.02
$C_{9,0}$	−8.11293	4.10262	−28.8307	−15.8577	−37.4355	2.39996
$C_{8,1}$	6.42682	5.20209	27.0642	−7.68852	−244.414	108.016
$C_{6,3}$	−121.649	114.854	534.622	368.480	3453.083	4508.33
$C_{5,4}$	2467.71	2217.473	5219.517	4876.258	6296.198	8169.887

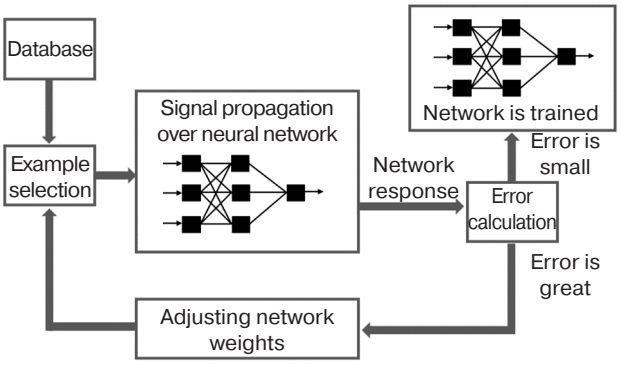


Fig. 1. Process of neural network learning

The neural network itself comprises a system based on many neurons, which receives information, performs simple calculations on it, and then passes it on. After receiving a weighted sum of input signals $net^{(j,1)}$ to its input, each neuron then passes this activation value $\sigma^{(j,1)}$ through the transfer function to obtain the following output values [10–15]:

$$net^{(j,1)} = \mathbf{w}^{(j,1)}\mathbf{x}^{\mathrm{T}} = w_0^{(j,1)} + \sum_{i=1}^{n} w_i^{(j,1)} x_i,$$

$$o^{(j,1)} = f\left(net^{(j,1)}\right),$$

where $\mathbf{w}^{(j,1)} = \left(w_0^{(j,1)}, w_1^{(j,1)}, ..., w_n^{(j,1)}\right), j = \overline{1, N_1}$ is the row vector of synaptic connections at the jth neuron input; $\mathbf{x}_i$ is the row vector of the ith input; N_1 is the number of neurons in the first hidden layer.

The activation function $f(net^{(j,1)})$ in a neural network plays an important role in determining the output of each neuron based on its input signal. This function allows the neuron to decide whether to activate and further transmit information or remain inactive. There are many different activation functions that can be used in neural networks, each having its own advantages and limitations. In the paper, the ReLU (Rectified Linear Unit) activation function is used; this comprises a simple non-linear function for transforming the input signal, zeroing out

all negative values, and keeping the positive values unchanged. Formally, the ReLU function is defined as follows:

$$o^{(j,1)} = \mathrm{ReLU}(x_i) = \max(0, x_i).$$

Learning this neural network consists in minimizing the error function $E(\overline{w})$ defined by the following expression:

$$E(\overline{w}) = \frac{1}{2}\sum_{l=1}^{N_{\text{out}}}\left(u_l - o_l\right)^2,$$

where u_l and o_l are the desired and actual state of neural network outputs of the lth neuron of the output layer, respectively; N_{out} is the number of neurons in the output layer.

Although there are currently several methods for minimizing the error function [2, 12, and 13], the greatest efficiency of the method is obtained when using the database standardization function. The standardization method (StandardScaler) is one of the data preprocessing methods used in machine learning for reducing all the original values of the dataset to the set of values from distribution with zero mean and standard deviation equal to 1. The standardization process consists of two steps. In the first step, the mean and standard deviation of each feature in the dataset are calculated. At the second step, each feature value is transformed by the following formula [14, 15]:

$$x_i = \frac{z_i - \overline{Z}}{\sigma_z},$$

where z_i is the original data value; $\overline{Z}$ and σ_z is the mean value and standard feature deviation, respectively.

The standardization method results in a standardized scale defining the place of each value in the dataset by measuring its deviation from the mean in standard deviation units to compare the data for use in machine learning. As the example, cumulant values resulted from standardization for QAM-64 and PSK-8 modulation and certain mismatches of frequency and initial signal phase are given in Table 2.

Table 2. Cumulant values resulted from standardization

Cumulant \ Modulation	$\Delta\omega = 0$ and $\Delta\varphi_0 = 0$		$\Delta\omega = 900$ Hz		$\Delta\varphi_0 = 0.04$ rad	
	QAM-64	PSK-8	QAM-64	PSK-8	QAM-64	PSK-8
$C_{2,0}$	−0.04020	−0.03519	0.09287	−0.07583	−0.51078	−0.50800
$C_{3,0}$	0.001785	−0.04103	−0.03306	0.03589	−0.00352	0.000098
$C_{2,1}$	0.02363	0.00961	0.01718	0.01703	−0.00368	−0.000024
$C_{4,0}$	−0.43586	−0.43379	−0.42396	−0.42372	−0.00605	0.00768
$C_{2,2}$	−0.43481	−0.43849	−0.42072	−0.42135	0.00736	0.00891
$C_{5,0}$	−0.00613	−0.00639	−0.00661	−0.00655	0.005990	0.006273
$C_{3,2}$	−0.00892	−0.00816	−0.00861	−0.00798	0.005992	0.006272
$C_{6,0}$	−0.00173	−0.00179	−0.00152	−0.00143	0.03657	0.036584
$C_{3,3}$	0.03518	0.03641	0.03381	0.03404	0.03717	0.03718
$C_{7,0}$	−0.00901	−0.00902	−0.00903	−0.009	−0.00052	−0.000533
$C_{6,1}$	−0.01062	−0.01063	−0.01065	−0.01065	−0.00049	−0.0005
$C_{4,3}$	−0.00693	−0.00696	−0.00698	−0.007	0.00054	0.000528
$C_{8,0}$	0.37488	0.37488	0.37485	0.37485	−0.03074	−0.03074
$C_{6,2}$	0.32239	0.32241	0.32237	0.32238	−0.03235	−0.03236
$C_{4,4}$	−0.38226	−0.38227	−0.38226	−0.38226	−0.34224	−0.34224
$C_{9,0}$	−0.00205	−0.00205	−0.00205	−0.00205	−0.00450	−0.004506
$C_{8,1}$	−0.00829	−0.00829	−0.00828	−0.00829	−0.00401	−0.00401
$C_{6,3}$	−0.02267	−0.02226	−0.02226	−0.02226	−0.33499	−0.33499
$C_{5,4}$	−0.37783	−0.37783	−0.37783	−0.37783	−0.33517	−0.33517

SIMULATION RESULTS

The multi-task neural network simulation is performed in the Python environment using Google Colab notebook [14]. For recognizing two groups of QAM-8 and APSK-16, QAM-64 and PSK-8 modulation, four databases are formed. Of these, the first two databases are formed for recognizing these types of modulation under conditions of carrier frequency offset, each base consisting of 12800 signals (800 signals for each offset value from the carrier frequency). Under conditions of initial phase offset, two databases each consisting of 16000 signals (800 signals for each value of the initial phase offset) are also formed. The simulation results of modulation type recognition under conditions of carrier frequency offset are shown in Figs. 2 and 3. The figures are presented in the form of tables, whose rows and columns correspond to the type of signal modulation and the carrier frequency offset. The cells show the results of modulation type recognition. For example, for Fig. 2, when recognizing QAM-8 signals with a zero-frequency shift (the first line in the figure is QAM-8 0), all 80 signals participating in the computer experiment are recognized correctly. When recognizing QAM-8 signal with a frequency shift of 1800 Hz (QAM-8 1800), 75 signals are recognized correctly, while 5 signals are mistakenly identified as APSK-16 1800.

It is clear from the figures that a multi-task neural network can be used not only to perform the recognition of modulation types, but also to determine the carrier frequency offset values. As described above, the accuracy of recognizing a certain type of modulation is understood as the probability of correct detection of this type of signal modulation among all the considered modulation types. In the simulation, this probability is estimated as a sample average, i.e., the ratio of the number of correctly identified signals with a given type of modulation to the total number of different signal realizations participating in the computer experiment. The recognition accuracy of QAM-8 and APSK-16 modulation with different $\Delta\omega$ values is 0.96. The results of recognition under conditions of initial phase offset are shown in Figs. 4 and 5.

From Figs. 3 and 5, it can be seen that the recognition accuracy of QAM-64 and PSK-8 modulation decreases at a large mismatch of both frequency and phase. This is due to the fact that several cumulant values at a large mismatch have unstable behavior for different signal realizations, while values of these several cumulant themselves differ slightly for these modulation types. However, use of the multi-task neural network provides high accuracy for estimating $\Delta\omega$ and $\Delta\varphi_0$ values.

Experimental results on recognizing the received signal with the unknown $\Delta\omega$ value are shown in Fig. 6. The experiment results in a recognition accuracy of 0.53 for QAM-64 modulation and 0.47 for PSK-8 modulation. The value of the carrier frequency offset $\Delta\omega$ equal to 600 Hz is determined with high confidence.

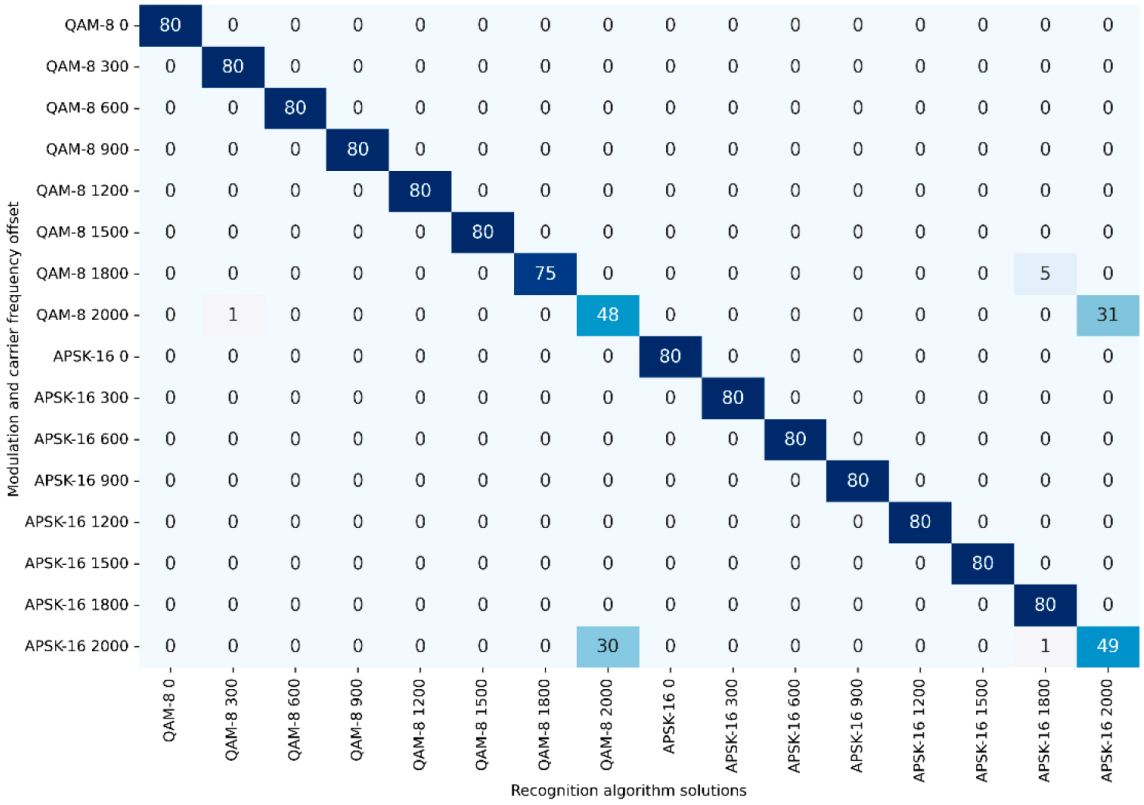


Fig. 2. Recognition results of QAM-8 and APSK-16 modulation at different $\Delta\omega$ values

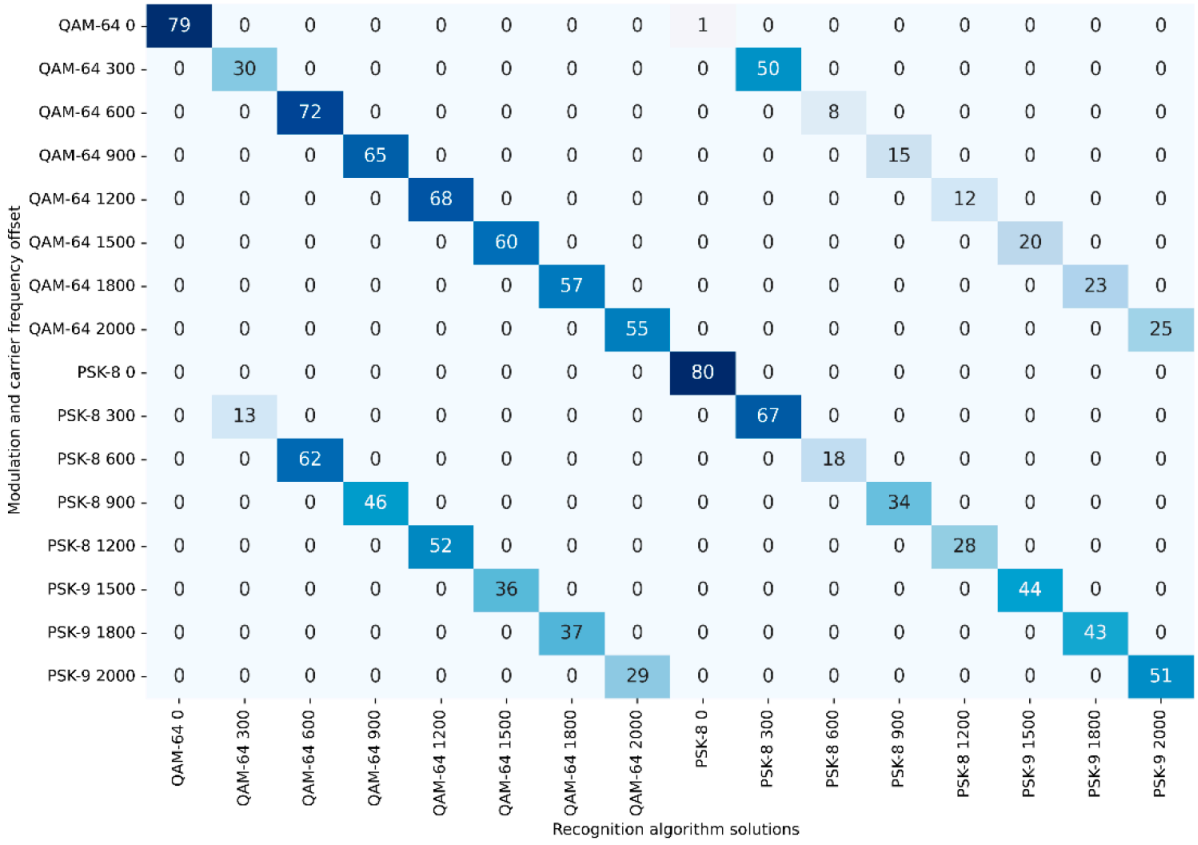


Fig. 3. Recognition results of QAM-64 and PSK-8 modulation at different Δω values

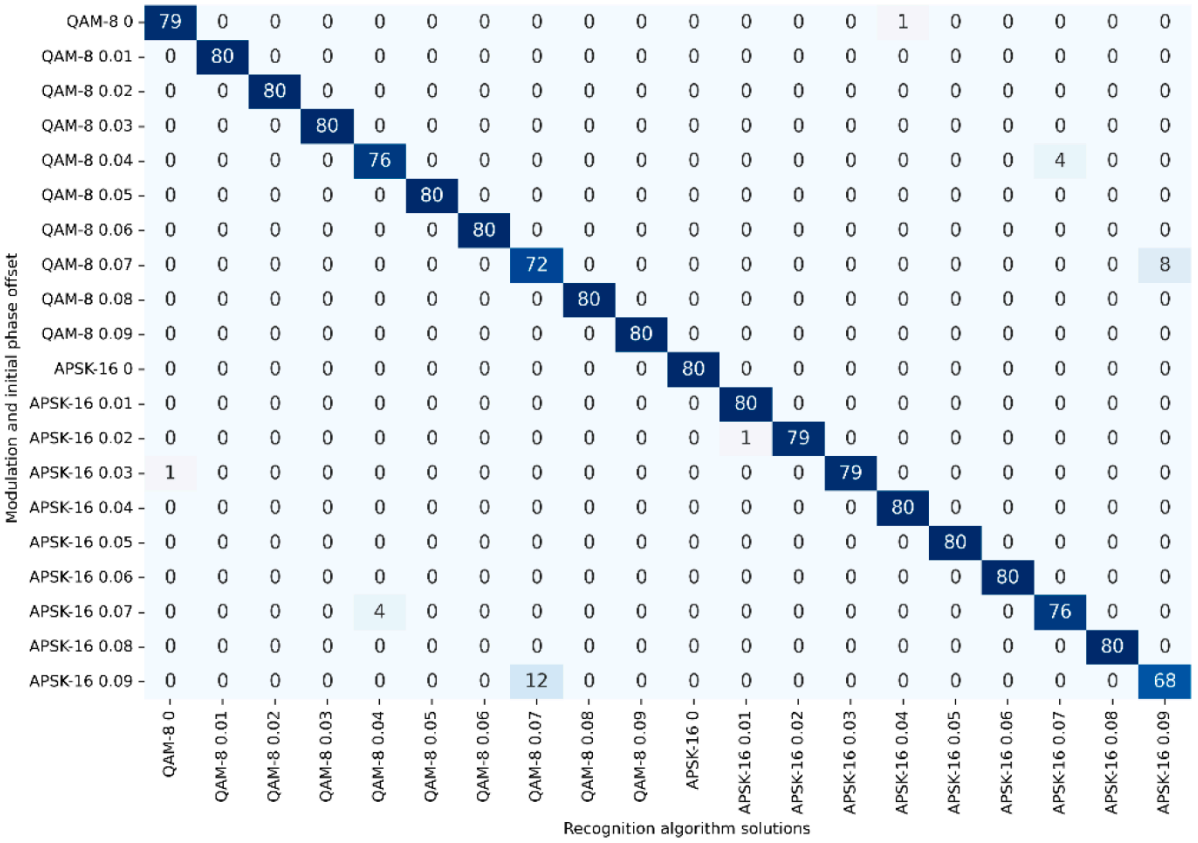


Fig. 4. Recognition results of QAM-8 and APSK-16 modulation at different $\Delta\varphi_0$ values

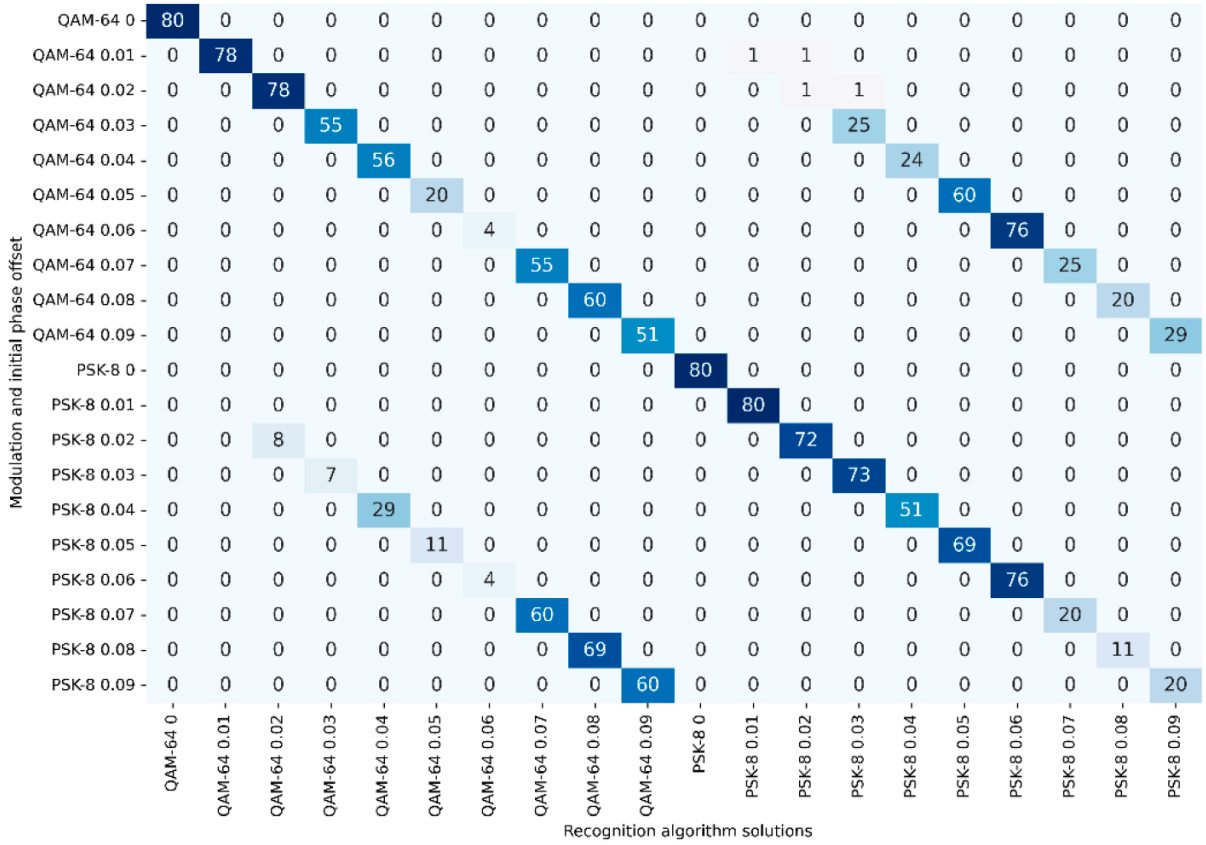


Fig. 5. Recognition results of QAM-64 and PSK-8 modulation at different $\Delta\varphi_0$ values

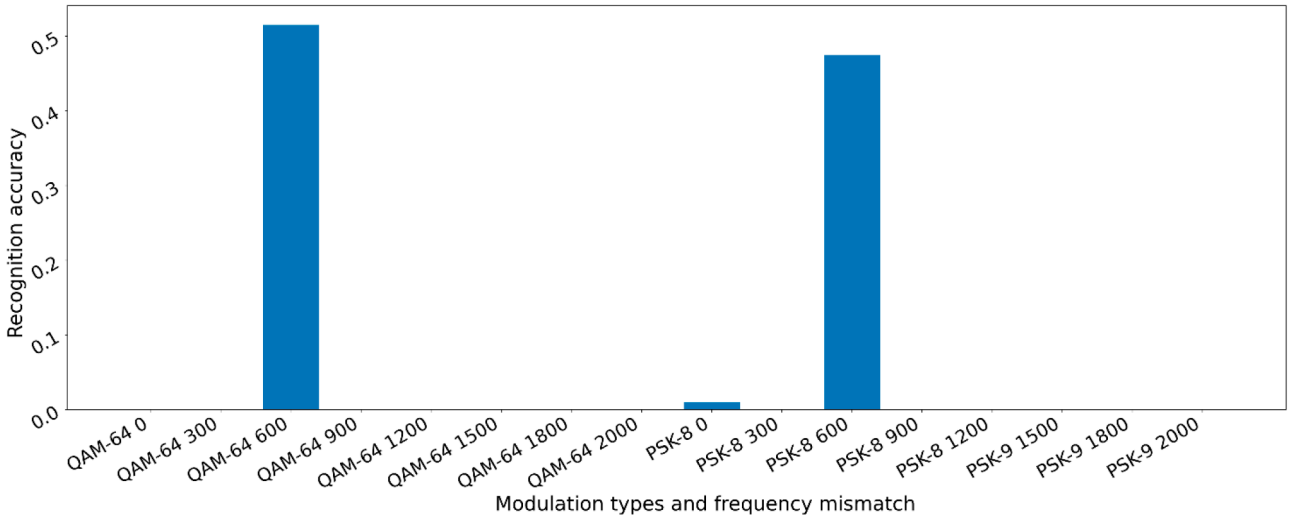


Fig. 6. Experimental result on recognizing the received signal modulation type

CONCLUSIONS

In the paper, the method for recognizing digital modulation types (QAM-8, APSK-8, QAM-64, and PSK-8) in the case of inaccurate knowledge of signal parameters, including carrier frequency and initial phase, is considered. The multi-task neural network is built using the StandardScaler algorithm for data standardization. The simulation results suggest that the multi-task neural network using cumulants as the information feature has the capacity

not only to recognize digital modulation types with high confidence in the case of defining the carrier frequency and initial phase values inaccurately, but also to estimate the associated values. In future, it is proposed to consider the recognition of modulation types at simultaneous uncertainty of carrier frequency and initial phase values.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Paramonov A.A., Nguyen M.V. Recognition of types of digital modulation of radio signals with multi-task neural network. *Vestnik vozdushno-kosmicheskoi oborony = Aerospace Defense Herald*. 2022;4(36):91–97 (in Russ.). Available from URL: https://www.elibrary.ru/item.asp?id=49815162
2. Paramonov A.A., Tikhonova O.V., Nguyen V.M. Recognition of digital modulation of radio signals using a multilayer neural network based on cumulative features. In: *Systems of computer mathematics and their applications: Proceedings of the 23rd International Scientific Conference.* Smolensk: SmolGU; 2022. Issue 23. P. 23–28 (in Russ.).
3. Nabilkov V.D., Priorov A.L., Dubov M.A. Using of the convolutional neural network CLDNN for classification of modulation types. In: *Digital Signal Processing and its Application* (*DSPA 2021*): *Reports 23rd International Conference*. Moscow: A.S. Popov Russian Scientific and Technical Society of Radio Engineering, Electronics and Communications; 2021. P. 228–231 (in Russ.). Available from URL: https://www.elibrary.ru/item.asp?id=45841831
4. Nguyen M.V., Miloradov G.A., Paramonov A.A. Convolutional neural network in the problem of recognizing digital modulation of radio signals. In: *Actual Problems and Prospects for the Development of Radio Engineering and Infocommunication Systems* (*Radioinfocom 2022*): *Collection of Scientific Articles based on the Materials of the 6th International Scientific and Practical Conference*. Moscow: MIREA – Russian Technological University; 2022. P. 181–185 (in Russ.). Available from URL: https://www.elibrary.ru/item.asp?id=49447332&pff=1
5. Avedyan E.D., Nhich D.V. To the selection of the best cumulants features in the recognition task of the digital modulation kind of the radio signals. *Informatizatsiya i svyaz' = Informatization and Communication*. 2015;4:11–15 (in Russ.). Available from URL: https://www.elibrary.ru/item.asp?id=24853422
6. Adjemov S.S., Klenov N.V., Tereshonok M.V., Chirov D.S. Methods for the automatic recognition of digital modulation of signals in cognitive radio systems. *Moscow Univ. Phys. Bull.*2015;70(6):448–456. https://doi.org/10.3103/S0027134915060028
   [Original Russian Text: Adjemov S.S., Klenov N.V., Tereshonok M.V., Chirov D.S. Methods for the automatic recognition of digital modulation of signals in cognitive radio systems. *Vestnik Moskovskogo Universiteta. Ser. 3. Fizika. Astronomiya*. 2015;6:19–27 (in Russ.). Available from URL: https://www.elibrary.ru/item.asp?id=25580690]
7. Grishin P.S., Shabanov A.V., Shcheglov A.V. Recognition of digital radio signal modulations using a multitasking convolutional neural network. In: *Intelligent Information Systems: Theory and Practice: A collection of scientific articles based on the materials of the First All-Russian Conference*. *Part 1*. Kursk: Kursk State University; 2020. P. 22–31 (in Russ.). Available from URL: https://elibrary.ru/KADKBX

СПИСОК ЛИТЕРАТУРЫ

1. Парамонов А.А., Нгуен М.В. Распознавание видов цифровой модуляции радиосигналов с использованием многозадачной нейронной сети. *Вестник воздушно-космической обороны*. 2022;4(36):91–97. URL: https://www.elibrary.ru/item.asp?id=49815162
2. Парамонов А.А., Тихонова О.В., Нгуен В.М. Распознавание видов цифровой модуляции радиосигналов с использованием многослойной нейронной сети по кумулятным признакам. *Системы компьютерной математики и их приложения: Материалы XXIII международной научной конференции*. Смоленск: Изд-во СмолГУ; 2022. Вып. 23. С. 23–28.
3. Набилков В.Д., Приоров А.Л., Дубов М.А. Использование сверточной нейронной сети CLDNN в задаче распознавания видов модуляции радиосигналов. *Цифровая обработка сигналов и ее применение* (*ЦОС – 2021*): *доклады XXIII Международной конференции*. М.: Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова; 2021. С. 228–231. URL: https://www.elibrary.ru/item.asp?id=45841831
4. Нгуен М.В., Милорадов Г.А., Парамонов А.А. Сверточная нейронная сеть в задаче распознавания цифровой модуляции радиосигналов. *Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем* (*Радиоинфоком – 2022*): *сборник научных статей по материалам VI Международной научно-практической конференции*. М.: МИРЭА – Российский технологический университет; 2022. С. 181–185. URL: https://www.elibrary.ru/item.asp?id=49447332&pff=1
5. Аведьян Э.Д., Дам В.Н. К выбору кумулянтных признаков в задаче распознавания видов цифровой модуляции радиосигналов. *Информатизация и связь*. 2015;4:11–15. URL: https://www.elibrary.ru/item.asp?id=24853422
6. Аджемов С.С., Кленов Н.В., Терешонок М.В., Чиров Д.С. Методы распознавания видов цифровой модуляций сигналов в когнитивных радиосистемах. *Вестник Московского Университета*. *Серия 3. Физика. Астрономия*. 2015;6:19–27. URL: https://www.elibrary.ru/item.asp?id=25580690
7. Гришин П.С., Шабанов А.В., Чеглов А.В. Распознавание цифровых модуляций радиосигналов с использованием многозадачной сверточной нейронной сети. *Интеллектуальные информационные системы: Теория и практика: сборник научных статей по материалам I Всероссийской конференции.* Часть 1. Курск: Курский государственный университет; 2020. С. 22–31. URL: https://elibrary.ru/KADKBX
8. Архипенков Д.В. Анализ параметров радиосигналов для идентификации источника излучения. *Доклады Белорусского государственного университета информатики и радиоэлектроники.* 2020;18(1):52–58. https://doi.org/10.35596/1729-7648-2020-18-1-52-58

8. Arkhipenkov D.V. Analysis of radio signal parameters for emission source identification. *Doklady Belorusskogo Gosudarstvennogo Universiteta Informatiki i Radioelektroniki* (*Doklady BGUIR*). 2020;18(1):52–58 (in Russ.). https://doi.org/10.35596/1729-7648-2020-18-1-52-58
9. Kendall M., Stuart A. *Teoriya raspredelenii* (*Distribution Theory*): transl. from Engl. Moscow: Nauka; 1966. 588 p. (in Russ.).
[Kendall M.G., Stuart A. *The Advanced Theory of Statistics. V. 1. Distribution Theory.* London; 1963. 433 p.]
10. Haykin S. *Neironnye seti: polnyi kurs* (*Neural networks: a complete course*): transl. from Engl. Moscow: Vil'yams; 2006. 1104 p. (in Russ.).
[Haykin S. *Neural Networks*. Upper Saddle River, NJ: Prentice Hall; 1999. 842 p.]
11. Elgendy M. *Deep Learning for Vision Systems*. Manning Publications Co; 2020. 480 p. ISBN 978-1-6172-9619-2
12. Geron A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc.; 2019. 856 p. ISBN 978-1-4920-3264-9
13. Finlay S. *Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies*. Relativistic; 2017. 150 p.
14. Raschka S., Mirjalili V. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing; 2019. 770 p. ISBN 978-1-7899-5575-0
15. Voronina V.V., Mikheev A.V., Yarushkina N.G., Svyatov K.V. *Teoriya i praktika mashinnogo obucheniya* (*Theory and Practice of Machine Learning*). Ul'yanovsk: UISTU; 2017. 290 p. (in Russ.). ISBN 978-5-9795-1712-4

9. Кендалл М., Стьюарт А. *Теория распределений*: пер. с англ. М.: Наука; 1966. 588 с.
10. Хайкин С. *Нейронные сети: полный курс*: пер. с англ. М.: Вильямс; 2006. 1104 с.
11. Elgendy M. *Deep Learning for Vision Systems*. Manning Publications Co; 2020. 480 p. ISBN 978-1-6172-9619-2
12. Geron A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc.; 2019. 856 p. ISBN 978-1-4920-3264-9
13. Finlay S. *Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies*. Relativistic; 2017. 150 p.
14. Raschka S., Mirjalili V. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing; 2019. 770 p. ISBN 978-1-7899-5575-0
15. Воронина В.В., Михеев А.В., Ярушкина Н.Г., Святов К.В. *Теория и практика машинного обучения*. Ульяновск: УлГТУ; 2017. 290 с. ISBN 978-5-9795-1712-4

About the authors

Aleksei A. Paramonov, Dr. Sci. (Eng.), Professor, Department of Radio Electronic Systems and Complexes, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: paramonov@mirea.ru. Scopus Author ID 57208923552, http://orcid.org/0000-0002-4537-4626

Van Minh Nguyen, Postgraduate Student, Department of Radio Electronic Systems and Complexes, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: nguyenminhhvkq1009@gmail.com. https://orcid.org/0009-0001-5062-7093

Minh Tuong Nguyen, Cand. Sci. (Eng.), Associate Professor, Department of Informatics, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: nguen_m@mirea.ru. https://orcid.org/0009-0002-7267-1121

Об авторах

Парамонов Алексей Анатольевич, д.т.н., профессор, кафедра радиоэлектронных систем и комплексов Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: paramonov@mirea.ru. Scopus Author ID 57208923552, http://orcid.org/0000-0002-4537-4626

Нгуен Ван Минь, аспирант, кафедра радиоэлектронных систем и комплексов Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: nguyenminhhvkq1009@gmail.com. https://orcid.org/0009-0001-5062-7093

Нгуен Минь Тыонг, к.т.н., доцент, кафедра информатики Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: nguen_m@mirea.ru. https://orcid.org/0009-0002-7267-1121

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 519.6
https://doi.org/10.32362/2500-316X-2023-11-4-59-71

RESEARCH ARTICLE

Modeling of spatial spread of COVID-19 pandemic waves in Russia using a kinetic-advection model

Vladimir V. Aristov [1, 2, @], Andrey V. Stroganov [1], Andrey D. Yastrebov [1]

[1] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[2] *Federal Research Center "Computer Science and Control", Russian Academy of Sciences, Moscow, 119333 Russia*
[@] *Corresponding author, e-mail: aristovvl@yandex.ru*

Abstract

Objectives. COVID-19 has a number of specific characteristics that distinguish it from past pandemics. In addition to the high infection rate, the high spread rate is due to the increased mobility of contemporary populations. The aim of the present work is to construct a mathematical model for the spread of the pandemic and identify patterns under the assumption that Moscow comprises the main source of viral infection in Russia. For this purpose, a two-parameter kinetic model describing the spatial spread of the epidemic is developed. The parameters are determined using theoretical constructions alongside statistical vehicle movement and population density data from various countries, additionally taking into account the development of the first wave on the examples of Russia, Italy and Chile with verification of values obtained from subsequent epidemic waves. This paper studies the development of epidemic events in Russia, starting from the third and including the most recent fifth and sixth waves. Our two-parameter model is based on a kinetic equation. The investigated possibility of predicting the spatial spread of the virus according to the time lag of reaching the peak of infections in Russia as a whole as compared to Moscow is connected with geographical features: in Russia, as in some other countries, the main source of infection can be identified. Moscow represents such a source in Russia due to serving as the largest transport hub in the country.

Methods. Mathematical modeling and data analysis methods are used.

Results. A predicted time lag between peaks of daily infections in Russia and Moscow is confirmed. Identified invariant parameters for COVID-19 epidemic waves can be used to predict the spread of the disease. The checks were carried out for the wave sequence for which predictions were made about the development of infection for Russia and when the recession following peak would occur. These forecasts for all waves were confirmed from the third to the last sixth waves to confirm the found pattern, which can be important for predicting future events.

Conclusions. The confirmed forecasts for the timing and rate of the recession can be used to make good predictions about the fifth and sixth waves of infection of the Omicron variant of the COVID-19 virus. Earlier predictions were confirmed by the statistical data.

Keywords: kinetic equation, COVID-19, wave propagation

• **Submitted:** 15.09.2022 • **Revised:** 02.02.2023 • **Accepted:** 15.05.2023

For citation: Aristov V.V., Stroganov A.V., Yastrebov A.D. Modeling of spatial spread of COVID-19 pandemic waves in Russia using a kinetic-advection model. *Russ. Technol. J.* 2023;11(4):59–71. https://doi.org/10.32362/2500-316X-2023-11-4-59-71

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Моделирование пространственного распространения волн пандемии COVID-19 в России на основе кинетико-переносного описания

В.В. Аристов [1, 2, @],
А.В. Строганов [1],
А.Д. Ястребов [1]

[1] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[2] *Федеральный исследовательский центр «Информатика и управление» Российской академии наук, Москва, 119333 Россия*
[@] *Автор для переписки, e-mail: aristovvl@yandex.ru*

Резюме

Цели. Пандемия COVID-19 обладает рядом важных особенностей по сравнению с прошлыми эпидемиями. Помимо высокой степени заражения, она имеет высокую скорость распространения за счет мобильности населения, связанной, в частности, с возросшей скоростью средств передвижения. Целью данной работы является построение математической модели распространения пандемии и выявление закономерностей в предположении, что основным источником вирусной инфекции в России является г. Москва. Для этого строится двухпараметрическая кинетическая модель, описывающая пространственное распространение эпидемии. Параметры находятся с помощью теоретических построений, оценок известных данных о статистике передвижения транспортных средств и плотности населения в различных странах, а также с учетом развития первой волны на примере России, Италии и Чили с проверкой значений для последующих эпидемических волн. Исследуется возможность предсказывать скорость пространственного распространения вируса по временно́му интервалу запаздывания достижения пика заражений в России по сравнению с Москвой. Это связано с географическими особенностями: в России, как и в некоторых других странах, можно выделить основной источник распространения инфекции. Таким источником в России выступает г. Москва – крупнейший в стране транспортный узел. Для реализации цели в настоящей работе изучается развитие эпидемических событий в России, начиная с 3-й, и вплоть до последних 5-й и 6-й волн.

Методы. Использованы методы математического моделирования и методы обработки статистических данных.

Результаты. Подтверждено, что величина запаздывания достижения пика заражений составляет в среднем 2.5 недели. Выявлена сохраняемость параметров для различных волн, поэтому модель обладает предсказательными возможностями. Проверки проводились для последовательности волн, для которых делались соответствующие предсказания о развитии заражения для России в целом и о том, когда произойдет спад. Данные прогнозы подтвердились для всех волн, начиная с 3-й, и вплоть до последней 6-й волны, что подтверждает найденную закономерность, важную для прогнозирования будущих событий.

Выводы. Прогнозы о начале и скорости выздоровления подтвердились, что дало возможность уверенно прогнозировать, в частности, протекание 5-й и 6-й волн пандемии, связанной с новым вирусным штаммом «омикрон». Предсказания, которые делались заранее, были проверены и получили подтверждение.

Ключевые слова: кинетическое уравнение, COVID-19, распространение волн

• **Поступила:** 15.09.2022 • **Доработана:** 02.02.2023 • **Принята к опубликованию:** 15.05.2023

Для цитирования: Аристов В.В., Строганов А.В., Ястребов А.Д. Моделирование пространственного распространения волн пандемии COVID-19 в России на основе кинетико-переносного описания. *Russ. Technol. J.* 2023;11(4):59–71. https://doi.org/10.32362/2500-316X-2023-11-4-59-71

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

As well as representing a social and medical challenge, the complexity of the COVID-19 pandemic phenomenon entailed the development of novel scientific approaches for its study. This led to an intensification of research activity involving numerous laboratories around the world. The contemporary development of information technology allows large volumes of data to be processed quickly. In addition to more complex mathematical models using powerful computer resources, simple models can also play a significant role in describing the spread of the virus. The development of the COVID-19 pandemic differs from previously known epidemics in a number of ways, some of which can be attributed to the increased rapidity of modern transportation patterns. The aim of this work is to simulate the processes of spatial spread of the pandemic according to a study of rapid virus transport scenarios rather than the diffusion-type equations used in traditional methods on the basis of slow contact processes (see [1–3]).

There are a number of works in which statistical methods are among those used to study the development of epidemics [4–14]. For the most part, susceptible–infected–recovered (SIR) and susceptible–exposed–infected–recovered (SEIR) models are used for studying spatial–local processes occurring in time. However, the conjunction with spatial development in such works is infrequent. Therefore, in the present work, we investigate only the spatial propagation of the epidemic.

For this purpose, a kinetic model is constructed based on model equations similar to those used to study various physical processes [15, 16].

The present work comprises a continuation, development, and generalization of our earlier research [17], in which common features of the method were determined by considering and studying the first and second pandemic waves. The developed model is applied to study all subsequent waves of the COVID-19 pandemic in Russia. Using the proposed kinetic approach, the nature of the contemporary pandemic spread it is investigated for the countries to which the one-dimensional model is applicable. For the first wave of the pandemic, processes were studied in Italy, Chile and Russia, while the study of the second wave was mainly limited to Russia. These countries are assumed to comprise major centers from which the spread of infection originated, which in turn determines the delay in disease development in individual regions and across the entire country. This assumption is used to make predictions for the unfolding of subsequent pandemic waves, as demonstrated by the spread of such waves in Russia.

It is possible to distinguish between two mechanisms of infection: transportable (studied in the present work) and contact, realized in the regions. The superposition of these two factors gives the sum of the number of infections. The first phase, corresponding to transmissible infection, lays the foundations for the next phase, consisting in the development of contact-related infection. We study the spread of carriers of the virus, which is associated with the use of various types of vehicles for transportation. Average travel speeds take into account differences between airplane, train, bus, and car transportation modes. Another model parameter consists in resistance to the advancement of virus carriers, which has the dimension of the frequency of disembarkation of passengers from vehicles during movement to their places of residence. This parameter mainly depends on population density. The same character of the course of the disease is subsequently assumed in different places. The third phase is associated with the spread of recovery, which is determined by reaching a maximum of new infections per day for each region.

The geographical features of Russia allow a simple one-dimensional model to be applied as for Italy and Chile (the pandemic spread was considered for these countries for the first wave). At the same time, taking into account the nature of infection for Russia, the main source in all waves is identified with Moscow (the regions to the east of the capital were studied). This forms a basis for judgements about the shift in the time of the beginning of recovery.

The third wave was associated with the emergence of the delta virus strain, which mostly entered Russian from India through Moscow airports. While the fourth wave was similar to the second in being mainly caused by people returning from summer vacations, in this case Moscow was also the main source of the new wave development. However, since other cities also contributed, the development of the processes was somewhat blurred in comparison with the development of the processes in the first and third waves. The fifth wave has its own characteristics, since the explosive nature of the infection with the Omicron strain can lead to some corrections in the spatial spread, accelerating this process to some extent.

The general patterns determined in the first wave and confirmed in the second wave remain unchanged due to the accepted accounting of infection associated with the movement of transport of infected passengers leading to a repetition of the pandemic wave spread pattern. Here, the average rate of spread depends not on the intensity of infection by new viruses, but the average speed of vehicles. Therefore, the model is also applied to describe subsequent pandemic waves. This also applies to the most recent sixth wave.

This study uses a one-dimensional kinetic model to predict and then verify the propagation patterns of the third, fourth, fifth, and sixth waves of the pandemic. A lag of the maximum infection in Russia as a whole compared to the maximum infection in Moscow by about 2.5 weeks was detected (taking into account the inevitable statistical error, the maximum is within two to three weeks). Thus, the timing of the pandemic wave in Russia can be judged from the development of the disease in Moscow. While infection values per day vary from wave to wave, some invariants of the waveform can be identified. Therefore, an attempt is made to find functions describing recurring forms of infection waves, which take into account different infection intensities for different strains—and, accordingly, different amplitude of fluctuations in the number of infected per day.

We considered whether the duration of the incubation period of infection affects the presented results. For the first wave, the period was seven to ten days, while for subsequent waves it was the same or less. However, this value is the same for Moscow and Russian regions, so there is practically no effect of this parameter on the lag time of infection.

DEVELOPED TRANSFER-KINETIC MODEL AND ITS ADJUSTMENT

Here is a brief description of the one-dimensional two-parameter advection model, taking into account the kinetic term:

$$\frac{\partial n}{\partial t} + U\frac{\partial n}{\partial x} = -\sigma n(t,x), \quad (1)$$

where t is time; x is distance; $n(t, x)$ is the density of moving virus carriers in the vehicle; U is the average speed of the vehicle; σ is the resistance factor (having the dimension of frequency) to the movement of infected elements mainly due to dropping off passengers from vehicles at places of residence.

The initial condition for the Cauchy problem is taken as:

$$n_0(x) = H(-x),$$

where $H(x)$ is the Heaviside function. This formulation of the problem means that in fact the virus carriers enter the studied area $x > 0$ through the boundary $x = 0$. The linear equation (1) is solved by standard methods; the analytical solution has the form:

$$n(t,x) = n_0(x - Ut)\mathrm{e}^{-\sigma\frac{x}{U}}. \quad (2)$$

We denote the density of virus carriers disembarking from the vehicle at a given point as $n_M(t, x)$. This density grows in the same way $n(t, x)$ decreases, hence, we can write as follows:

$$\frac{dn}{dt} + \frac{dn_M}{dt} = 0,$$
$$\frac{dn}{dt} = \frac{\partial n}{\partial t} + U\frac{\partial n}{\partial x},$$
$$\frac{dn_M}{dt} = \frac{\partial n_M}{\partial t}.$$

Note that the total and partial derivatives of n differ from each other due to the progression of elements (virus carriers) with the speed U. For n_M, these derivatives are coincident, since the elements with this density are stationary.

Given (1), the equation for $n_M(t, x)$ takes the following form:

$$\frac{\partial n_M}{\partial t} = \sigma n. \quad (3)$$

The solution to this equation is written as follows

$$n_M(t,x) = n_{M_0}(x) + \int_0^t \sigma n(\tau,x)d\tau.$$

By substituting expression (2) and the initial conditions and $n_{M_0}(x) = 0$ and $n_0(x) = H(-x)$, we obtain an expression for $n_M(t, x)$

$$n_M(t,x) = \frac{\sigma}{U}(Ut - x) \cdot H(Ut - x)e^{-\sigma\frac{x}{U}}. \qquad (4)$$

In a large series of calculations, it was sometimes more convenient to use a numerical approach in place of an analytical expression; here, a simple Courant–Isaacson–Rees scheme is applied.

The question of specifying the parameters U and σ is important. The estimation of the first parameter takes the value of the second to be equal to that obtained in our previous works. Using the values of the first of these parameters obtained in the study of the first and second waves, predictions are made about the speeds of the pandemic spread and convalescence waves, respectively. The theoretically estimated average generalized speed of movement vehicle is compared it with the real data on the spread of the disease; here, it is deemed acceptable to make some corrections to the resulting speed parameter for subsequent waves. We consider four main modes of transportation: car, bus, train, and airplane (while bicycles, motorcycles, scooters, and walking also contribute, for moving a large numbers of passengers over appreciable distances, these comprise the main modes of transport). It is necessary to introduce some weight averaging, taking into account the relative proportions of passengers using one or another vehicle type.

More accurate quantitative estimates were also obtained for the average weighted speed of trains. According to the Russian Railways press release[1], $9.614 \cdot 10^8$ passengers were transported by short-distance trains in Russia in 2021, who travelled a total of $2.9 \cdot 10^{10}$ km. Long-distance trains carried $9.2 \cdot 10^7$ passengers, who traveled a total of $7.44 \cdot 10^{10}$ km. Thus, on average, each passenger travelled 30.16 km on short distance trains and 808.7 km on long distance trains. According to the Ministry of Transport of the Russian Federation[2] the average speed of trains in Russia in 2021 is in the range from 57 to 65 km/h, i.e., passengers of both short-distance and long-distance trains on average reached their destination in less than a day. Since the average distances obtained earlier can be assumed to be covered in a day, the average speed of short-distance trains is taken to be 30.16 km/day, and long-distance trains—808.7 km/day. As a result, the average speed of trains in Russia in 2021 is

$$U = \frac{9.614}{9.614 + 0.92} \cdot 30.16 + \frac{0.92}{9.614 + 0.92} \cdot 808.7 = 98.15 \text{ km/day}.$$

One would expect the real data to be close to this value, since this was the case for all of the studied pandemic waves. Thus, the approximation of the weighted average speed of the vehicle and the speed of the epidemic spread in the first few days is true for all waves. Differences of 10–15% lie within statistical error, confirming the assumed possibility of using the U parameter in the model.

Note that some semi-empirical calculations based on an analysis of annual passenger turnover values for other vehicles confirm the above estimates, but we do not cite them because they require more careful consideration, which is not the purpose of this article.

SPREAD OF THE THIRD WAVE OF THE COVID-19 PANDEMIC IN RUSSIA

The results obtained in [17] are used to analyze the development of subsequent waves. First of all, let us consider the character of the third-wave propagation. Firstly, in [17] the data of pandemic development for Moscow are given. Once the maximum of infections for a day is reached, the beginning of recovery is judged by the calendar number corresponding to the maximum. A prediction is made as to the possible day of reaching such a maximum and the beginning of recovery for Russia. This hypothesis is then tested.

According to the data processing in [17], the value of the parameter of the average speed of vehicles was obtained. This speed was estimated as U = 75–90 km/day, which is close to the above estimates of average speed with an accuracy of 10% for the value of 90–100 km/day.

The time of movement of the wave from Moscow to the center of masses of the Russian population can be calculated. According to the density distribution graph given in [11], this point approximately corresponds to a distance of 1000–1200 km from Moscow; therefore, a significant outbreak in Moscow can be expected to affect the whole of Russia in about two weeks. However, since this is a relatively crude estimate, such a lag can in practice be expected to extend for up to three weeks. Then, for example, the maximum infection rate per day in Moscow appears about two weeks earlier than the infection peak in Russia as a whole. Since this characteristic point is interpreted as the beginning of recovery, the recovery wave also has a corresponding lag.

[1] More than 1 bn passengers carried on the Russian Railways network in 2021 | Press releases | Company (rzd.ru). https://company.rzd.ru/ru/9397/page/104069?id=269758. Accessed December 21, 2022 (in Russ.).

[2] Average speed of passenger trains in Russia may grow to 65 km/h by 2031. Ministry of Transport of the Russian Federation. https://mintrans.gov.ru/press-center/branch-news/595. Accessed December 21, 2022 (in Russ.).

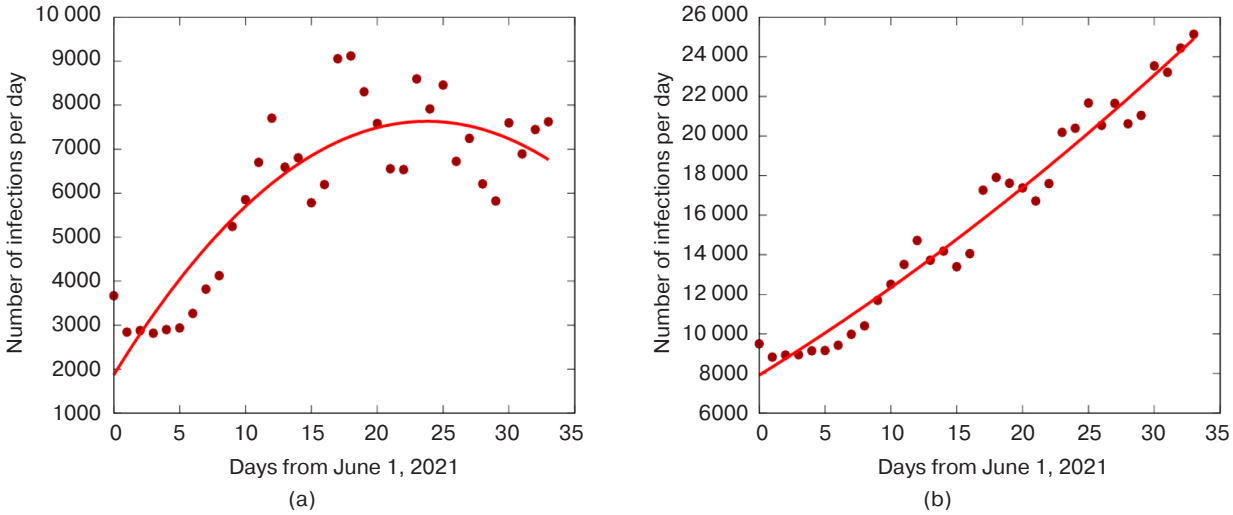


Fig. 1. Number of infections per day for Moscow (a) and Russia (b) in June 2021

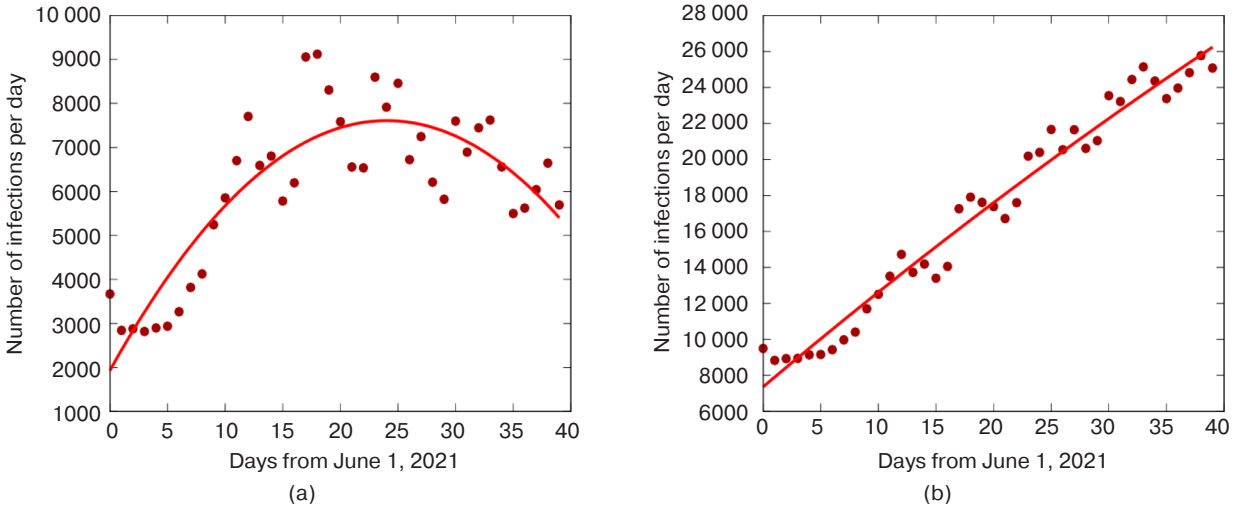


Fig. 2. Number of infections per day for Moscow (a) and Russia (b) in June–July 2021

The diagrams (Fig. 1) concerning the month of June 2021 (the data until June 29th) are constructed on the basis of the data on the number of infections in Moscow and Russia as a whole. It can be seen that the maximum of infection for Moscow passed at the end of the month (according to the applied mean square approximation), while for Russia there was only an increase in the number of infections per day. Therefore, we could expect that the maximum infection in Russia would be reached by the middle of July.

Figure 2 shows infection data for Moscow and Russia as of July 11, which is 12 days later than the previous graphs in Fig. 1. We can see a clearly formed "hump" of infections in Moscow, while in Russia as a whole, the maximum is only beginning to emerge.

Figure 3 shows the expected development parabola of events (number of infections in Russia) based on the received data up to July 15, 2021. So, it seems that the parabola line would have to be clarified, since only an extrapolation is given.

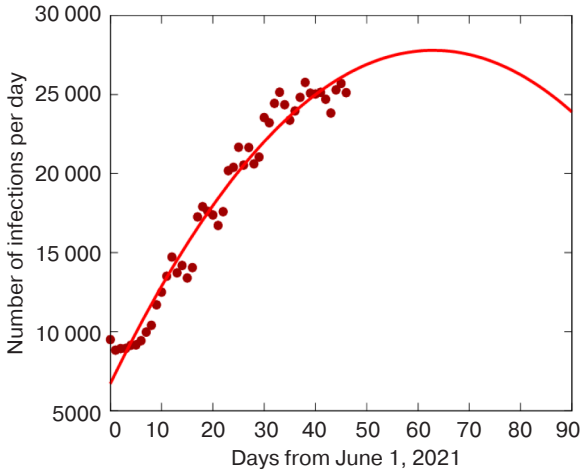


Fig. 3. Expected number of infections for the third wave for Russia based on the received data until mid-July 2021

We can conclude that the predictions turned out to be justified. The maximum for Moscow was reached by June 25. Therefore, we assumed that the maximum

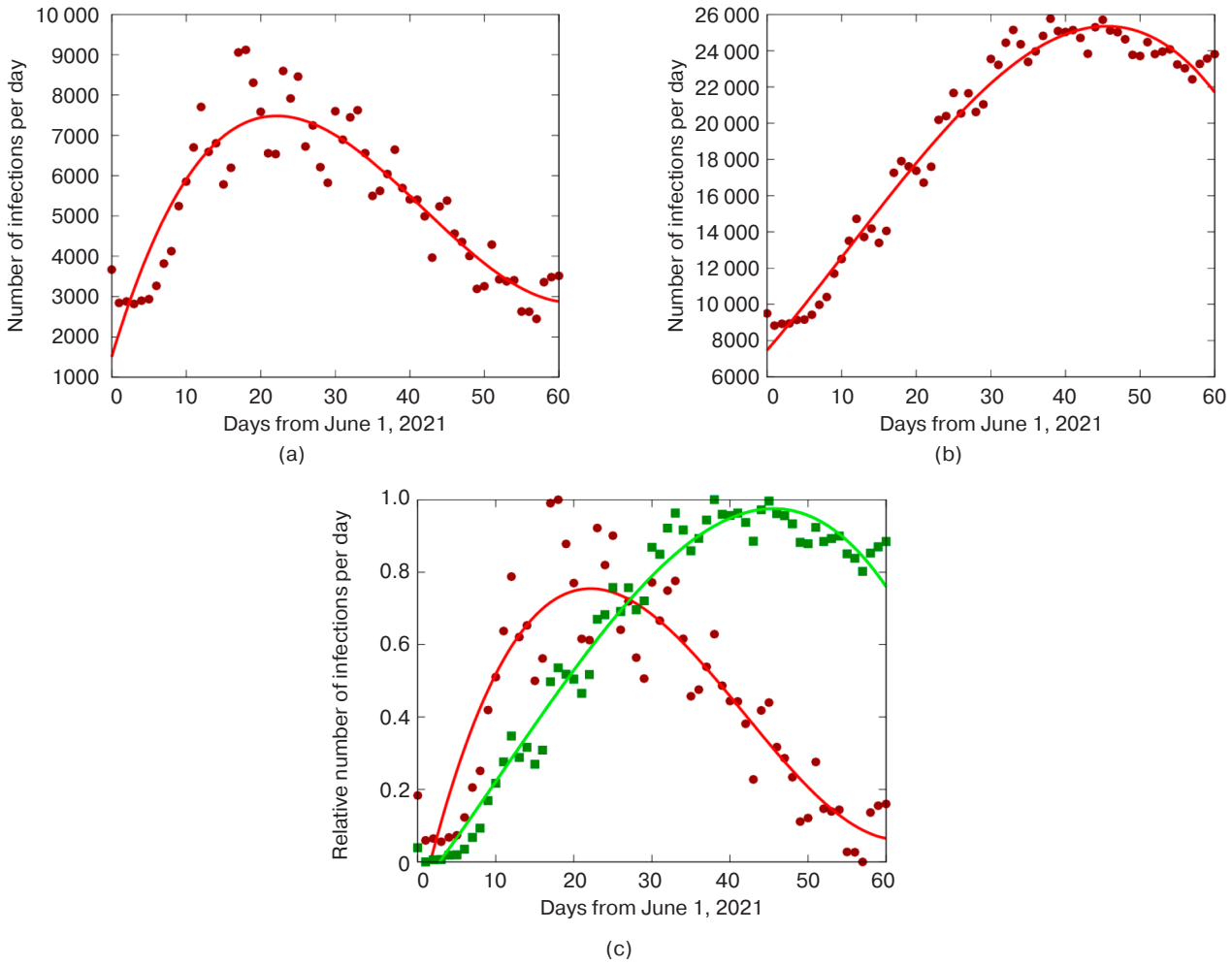


Fig. 4. Real data and averaging lines by LSM: total number of infected for Moscow (a) and for Russia as a whole (b); relative number of infections for Moscow and for Russia as a whole (c)

for Russia should be reached (taking into account the parameters obtained from the study of previous pandemic waves) in about two to three weeks, i.e., by mid-July. The curves constructed using the least squares method (LSM) corresponded to the forecast. But it is interesting that "peaks," i.e., absolute maximums of infections in Moscow and Russia are displaced by this very value: they occur around June 25 and July 15 respectively (generally speaking, such values are not sufficiently representative due to not very representative statistics and random outliers). As time passes and new data become available each day, the LSM parabola shifts somewhat to the right, a trend that has been noted before. But the magnitude of the lag remains the same: the maxima are now July 5 and July 25, respectively.

For the third wave, the shift in the time of onset of recovery-maximum infection for Moscow and Russia as a whole is three weeks. The maximum for Moscow in Fig. 4 corresponds approximately to the beginning to the middle of the third decade of June, while the maximum for Russia as a whole corresponds approximately to the middle of July. The graphs in Fig. 4c present relative values of infections, which were calculated according to the formula $A_{rel} = (A - A_{min})/(A_{max} - A_{min})$, where A_{max} and A_{min} mean the maximum and minimum of this value, respectively (Figs. 4a and 4b).

DEVELOPMENT OF THE FOURTH WAVE OF THE PANDEMIC

The regularities identified for the previous waves were used to predict the behavior of the fourth wave. In general, the predictions were confirmed as applying both to the nature of the infection curves and to the shifts of Russia relative to Moscow.

For the fourth wave, the dates of maximum infections for Moscow and Russia as a whole were late October and mid-November, separated by two to three weeks (Fig. 5). Moreover, this shift for the fourth wave is somewhat less than for the third wave, which can be explained by the return of patients from vacations not only through Moscow, but also through other cities.

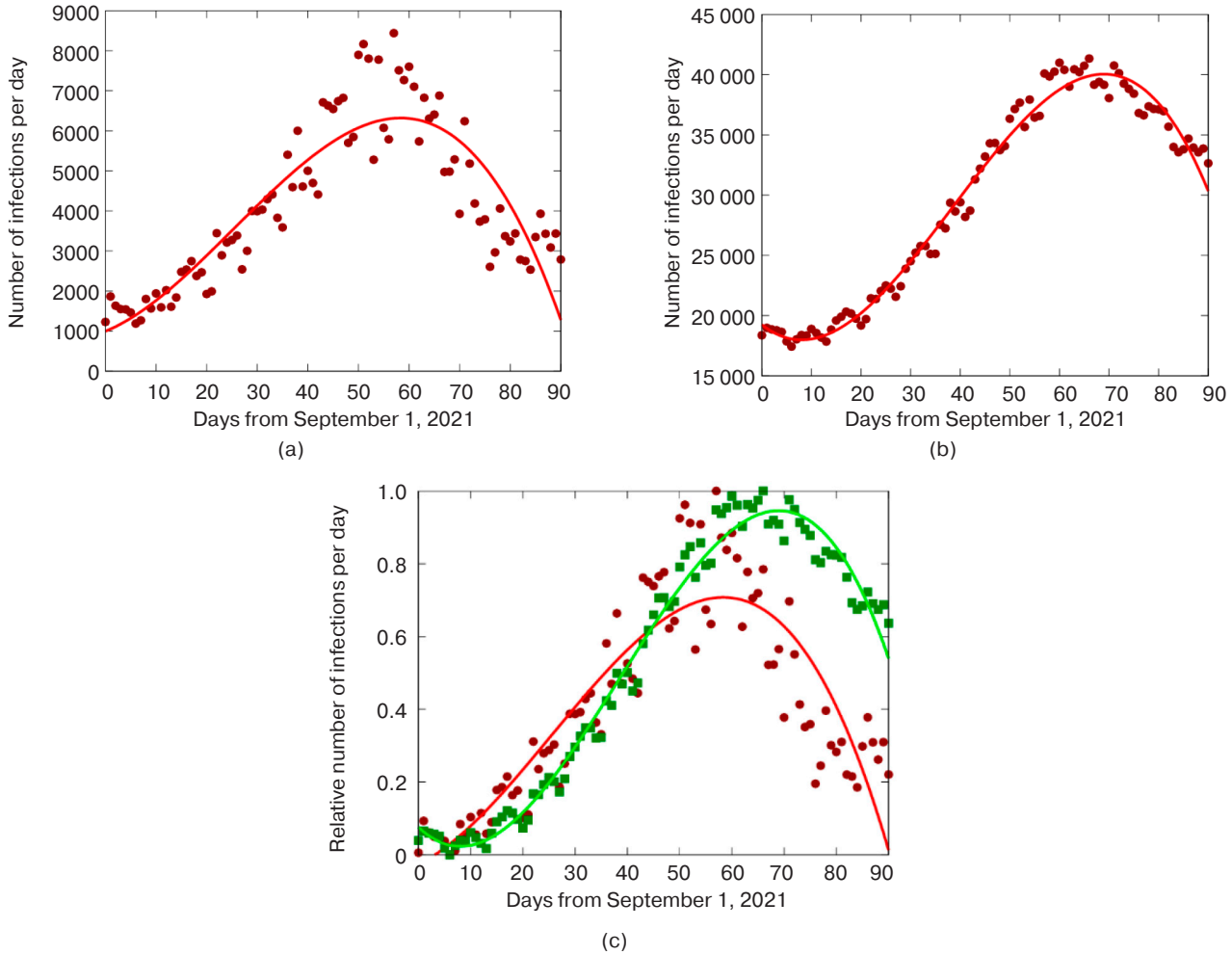


Fig. 5. Real data and averaging lines by LSM, total number of infected for Moscow (a) and for Russia (b), relative number of infections for Moscow and for Russia (c)

MODEL VALIDATION FOR THE FIFTH WAVE OF THE PANDEMIC

Similar processes for the fifth wave are considered. Note that on February 10, 2022, the method of counting sick people changed, which increased the readings, so a correction factor was added to the calculations. Due to the new Omicron strain, the infection was more intense, but the spatial spread was not expected to be much affected. It was assumed that since the maximum for Moscow corresponded to approximately February 1, the maximum for Russia would be around February 14.

The development of this wave is characterized by its own peculiarities. Here possible sharp outbreaks of local infection, particularly in Moscow, correspond to the greater infectability of this strain. Thus, Fig. 6, which shows the number of infections in the first three weeks of the new wave, we can see that the number of cases per day sharply increased starting from January 18. Thus, the expected manifestation of the disease for Russia as a whole involves a two- to three-week lag, which was indeed from around February 3.

The graph in Fig. 6 gives the number of cases per day for Russia without taking into account the Moscow figures.

Based on the results of January, the predicted decline in infection (the beginning of recovery) for Russia occurred by mid-February (Fig. 7). However, the question as to whether the speed of spread of the spatial wave is affected by the nature of the virus remains open. Figure 7 shows graphs of absolute increments of infections per day, as well as more indicative relative values of infections.

The build-up of relative values can help to identify universal properties of the model.

Figure 8 shows the results of the number of infections per day based on actual data as of mid-February. As expected, the maximum of infections across Russia as a whole is formed by mid-February. We note a certain outlier in the data associated with the official recalculation of statistics. The forecast of a shift of 12–14 days was confirmed.

The distributions for Moscow and Russia at the end of February were also built-up (Fig. 9). The data indicate

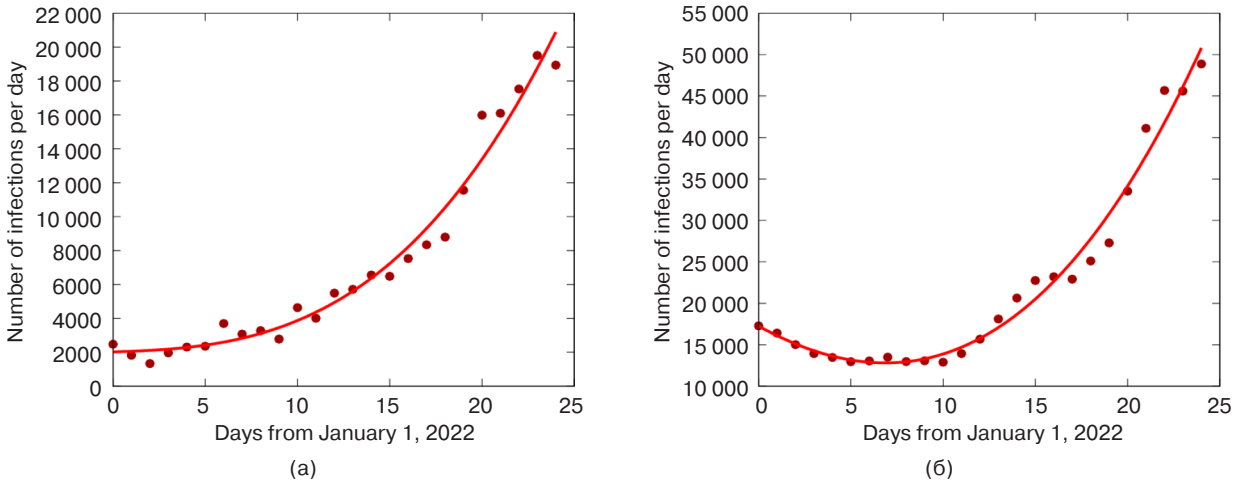


Fig. 6. Number of infections (absolute value) per day in Moscow (a) and in Russia (b)

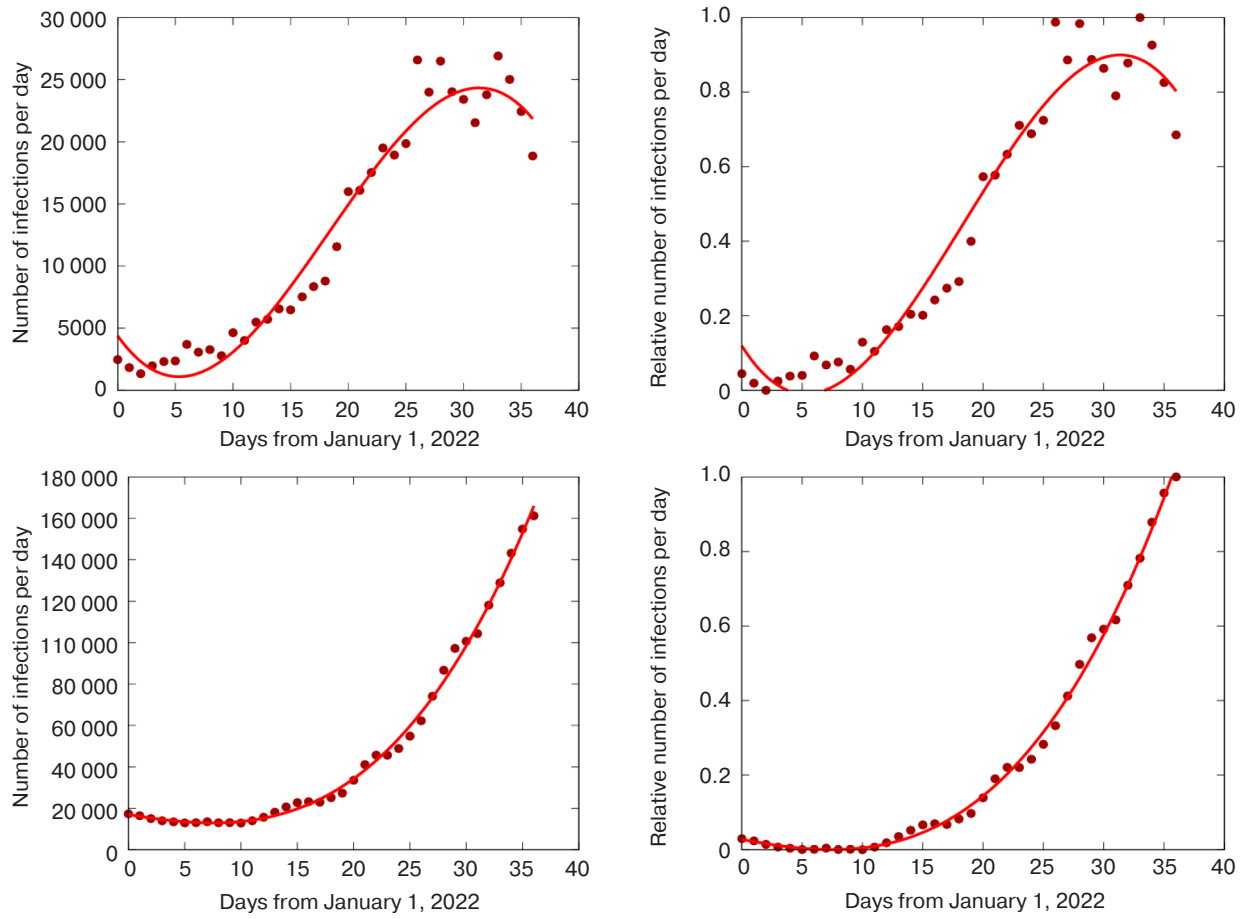


Fig. 7. Number of infections per day from January 1 to February 7, 2022, upper figures refer to Moscow, lower figures refer to Russia excluding Moscow figures

that the curves of the fifth wave have already been formed. The lines resemble those presented in Figs. 7 and 8.

Thus, the predictions for the fifth wave made in late January are confirmed by real data.

It is useful to compare the curves for the relative values in the fourth and fifth pandemic waves (Fig. 10). Let us note some differences in the character of the lines, which is also related to the properties of the LSM used in the averaging. For the fourth wave, the smooth graph for Russia has a noticeably higher maximum than the graph for Moscow, which is associated with a relatively slower decline in the number of infections in Russia compared

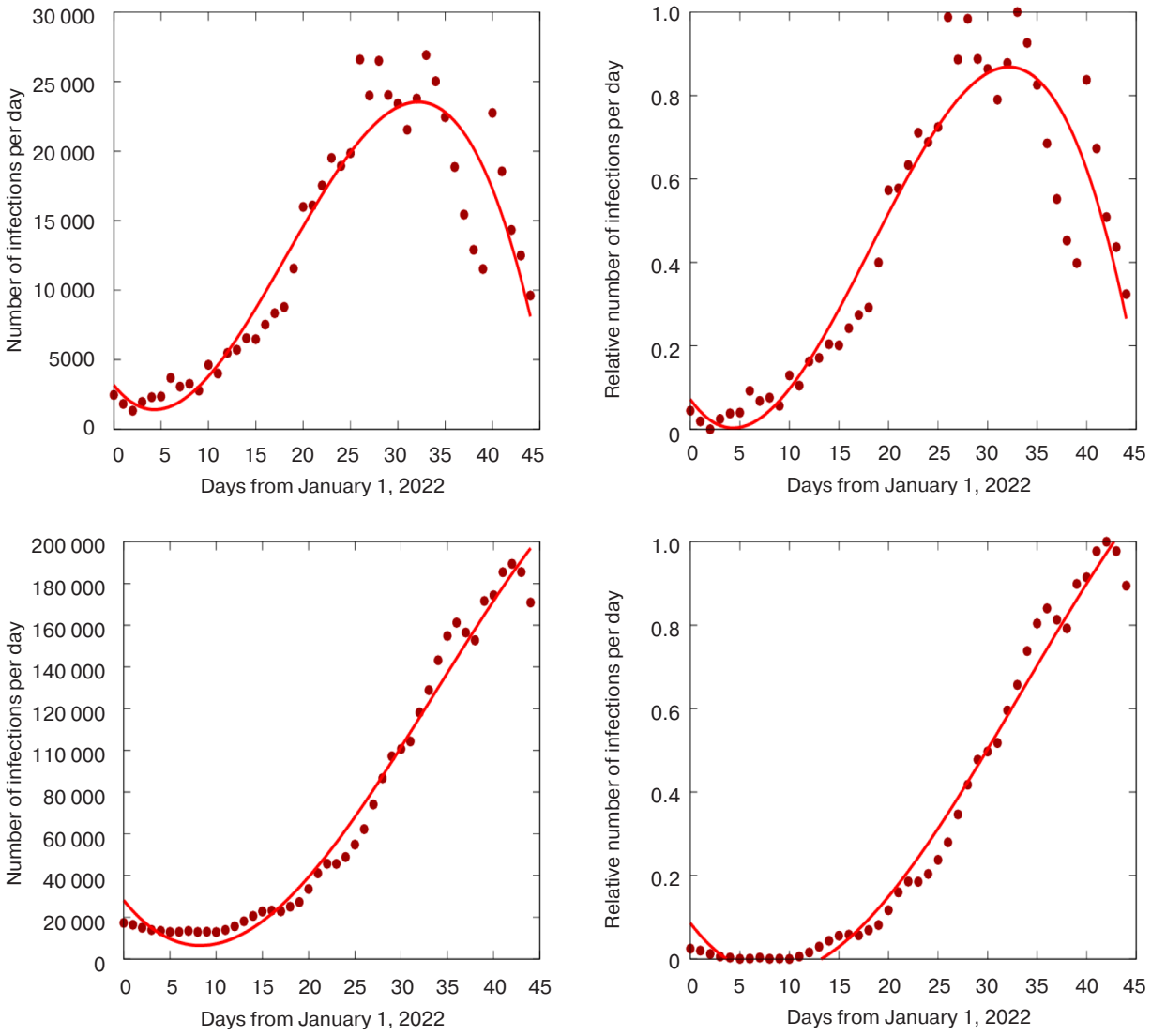


Fig. 8. Pandemic development in Moscow (top) and Russia without Moscow indicators (bottom) from January 1 to February 15, 2022

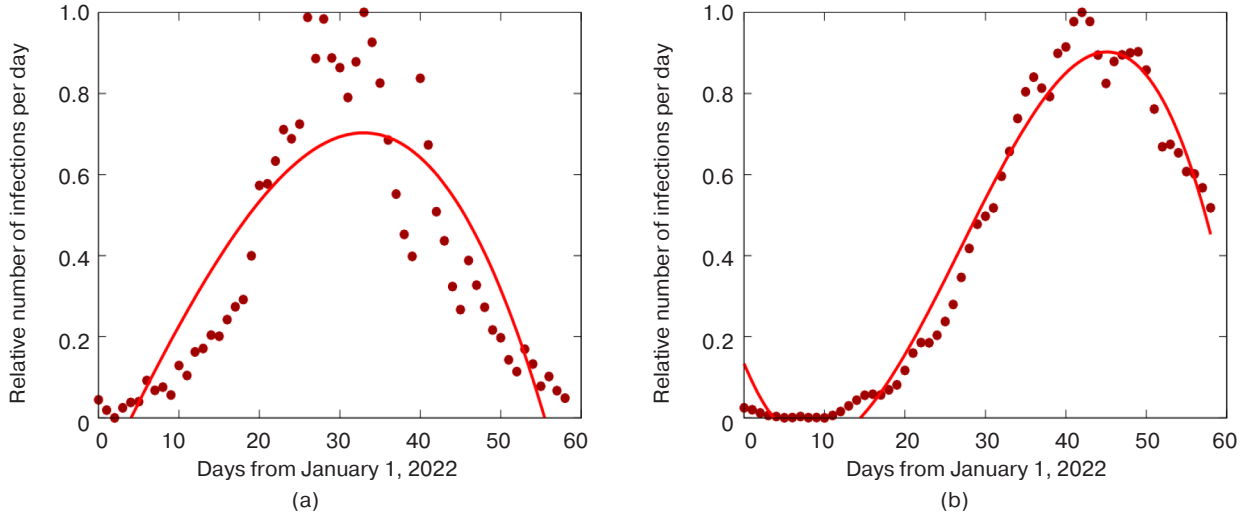


Fig. 9. Development of the pandemic in Moscow (a) and Russia without Moscow indicators (b) up to early March

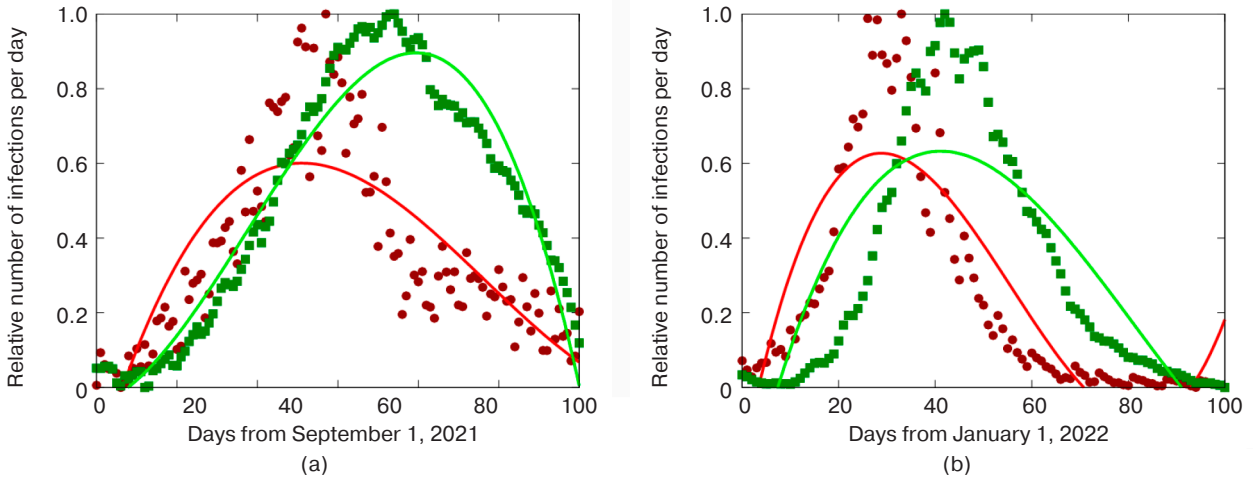


Fig. 10. Distribution of the number of infections for the fourth (a) and fifth (b) pandemic waves in Moscow (red lines) and Russia without regard to Moscow indicators (green lines)

to Moscow after the peak of infections of the fourth wave. For the fifth wave, the plots constructed with the help of LSM have a similar amplitude. However, the time shifts of the local maxima for these different waves are approximately the same.

SEARCH FOR COMMONALITIES IN THE FORMS OF PANDEMIC WAVES AND THEIR USE

Along with the values determining the shift of maximum waves in different spatial points of the country, we can also try to identify patterns in the form of different pandemic waves in individual points—e.g., in the city of Moscow—taking into account changes in the nature of infection in new epidemic waves.

The construction of some forms realized in successive pandemic waves is an important task, which partly overlaps with the topic of works [1–3].

To do this, we use the patterns obtained earlier. Figure 11 shows the expected infection curves in the third wave, taking into account the data on the second and the first waves. We can judge from them how accurate the forecast will be. A curve based on the method of least squares was plotted according to the real data for the second pandemic wave. This line is then moved to the right side of the figure and superimposed on the beginning of the infection points in the third wave. This sets the prediction of the maximum infection that was expected by about July 15, 2021. Here we also plotted the LSM curve using the new data obtained for the third wave of the pandemic. According to this graph, the expected maximum should also be reached around July 15. Figure 11 shows the corresponding lines; here however, the points are plotted up to July 15. We can conclude that these obtained curves approximately correspond to the forecasts. Although the shape of the curves is different, they have certain features in common.

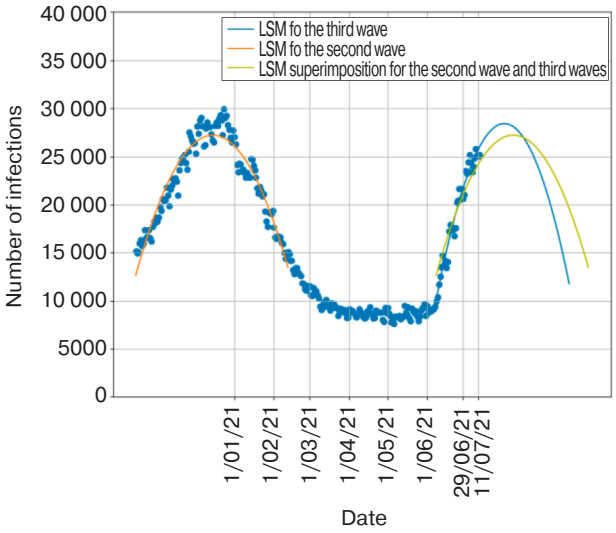


Fig. 11. Superimposition of the curve obtained for the second wave on the expected curve of the third wave up to June 15

Figure 12 shows the curves for the relative number of infections as a function of time. Real data with characteristic statistical outliers (a) and lines averaged by the LSM method (b) are shown. These are shifted so that the maxima correspond to zero in time. For each wave, the day with the maximum number of infections and 60 days before and after (except the fifth wave) are taken. In the scaling for each time interval, the minimum turned out to be 0, while the maximum was 1. Here, while the wave profiles correspond fairly closely, we note the exception of the fifth wave, which may be attributed to the strong intensity of infection spread.

The identified similarity of wave forms indicates a certain universalism in the development of this infection in Russia, reflecting the situation in Moscow, allowing the cautious prediction of subsequent waves, taking into account the varying intensity of strains and vaccination.

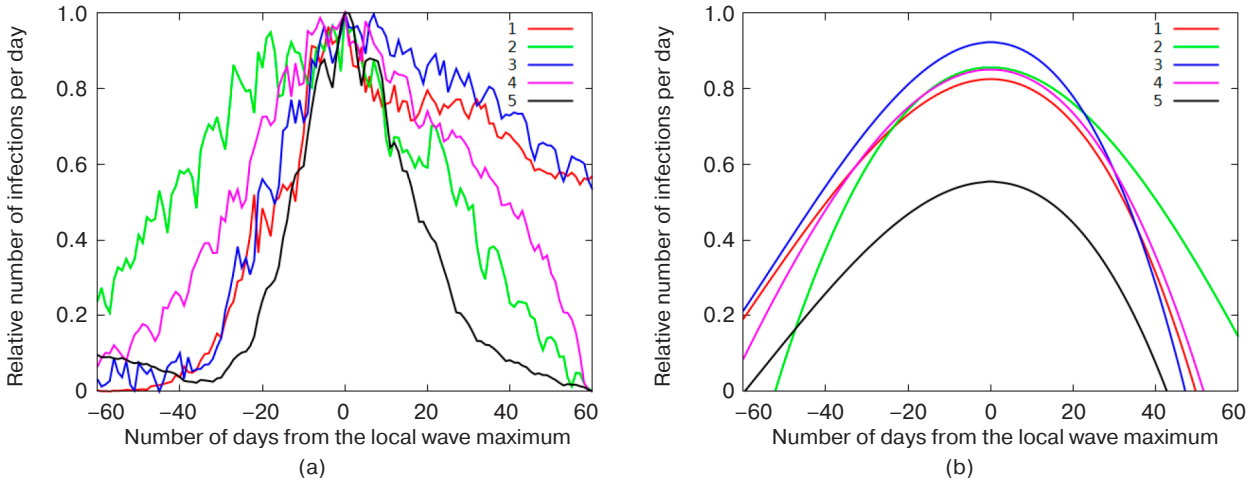


Fig. 12. Real data for all five waves (a) and smoothing curves for them of the third degree of LSM (b)

CONCLUSIONS

The present work has demonstrated the suitability of the previously proposed model to describe the propagation of modern pandemic waves over the territory of Russia. Despite differences in the features of the succeeding pandemic waves, the basic regularities apply. Thus, the model can be used to predict the parameters of future possible waves. Assumptions made in August 2022 for the new sixth wave were fully confirmed: the sixth wave developed similarly to the previous ones, including the lag of about 2.5 weeks between Moscow and Russia as a whole. The maximum infection rate per day was reached by the middle of the third decade of August, while that for Russia as a whole was clearly recorded by the middle of September.

At present, the authors are studying a two-dimensional problem in terms of the corresponding numerical scheme, which will allow other main centers of infection to be taken into account, primarily Saint Petersburg.

Authors' contribution. All authors equally contributed to the present work.

REFERENCES

1. Acioli P.H. Diffusion as a first model of spread of viral infection. *Am. J. Phys*. 2020;88(8):600–604. https://doi.org/10.1119/10.0001464
2. Bärwolff G. A Local and Time Resolution of the COVID-19 Propagation—A Two-Dimensional Approach for Germany Including Diffusion Phenomena to Describe the Spatial Spread of the COVID-19 Pandemic. *Physics.* 2021;3(3):536–548. https://doi.org/10.3390/physics3030033
3. Duffey R.B. Infection Waves in Pandemics and Risk Prediction: Physical Diffusion Theory and Data Comparisons. *Journal of Risk Analysis and Crisis Response.* 2021;11(2):67–74. https://doi.org/10.2991/jracr.k.210609.001
4. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A.* 1927;115(772):700–721. https://doi.org/10.1098/rspa.1927.0118
5. Anderson R.M., May R.M. *Infectious Diseases of Humans*. Oxford, UK: Oxford University Press; 1992. 768 p.
6. Van Kampen N.G. *Stochastic Processes in Physics and Chemistry*. Amsterdam, The Netherlands: Elsevier North-Holland; 1981. 478 p.
7. Ivorra B., Ferrández M.R., Vela-Pérez M., Ramos A.M. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Commun. Nonlinear Sci. Numer. Simul.* 2020;88:105303. https://doi.org/10.1016/j.cnsns.2020.105303
8. Gross B., Zheng Z., Liu S., Chen X., Sela A., Li J., Li D., Havlin S. Spatio-temporal propagation of COVID-19 pandemics. *EPL* (*Europhysics Letters*). 2020;131(5):58003. https://doi.org/10.1209/0295-5075/131/58003
9. Ramaswamy H., Oberai A.A., Yortsos Y.C. A comprehensive spatial-temporal infection model. *Chem. Eng. Sci.* 2021;233(100):116347. https://doi.org/10.1016/j.ces.2020.116347
10. Maier B., Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science.* 2020;368(6492):742–746. https://doi.org/10.1126/science.abb4557
11. Bontempi E., Vergalli S., Squazzoni F. Understanding COVID-19 diffusion requires an interdisciplinary, multi-dimensional approach. *Environ. Res*. 2020;188:109814. https://doi.org/10.1016/j.envres.2020.109814

12. Neipel J., Bauermann J., Bo S., Harmon T., Jülicher F. Power-law population heterogeneity governs epidemic waves. *PLoS ONE* 2020;15(10):e0239678. https://doi.org/10.1371/journal.pone.0239678
13. Saffary T., Adegboye O.A., Gayawan E., Elfaki F., Kuddus M.A., Saffary R. Analysis of COVID-19 Cases' Spatial Dependence in US Counties Reveals Health Inequalities. *Front Public Health*. 2020;12(8):579190. https://doi.org/10.3389/fpubh.2020.579190
14. Ascani A., Faggian A., Montresor S. The geography of COVID-19 and the structure of local economies: The case of Italy. *J. Regional Sci.* 2021;61(2):407–441. https://doi.org/10.1111/jors.12510
15. Prigogine I., Herman R. *Kinetic Theory of Vehicular Traffic*. New York, USA: Elsevier; 1971. 100 p.
16. Aristov V.V., Ilyin O.V. Kinetic Models for Historical Processes of Fast Invasion and Aggression. *Phys. Rev. E.* 2015;91(4):04286. https://doi.org/10.1103/PhysRevE.91.042806
17. Aristov V.V., Stroganov A.V., Yastrebov A.D. Application of a Kinetic Model for Studying the Spatial Spread of COVID-19. *Dokl. Phys.* 2021;66(5):129–133. https://doi.org/10.1134/S1028335821050013
[Original Russian Text: Aristov V.V., Stroganov A.V., Yastrebov A. D. Application of a Kinetic Model for Studying the Spatial Spread of COVID-19. *Doklady Rossiiskoi Akademii Nauk. Fizika, Tekhnicheskie nauki*. 2021;498(1):27–32 (in Russ.). https://doi.org/10.31857/S2686740021030020]

About the authors

Vladimir V. Aristov, Dr. Sci. (Phys.-Math.), Professor, Department of Higher Mathematics, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia); Chief Researcher, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences (44/2, Vavilova ul., Moscow, 119333 Russia). E-mail: aristovvl@yandex.ru. Scopus Author ID 35517535600, RSCI SPIN-code 1481-8088, https://orcid.org/0000-0003-2568-3453

Andrey V. Stroganov, Cand. Sci. (Phys.-Math.), Assistant Professor, Department of Higher Mathematics, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: savthe@gmail.com. Scopus Author ID 36667697700. https://orcid.org/0000-0001-6487-6093

Andrey D. Yastrebov, Postgraduate Student, Department of Higher Mathematics, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: andr.yast711@gmail.com. Scopus Author ID 57314418000, https://orcid.org/0000-0003-0072-6226

Об авторах

Аристов Владимир Владимирович, д.ф.-м.н., профессор, кафедра высшей математики Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78); главный научный сотрудник, ФГУ Федеральный исследовательский центр «Информатика и управление» Российской академии наук (119333, Россия, Москва, ул. Вавилова, д. 44/2). E-mail: aristovvl@yandex.ru. Scopus Author ID 35517535600, SPIN-код РИНЦ 1481-8088, https://orcid.org/0000-0003-2568-3453

Строганов Андрей Валентинович, к.ф.-м.н., доцент, кафедра высшей математики Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78), E-mail: savthe@gmail.com. Scopus Author ID 36667697700, https://orcid.org/0000-0001-6487-6093

Ястребов Андрей Дмитриевич, аспирант, кафедра высшей математики Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78), E-mail: andr.yast711@gmail.com. Scopus Author ID 57314418000, https://orcid.org/0000-0003-0072-6226

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 519.857
https://doi.org/10.32362/2500-316X-2023-11-4-72-83

RESEARCH ARTICLE

Combined approximation algorithms for interactive design of road routes in CAD

Dmitry A. Karpov,
Valery I. Struchenkov @

MIREA – Russian Technological University, Moscow, 119454 Russia
@ *Corresponding author, e-mail: str1942@mail.ru*

Abstract

Objectives. The aim of the work is to create algorithms for approximating a sequence of points on a plane by arcs of clothoids and circles. Such a problem typically arises in the design of railroad and highway routes. The plan (projection onto a horizontal plane) of the road route is a curve (spline) consisting of a repeating bundle of elements "straight line + clothoid arc + circle arc + clothoid arc + ...". Such a combination of elements provides continuity not only for the curve and its tangent, but also for the curvature. Since the number of spline elements is not known in advance, and their parameters are subject to restrictions, there is no mathematically consistent algorithm for this problem. The two-stage scheme for solving the problem is developed at RTU MIREA only for a spline with lines and circles (i.e., without clothoid elements). At the first stage, the scheme uses dynamic programming to determine the number of spline elements. At the second stage, the scheme optimizes parameters of the spline using nonlinear programming. This scheme has yet to be implemented for a spline with clothoids due to a significantly more complicated nature of this problem. Therefore, the design of route plans in existing computer aided design (CAD) systems is carried out in interactive mode using iterative selection of elements. In this regard, it makes sense to develop mathematically consistent algorithms for element-by-element approximation.

Methods. The problem of element-by-element approximation by a circle and a clothoid is formalized as a low-dimensional non-linear programming problem. The objective function is the sum of squared deviations from the original points. Since a clothoid can only be represented in Cartesian coordinates by power series, there are difficulties in calculating the derivatives of the objective function with respect to the desired parameters of the spline elements. The proposed mathematically consistent algorithm for calculating these derivatives is based on the integral representation of the Cartesian coordinates of the points of the clothoid as functions of its length.

Results. A mathematical model and algorithms have been developed for approximating a sequence of points on a plane by clothoids and circles using the method of nonlinear programming. A second-order algorithm is implemented with the calculation and inversion of the matrix of second derivatives (Hesse matrix).

Conclusions. For approximation by circles and clothoids using nonlinear programming, it is not necessary to have an analytical expression of the objective function in terms of the required variables. The proposed algorithms make it possible to calculate not only the first, but also the second derivatives in the absence of such expressions.

Keywords: route plan, spline, non-linear programming, clothoid, objective function, gradient, Hessian matrix

• **Submitted:** 20.12.2022 • **Revised:** 12.04.2023 • **Accepted:** 14.05.2023

For citation: Karpov D.A., Struchenkov V.I. Combined approximation algorithms for interactive design of road routes in CAD. *Russ. Technol. J.* 2023;11(4):72–83. https://doi.org/10.32362/2500-316X-2023-11-4-72-83

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Комбинированные алгоритмы аппроксимации для интерактивного проектирования дорожных трасс в системах автоматизированного проектирования

Д.А. Карпов,
В.И. Струченков @

МИРЭА – Российский технологический университет, Москва, 119454 Россия
@ *Автор для переписки, e-mail: str1942@mail.ru*

Резюме

Цели. Цель работы состоит в создании алгоритмов аппроксимации последовательности точек на плоскости дугами клотоид и окружностей. Такая задача возникает в проектировании трасс железных и автомобильных дорог. План (проекция на горизонтальную плоскость) трассы дороги – это кривая (сплайн), состоящая из повторяющейся связки элементов «прямая + дуга клотоиды + дуга окружность + дуга клотоиды + …». Такая комбинация элементов обеспечивает непрерывность не только кривой и касательной к ней, но и кривизны. Поскольку число элементов сплайна заранее неизвестно, а на их параметры накладываются ограничения, для этой задачи пока не опубликовано математически корректного алгоритма. Разработанная в РТУ МИРЭА двухэтапная схема решения задачи с определением числа элементов сплайна с помощью динамического программирования на первом этапе и оптимизацией его параметров с применением нелинейного программирования на втором, реализована только для сплайна с прямыми и окружностями (без клотоид). Ее реализация для сплайна с клотоидами много сложнее и пока не выполнена в силу ряда причин. В действующих системах автоматизированного проектирования (САПР) проектирование плана трассы выполняется в интерактивном режиме с последовательным подбором элементов. В этой связи имеет смысл разработка математически корректных алгоритмов поэлементной аппроксимации.

Метод. Задача поэлементной аппроксимации окружностью или клотоидой формализована как задача нелинейного программирования малой размерности. Целевая функция – сумма квадратов отклонений от исходных точек. Поскольку клотоида в декартовых координатах представляется степенными рядами, возникают трудности вычисления производных целевой функции по искомым параметрам элементов сплайна. Предложен математически корректный алгоритм вычисления этих производных на основе интегрального представления декартовых координат точек клотоиды как функций ее длины.

Результаты. Разработаны математическая модель и алгоритмы аппроксимации последовательности точек на плоскости клотоидой и окружностью с применением метода нелинейного программирования. Реализован алгоритм второго порядка с вычислением и обращением матрицы вторых производных (матрица Гессе).

Выводы. Для аппроксимации окружностью и клотоидой с применением нелинейного программирования необязательно иметь аналитическое выражение целевой функции через искомые переменные. Предложенные алгоритмы позволяют вычислять не только первые, но и вторые производные в отсутствие таких выражений.

Ключевые слова: план трассы, сплайн, нелинейное программирование, клотоида, целевая функция, градиент, матрица Гессе

• **Поступила:** 20.12.2022 • **Доработана:** 12.04.2023 • **Принята к опубликованию:** 14.05.2023

Для цитирования: Карпов Д.А., Струченков В.И. Комбинированные алгоритмы аппроксимации для интерактивного проектирования дорожных трасс в системах автоматизированного проектирования. *Russ. Technol. J.* 2023;11(4):72–83. https://doi.org/10.32362/2500-316X-2023-11-4-72-83

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The problem of approximating a curve defined by a sequence of points in the plane by circles and clothoids is much more complicated than the widely-used linear or parabolic regression approach. In the case of search for an approximating circle in Cartesian coordinates, the problem is reduced to a nonlinear system of equations. In the case of a clothoid, it is impossible to obtain even this, since such a curve cannot be represented in the form $y = f(x)$. A mathematically consistent algorithm for approximation by a clothoid has not yet been found in the literature.

Instead, current CAD solutions either use the method of enumeration of variants assigned by the designer or approximation by the involute method. The latter was proposed in the pre-computer era [1] as a means of calculating the shifts of a route plan to bring it to a given design position. During reconstruction of a railroad route plan, a design position is unknown. When changing plan parameters such as lengths of transition curves (clothoid arcs) or radii of circular curves, the design position is set in one way or another by the designer. After that, the computer calculates all design parameters of the route plan and shifts of the existing route. The involute method was also used for approximate solution of approximation problem instead of its exact solution in Cartesian coordinates[1]. This method has been successfully used to design a route plan as a whole within the railway haul back in the 1980s on such unsuitable for creating CAD computers as Minsk 32 and EU 1033.

The avoidance of Cartesian coordinates seemed to be forced. At the same time, it was known that the error of the method due to the presence of large angles and small radii can be very significant. This is especially true for heavily dislocated curves and correspondingly large shifts or when designing reconstruction with significant changes in the route plan parameters.

The involute method was also used in the 1980s for optimizing the horizontal alignment of new railroads in stressed sections [2, 3]. Here, the route plan was represented as a broken line to find the number of elements and their approximate location, which was then transformed to the required shape using this method with subsequent optimization of the resulting spline parameters [3]. Despite the disadvantages of the involute method, its wide application is explained by the fact that in contrast to Cartesian coordinates, it uses single-valued functions having simple analytical expressions: a parabola of the second order instead of a circle and a cubic parabola in the place of a clothoid.

Various heuristic algorithms were proposed in the works of authors [4–12]. Initially, these were based on analyzing initial broken line characteristics obtained after connecting adjacent survey points by line segments, such as rotation angles in the vertices of the broken line and curvature graphs. Subsequently, the use of genetic algorithms became more common. In Russia, various programs are used in current CAD systems to facilitate element-by-element selection and evaluation of trace variants. This approach has been most successfully realized by the Topomatic[2] company in the *Robur* system, which uses an existing semiautomatic trace plan parameter selection method based on curvature graphs. Although there have been claims to have solved the problem of designing railroad track plan reconstruction, the substantiating algorithms have yet to be published in full. Among recently published heuristic algorithms, the ones presented in the works of Chinese professor Hao Pu and colleagues [13–15] should be noted.

It is hard to believe in the existence of an optimal solution for several circular and transitive

[1] *Methodical recommendations for the calculation of composite curves of the railroad route plan*. Moscow: All-Russian Research Institute of Transport Construction; 1985. 26 p. (in Russ.).

[2] Topomatic Robur product documentation “Path selection via curvature graph.” http://help.topomatic.ru/v6/doku.php?id=rail:rail:tasks:selection_path:start. Accessed April 18, 2023 (in Russ.).

curves (clothoids) if there is no optimal solution algorithm for a single clothoid.

As with the general design of plan- and longitudinal road profiles, the problem under consideration is characterized by the lack of a known optimal solution for complex cases. In the absence of designer interest in the achievement of such a solution, anything plausible that the computer gives out can be called "optimal". This represents the fundamental difference between this problem and those associated with modeling geometric shapes of roads [16].

Nevertheless, the development of mathematically substantiated algorithms of approximation by a circle and a clothoid remains both theoretically and practically relevant, since the result can be useful not only in road design. The up-to-date level of computer technology allows solving this problem in a reasonable time on publicly available computers and without the use of palliative algorithms.

The aim of this study is to present mathematically substantiated algorithms for approximation by a circle and a clothoid in Cartesian coordinates by means of nonlinear programming algorithms that use the involute method only to obtain an initial approximation, followed by an optimization of the parameters used to determine the position of the circle or clothoid.

PROBLEM STATEMENT

For a given sequence of points in the plane find a clothoid (circle) such that the sum of squares of distances h_i (i = 1, 2, …, n) from the given points to the clothoid (circle) is minimal. The distances are calculated using the normal line from a given point to the clothoid (circle).

The initial point of the desired curve is given. The direction of the tangent to the desired curve at the initial point and the minimum and maximum radius of curvature of the desired curve can also be specified.

Here, the objective function is

$$F(\mathbf{h}) = \frac{1}{2}\sum_{1}^{n} h_i^2 \to \min. \qquad (1)$$

Here, $\mathbf{h}(h_1, h_2, ..., h_n)$ is the vector of unknowns, while n is their number.

INVOLUTE METHOD

Unlike other curves (parabolas, circles, sinusoids, etc.), one cannot speak of an involute without specifying another curve (evolute) that generates the involute. Various definitions of involute can be found in the literature. In mathematics, an involute is a *curve* for which a given evolute is the locus of curvature centers [17]. Consequently, the normal line at each point of an involute is a tangent to the evolute. This has to be a tangent rather than a secant because there can be only one center of curvature at each point of a curve. In this context, an involute is a curve described by the end of a flexible, inextensible thread coiled from an evolute (e.g., a circle).

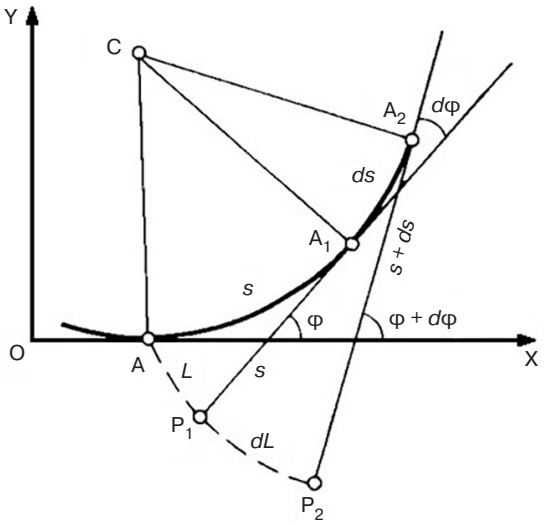


Fig. 1. Construction of the involute of a circle. *L* is the involute length

Figure 1 depicts the construction of the involute of a circle. Here, A is the initial point on the evolute, A_1 and A_2 are its new positions on the evolute. They correspond to the lengths of arcs counted from the initial point, s and $s + ds$. These lengths are tangentially set in the direction opposite to the motion from the initial point. The points P_1 and P_2 are obtained on the involute, respectively. The arc lengths of the involute are AP_1 and AP_2. At full rotation ($\varphi = 2\pi$) the length of the involute $E = 2\pi R$, where R is the radius of the circle. Thus, it is not the initial point A that is fixed, but its new positions on the evolute, in which tangents are constructed and on which the arcs of the involute are unfolded. *The involute of a circle is an unfolding spiral.*

However, in [1, 18, 19] a different treatment of the circle involute concept and the means of its construction is stated (Fig. 2). Here, the point A of the beginning of circular curve is fixed and for each point of the circle (a) the arc length from the initial point to the current one (Aa) is unfolded by a tangent from the point A. Thus, a *separate* involute is obtained (aa', bb' ...) for each point of the circle. Although the curve aa' is not a circle involute, its length is equal to the length of the corresponding section of involute. Hence, the confusion of the terms arose.

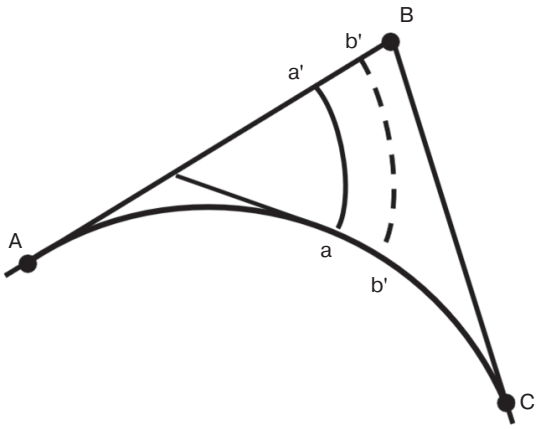


Fig. 2. Construction of the circle involute according to [1, 18, 19]

In [1, 18], circle involutes are *curves* constructed in a described way. But in [19, p. 243, Fig. 7.20], we can find: "*Involute* (italics of the authors) is the *length* of the arc aa', which will be described by the end of an inextensible thread, stretched on the circle Aa and fixed at the point A, when the thread is straightened."

If we assume that the involute is not the *curve* discussed above, but the *arc length* aa', then this length is $K^2/(2R)$, where K is the arc length Aa, R is the radius of the circle, while the circle involute is not a spiral in Cartesian coordinates, but a parabola of the second degree as a function of the evolute length.

If we follow the method of construction of the circle involute adopted in [1, 18, 19], then for each point of the circle (evolute) we will get a different involute, but they will all end on a tangent to the circle at the initial point, which is fixed. The true circle involute should be constructed as shown in Fig. 1.

From Fig. 1 it follows that $dL = \varphi(s)ds$ since $dL = sd\varphi = sds/R = \varphi ds$, and further:

$$L(S) = \int_0^S \varphi(s)ds. \qquad (2)$$

Here S is the length of the evolute from the initial point, while $\varphi(s)$ is the dependence of the angle of the tangent to the evolute with the OX axis (or any given direction, which is not fundamental) on s, referred to as the angle diagram. The term "involute" was defined for smooth curves, and Eq. (2) is true for all such curves. However, in our problem (1), the initial curve is a broken line, and we cannot apply the above definitions of involute to it—i.e., consider the broken line as an involute. However, Eq. (2) can be generalized if we consider a curve whose length is calculated through the angle diagram of an evolute by Eq. (2), i.e., the length of an involute is the area of the angle diagram as a function of length of the original curve (particularly, a polyline). Thus, the involute itself is of no interest for the problem in question. Here, it is significant that the difference of lengths of involutes of two evolutes having a common point is, under some additional conditions, approximately equal to the distance between the evolutes along the normal. In our problem, the first evolute is a broken line connecting the approximated points, while the second is a design curve (a circle or a clothoid). Hence the involute method, which consists in the following:

1. We calculate the angles of the original broken line with OX axis and the values of the involute length L_i of the original broken line for survey points $i = 1, 2, \ldots, n$.
   $L_1 = 0$, further

$$L_i = \sum_{j=2}^{i} \varphi_j s_j. \qquad (3)$$

Here φ_j is the angle with OX axis of the jth leg of the broken line (from point $j-1$ to point j), s_j is the length of this leg.

For a circle, the rotation angle depends linearly on the arc length, so the length of the design involute $L_{des}(S)$ is a second-order parabola.

2. If the design curve is given, the difference $L_{des}(S) - L_{init}(S)$ approximates the desired shift h_j, where j and S must match each other.
3. If we want to solve the approximation problem, the unknown coefficients of the desired parabola (a and b) are determined by the least square method.

For a circle, the involute *length* $L_{des}(S) = aS^2 + bS$, where S, as before, is the arc length from a given initial point. The meaning of the required parameters is as follows: $a = 1/(2R)$ and b is the angle with OX axis of the tangent to the circle at the initial point.

We obtain the problem:

$$F(a,b) = \sum_{i=2}^{n} \left(aS_i^2 + bS_i - L_i \right)^2 \to \min. \qquad (4)$$

Here S_i is the length of the broken line from the first point to the ith survey point ($i = 2, \ldots, n$), $S_1 = 0$.

Conditions $\frac{\partial F}{\partial a} = 0$ and $\frac{\partial F}{\partial b} = 0$ give a system of two linear equations, from which we obtain a and b; then $R = 1/(2a)$. Since the angle b can be given, only a remains unknown.

When approximating with a clothoid, the following options are possible: the clothoid connects a line to a circle; a circle to a line; two circles of various radii.

If the lengths of the contiguous clothoids are significantly shorter than the lengths of the circular

curves, it is important to select the position of the circle, and only subsequently fit the clothoids. However, the opposite situation is also possible, where the clothoid is an independent element. For example, in highway design, one finds the term "clothoidal tracing" [20]. In this case, the route should not be thought of as consisting of only clothoids and straight lines, but clothoids prevail in the share by length.

By definition, the clothoid curvature $\sigma = \sigma_0 + kS$, where σ_0 is the curvature at the initial point (at length $S = 0$), k is the curvature change rate. The current angle of the tangent with OX axis is $\varphi = \varphi_0 + \sigma_0 S + kS^2/2$, where φ_0 is the initial angle of the tangent with OX axis. Finally, the length of the clothoid involute is equal to

$$L = \varphi_0 S + \sigma_0 S^2/2 + kS^3/6.$$

Clothoid approximation by the involute method in the general case is reduced to the problem:

$$F(\varphi_0, \sigma_0, k) = \\ = \sum_{i=2}^{n} \left(\varphi_0 S_i + \frac{\sigma_0 S_i^2}{2} + \frac{kS_i^3}{6} - L_i \right)^2 \to \min. \quad (5)$$

Here S_i and L_i have the same meaning as in the circle approximation problem (4). One or two variables can be fixed. In any case the problem (5) is solved simply. By differentiating (5) by the desired variables and equating the derivatives to zero, we obtain a system of linear equations (or one equation if two unknown variables are fixed). After solving the system, one should calculate the curvature at the end point of the clothoid $\sigma_n = \sigma_0 + kS_n$.

If $R_n = 1/\sigma_n$ appears outside the admissible limits, we should take $R_n = R_{\text{lim}}$ and calculate $k = (1/R_{\text{lim}} - \sigma_0)/S_n$ at a given σ_0. Then we obtain the unknown φ_0 by substituting the found k in (5) and solving the problem with one unknown variable. If σ_0 is not given, we substitute $\sigma_0 = \sigma_n - kS_n$ in (5) and solve the problem (5) with the remaining unknown variables and the found σ_n.

Thus, it is very easy to use the involute method for approximation by both a circle and clothoid. Unfortunately, the accuracy of the method may be insufficient. To test this statement, the Cartesian coordinates of the ends of the chords of a *given* length on a *given* circle were calculated as the coordinates of the initial survey points. Using these coordinates, the chord angles with the OX axis and involute lengths were calculated by Eq. (3). Then the problem (4) was solved. Obviously, its solution is the variables corresponding to the initial circle. In this case the deviations of all initial points from the approximating circle should be equal to zero. However, this was not the case. At circle radius $R = 500$ m or more, chords $l_x = 20$ m, and circle length $S \le 500$ m (i.e., at rotation angles less than 1 rad), the deviations of the obtained circle from the original were less than 0.01 m. However, at $R = 200$ m, $l_x = 20$ m, $S = 200$ m, the radius determined by the involute method was 199.9167 m instead of 200 m and the maximum deviation D_{max} was 0.0383 m instead of 0. At $S = 400$ m (rotation angle of 2 rad) for the same circle, $D_{\text{max}} = 0.1180$ m instead of 0. In all cases, the length differences of design and initial involutes of the approximated circle in all points of survey were equal to zero. At $l_x = 10$ m, the accuracy of the method is significantly higher. Thus, at a division by 10 m we get $R = 299.9861$ and $D_{\text{max}} = 0.00297$ if $R = 300$ m, $S = 200$ m, while at a division by 20 m, we obtain $R = 299.9444$ m and $D_{\text{max}} = 0.01190$ m.

Similar calculations were performed with respect to the use of the involute method to approximate the clothoid. For compressing them, arcs of equal length were used, rather than chords, simplifying the calculation of Cartesian coordinates of points at the ends of arcs. These points were treated as survey points, chord lengths were calculated, and, as for circles, angles and involute lengths L_i were calculated using Eq. (3). Then, the problem (5) was solved with $\sigma_0 = 0$ and the unknown variables φ_0 and k.

As one would expect, the involute method works well for small values of the parameter k and short clothoids. Otherwise, the results are unsatisfactory.

Example 1. A clothoid of length 400 m is divided into 20 equal parts every 20.0 m; $k = 3.333333 \cdot 10^{-5}$. We obtain $\varphi_0 = -0.00027412$ instead of 0 and $k = 3.337930 \cdot 10^{-5}$ instead of the original $k = 3.333333 \cdot 10^{-5}$.

Approximation of the involute length as a function of length of a broken line by the cubic parabola (solution of problem 5) is performed with deviations not more than 0.041 m. However, the maximum deviation D_{max} of the initial points from the obtained clothoid along the normal line to it equal to 0.255 m.

Example 2. Same problem, but the length of the clothoid is 200 m. Approximation with the cubic parabola is performed with deviations not more than 0.001 m. We obtain $\varphi_0 = -1.839272551 \cdot 10^{-5}$ instead of 0 and $k = 3.334527 \cdot 10^{-5}$ instead of the original $k = 3.333333 \cdot 10^{-5}$, $D_{\text{max}} = 0.048$ m.

It should be noted that these calculations show not so much how the involute method finds an optimum, but rather how it deviates from it. For a "heavily hit" initial route, the deviations from the optimum of the solutions obtained by the involute method may be significantly higher, especially for small radii and large rotation angles. However, in any case, these solutions can be used as initial approximations for the optimal approximation by a circle.

OPTIMIZATION OF THE INITIAL APPROXIMATION WHEN APPROXIMATING BY A CIRCLE

Let us assume that the coordinates of points to be approximated by the circle and its initial point A are given. Although we also usually fix the direction of the tangent at this point (angle α with OX axis), we will assume without loss of generality that this angle and the circle radius R are unknown. Coordinates of the circle center (Fig. 3) are

$$x_c = x_a - R \cdot \sin\alpha \text{ and } y_c = y_a + R \cdot \cos\alpha. \quad (6)$$

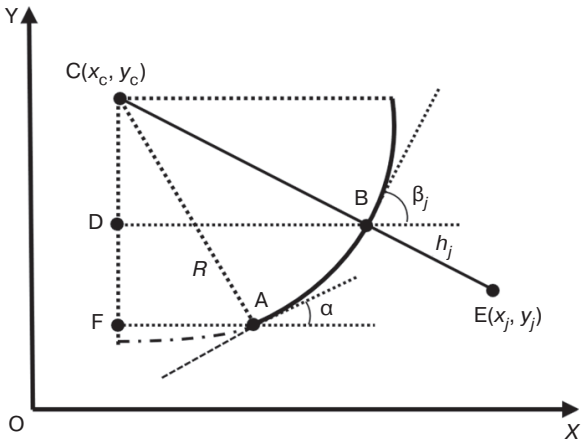


Fig. 3. Calculation of the derivatives of deviations h_j from the circle

Deviation h_j (BE in Fig. 3) of an arbitrary survey point E with coordinates x_j, y_j from the circle is

$$h_j = \sqrt{(x_j - x_c)^2 + (y_j - y_c)^2} - R. \quad (7)$$

Using (6) and (7), we obtain the derivatives:

$$\frac{\partial h_j}{\partial R} = \cos(\beta_j - \alpha) - 1 \text{ and } \frac{\partial h_j}{\partial \alpha} = R\sin(\beta_j - \alpha).$$

Here, β_j is the angle of the tangent to the circle at point B.

Then we calculate the gradient of the objective function (1):

$$\begin{aligned} \frac{\partial F}{\partial R} &= \sum_{j=1}^{n} h_j \cdot \frac{\partial h_j}{\partial R} = \sum_{j=1}^{n} h_j(\cos(\beta_j - \alpha) - 1), \\ \frac{\partial F}{\partial \alpha} &= \sum_{j=1}^{n} h_j \cdot \frac{\partial h_j}{\partial \alpha} = R\sum_{j=1}^{n} h_j \sin(\beta_j - \alpha). \end{aligned} \quad (8)$$

The problem of approximation by a circle with objective function (1) is reduced to a two-dimensional minimization problem with restriction on R and initial approximation obtained by the method of involutes as a result of solving problem (4). Here, the presence of a good initial approximation is especially important, since there is no reason to believe that the problem is one-extremal.

Various methods can be used to solve the problem [21–23], in particular gradient or coordinate descent, changing alternately R and α.

Note, that if the initial point A and angle α are fixed, setting one more point of the circle (e.g., the end point) defines the circle uniquely. The enumeration of several such points and calculation of the value of the objective function for each of them may be sufficient, especially for large radii, when the involute method gives acceptable results, which can be improved and confirmed.

OPTIMIZATION OF THE INITIAL APPROXIMATION WHEN APPROXIMATING BY A CLOTHOID

As in the approximation by a circle, the main problem when approximating by clothoid is to calculate the derived distances of the given survey points from the clothoid by the parameters defining it. Let us show how this problem is solved by the example of the transition from a line to a circle using a clothoid.

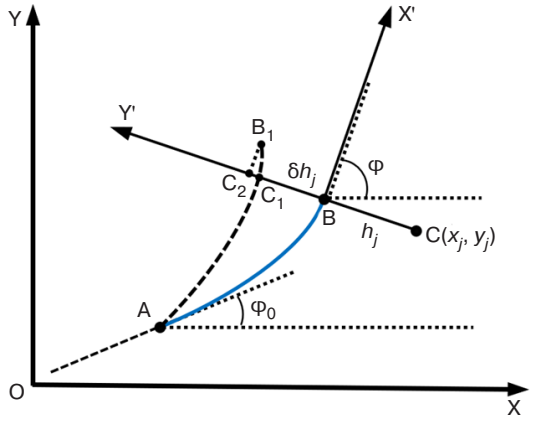


Fig. 4. To the calculation of the derivatives of the deviations h_j from the clothoid

In Fig. 4, AB is the initial position of the clothoid; AB_1 is its new position when one of the clothoid parameters is changed; φ_0 and φ are the angles of the tangent with the OX axis; $CB = h_j$ is the deviation of the survey point C from the clothoid; BC_1 is its increment.

Let us denote the incremental coordinates of point B at its transition to a new position B_1 due to a change in any of the parameters of the clothoid by δx_B and δy_B. In the coordinate system with a center in point B and axes directed along the tangent and normal to the clothoid, respectively, the coordinate y' of the point B_1, i.e., $BC_2 = \delta y_B\cos\varphi - \delta x_B\sin\varphi$. The value C_1C_2 can be neglected

compared to C_1B, and the linear part of the increment can be assumed to be

$$\delta h_j = \delta y_B \cos\varphi - \delta x_B \sin\varphi. \quad (9)$$

Since the initial point A is fixed, the parameters determining the position of the clothoid—and hence the quality of the approximation—are in the general case: φ_0 is the angle of the tangent with the axis OX at the initial point (at $L = 0$), σ_0 is the initial curvature; k is the clothoid parameter (the curvature change rate, i.e., the curvature derivative over the length).

In the case of the transition from a line to a circle $\sigma_0 = 0$, only two variables φ_0 and k remain. In fact, the problem is reduced to the calculation of derived coordinates of points of the clothoid by these variables. Coordinates of intersection points of normals from each survey point with the clothoid and the length of the clothoid L_j from the initial point to the intersection points are calculated before the derivatives are calculated using the iterative algorithm along with the tangent angles with the OX axis [24].

Since the Cartesian coordinates of the clothoid points as functions of its lengths are expressed by power series, we will use their integral representation:

$$\begin{aligned} x_B &= x_A + \int_0^L \cos\left(\varphi_0 + \frac{kt^2}{2}\right) dt, \\ y_B &= y_A + \int_0^L \sin\left(\varphi_0 + \frac{kt^2}{2}\right) dt. \end{aligned} \quad (10)$$

By differentiating (10), we obtain

$$\frac{\partial x_B}{\partial \varphi_0} = -\int_0^L \sin\left(\varphi_0 + \frac{kt^2}{2}\right) dt = -(y_B - y_A),$$

$$\frac{\partial y_B}{\partial \varphi_0} = \int_0^L \cos\left(\varphi_0 + \frac{kt^2}{2}\right) dt = (x_B - x_A),$$

$$\frac{\partial x_B}{\partial k} = -\int_0^L \sin\left(\varphi_0 + \frac{kt^2}{2}\right) \frac{t^2}{2} dt =$$

$$= \frac{1}{2k} \int_0^L td \cos\left(\varphi_0 + \frac{kt^2}{2}\right) = \frac{L\cos\varphi - (x_B - x_A)}{2k},$$

$$\frac{\partial y_B}{\partial k} = \int_0^L \cos\left(\varphi_0 + \frac{kt^2}{2}\right) \frac{t^2}{2} dt =$$

$$= \frac{1}{2k} \int_0^L td \sin\left(\varphi_0 + \frac{kt^2}{2}\right) = \frac{L\sin\varphi - (y_B - y_A)}{2k}.$$

Using (9), we obtain the required derivatives of the survey point deviations from the clothoid:

$$\begin{aligned} \frac{\partial h_j}{\partial \varphi_0} &= (x_B - x_A)\cos\varphi + (y_B - y_A)\sin\varphi, \\ \frac{\partial h_j}{\partial k} &= \frac{(x_B - x_A)\sin\varphi - (y_B - y_A)\cos\varphi}{2k}. \end{aligned} \quad (11)$$

Here A is the initial point of the clothoid; h_j is the deviation of an arbitrary jth survey point from the clothoid along the normal line to it; B is the point where this normal line intersects the clothoid; φ is the angle of the tangent at this point with OX axis; L is the length of the clothoid from point A to point B.

Formulas (11) allow us to find the gradient of the function (1)

$$\frac{\partial F}{\partial k} = \sum_{j=1}^{n} h_j \cdot \frac{\partial h_j}{\partial k}; \quad \frac{\partial F}{\partial \varphi_0} = \sum_{j=1}^{n} h_j \cdot \frac{\partial h_j}{\partial \varphi_0}. \quad (12)$$

Further, it is possible to use gradient methods of optimization, e.g., the method of conjugate gradients [22–24]. However, Eqs. (10)–(12) allow a calculation of the matrix of second derivatives of the objective function **G** (Hessian matrix) to apply the more effective second-order method [23, 24]:

$$\begin{aligned} \frac{\partial^2 F}{\partial k^2} &= \sum_{j=1}^{n} \left(\left(\frac{\partial h_j}{\partial k} \right)^2 + h_j \cdot \frac{\partial^2 h_j}{\partial k^2} \right) = G_{11}, \\ \frac{\partial^2 F}{\partial k \partial \varphi_0} &= \sum_{j=1}^{n} \left(\frac{\partial h_j}{\partial \varphi_0} \cdot \frac{\partial h_j}{\partial k} + \frac{\partial^2 h_j}{\partial k \partial \varphi_0} \right) = G_{12} = G_{21}, \\ \frac{\partial^2 F}{\partial \varphi_0^2} &= \sum_{j=1}^{n} \left(\left(\frac{\partial h_j}{\partial \varphi_0} \right)^2 + h_j \cdot \frac{\partial^2 h_j}{\partial \varphi_0^2} \right) = G_{22}. \end{aligned} \quad (13)$$

Calculation of the Hessian matrix is reduced to the calculation of the second derivatives of the deviations h_j of the approximated points from the clothoid by the desired variables φ_0 and k, since the first derivatives have already been calculated.

In accordance with (9)

$$\frac{\partial^2 h_j}{\partial k^2} = -\sin\varphi \cdot \frac{\partial^2 x_B}{\partial k^2} + \cos\varphi \cdot \frac{\partial^2 y_B}{\partial k^2} =$$

$$= \sin\varphi \int_0^L \cos\left(\varphi_0 + \frac{kt^2}{2}\right) \frac{t^4}{4} dt -$$

$$- \cos\varphi \int_0^L \sin\left(\varphi_0 + \frac{kt^2}{2}\right) \frac{t^4}{4} dt.$$

Integrating by parts and using the previously calculated integrals, we obtain

$$\frac{\partial^2 h_j}{\partial k^2} = \frac{1}{4k}\left\{L^3 + \frac{3}{k}(\cos\varphi(y_B - y_A) - \sin\varphi(x_B - x_A))\right\},$$
$$\frac{\partial^2 h_j}{\partial \varphi_0^2} = -\sin\varphi \cdot \frac{\partial^2 x_B}{\partial \varphi_0^2} + \cos\varphi \cdot \frac{\partial^2 y_B}{\partial \varphi_0^2} =$$
$$= \sin\varphi(x_B - x_A) - \cos\varphi(y_B - y_A),$$
$$\frac{\partial^2 h_j}{\partial k \partial \varphi_0} = \sin\varphi \int_0^L \cos\left(\varphi_0 + \frac{kt^2}{2}\right)\frac{t^2}{2}dt - \quad (14)$$
$$-\cos\varphi \int_0^L \sin\left(\varphi_0 + \frac{kt^2}{2}\right)\frac{t^2}{2}dt =$$
$$= \frac{1}{2k}\left\{L - \sin\varphi(y_B - y_A) - \cos\varphi(x_B - x_A)\right\}.$$

Using the derivatives obtained, we calculate the Hessian matrix and its inverse $\mathbf{G}^{-1}$.

Let us denote: $\mathbf{x}(\varphi_0, k)$ is the vector of the unknown variables, $\mathbf{g}\left(\frac{\partial F}{\partial \varphi_0}, \frac{\partial F}{\partial k}\right)$ is the gradient. For the initial approximation obtained by the involute method (zero iteration) these are $\mathbf{x^0}$ and $\mathbf{g^0}$. Let us calculate the next iteration point:

$$\mathrm{x}^1 = \mathrm{x}^0 - \mathrm{G}^{-1} \cdot \mathrm{g}^0. \quad (15)$$

For a positively defined Hessian matrix, this is the point of minimum of the objective function when decomposed into a Taylor series and restricted to the second derivatives. In the general case, one step to the point of minimum on the ray $\mathbf{x}^0 - \lambda\mathbf{G}^{-1} \cdot \mathbf{g}^0$ is insufficient. In the quadratic problem, the minimum is reached at $\lambda = 1$. In general case at $\lambda = 1$, we obtain the minimum point not of the original function, but of the approximating quadratic form with the Hessian matrix, which is not the same thing. Therefore, $\lambda = 1$ should be regarded only as an approximate value: the exact value should be sought by solving the problem of one-dimensional optimization of the function $F(\mathbf{x})$ on the ray $\mathbf{x}^0 - \lambda\mathbf{G}^{-1} \cdot \mathbf{g}^0$, considering it as a function of the single parameter λ.

At the resulting point of minimum in the direction, we again have to calculate the gradient and the matrix inverse to the Hessian matrix, and so on.

In general, the algorithm of the clothoid approximation consists of the following steps.

1. Construct the angular diagram of the original broken line by survey points.
2. Calculate the involute length of the broken line sequentially by the survey points by the formula (3).
3. Solve the problem of approximation of the obtained broken line. In general case, approximation by cubic parabola in the presence of three unknown parameters of clothoid or by square parabola in the presence of two parameters, as considered above. In this case, the system of no more than three linear equations is solved.
4. For a clothoid corresponding to the obtained solution, a special iterative algorithm determines the intersection points with the normals from each survey point, the angles of the tangents at these points with the OX axis, the corresponding lengths from the initial point of the clothoid to each of them, and the deviations h_j of the survey points from the clothoid.
5. The first and second derivatives of the deviations h_j are calculated for the required parameters of the clothoid (12), (13).
6. The gradient of the objective function (14) is calculated and the conditions of the end of counting are checked (for example, the smallness of the gradient norm). If the count terminating conditions are not satisfied, then:
7. Hessian matrix and its inverse are calculated.
8. The point of minimum of corresponding quadratic form (15) is determined, the problem of one-dimensional optimization, i.e., correction of step in search direction, is solved, and transition to new iteration point and further with new values of unknowns to item 4 is performed.

To adjust and verify the clothoid approximation algorithm, we first used the results of the test problems using the involute method, for which the optimum was known, but the involute solutions were unsatisfactory.

In the involute method Example 1 discussed above, the maximum deviation D_{max} = 0.255 m instead of 0, φ_0 = −0.00027412 instead of 0 and $k = 3.337930 \cdot 10^{-5}$ instead of $k = 3.333333 \cdot 10^{-5}$. Objective function F^0 = 0.10298. At the point of minimum of the quadratic form (15), $F = 1.08524 \cdot 10^{-7}$, $\varphi_0 = 3.099241 \cdot 10^{-7}$, $k = 3.333335 \cdot 10^{-5}$. After the first iteration $F = 3.70261 \cdot 10^{-15}$, $\varphi_0 = -2.422261 \cdot 10^{-14}$, $k = 3.333333333 \cdot 10^{-5}$. No deviations exceeded 0.00006 m.

In Example 2, the same problem was solved, but with a clothoid length of 200 m. By the involute method, $\varphi_0 = -1.839272551 \cdot 10^{-5}$ instead of 0 and $k = 3.334527 \cdot 10^{-5}$ instead of $k = 3.333333 \cdot 10^{-5}$ were obtained. D_{max} = 0.048 m. At the minimum point of the quadratic form (15) $F = 1.80438 \cdot 10^{-14}$, $\varphi_0 = -9.592883457 \cdot 10^{-10}$, $k = 3.333334 \cdot 10^{-5}$, D_{max} = 0.0002 m. The iteration was interrupted due to achieving the required gradient accuracy.

CONCLUSIONS

New possibilities offered by contemporary publicly available computers, along with the theory and methods of computer development of design

solutions, are far from being fully utilized in existing linear structure CAD solutions, which are based on ideas from more than 50 years ago. Transition to the development and introduction of intelligent CAD, in which design solutions are given by computer as a result of optimization problem solving, is already possible in many design problems. However, because of lack of consumer interest in creating such systems and the high labor and funds for their development in Russia, such transition is unlikely in the near future. The most advanced in this respect are the mentioned works of Chinese scientists. However, existing CAD systems can be improved by applying optimization programs in the interactive process of making design decisions. Thus, the approximation algorithms outlined in the article can be applied instead of manually assigned solutions. The outlined clothoid approximation algorithm may be useful also when solving problems not related to roadway design.

The closest task is the generalization of the outlined method for calculating the derivatives of the objective function by the parameters determining the curve as a whole (rather than a single clothoid), in the absence of analytical expression of this function through the required parameters. We are talking about the transition from the solution of the problem of approximation by a single clothoid to spline approximation by the sequence of several bundles of "straight line + clothoid + circle + clothoid +...".

Authors' contribution. All authors equally contributed to the present work.

REFERENCES

1. Gorinov A.V. *Izyskaniya i proektirovanie zheleznykh dorog* (*Research and Design of Railways*). V. 2. Moscow: Transzheldorizdat; 1961. 338 p. (in Russ.).
2. Sheidwasser D.M. Optimization of railroad tracks on busy tracks. In: *Avtomatizatsiya proektirovaniya ob'ektov transportnogo stroitel'stva* (*Automation of the Design of Transport Construction Objects*). Moscow: Transport; 1986. P. 16–29 (in Russ.).
3. Struchenkov V.I., Sheidwasser D.M. Optimization on a computer of the route of a new railway on stressful passages. *Transportnoe stroitel'stvo = Transport Construction*. 1987;3:7–9 (in Russ.).
4. Jha M.K., McCall C., Schonfeld P. Using GIS, genetic algorithms, and visualization in highway development. *Computer-Aided Civil and Infrastructure Engineering*. 2001;16(6):399–414. https://doi.org/10.1111/0885-9507.00242
5. Jha M.K., Schonfeld P. A highway alignment optimization model using geographic information systems. *Transp. Res. Part A. Policy Pract*. 2004;8(6):455–481. https://doi.org/10.1016/j.tra.2004.04.001
6. Jong J.C., Jha M.K., Schonfeld P. Preliminary highway design with genetic algorithms and geographic information systems. *Computer-Aided Civil and Infrastructure Engineering*. 2000;15(4):261–271. https://doi.org/10.1111/0885-9507.00190
7. Kang M.W., Schonfeld P., Yang N. Prescreening and repairing in a genetic algorithm for highway alignment optimization. *Computer-Aided Civil and Infrastructure Engineering*.2009;24(2):109–119.https://doi.org/10.1111/j.1467-8667.2008.00574.x
8. Pushak Y., Hare W., Lucet Y. Multiple-path selection for new highway alignments using discrete algorithms. *Eur. J. Oper. Res*. 2016;248(2):415–427. https://doi.org/10.1016/j.ejor.2015.07.039
9. Sarma K.C., Adeli H. Bilevel parallel genetic algorithms for optimization of large steel structures. *Computer Aided Civil and Infrastructure Engineering*. 2001;16(5):295–304. https://doi.org/10.1111/0885-9507.00234

СПИСОК ЛИТЕРАТУРЫ

1. Горинов А.В. *Изыскания и проектирование железных дорог*. Т. 2. М.: Трансжелдориздат; 1961. 338 с.
2. Шейдвассер Д.М. Оптимизация трасс железных дорог на напряженных ходах. В кн.: *Автоматизация проектирования объектов транспортного строительства*. М.: Транспорт; 1986. С. 16–29.
3. Струченков В.И., Шейдвассер Д.М. Оптимизация на ЭВМ трассы новой железной дороги на напряженных ходах. *Транспортное строительство*. 1987;3:7–9.
4. Jha M.K., McCall C., Schonfeld P. Using GIS, genetic algorithms, and visualization in highway development. *Computer-Aided Civil and Infrastructure Engineering*. 2001;16(6):399–414. https://doi.org/10.1111/0885-9507.00242
5. Jha M.K., Schonfeld P. A highway alignment optimization model using geographic information systems. *Transp. Res. Part A. Policy Pract*. 2004;8(6):455–481. https://doi.org/10.1016/j.tra.2004.04.001
6. Jong J.C., Jha M.K., Schonfeld P. Preliminary highway design with genetic algorithms and geographic information systems. *Computer-Aided Civil and Infrastructure Engineering*. 2000;15(4):261–271. https://doi.org/10.1111/0885-9507.00190
7. Kang M.W., Schonfeld P., Yang N. Prescreening and repairing in a genetic algorithm for highway alignment optimization. *Computer-Aided Civil and Infrastructure Engineering*.2009;24(2):109–119.https://doi.org/10.1111/j.1467-8667.2008.00574.x
8. Pushak Y., Hare W., Lucet Y. Multiple-path selection for new highway alignments using discrete algorithms. *Eur. J. Oper. Res*. 2016;248(2):415–427. https://doi.org/10.1016/j.ejor.2015.07.039
9. Sarma K.C., Adeli H. Bilevel parallel genetic algorithms for optimization of large steel structures. *Computer Aided Civil and Infrastructure Engineering*. 2001;16(5):295–304. https://doi.org/10.1111/0885-9507.00234
10. Shafahi Y., Bagherian M. A customized particle swarm method to solve highway alignment optimization problem. *Computer-Aided Civil and Infrastructure Engineering*. 2013;28(1):52–67. https://doi.org/10.1111/j.1467-8667.2012.00769.x

10. Shafahi Y., Bagherian M. A customized particle swarm method to solve highway alignment optimization problem. *Computer-Aided Civil and Infrastructure Engineering*. 2013;28(1):52–67. https://doi.org/10.1111/j.1467-8667.2012.00769.x
11. Bosurgi G., D'Andrea A. A polynomial parametric curve (PPC-curve) for the design of horizontal geometry of highways. *Computer-Aided Civil and Infrastructure Engineering*. 2012;27(4):303–312. https://doi.org/10.1111/j.1467-8667.2011.00750.x
12. Cerf R. *The quasispecies regime for the simple genetic algorithm with roulette wheel selection*. arXiv:1506.0981v2. https://doi.org/10.48550/arXiv.1506.09081
13. Pu H., Li W., Schonfeld P., et al. A Method for Automatically Recreating the Horizontal Alignment Geometry of Existing Railways. *Computer-Aided Civil and Infrastructure Engineering*. 2019;34(1):71–94. https://doi.org/10.1111/mice.12392
14. Li W., Zhen S., Schonfeld P., et al. Recreating Existing Railway Horizontal Alignments Automatically Using Overall Swing Iteration. *J. Transport. Eng. Part A: Systems*. 2022;148(8). https://doi.org/10.1061/JTEPBS.0000691
15. Pu H., Fu H., Schonfeld P., et al. Modelling and optimization of constrained alignments for existing railway reconstruction. *Int. J. Rail Transportat*. 2023;11(3):428–447. https://doi.org/10.1080/23248378.2022.2081878
16. Sal'kov N.A. *Modelirovanie geometricheskikh form avtomobil'nykh dorog*: Monografiya (*Modeling Geometric Shapes of Highways*: Monograph). Moscow: INFRA-M; 2019. 162 p. (in Russ.).
17. Smirnov V.I. *Kurs vysshei matematiki* (*Higher Mathematics Course*). V. 2. Moscow: Nauka; 1979. 479 p. (in Russ.).
18. Gorinov A.V., Kantor I.I., Kondratenko A.P., Turbin I.V. *Izyskaniya i proektirovanie zheleznykh dorog* (*Research and Design of Railways*). Moscow: Transport; 1979. 319 p. (in Russ.).
19. Kantor I.I. *Izyskaniya i proektirovanie zheleznykh dorog* (*Research and Design of Railways*). Moscow: Akademkniga; 2003. 288 p. (in Russ.).
20. Fedotov G.A., Pospelov I.I. *Izyskaniya i proektirovanie avtomobil'nykh dorog* (*Research and Design of Highways*). V. 1. Moscow: Vysshaya shkola; 2009. 650 p. (in Russ.).
21. Gill Ph.E., Murray W., Wright M.H. *Prakticheskaya optimizatsiya* (*Practical Optimization*): transl. from Engl. Moscow: Mir; 1985. 509 p. (in Russ.).
[Gill Ph.E., Murray W., Wright M.H. *Practical Optimization*. London: Academic Press; 1981. 402 p.]
22. Kochenderfer M.D., Wheeler T.A. *Algoritmy optimizatsii* (*Algorithms for Optimization*). Moscow: Vil'yams; 2020. 528 p. (in Russ.).
[Kochenderfer M.D., Wheeler T.A. *Algorithms for Optimization*. London: MIT Press; 2019. 520 p.]
23. Chernorutskii I.G. *Metody optimizatsii. Komp'yuternye tekhnologii* (*Methods of optimization. Computer technologies*). St. Petersburg: BHV-Petersburg; 2011. 329 p. (in Russ.).
11. Bosurgi G., D'Andrea A. A polynomial parametric curve (PPC-curve) for the design of horizontal geometry of highways. *Computer-Aided Civil and Infrastructure Engineering*. 2012;27(4):303–312. https://doi.org/10.1111/j.1467-8667.2011.00750.x
12. Cerf R. *The quasispecies regime for the simple genetic algorithm with roulette wheel selection*. arXiv:1506.0981v2. https://doi.org/10.48550/arXiv.1506.09081
13. Pu H., Li W., Schonfeld P., et al. A Method for Automatically Recreating the Horizontal Alignment Geometry of Existing Railways. *Computer-Aided Civil and Infrastructure Engineering*. 2019;34(1):71–94. https://doi.org/10.1111/mice.12392
14. Li W., Zhen S., Schonfeld P., et al. Recreating Existing Railway Horizontal Alignments Automatically Using Overall Swing Iteration. *J. Transport. Eng. Part A: Systems*. 2022;148(8). https://doi.org/10.1061/JTEPBS.0000691
15. Pu H., Fu H., Schonfeld P., et al. Modelling and optimization of constrained alignments for existing railway reconstruction. *Int. J. Rail Transportat*. 2023;11(3):428–447. https://doi.org/10.1080/23248378.2022.2081878
16. Сальков Н.А. *Моделирование геометрических форм автомобильных дорог:* монография. М.: ИНФРА-М; 2019. 162 с.
17. Смирнов В.И. *Курс высшей математики*. Т. 2. М.: Наука; 1967. 479 с.
18. Горинов А.В., Кантор И.И., Кондратенко А.П., Турбин И.В. *Изыскания и проектирование железных дорог.* М.: Транспорт; 1979. 319 с.
19. Кантор И.И. *Изыскания и проектирование железных дорог*. М.: ИКЦ «Академкнига»; 2003. 288 с.
20. Федотов Г.А., Поспелов И.И. *Изыскания и проектирование автомобильных дорог*. Кн. 1. М.: Высшая школа; 2009. 650 с.
21. Гилл Ф., Мюррей У., Райт М. *Практическая оптимизация*: пер. с англ. М.: Мир; 1985. 509 с.
22. Кохендерфер М.Д., Уилер Т.А. *Алгоритмы оптимизации*. М.: Вильямс; 2020. 528 с.
23. Черноруцкий И.Г. *Методы оптимизации. Компьютерные технологии*. СПб.: БХВ-Петербург; 2011. 329 с.
24. Струченков В.И. Новый алгоритм поэлементного расчета трасс в САПР линейных сооружений. *Информационные технологии.* 2015;21(4):271–276.

24. Struchenkov V.I. New algorithm for perelement calculation of line structures routes. *Informacionnye tekhnologii = Information Technologies*. 2015;21(4): 271–276 (in Russ.).

About the authors

Dmitry A. Karpov, Cand. Sci. (Eng.), Head of the General Informatics Department, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: karpov@mirea.ru. RSCI SPIN-code 2619-7100, https://orcid.org/0000-0003-3734-7182

Valery I. Struchenkov, Dr. Sci. (Eng.), Professor, General Informatics Department, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: str1942@mail.ru. RSCI SPIN-code 4581-4698, https://orcid.org/0000-0002-9801-7454

Об авторах

Карпов Дмитрий Анатольевич, к.т.н., заведующий кафедрой общей информатики Института искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: karpov@mirea.ru. SPIN-код РИНЦ 2619-7100, https://orcid.org/0000-0003-3734-7182

Струченков Валерий Иванович, д.т.н., профессор, кафедра общей информатики Института искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: str1942@mail.ru. SPIN-код РИНЦ 4581-4698, https://orcid.org/0000-0002-9801-7454

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 621.372.8
https://doi.org/10.32362/2500-316X-2023-11-4-84-93

RESEARCH ARTICLE

Models of waveguides combining gradient and nonlinear optical layers

Sergey E. Savotchenko [@]

V.G. Shukhov Belgorod State Technological University, Belgorod, 308012 Russia
[@] *Corresponding author, e-mail: savotchenkose@mail.ru*

Abstract

Objectives. Theoretical studies of the waveguide properties of interfaces between nonlinear optical and graded-index media are important for application in optoelectronics. Waveguides combining layers with different optical properties seem to be the most promising, since they can be matched to optimal characteristics using a wide range of control parameters. The paper aims to develop a theory of composite optically nonlinear graded-index waveguides with an arbitrary profile, within which it is possible to obtain exact analytical expressions for surface waves and waveguide modes in an explicit form. The main feature of the theory proposed in this paper is its applicability for describing surface waves and waveguide modes, in which the field is concentrated inside the gradient layer and does not exceed its boundary, avoiding contact with the nonlinear layer.

Methods. Analytical methods of the theory of optical waveguides and nonlinear optics are used.

Results. A theoretical description of the waveguide properties of the interface between two media having significantly different optical characteristics is carried out. The formulated model of a plane waveguide is applicable to media having an arbitrary spatial permittivity profile. An analytical expression describing a surface wave propagating along the interface between a medium having stepwise nonlinearity and a gradient layer with an arbitrary permittivity profile is obtained. Additionally, analytical expressions for surface waves propagating along the interface between a medium with Kerr nonlinearity (both self-focusing and defocusing), as well as graded-index media characterized by exponential and linear permittivity profiles, are obtained.

Conclusions. The proposed theory supports a visual description in an explicit analytical form of a narrowly localized light beam within such waveguides. It is shown that by combining different semiconductor crystals in a composite waveguide, it is possible to obtain a nonlinear optical layer on one side of the waveguide interface and a layer with a graded-index dielectric permittivity profile on the other.

Keywords: nonlinear optics, nonlinear waves, optical nonlinearity, Kerr nonlinearity, optical waveguide, graded-index waveguide

• **Submitted:** 13.12.2022 • **Revised:** 17.02.2023 • **Accepted:** 18.05.2023

For citation: Savotchenko S.E. Models of waveguides combining gradient and nonlinear optical layers. *Russ. Technol. J.* 2023;11(4):84–93. https://doi.org/10.32362/2500-316X-2023-11-4-84-93

Financial disclosure: The author has no a financial or property interest in any material or method mentioned.

The author declares no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Модели волноводов, сочетающих градиентные и нелинейно-оптические слои

С.Е. Савотченко [@]

Белгородский государственный технологический университет им. В.Г. Шухова, Белгород, 308012 Россия
[@] *Автор для переписки, e-mail: savotchenkose@mail.ru*

Резюме

Цели. Теоретические исследования волноводных свойств границ раздела нелинейно-оптических и градиентных сред являются важными для использования в оптоэлектронике. Комбинированные волноводы, сочетающие слои с различными оптическими свойствами, представляются наиболее перспективными, поскольку для них можно подобрать оптимальные характеристики с помощью широкого ряда управляющих параметров. Цель работы – разработка теории композитных оптически-нелинейных градиентных волноводов с произвольным профилем, в рамках которой возможно получение точных аналитических выражений для поверхностных волн и волноводных мод в явном виде. Основной особенностью предлагаемой в данной работе теории является то, что она применима для описания поверхностных волн и волноводных мод, поле в которых сосредоточено внутри градиентного слоя и не выходит за его границу, не контактирующую с нелинейным слоем.

Методы. Использованы аналитические методы теории оптических волноводов, нелинейной оптики.

Результаты. Проведено теоретическое описание волноводных свойств границы раздела двух сред с принципиально различными оптическими характеристиками. Сформулированная модель плоского волновода применима для сред с произвольным распределением пространственного профиля диэлектрической проницаемости. Получено аналитическое выражение, описывающее поверхностную волну, распространяющуюся вдоль границы раздела среды со ступенчатой нелинейностью и градиентного слоя с произвольным профилем диэлектрической проницаемости. Также получены аналитические выражения для поверхностных волн, распространяющихся вдоль границы раздела среды с керровской нелинейностью (как самофокусирующей, так и дефокусирующей) с градиентными средами, характеризующимися экспоненциальным и линейным профилями диэлектрической проницаемости.

Выводы. Предложенная теория позволяет наглядно описать в явном аналитическом виде узко локализованные световые потоки в таких волноводах. Показано, что сочетание различных полупроводниковых кристаллов в композитном волноводе позволяет получить с одной стороны от волноведущего интерфейса нелинейно-оптический слой, а с другой – слой с градиентным профилем диэлектрической проницаемости.

Ключевые слова: нелинейная оптика, нелинейные волны, оптическая нелинейность, керровская нелинейность, оптический волновод, градиентный волновод

• **Поступила:** 13.12.2022 • **Доработана:** 17.02.2023 • **Принята к опубликованию:** 18.05.2023

Для цитирования: Савотченко С.Е. Модели волноводов, сочетающих градиентные и нелинейно-оптические слои. *Russ. Technol. J.* 2023;11(4):84–93. https://doi.org/10.32362/2500-316X-2023-11-4-84-93

Прозрачность финансовой деятельности: Автор не имеет финансовой заинтересованности в представленных материалах или методах.

Автор заявляет об отсутствии конфликта интересов.

INTRODUCTION

Waveguide structures are typically based on differences in the optical properties of their layers [1, 2]. In optoelectronics, waveguides whose layers have constant refractive index values [3], those having a variable spatial distribution profile (graded-index) [4], as well as those providing a nonlinear optical response when the refractive index depends on the light flux intensity [5], are widely used. Combined waveguides combining such layers appear to be the most promising for optoelectronics, since optimal characteristics can be selected for them using a wide range of control parameters [6, 7].

In recent years, theoretical studies on the waveguide properties of the interfaces between nonlinear-optical and graded-index media have intensified [8, 9]. In particular, waveguide modes having a layered structure consisting of a gradient layer with linear [10, 11] and exponential [12, 13] profiles in contact with the nonlinear Kerr medium have been obtained. Our recent works present the results of theoretical studies on the waveguide properties of structures combining pairs of different gradient and nonlinear media. In particular, linear [14–19], parabolic [20, 21], and exponential profiles [22, 23] of the refractive index/dielectric permittivity of media in contact with media of different nonlinearities, such as stepwise [14, 22], Kerr [20] and its generalizations [16–18, 20, 21], as well as those of a photorefractive diffusive type [21, 23], are considered. Symmetric three-layer structures consisting of a linear-gradient layer in a medium with Kerr nonlinearity [24] and in a medium with photorefractive nonlinearity [25] are also described.

The paper proposes the theoretical description of surface waves propagating along a flat media interface, one of which has a nonlinear response, while the other is characterized by a spatial gradient of dielectric permittivity. The novelty of the work lies in the proposed theory of such composite optically nonlinear graded-index waveguides with an arbitrary profile, within the framework of which analytical expressions for surface waves can be obtained. The main feature of this theory is its applicability to the description of surface waves and waveguide modes, in which the field is concentrated inside the gradient layer and does not go beyond its boundary not contacting with a nonlinear layer. The proposed theory permits a clear description in an explicit analytical form of the narrowly localized light fluxes in such waveguides. Examples of semiconductor materials that could be used for designing waveguides of this type are also considered.

1. THEORY AND RESEARCH METHODS

A flat interface between two nonmagnetic media of fundamentally different optical types is considered as a simple waveguide structure. In particular, a medium with a smooth change in the spatial profile of the refractive index/dielectric permittivity (i.e., $n = n(x)$ or $\varepsilon = \varepsilon(x)$, respectively; x being the spatial coordinate in the direction perpendicular to the contact plane $x = 0$) is on one side of the boundary, and a non-linear optical medium wherein the refractive index/permittivity depends on light intensity I (i.e., $n = n(I)$ or $\varepsilon = \varepsilon(I)$, $I = |E|^2$; E being the amplitude of electric field strength) is on the other side. It is assumed that the considered media have no dielectric losses; consequently, all their optical characteristics are valid values.

The origin is located in the yz plane such that the x-axis is perpendicular to the interface located in the $x = 0$ plane. We shall consider a transverse wave propagating along the interface with a non-zero component of the electric field strength: $E_y = \psi(x)\exp(i\beta z - i\omega t)$, where $\beta = kn$ is the propagation constant; $n = ck/\omega$ is the effective refractive index; c is the speed of light in vacuum; ω is a frequency; $k = 2\pi/\lambda$ is a wave number; λ is a wavelength; $\psi(x)$ is the spatial distribution of electric field strength in the direction transverse to the interface. As is well known [1, 2], function $\psi(x)$ obeys the stationary equation (magnetic permeability is assumed to be equal to one):

$$\psi''(x) + \{\varepsilon(x, I) - n^2\}k^2\psi(x) = 0, \quad (1)$$

where the dielectric permittivity of the waveguide system is written in the following form:

$$\varepsilon(x, I) = \begin{cases} \varepsilon_G(x), x < 0, \\ \varepsilon_N(I), x > 0. \end{cases} \quad (2)$$

For definiteness, the graded-index medium is located in the left half-space and characterized by dielectric permittivity $\varepsilon_G(x)$, while the optically nonlinear medium is located in the right half-space and characterized by dielectric permittivity $\varepsilon_N(I)$.

If the transverse field distribution is represented as:

$$\psi(x) = \begin{cases} \psi_G(x), x < 0, \\ \psi_N(x), x > 0, \end{cases} \quad (3)$$

then, considering (2), Eq. (1) splits into the following two:

$$\psi_G''(x) + \{\varepsilon_G(x) - n^2\}k^2\psi_G(x) = 0, x < 0, \quad (4)$$

$$\psi_N''(x) + \{\varepsilon_N(I) - n^2\}k^2\psi_N(x) = 0, x > 0. \quad (5)$$

Equations (4) and (5) should be supplemented with boundary conditions describing the continuity requirement for the field components at the interface:

$$\psi_N(+0) = \psi_G(-0), \quad \psi'_N(+0) = \psi'_G(-0), \tag{6}$$

as well as boundedness at infinity: $|\psi(x)| \to 0$, $|x| \to \infty$.

Since the localization of the light beam energy along the waveguide layer represents an important task, the focus is on such waves for whom the depth of field penetration is less than the thickness of the gradient layer. It has been shown in a number of our works [14–22] that, in certain intervals of the waveguide system parameters, the field amplitude at distances of the order of the gradient layer thickness may be much smaller than the amplitude at the interface and can therefore be neglected. Consequently, there is no need to use the boundary conditions at the distance of the gradient layer thickness in this case; the boundary conditions at the interface are sufficient for use when describing narrowly localized surface waves.

The solution of Eq. (4) may be written in the following form:

$$\psi_G(x) = \psi_0 \frac{F(g(x))}{F(g(0))}, \tag{7}$$

where ψ_0—field amplitude at the interface; $F(g)$—special function with an auxiliary argument $g(x)$ solving Eq. (4) analytically for a given permittivity profile $\varepsilon_G(x)$ and satisfying the following equation:

$$F''g' + F'g'' + \{\varepsilon_G(x) - n^2\}k^2F = 0. \tag{8}$$

The explicit form of dependence $g(x)$ is related to the substitution of variables when reducing Eq. (8) to the form for which it is known that the exact solution is expressed through some special function. For example, in the case of linear replacement $g(x) = ax + b$ in (8), it would be $g' = a$ and $g'' = 0$; then, it would be simplified as $aF'' + \{\varepsilon_G(x) - n^2\}k^2F = 0$.

When choosing to write the solution in the form (7), it is necessary to consider the requirement at infinity: $|F(g(x))| \to 0$, $|x| \to \infty$.

We shall briefly summarize the basic provisions of the theory of the composite waveguides under consideration. The waveguide comprises an ultrathin boundary separating the medium with a dielectric permittivity graded-index and a nonlinear optical medium. The transverse wave is localized so that the field does not extend beyond the boundary of the gradient layer. The electric field distribution in the direction transverse to the interface is determined from the stationary one-dimensional nonlinear equation to satisfy the conjunction condition at the interface and the disappearance at infinity condition.

The explicit form of the solution of Eq. (5) depends on selecting the nonlinearity model of the medium. We shall further consider the waves arising in such waveguides with profiles narrowly localized along the interface of the media using the examples of Kerr and stepwise nonlinearities and the graded-index exponential profile.

2. RESULTS AND DISCUSSION

2.1. Kerr nonlinearity

The most common form of nonlinear response of an optical system is the Kerr response, for which the dielectric permittivity depends linearly on the light intensity:

$$\varepsilon_N(I) = \varepsilon_{0N} + \alpha I, \tag{9}$$

where ε_{0N}—unperturbed dielectric constant (positive); α—Kerr nonlinearity coefficient whose positive and negative value corresponds to the self-focusing medium and defocusing one, respectively.

Then, considering dielectric permittivity (9), Eq. (5) may be written in the following form:

$$\psi''_N(x) - q_N^2\psi_N(x) + \alpha k^2\psi_N^3(x) = 0, \tag{10}$$

where $q_N^2 = k^2(n^2 - \varepsilon_{0N})$.

The solution of nonlinear Eq. (10) at $n^2 > \varepsilon_{0N}$, satisfying the condition $|\psi(x)| \to 0$, $|x| \to \infty$, may be written as follows:

$$\psi_N(x) = \begin{cases} \sqrt{\frac{2}{\alpha}} \cdot \frac{q_N}{k\mathrm{ch}q_N(x - x_N)}, & \alpha > 0, \\ \sqrt{\frac{2}{|\alpha|}} \cdot \frac{q_N}{k\mathrm{sh}q_N(x - x_N)}, & \alpha < 0, \end{cases} \tag{12}$$

where the x_N value characterizes the position of the "soliton" center and is determined from the boundary conditions (6). Substituting (7) and (12) into (6), the following may be written:

$$x_N = \begin{cases} \frac{1}{k\sqrt{n^2 - \varepsilon_{0N}}} \cdot \mathrm{arctg}\left(\frac{\varepsilon_{G\,eff}}{n^2 - \varepsilon_{0N}}\right)^{1/2}, & \alpha > 0, \\ \frac{1}{k\sqrt{n^2 - \varepsilon_{0N}}} \cdot \mathrm{arcctg}\left(\frac{\varepsilon_{G\,eff}}{n^2 - \varepsilon_{0N}}\right)^{1/2}, & \alpha < 0, \end{cases} \tag{13}$$

where the effective permittivity of the gradient layer is introduced, as follows:

$$\varepsilon_{G\,eff} = \frac{F'(g(0))}{F(g(0))} \cdot \frac{g'(0)}{k}. \tag{14}$$

Similarly, after substituting (7) and (12) into (6), the field intensity at the interface is obtained as follows:

$$I_0 = \psi_0^2 = \begin{cases} 2(n^2 - \varepsilon_{0N} - \varepsilon_{G\,eff})/\alpha, & \alpha > 0, \\ 2(\varepsilon_{G\,eff} + \varepsilon_{0N} - n^2)/|\alpha|, & \alpha < 0. \end{cases} \quad (15)$$

As an example, the exponential dielectric permittivity profile is considered:

$$\varepsilon_G(x) = \varepsilon_e + (\varepsilon_0 - \varepsilon_e)e^{2x/h}, \quad (16)$$

where ε_0 and ε_e are dielectric constants (positive) at the interface and at the end of the gradient layer of characteristic thickness h, respectively.

Substituting the profile (16) into Eq. (4), the following may be written:

$$\psi_G''(x) + (V/a)^2 \exp(2x/a) - q^2\}\psi_G(x) = 0, \quad (17)$$

where $V^2 = a^2k^2(\varepsilon_0 - \varepsilon_e)$, $q^2 = k^2(n^2 - \varepsilon_e)$.

In this case, $F(g(x)) = J_\nu(g(x))$, where $J_\nu(g)$ is the Bessel function of the first kind, $g(x) = Ve^{x/a}$; then Eq. (7) defining the solution of Eq. (17) at $n^2 > \varepsilon_e$ and satisfying the condition $|\psi(x)| \to 0$, $x \to -\infty$ may be written in the following form:

$$\psi_G(x) = \psi_0 J_{aq}(Ve^{x/a})/J_{aq}(V). \quad (18)$$

The effective permittivity of the exponential gradient layer is written as follows:

$$\varepsilon_{G\,eff} = \frac{J'_{aq}(V)}{J_{aq}(V)} \cdot \frac{V}{ak}. \quad (19)$$

Thus, using Eqs. (3), (12), and (18), the wave propagating along the interface of the self-focusing Kerr nonlinear medium and the exponential gradient layer may be written in the following form:

$$\psi(x) = \begin{cases} \sqrt{\frac{2(n^2 - \varepsilon_{0N} - \varepsilon_{G\,eff})}{\alpha}} \cdot \frac{J_{aq}(Ve^{x/a})}{J_{aq}(V)}, & x < 0, \\ \sqrt{\frac{2}{\alpha}} \cdot \frac{q_N}{k\mathrm{ch}q_N(x - x_N)}, & x > 0, \end{cases} \quad (20)$$

where the dielectric permittivity of the exponential gradient layer is defined by Eq. (19).

As a second example, the linear dielectric permittivity profile is considered:

$$\varepsilon_G(x) = \varepsilon_0 + (\varepsilon_0 - \varepsilon_e)(x/h). \quad (21)$$

where h is the layer thickness in the case of a linear function.

Substituting profile (21) into Eq. (4), the following may be written:

$$\psi_G''(x) + \{\varepsilon_0 - n^2 + (\varepsilon_0 - \varepsilon_e)(x/h)\}k^2\psi_G(x) = 0. \quad (22)$$

In this case, $F(g(x)) = \mathrm{Ai}(g(x))$, where $\mathrm{Ai}(g)$ is the Airy function of the first kind, $g(x) = -x/x_G + \delta$ and $x_G = \{a/k_0^2(\varepsilon_0 - \varepsilon_e)\}^{1/3}$, $\delta = -(\varepsilon_0 - n^2)h/x_G(\varepsilon_0 - \varepsilon_e)$. Then Eq. (7) defining the solution of Eq. (22) at $\varepsilon_e < n^2 < \varepsilon_0$ and satisfying the condition $|\psi(x)| \to 0$, $x \to -\infty$ may be written in the following form:

$$\psi_G(x) = \psi_0 \mathrm{Ai}(-x/x_L + \delta)/\mathrm{Ai}(\delta). \quad (23)$$

The effective permittivity of the linear gradient layer is written as follows:

$$\varepsilon_{G\,eff} = -\frac{1}{kx_L} \cdot \frac{\mathrm{Ai}'(\delta)}{\mathrm{Ai}(\delta)}. \quad (24)$$

Thus, using Eqs. (3), (12), and (24), the wave localized along the interface of the self-focusing Kerr nonlinear medium and the linear gradient layer may be written in the following form:

$$\psi(x) = \begin{cases} \sqrt{\frac{2(n^2 - \varepsilon_{0N} - \varepsilon_{G\,eff})}{\alpha}} \cdot \frac{\mathrm{Ai}(-x/x_L + \delta)}{\mathrm{Ai}(\delta)}, & x < 0, \\ \sqrt{\frac{2}{\alpha}} \cdot \frac{q_N}{k\mathrm{ch}q_N(x - x_N)}, & x > 0, \end{cases} \quad (25)$$

where the dielectric permittivity of the exponential gradient layer is determined by Eq. (24).

2.2. Stepwise nonlinearity

Let now a nonlinear medium whose dielectric permittivity is described by the step function (the model of the simplest nonlinear medium [26] or the "sharp step" model [27]) contact the gradient layer:

$$\varepsilon_N(|E|) = \begin{cases} \varepsilon_1, |E| < E_s, \\ \varepsilon_2, |E| > E_s, \end{cases} \quad (26)$$

where E_s is the field value; when reached, the abrupt (instantaneous) switching from one value of the dielectric constant ε_1 to another ε_2 ($\varepsilon_2 > \varepsilon_1$) occurs.

Thus, near the contact in the nonlinear medium where $|E| > E_s$, there is a region (near-surface domain) of width x_s in which the dielectric constant has value ε_2, while outside of it where $|E| < E_s$, the dielectric constant has value ε_1. The position of the boundary of the near-surface domain x_s is determined by the following conditions:

$$\psi_N(x_s + 0) = \psi_N(x_s - 0) = E_s, \quad \psi'_N(x_s + 0) = \psi'_N(x_s - 0). \quad (27)$$

As shown in [26], in the stepwise nonlinearity model, Eq. (5) with permittivity (26) decomposes into two:

$$\psi_N''(x) - (n^2 - \varepsilon_1)k^2\psi_N(x) = 0, \ |E| < E_s, \quad (28)$$

$$\psi_N''(x) + (\varepsilon_2 - n^2)k^2\psi_N(x) = 0, \ |E| > E_s. \quad (29)$$

The solution of Eq. (28) at $n^2 > \varepsilon_1$ is written as follows:

$$\psi_N(x) = E_s e^{-q_1(x-x_s)}, \quad (30)$$

where $q_1^2 = (n^2 - \varepsilon_1)k^2$, while the solution of Eq. (28) at $n^2 < \varepsilon_2$ is written in the following form:

$$\psi_N(x) = \psi_0 \cos(p_2(x - x_m))/\cos(p_2 x_m), \quad (31)$$

where $p_2^2 = (\varepsilon_2 - n^2)k^2$, while values x_s, x_m are determined from the boundary conditions.

Substituting solutions (7), (30), and (31) into boundary conditions (6) and (27), the following may be written:

$$x_m = \frac{1}{k\sqrt{\varepsilon_2 - n^2}} \cdot \operatorname{arctg}\left(\frac{\varepsilon_{G\,eff}}{\varepsilon_2 - n^2}\right)^{1/2}, \quad (32)$$

$$x_s = x_m + q_1 / p_2^2, \quad (33)$$

$$I_0 = \psi_0^2 = E_s^2 \cdot \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_{G\,eff} + \varepsilon_2 - n^2}, \quad (34)$$

where the dielectric permittivity of the exponential gradient layer is determined by Eq. (19).

Thus, the surface wave propagating along the interface between the medium with stepwise nonlinearity and a gradient layer with an arbitrary profile may be written in the following form:

$$\psi(x) = E_s \cdot \begin{cases} \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_{G\,eff} + \varepsilon_2 - n^2}\right)^{1/2} \frac{F(g(x))}{F(g(0))}, & x < 0, \\ \left(\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 - n^2}\right)^{1/2} \cos(p_2(x - x_m)), & 0 < x < x_s, \\ e^{-q_1(x-x_s)}, & x > x_s, \end{cases} \quad (35)$$

where the position of the near-surface domain boundary is determined by Eq. (33).

In [22], the case of contact between a medium with abrupt nonlinearity and a medium with an exponential dielectric permittivity profile is considered. In this work, the impact of optical parameters of the corresponding waveguide structures on the wave profiles and their controlled localization is analyzed (and illustrated) in detail.

2.3. Discussion

The resulting Eq. (20) for surface waves differs from the one given in [14]. Wave (20) can propagate at the effective refractive index (propagation constant) varying arbitrarily in the admissible range of values. In [14], the authors present and analyze the dispersion equation linking the effective refractive index and the optical characteristics of the waveguide, i.e., its value is fixed at given values of the waveguide characteristics. In particular, they analyze the dependence of the effective refractive index n on the nonlinearity parameter αI_0 being the product of the Kerr nonlinearity coefficient on the field intensity at the media interface.

As is known, the propagation constant is related to the angle of incidence of the beam exciting the surface wave, which can vary under experimental conditions. Therefore, such a parameter should be considered as controlling, i.e., varying in experiments on the selected waveguide structure. In the work, Eq. (15) determining the dependence of the nonlinearity parameter αI_0 on the effective refractive index n is obtained. In our opinion, this dependence is more practical from the experimental point of view.

The difference between the obtained Eq. (25) describing the wave narrowly localized along the interface of a self-focusing Kerr nonlinear medium and the linear gradient layer should also be emphasized. The surface wave profile obtained in [10, 12] transverse to the interface of such media can go beyond the gradient layer. The developed theory is suitable for describing waves whose spatial intensity distribution is completely concentrated in the gradient layer and does not go beyond its limits.

The following are examples of materials on which composite waveguides combining nonlinear optical layers and gradient layers with a spatial distribution of optical characteristics can be based.

In nonlinear optical crystals, localization of the light beam occurs due to the nonlinear response of the medium. Under certain conditions, field localization occurs not only in a medium with self-focusing nonlinearity, but also with defocusing nonlinearity [28].

The dielectric permittivity—or square of the refractive index—depends on the square of the electric field strength, i.e., linearly on the light intensity (Kerr nonlinearity), which is characteristic of KDP

and $LiNbO_3$ crystals at certain wavelengths. This form of nonlinear response is observed in multilayer Co/TiO_2 nanocomposite films in a wavelength range of 400–1000 nm at temperatures of 10–50°C [29]. In [30], the Kerr effect strengthening in cobalt-based thin films in the same wavelength range is noted.

It is indicated in [31, 32] that a high-intensity light beam can change the optical characteristics of crystals in narrow regions along the direction of its propagation. As a result, the formation of a near-surface layer with optical characteristics different from the rest of the crystal is observed. Such changes are caused by the nonlinear response of the crystal depending on the intensity of the light beam propagating along its surface. The theoretical description of such phenomena is based on models such as "sharp" stepwise nonlinearity [26, 33], "smooth" stepwise nonlinearity [27], and saturable nonlinearity in various formulations [34, 35].

As noted in Section 2.2, the model of "sharp" stepwise nonlinearity describes the change in dielectric permittivity abruptly from one constant value to another when the intensity of the light beam reaches a certain threshold value. Within the framework of such a model, many authors have been able to obtain in explicit analytical form the results applied to the theoretical description of the features of surface wave propagation [26], self-reflection [33], self-localized optical pulses [27], and optical bistabilities [36]. This dependence can be considered as a limiting case of the model of "smooth" stepwise nonlinearity in the case of a sharp increase in dielectric permittivity at a small increase in the light beam intensity.

It is shown in [37] that the change in the dielectric permittivity of a semiconductor with exciton-exciton interaction in a certain spectral range may occur quite abruptly. The authors [38, 39] noted that such behavior can be observed in semiconductor crystals such as CdS and CdSe crystals with low biexciton binding energy within the range of approximately 0.5–3.0 MeV. The authors explain this phenomenon by the fact that coherent photons passing through the semiconductor film excite coherent excitons with the same wave vector- and phase values as the photons. This optical interaction results in biexcitons determining both the polarization of the crystal and the concentration of quasiparticles responsible for its optical properties. Optical nonlinear effects can be observed at relatively low intensities of the incident light beam.

It should be noted that some media under intensive illumination undergo rapid changes in their optical properties: in particular, in semiconductor-doped CdSSe and Schott OG 550 glasses [40, 41], ion-doped $GdAlO_3:Cr^{3+}$ crystals [42], and thin films formed from the photochromic protein bacteriorhodopsin [43]. A practically discontinuous change of the refractive index from one value to another depending on the intensity of the light beam was observed in them in the limit at small relaxation times.

On the other hand, many semiconductor heterostructures used in modern optoelectronics exhibit a dependence of the refractive index or dielectric function on the spatial distance [3]. In particular, the refractive indices of GaAs/GaAlAs [44, 45], InGaAs/InAlAs [46], and InGaAsP/InP [47] semiconductor photonic crystals are described by spatially graded-index profiles. Therefore, these materials can be referred to as optically graded-index media. It should be noted that the use of gallium arsenide, as well as heterostructures based on it such as Ga1–*x*Al*x*As, Ga1–*x*In*x*As and Ga1–*x*Al*x*N, Ga1–*x*In*x*N, seems very promising due to their enhanced radiation resistance [48] in comparison with other crystals used in semiconductor optoelectronics.

Spatial distributions of refractive indices are often created by implanting ions into glasses, which induce ion-exchange processes creating stress induction near the surface due to the large difference between the ionic radii of the exchanged ions. For example, graded refractive index profiles are formed in BK7 glasses by introducing K^+-Na^+ ions and in lime-sodium glass by introducing Ag^+-Na^+ ions [49]. It should be noted that the profiles obtained by thermal diffusion of metal ions into the glass substrate are close to the exponential profile [50].

Thus, by combining different semiconductor crystals in a composite waveguide, it becomes possible to obtain a nonlinear optical layer on one side of the waveguide interface and a layer with a graded dielectric permittivity profile on the other.

CONCLUSIONS

In the present work, a theoretical description of waveguide properties of the interface between two media with fundamentally different optical characteristics is presented. The formulated flat waveguide model is applicable to media having an arbitrary distribution of the spatial dielectric permittivity profile.

An analytical expression describing the surface wave propagating along the interface between the medium with a stepwise nonlinearity and the gradient layer with an arbitrary dielectric permittivity profile is obtained. Analytical expressions for surface waves propagating along the interface of the medium with Kerr nonlinearity (both self-focusing and defocusing) with graded-index media characterized by exponential and linear dielectric permittivity profiles are also presented.

The above analysis of materials shows that it is possible to select semiconductor crystals on which composite waveguides combining nonlinear optical layers and gradient layers with spatial distribution of optical characteristics may be based.

The results obtained in this work may be of value in designing elements of optical devices based on the possibility to control the localization of light beams along the waveguiding surfaces of the interface of contacting media.

Acknowledgments

The study was performed using the equipment of the High Technology Center at the V.G. Shukhov Belgorod State Technological University.

REFERENCES

1. Adams M.J. *An Introduction to Optical Waveguides*. Chichester: Wiley; 1981. 401 p.
2. Chen C.-L. *Foundations for Guided-Wave Optics*. New York: John Wiley & Sons Inc.; 2005. 462 p. https://doi.org/10.1002/0470042222
3. Dragoman D., Dragoman M. *Advanced Optoelectronic devices*. Berlin: Springer; 1999. 424 p.
4. Bednarik M., Cervenka M. Electromagnetic waves in graded-index planar waveguides. *J. Opt. Soc. Am. B*. 2020;37(12):3631–3643. https://doi.org/10.1364/JOSAB.408679
5. Čada M., Qasymeh M., Pištora J. Optical Wave Propagation in Kerr Media. In: *Wave Propagation. Theories and Applications*. London: IntechOpen; 2013. P. 175–192. http://doi.org/10.5772/51293
6. Shvartsburg A.B., Maradudin A. *Waves in Gradient Metamaterials*. Singapore: World Scientific; 2013. 339 p. https://doi.org/10.1142/8649
7. Al-Bader S.J., Jamid H.A. Graded-index optical waveguides with nonlinear cladding. *J. Opt. Soc. Am. A*. 1988;5(3):374–379. https://doi.org/10.1364/JOSAA.5.000374
8. Taya S.A., Kullab H.M., Qadoura I.M. Dispersion properties of slab waveguides with double negative material guiding layer and nonlinear substrate. *J. Opt. Soc. Am. B*. 2013;30(7):2008–2013. https://doi.org/10.1364/JOSAB.30.002008
9. Almawgani A.H.M., Taya S.A., Hussein A.J., Colak I. Dispersion properties of a slab waveguide with a graded-index core layer and a nonlinear cladding using the WKB approximation method. *J. Opt. Soc. Am. B*. 2022;39(6):1606–1613. https://doi.org/10.1364/JOSAB.458569
10. Hussein A.J., Nassar Z.M., Taya S.A. Dispersion properties of slab waveguides with a linear graded-index film and a nonlinear substrate. *Microsyst. Technol*. 2021;27(7):2589–2594. https://doi.org/10.1007/s00542-020-05016-z
11. Taya S.A., Hussein A.J., Colak I. An exact solution of a slab waveguide dispersion relation with a linear graded-index guiding layer (TM case). *Microsyst Technol*. 2022;28(22):1213–1219. https://doi.org/10.1007/s00542-022-05281-0
12. Taya S.A., Hussein A.J., Ramahi O.M., Colak I., Chaouche Y.B. Dispersion curves of a slab waveguide with a nonlinear covering medium and an exponential graded-index thin film (transverse magnetic case). *J. Opt. Soc. Am. B*. 2021;38(11):3237–3243. https://doi.org/10.1364/JOSAB.439034
13. Hussein A.J., Taya S.A., Vigneswaran D., Udiayakumar R., Upadhyay A., Anwar T., Amiri I.S. Universal dispersion curves of a planar waveguide with an exponential graded-index guiding layer and a nonlinear cladding. *Results in Physics*. 2021;20:103734. https://doi.org/10.1016/j.rinp.2020.103734
14. Savotchenko S.E. The surface waves propagating along the contact between the layer with the constant gradient of refractive index and photorefractive crystal. *J. Opt.* 2022;24(4):045501. https://doi.org/10.1088/2040-8986/ac51e9
15. Savotchenko S.E. The composite planar waveguide structure consisting of the linearly graded-index layer and the nonlinear layer formed with an increasing the electric field. *Optik*. 2022;252:168542. https://doi.org/10.1016/j.ijleo.2021.168542
16. Savotchenko S.E. Light localization in a linearly graded-index substrate covered by intensity dependent nonlinear self-focusing cladding. *J. Opt*. 2022;24(6):065503. https://doi.org/10.1088/2040-8986/ac6bab
17. Savotchenko S.E. Discrete spectrum of waveguide modes of a linearly graded-index film introduced into a medium with a stepwise nonlinearity. *Optik*. 2023;281(6):170835. https://doi.org/10.1016/j.ijleo.2023.170835
18. Savotchenko S.E. Guided waves in a graded-index substrate covered by an intensity-dependent defocusing nonlinear medium. *Appl. Phys. B: Lasers and Optics*. 2022;128(8):153. https://doi.org/10.1007/s00340-022-07872-1
19. Savotchenko S.E. Nonlinear surface waves propagating along the contact between the graded-index layer and the medium with near surface layer where Kerr nonlinearity disappears with increasing light intensity. *Optik*. 2023;272:170373. https://doi.org/10.1016/j.ijleo.2022.170373
20. Savotchenko S.E. Surface waves propagating along the interface between a parabolic graded-index medium and a self-focusing nonlinear medium: exact analytical solution. *J. Opt*. 2022;24(10):105501. https://doi.org/10.1088/2040-8986/ac8e80
21. Savotchenko S.E. Surface waves propagating along the interface between parabolic graded-index medium and photorefractive crystal with diffusion nonlinearity. *Phys. B: Condensed Matter*. 2023;648(2):414434. https://doi.org/10.1016/j.physb.2022.414434
22. Savotchenko S.E. Surface waves propagating along the interface separating an exponential graded-index medium and the medium with a step change in the dielectric constant. *Optik*. 2022;271(12):170092. https://doi.org/10.1016/j.ijleo.2022.170092
23. Savotchenko S.E. Waveguide properties of interface separating a photorefractive crystal with diffusion nonlinearity and an exponential graded-index medium. *Phys. Lett. A*. 2022;455(12):128516. https://doi.org/10.1016/j.physleta.2022.128516
24. Savotchenko S.E. New types of transverse electric nonlinear waves propagating along a linearly graded-index layer in a medium with Kerr nonlinearity. *Opt. Quant. Electron*. 2023;55(1):74. https://doi.org/10.1007/s11082-022-04323-1

25. Savotchenko S.E. Temperature controlled waveguide properties of the linearly graded-index film in semiconductor crystal with the photorefractive nonlinearity. *Appl. Phys. B: Lasers and Optics*. 2023;129(1):7. https://doi.org/10.1007/s00340-022-07950-4
26. Khadzhi P.I., Fedorov L.V., Torstveit S. Nonlinear surface waves for the simplest model of nonlinear medium. *Phys. Tech. Lett*. 1991;61:110–113.
27. Kaplan A.E. Multistable self-trapping of light and multistable soliton pulse propagation. *IEEE J. Quant. Electron.* 1985;21(9):1538–1543. https://doi.org/10.1109/JQE.1985.1072828
28. Kartashov Y.V., Malomed B.A., Torner L. Solitons in nonlinear lattices. *Rev. Mod. Phys*. 2011;83(1):247–305. http://doi.org/10.1103/RevModPhys.83.247
29. Laudyn U.A., Rutkowska K.A., Rutkowski R.T., Karpierz M.A., Woliński T.R., Wójcik J. Nonlinear effects in photonic crystal fibers filled with nematic liquid crystals. *Cent. Eur. J. Phys.* 2008;6(3):612–618. https://doi.org/10.2478/s11534-008-0096-z
30. Polyakov V.V., Polyakova K.P., Seredkin V.A., Patrin G.S. The enhanced magneto-optical Kerr effect in Co/TiO_2 multilayer films. *Tech. Phys. Lett.* 2012;38(10):921–923. https://doi.org/10.1134/S1063785012100227
31. Jarque E.C., Malyshev V.A. Nonlinear reflection from a dense saturable absorber: from stability to chaos. *Opt. Commun*. 1997;14291(3):66–70. https://doi.org/10.1016/S0030-4018(97)00275-7
32. Schuzgen A., Peyghambarian N., Hughes S. Doppler Shifted Self Reflection from a Semiconductor. *Phys. Stat. Sol.* (*B*). 1999;206(1):125–130. https://doi.org/10.1002/(SICI)1521-3951(199803)206:1<125::AID-PSSB125>3.0.CO;2-8
33. Lyakhomskaya K.D., Khadzhi P.I. Self-reflection effect in naïve model of nonlinear media. *Tech. Phys*. 2000;45(11):1457–1461. https://doi.org/10.1134/1.1325030 [Original Russian Text: Lyakhomskaya K.D., Khadzhi P.I. Self-reflection effect in naïve model of nonlinear media. *Zhurnal Tekhnicheskoi Fiziki.* 2000;70(11):86–90 (in Russ.).]
34. Christian J.M., McDonald G.S., Chamorro-Posada P. Bistable Helmholtz bright solitons in saturable materials. *J. Opt. Soc. Am. B*. 2009;26(12):2323–2330. https://doi.org/10.1364/JOSAB.26.002323
35. Korovai O.V. Nonlinear s-polarized quasi-surface waves in the symmetric structure with a metamaterial core. *Phys. Solid State.* 2015;57(7):1456–1462. https://doi.org/10.1134/S1063783415070197
36. Enns R.H., Rangnekar S.S., Kaplan A.E. Bistable-soliton pulse propagation: Stability aspect. *Phys. Rev. A*. 1987;36(3):1270–1279. https://doi.org/10.1103/PhysRevA.36.1270
37. Khadzhi P.I., Rusanov A.M., Gaivan S.L. Cavity-free optical bistability of a thin semiconductor film in the exciton region of the spectrum. 1999;29(6):539–541. https://doi.org/10.1070/QE1999v029n06ABEH001526
38. Khadzhi P.I., Gaivan S.L. Nonlinear interaction of an ultrashort light pulse with a thin semiconductor film under conditions of two-photon excitation of biexcitons. *Quantum Electron.* 1995;25(9):897–900. https://doi.org/10.1070/QE1995v025n09ABEH000497
39. Corovai A.V., Khadzhi P.I. Optical properties of a semiconductor upon two-photon excitation of biexcitons by a powerful pump pulse and one-photon probing in the M band. *Quantum Electron.* 2001;31(10):937–939. https://doi.org/10.1070/QE2001v031n10ABEH002080
40. Roussignol P., Ricard D., Flytzanis C. Nonlinear optical properties of commercial semiconductor-doped glasses. *Appl. Phys. A*. 1987;44:285–292. https://doi.org/10.1007/BF00624594
41. Vanhaudenarde A., Trespidi M., Frey R. Refractive-index changes during photodarkening in semiconductor-doped glasses. *J. Opt. Soc. Am. B.* 1994;11(8):1474–1479. https://doi.org/10.1364/JOSAB.11.001474
42. Catunda T., Cury L.A. Transverse self-phase modulation in ruby and $GdAlO_3:Cr^{+3}$ crystals. *J. Opt. Soc. Am. B*. 1990;7(8):1445–1455. https://doi.org/10.1364/JOSAB.7.001445
43. Wang S.Q., Wang X., Birge R., Downie J.D., Timucin D., Gary C. Propagation of a Gaussian beam in a bacteriorhodopsin film. *J. Opt. Soc. Am. B.* 1998;15(5):1602–1609. https://doi.org/10.1364/JOSAB.15.001602
44. Mendoza-Alvarez J.G., Nunes F.D., Patel N.B. Refractive index dependence on free carriers for GaAs. *J. Appl. Phys*. 1980;51(8):4365-4367. https://doi.org/10.1063/1.328298
45. Ravindran S., Datta A., Alameh K., Lee Y.T. GaAs based *long-wavelength* microring resonator optical switches utilising bias assisted carrier-injection induced refractive index change. *Opt. Express.* 2012;20(14):15610–15627. https://doi.org/10.1364/OE.20.015610
46. Zucker J.E., Chang T.Y., Wegener M., Sauer N.J., Jones K.L., Chemla D.S. Large refractive index changes in tunable-electron-density InGaAs/InAlAs quantum wells. *IEEE Photon. Technol. Lett*. 1990;2(1):29–31. https://doi.org/10.1109/68.47032
47. Ishida K., Nakamura H., Matsumura H. InGaAsP/InP optical switches using carrier induced refractive index change. *Appl. Phys. Lett*. 1987;50(3):141–142. https://doi.org/10.1063/1.97695
48. Vigdorovich E.N. Radiation resistance of epitaxial structures based on GaAs. *Russ. Technol. J*. 2019;7(3):41–49 (in Russ.). https://doi.org/10.32362/2500-316X-2019-7-3-41-49
49. Karasiński P., Rogoziński R. Influence of refractive profile shape on the distribution of modal attenuation in planar structures with absorption cover. *Opt. Commun*. 2007;269(1):76–88. https://doi.org/10.1016/j.optcom.2006.07.067
50. Shutyi A., Sementsov D., Kazakevich A.V., Sannikov D. Waveguide regimes of a graded-index planar waveguide with cladding. *Tech. Phys*. 1999;44(1):1329–1333. https://doi.org/10.1134/1.1259518

About the author

Sergey E. Savotchenko, Dr. Sci. (Phys.-Math.), Associate Professor, Professor, High Mathematics Department, V.G. Shukhov Belgorod State Technological University (46, Kostyukova ul., Belgorod, 308012 Russia). E-mail: savotchenkose@mail.ru. Scopus Author ID 6603577988, RSCI SPIN-code 2552-4344, https://orcid.org/0000-0002-7158-9145

Об авторе

Савотченко Сергей Евгеньевич, д.ф.-м.н, доцент, профессор кафедры высшей математики, ФГБОУ ВО «Белгородский государственный технологический университет им. В.Г. Шухова» (308012, Россия, Белгород, ул. Костюкова, д. 46). E-mail: savotchenkose@mail.ru. Scopus Author ID 6603577988, SPIN-код РИНЦ 2552-4344, https://orcid.org/0000-0002-7158-9145

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 621.391:53.08
https://doi.org/10.32362/2500-316X-2023-11-4-94-104

RESEARCH ARTICLE

Restoration of a blurred photographic image of a moving object obtained at the resolution limit

Victor B. Fedorov [@], Sergey G. Kharlamov, Anatoly I. Starikovskiy

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: feodorov@mirea.ru

Abstract

Objectives. When processing images of the Earth's surface obtained from satellites, the problem of restoring a blurry image of a moving object is of great practical importance. The aim of this work is to study the possibility of improving the quality of restoration of blurry images obtained at the limit of the resolution of the camera.

Methods. Digital signal processing methods informed by the theory of incorrect and ill-conditioned problems were used.

Results. The proposed method for restoring a blurred photographic image of a moving object differs from traditional approaches in that the discrete convolution equation, to which the problem of restoring a blurred image is reduced, is obtained by approximating the corresponding integral equation based on the Kotelnikov interpolation series rather than on the traditional basis of the quadrature formula. In the work, formulas are obtained for calculating the kernel of the convolution obtained using the Kotelnikov interpolation series. The discrete convolution inversion problem, which belongs to the class of ill-posed problems, requires regularization. Results of traditional approaches to restoring blurred images using the quadrature formula with Tikhonov regularization and the proposed method based on the Kotelnikov interpolation series are compared. Although the quality of the blurred image restoration is almost the same in both cases, in the quadrature formula the blur value is expressed as an integer number of pixels, while, when using the Kotelnikov series, this value can also be specified in fractions of a pixel.

Conclusions. The expediency of discretizing the convolution describing the image distortion of the blur type on the basis of the Kotelnikov interpolation series when processing a blurred image obtained at the limit of the resolution of the camera is demonstrated. In this case, the amount of blur can be expressed in fractions of a pixel. This situation typically arises when processing satellite photography of the Earth's surface.

Keywords: blurred image of a moving object, resolution, image restoration, Tikhonov regularization, inverse problem, Kotelnikov interpolation series, edge effect

• **Submitted:** 05.12.2022 • **Revised:** 09.03.2023 • **Accepted:** 06.06.2023

For citation: Fedorov V.B., Kharlamov S.G., Starikovskiy A.I. Restoration of a blurred photographic image of a moving object obtained at the resolution limit. *Russ. Technol. J.* 2023;11(4):94–104. https://doi.org/10.32362/2500-316X-2023-11-4-94-104

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Восстановление смазанного фотографического изображения движущегося объекта, получаемого на пределе разрешающей способности

В.Б. Федоров [@], **С.Г. Харламов**, **А.И. Стариковский**

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: feodorov@mirea.ru*

Резюме

Цели. Задача восстановления смазанного изображения движущегося объекта имеет большое практическое значение, в частности, при обработке изображений поверхности Земли, получаемых со спутников. Целью работы является исследование возможности повышения качества восстановления смазанных изображений, получаемых на пределе разрешающей способности фотоаппарата.

Методы. Использованы методы цифровой обработки сигналов, методы теории некорректных и плохо обусловленных задач.

Результаты. Предложен метод восстановления «смазанного» фотографического изображения движущегося объекта, отличающийся от традиционных подходов тем, что уравнение дискретной свертки, к решению которого сводится задача восстановления смазанного изображения, получается путем аппроксимации соответствующего интегрального уравнения на основе интерполяционного ряда Котельникова, а не на основе квадратурной формулы, как это делается традиционно. В работе получены формулы для вычисления ядра свертки, получаемой с применением интерполяционного ряда Котельникова. Как известно, задача обращения дискретной свертки относится к классу некорректных задач и требует регуляризации. Дано сравнение результатов восстановления смазанных изображений (с использованием регуляризации по Тихонову), осуществляемого как традиционным путем, т.е. с применением квадратурной формулы, так и предлагаемым способом, основывающимся на интерполяционном ряде Котельникова. Показано, что качество восстановления смазанного изображения в обоих случаях получается практически одинаковым. Однако использование квадратурной формулы предполагает, что величина «смаза» выражена целым числом пикселей, в то время как в случае использования ряда Котельникова эта величина может задаваться и долями пикселя.

Выводы. Показано, что дискретизацию свертки, описывающей искажение изображения типа «смаз», целесообразно осуществлять на основе интерполяционного ряда Котельникова в случае, когда осуществляется обработка смазанного изображения, получаемого на пределе разрешающей способности фотоаппарата. Это обусловлено тем, что в этом случае величина «смаза» может составлять доли пикселя. Такая ситуация характерна, например, для спутниковой фотосъемки поверхности Земли.

Ключевые слова: «смаз» изображения движущегося объекта, разрешающая способность, восстановление изображения, тихоновская регуляризация, обратная задача, интерполяционный ряд Котельникова, краевой эффект

• **Поступила:** 05.12.2022 • **Доработана:** 09.03.2023 • **Принята к опубликованию:** 06.06.2023

Для цитирования: Федоров В.Б., Харламов С.Г., Стариковский А.И. Восстановление смазанного фотографического изображения движущегося объекта, получаемого на пределе разрешающей способности. *Russ. Technol. J.* 2023;11(4):94–104. https://doi.org/10.32362/2500-316X-2023-11-4-94-104

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The problem of restoring a blurred image of a moving object has been quite well studied[1] [1–14]. Thus, in the case when the velocity of the moving object is known a priori, its solution is reduced to an inversion of the discrete convolution. This is obtained by approximating the corresponding integral convolution by replacing the integral with a quadrature formula (typically the trapezium quadrature formula). However, another possibility, proposed in this work, is based on the replacement of the integrand by its Kotelnikov interpolation series. In this case, the integrand function expresses the dependence of the brightness of the image point on its coordinates.

If we consider an image whose smallest details are significantly larger than the pixel size, then, when restoring it after blurring, the value of blurring can be set to the accuracy of the pixel size, i.e., expressed by an integer number of pixels. In this case, as we will show, the result of restoration is almost independent of the way in which the convolution was discretized. However, this is not the case when it is necessary to restore a blurred image obtained at the resolution limit of the camera, i.e., when the smallest details of the image are close to the pixel size. In this case, it may be necessary to express the amount of blur in fractions of a pixel in order to maximize resolution. In practice, this may imply the need to interpolate the blurred image prior to its restoration, i.e., to virtually represent it as if by smaller pixels. By doing this, it will be possible to express the exact value of blur by a whole number of these smaller pixels and then apply a traditional restoration method.

However, such approaches are associated with additional consumption of computer memory. For example, if the pixel size has to be reduced by a factor of 10, then the size of the corresponding image array will increase by a factor of 100.

The proposed method of integral convolution approximation based on the Kotelnikov interpolation series allows blur to be specified in fractions of a pixel: consequently, no interpolation of the image implying additional consumption of computer memory is required.

METHOD

Basic integral equation

A basic integral equation linking the brightness of points of the blurred image $Q(x, y)$ with the brightness of points of the restored image $P(x, y)$ has the form:

$$Q(x,y) = w^2 \int_0^T P(x - v_x t; y - v_y t)dt,$$

where (x, y) are Cartesian coordinates of the current point; (v_x, v_y) are Cartesian coordinates of the image velocity on the CCD matrix surface (by assumption all image points move with the same velocity); w is the pixel size; T is the exposure time [1–3].

If the vertical component of velocity is absent, then we have a Fredholm integral equation of the first kind of convolution type:

$$Q(x) = \frac{w^2}{v_x} \int_0^{v_x T} P(x - \xi)d\xi. \quad (1)$$

Discretization of the convolution based on the quadrature formula

By replacing the integral in Eq. (1) with the trapezium quadrature formula, we obtain

$$Q(x) \approx \frac{w^2}{v_x} \sum_{k=0}^{n} c_k P(x - kw),$$

where $c_0 = c_n = 0.5$, $c_1 = \ldots = c_{n-1} = 1$ are the coefficients of the quadrature formula; $n = v_x/w$ is the amount of image shift during exposure time, expressed as an integer number of pixels; $k \in \{0, \ldots, n\}$.

Or, since within a single pixel the brightness must be considered constant, we have

$$Q(m) \approx \sum_{k=0}^{\min\{n,m\}} \frac{w^2}{v_x} c_k P(m - k), \quad (2)$$

where m is the pixel number; $m \in \{0, \ldots, M - 1\}$; M is the number of pixels in one row, $k \in \{0, \ldots, m\}$; $Q[m] = Q(mw)$; $P[m] = P(mw)$.

Then the equality in (2) will be considered as exact.

Convolution (2) can also be written in matrix form

$$\mathbf{Q} = \mathbf{A} \cdot \mathbf{P}, \quad (3)$$

where, for example, if $M = 8$ and $n = 3$, the Toeplitz matrix $\mathbf{A}$ would have the form:

$$\mathbf{A} = \frac{w^2}{v_x} \begin{bmatrix} 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.5 & 1.0 & 1.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.5 & 1.0 & 1.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.5 & 1.0 & 1.0 & 0.5 \end{bmatrix}.$$

[1] Gruzman I.S., Kirichuk V.S., Kosykh V.P., Peretyagin G.I., Spector A.A. *Digital Image Processing in Information Systems*: Tutorial. Novosibirsk: NSTU; 2002. 352 p. (in Russ.).

Convolution discretization based on the Kotelnikov interpolation series

Assuming that the maximum spatial frequency in the Fourier transform of the image $P(x)$ does not exceed $1/2w$, the subintegral function in (1) can be represented by the Kotelnikov interpolation series:

$$P(x) = \sum_{k=-\infty}^{+\infty} P[k]sinc\left(\frac{x}{w} - k\right).$$

Inserting this expression into (1), we obtain

$$Q(x) = \frac{w^2}{v_x} \sum_{k=-\infty}^{+\infty} P[k] \int_0^{v_x T} sinc\left(\frac{x}{w} - \frac{\xi}{w} - k\right) d\xi$$

or

$$Q[m] = \sum_{k=-\infty}^{+\infty} P[k]a[m-k] = \sum_{k=-\infty}^{+\infty} a[k]P[m-k].$$

However, since the horizontal dimensions of the image are limited to M pixels, we arrive at an approximate equality (which we will further assume to be exact):

$$Q[m] \approx \sum_{k=0}^{M-1} P[k]a[m-k], \quad (4)$$

where the kernel of this discrete convolution is defined by the formula

$$a[k] = \frac{w^2}{v_x \pi}\left(Si\left(\pi\left(\frac{v_x T}{w} - k \right)\right) + Si(\pi k)\right). \quad (5)$$

Corresponding diagrams are shown in Fig. 1.

As we can see from the above diagrams, the convolution kernel (4) approaches the convolution kernel (2) as the amount of image displacement for the exposure time increases. In the case of convolution (4), the shift value can be expressed not only by an integer number of pixels, but also by fractions of a pixel. In the case of convolution (2), however, this value is always expressed as an integer number of pixels.

Convolution (4) can also be written in matrix form:

$$\mathbf{Q} = \mathbf{A} \cdot \mathbf{P}. \quad (6)$$

However, here, unlike case (3), the Toeplitz matrix **A** will no longer be a lower triangular matrix (it will not have zero diagonals).

Simulation of a blurred image (direct problem solving)

For numerical experiments on blurred image restoration, it is necessary to have the images blurred in the right way. It is possible to obtain such images using relations (2) or (4), considering function $P[m]$ as given and function $Q[m]$ as unknown. The mentioned relations can be reduced to a cyclic convolution, which is effectively computed (and reversed) on the basis of the fast Fourier transform (FFT). For example, Figs. 2 and 3 show the original image and the horizontally blurred 20 pixels image obtained in this way, respectively.

When simulating a blurred image, the essential point is the presence or absence of transient edges in the blurred image. In the horizontally blurred image shown in Fig. 3, such transient edges (left and right) are present. Each of these transient edges comprises a strip of width

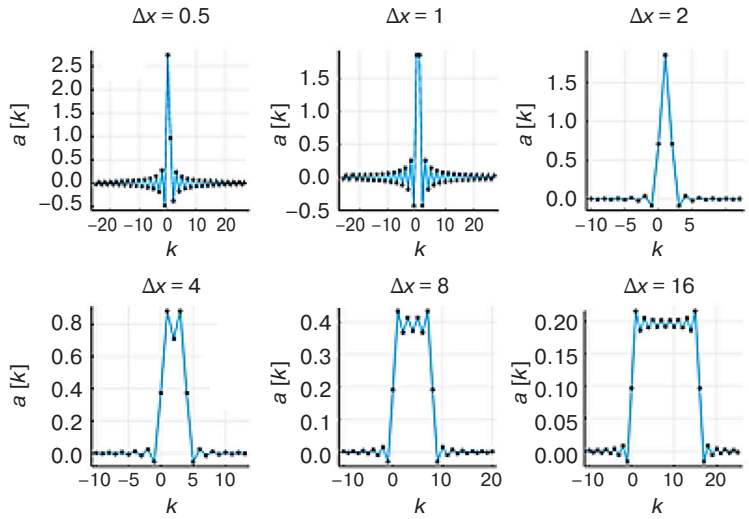


Fig. 1. Diagram of the kernel of the discrete convolution (4) obtained using the Kotelnikov interpolation theorem and corresponding to different values of the image blur $\Delta x = vxT/w$

Fig. 2. Source image

Fig. 3. Simulated blurred image with uncropped edges (horizontal blur size—20 pixels)

equal to the amount of image displacement, whose brightness gradually decreases to zero.

However, since the prototype of the obtained image is like an infinite ribbon with no edges at both ends, the real blurred photographic image will have no such edges. Therefore, in order to conform the obtained blurred image to reality, its edges should be cropped.

RESULTS

Results of blurred image restoration without the application of regularization

Reversal of the simulated blurred image, as well as obtaining the solution of the direct problem, is feasible on the basis of FFT. Thus, a discrete convolution of the form (2) or (4) can be represented as a cyclic convolution by augmenting each of the finite sequences included in the convolution with the required number of zeros

$$\mathbf{Q}[m] = \sum_{k=0}^{M-1} \mathbf{P}[k]\mathbf{a}[m-k], \qquad (7)$$

where $\mathbf{Q}$, $\mathbf{P}$, and $\mathbf{a}$ are arrays of length equal to length($\mathbf{P}$) + + length($\mathbf{a}$) − 1 obtained from the corresponding arrays Q, P, a (included in (2) and (4)) by adding the required number of zeros.

Then, to reverse this convolution a so-called inverse filter can be used, whose effect in the frequency domain is as follows:

$$\mathcal{F}[\mathbf{P}] = \frac{1}{\mathcal{F}[\mathbf{a}]}\mathcal{F}[\mathbf{Q}],$$

where $\mathcal{F}[\cdot]$ denotes the discrete Fourier transform; $1/\mathcal{F}[\mathbf{a}]$ is the so-called transfer function of the inverse filter.

If in this case the transient edges of the simulated blurred image are not cut off, the restoration result will be almost perfect even without the use of regularization. An example of an image restored in this way is shown in Fig. 4. Experiments show that, in order to obtain this result, the length of tails of the Kotelnikov kernel should be not less than 3 in cases where the edges of a blurred image $\Delta x = \frac{v_x T}{w} = 20$ remain uncropped. The length of tails refers to the number of counts of the kernel with negative indices, which is also equal to the number of counts with indices greater than Δx.

However, as already noted, the real blurred image is distinguished by the fact that it has no transient edges. If, in accordance with reality, the resulting transient edges are cut off, the result of restoration (without regularization) will be extremely poor. A corresponding example of the result of restoration is shown in Fig. 5.

It is possible to try to correct the situation by performing a preliminary restoration of the cropped edges. Such restoration can simply entail the addition of two vertical bars equal to the width of the blur to the left and right of the cropped image, whose pixel brightness of the horizontal rows will increase/decrease (by

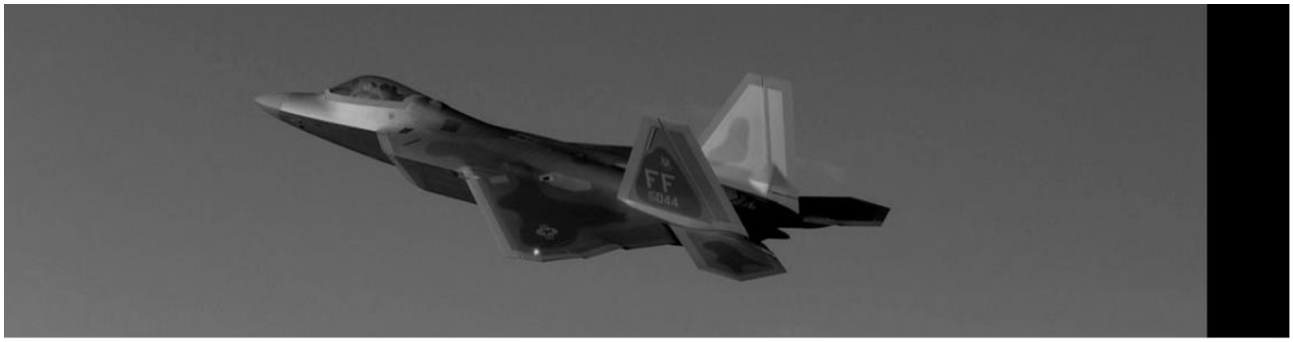


Fig. 4. Blurred image with uncropped edges restored with an inverse filter using the Kotelnikov kernel and without regularization

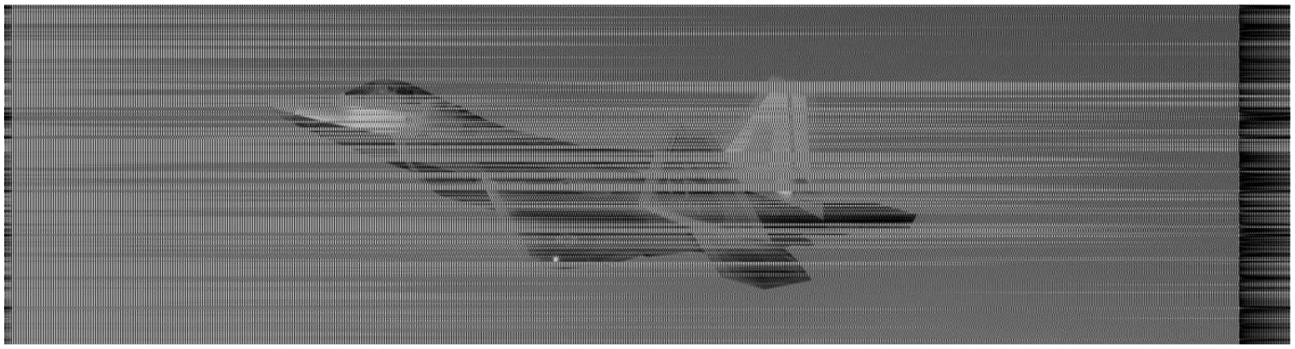


Fig. 5. Blurred image with uncropped edges restored with an inverse filter using the trapezium quadrature formula and no regularization

continuity) according to a linear law. Figure 6 shows an example of a simulated blurred image with cropped transient edges, while Figs. 7 and 8 depict the results of its restoration after the restoration of the cropped edges.

As can be seen, the result of restoration by reversing the discrete convolution with the Kotelnikov kernel turned out to be much better. However, as shown by numerical experiments, the quality of the restoration significantly depends on how many counts in the tails of the Kotelnikov kernel, defined by formula (5), are taken into account. In this case, only counts with indices from 0 to 20 (with a blur value of 20 pixels) were taken into account, i.e., the tails of the kernel were completely discarded. In general, we cannot claim that the treatment of the discrete convolution with the Kotelnikov kernel without the application of regularization in all cases guarantees a satisfactory result.

As can be seen from the examples, although edge restoration slightly improves the result of image restoration, its quality is still generally unsatisfactory. The reason is that the inversion of discrete convolutions (2) or (4) can be considered as a solution of a system of linear algebraic equations (SLAE) of the form (3) or (6) respectively.

Errors of restoration (and moreover the complete absence of such restoration) should be considered as

Fig. 6. Simulated blurred image with cropped transient edges at a horizontal blur value of 20 pixels

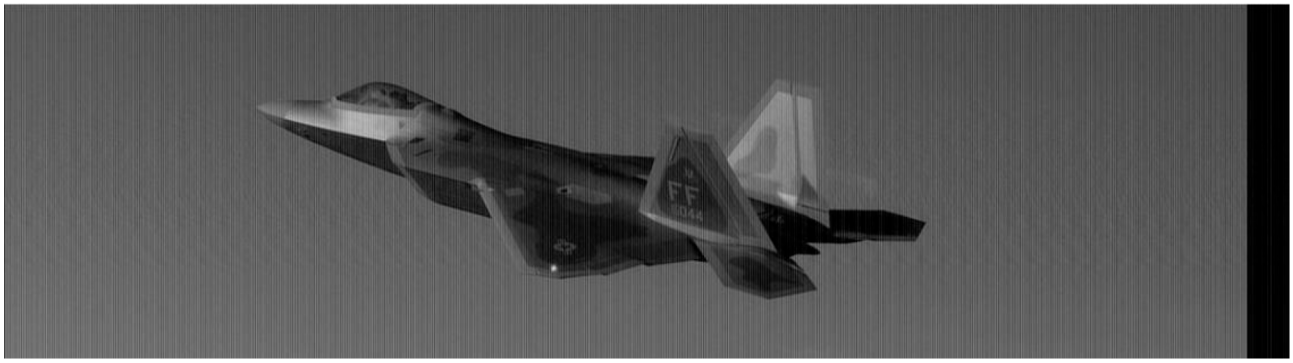


Fig. 7. Blurred image with cropped edges restored with an inverse filter using the Kotelnikov kernel and no regularization

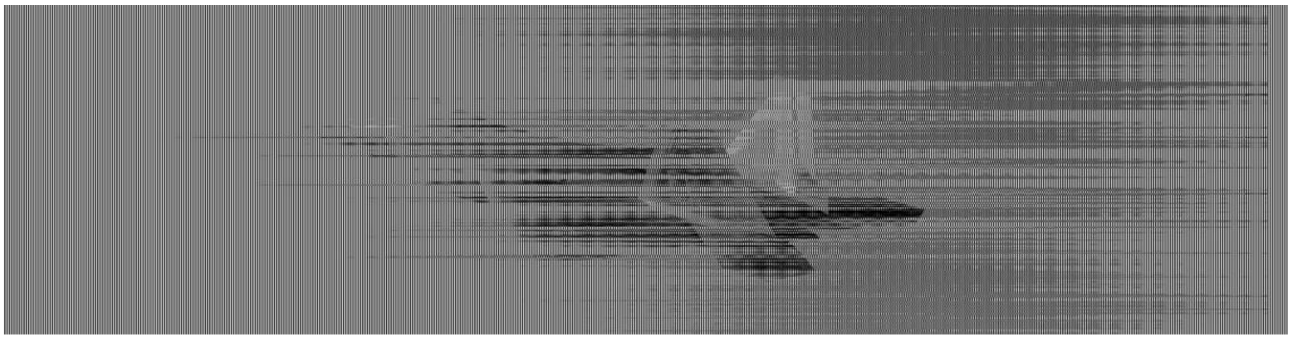


Fig. 8. Blurred image with cropped edges restored with an inverse filter using the trapezium quadrature formula and no regularization

errors in the setting of initial data (i.e. errors in several extreme values $Q[0]$, …, $Q[M-1]$). The estimate of the relative error of the obtained solution is known to be proportional to the relative error of the right part of SLAE with the proportionality coefficient equal to the conditionality number of SLAE matrix. This number, whose value depends both on the matrix size and on the value of blur, is typically quite large. Thus, even a small relative error of initial data can lead to very significant relative errors of the obtained solution. To mitigate this phenomenon, the Tikhonov regularization method is usually used in the convolution reversal [1–3].

Results of blurred image restoration by inverse filtering with Tikhonov regularization

The image restoration problem is known to be very sensitive to errors of initial data (the solution of the Fredholm 1-grade integral equation (1) belongs to the class of incorrectly posed problems). Therefore, in the presence of any significant errors in the initial data, the convolution reversal requires the application of regularization.

The transfer function of the Tikhonov regularized discrete inverse filter has the form:

$$\frac{A[m]}{|A[m]|^2 + \alpha W^{2p}[m]},$$

where $A = \mathcal{F}[\mathbf{a}]$, W^{2p} is the stabilizing function of the order $p = 0, 1, 2, \ldots$; α is the regularization parameter ($\alpha \geq 0$, with $\alpha = 0$ there is no regularization; this parameter can be chosen, for example, experimentally); array a is defined in Eq. (7) [1–3].

Stabilizing function of zero order is $W^0 = \mathcal{F}([1, 0, \ldots, 0])$, i.e., identically equal to 1. The first-order regularizing function is defined as $W^2 = \mathcal{F}([1, -2, 1, 0, \ldots, 0])$, i.e., the Fourier transform of the 2nd-order finite-difference filter. Regularizing functions of higher orders are the degrees of W^2. The results of restoration with edge restoration and application of regularization of horizontally blurred image by 20 pixels are shown in Figs. 9 and 10.

As we can see from the obtained results, the quality of the blurred image restoration is equally good when using the quadrature formula as in the case of using the Kotelnikov series.

Results of numerical experiments on image restoration at blur values expressed in fractions of a pixel

Let us now consider the situation when we need to restore a blurred image obtained at the resolution limit of the camera, i.e., when the instantaneous images of the smallest details of the moving image are close to the size of a pixel. In this case, we will consider a test image

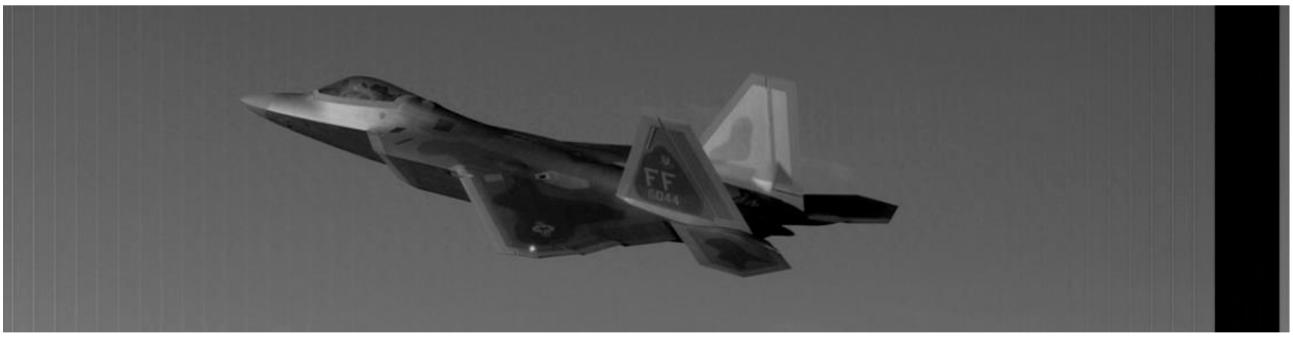


Fig. 9. Blurred image after cropped edge restoration, restored via inverse filter using the Kotelnikov kernel ($\alpha = 10^{-3}$, $p = 1$)

Fig. 10. Blurred image after cropped edge restoration, restored using an inverse filter using the trapezium quadrature formula ($\alpha = 10^{-2}$, $p = 1$)

consisting of parallel strips of decreasing width down to the size of one pixel (in the image space) as a model of the subject. Figure 11 shows a momentary image of such a test image (taking into account the finiteness of pixel size of the camera CCD matrix). Figure 12 shows a horizontally blurred image by 5.5 pixels of the same test image.

Figure 13 shows the restored blurred image with uncropped edges using the inverse filter using the Kotelnikov kernel and no regularization. To obtain the best restoration result for a given value of blur of the image, the tail length of the Kotelnikov kernel should be chosen experimentally. Thus, the result shown in Fig. 13 is obtained with a tail length equal to 2.

As we can see from comparison of Fig. 14 and 15, the contrast of the restored blurred picture of the test image using the Kotelnikov kernel is higher than when using the quadrature formula. This is not surprising, since the value of blur is equal to 5.5 pixels: when using the quadrature formula, only the shift by an integer number of pixels can be taken into account. Thus, in case of using the quadrature formula, the inverse filter has been adjusted to a blur value equal not to 5.5, but to 6 (because when using the quadrature formula there is no possibility to consider the fractional part of the blur value).

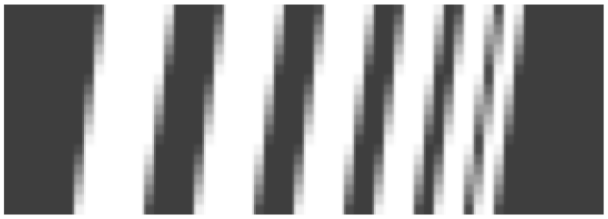


Fig. 11. Instantaneous (unblurred) picture of the test image, obtained at the resolution limit

Fig. 12. Horizontally blurred image of the world by 5.5 pixels (the edges are not cropped)

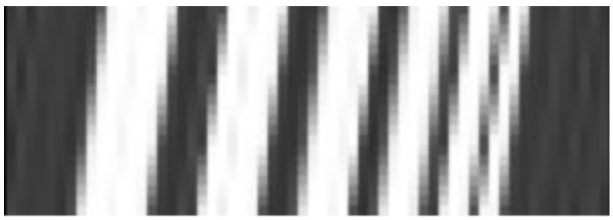


Fig. 13. Blurred picture of a test image with uncropped edges, restored with an inverse filter using the Kotelnikov kernel and no regularization

Fig. 14. Result of the cropped blurred picture restoration of the test image with the inverse filter with the Kotelnikov kernel with preliminary restoration of the cropped edges ($\alpha = 10^{-3}$, $p = 1$)

Fig. 15. Result of the cropped blurred picture restoration of the test image with uncropped edges using the inverse filter using the trapezoid quadrature formula with preliminary restoration of the cropped edges ($\alpha = 10^{-2}$, $p = 1$)

CONCLUSIONS

Results of numerical experiments show that, in cases when the blur value is expressed by a whole number of pixels, the proposed approach based on the Kotelnikov kernel works as well as the traditional technique based on the use of quadrature formulas. However, the use of the Kotelnikov series will have an advantage if the blur value of the image is expressed in fractions of pixels, because in this case, to fully account for a priori information about the blur value, there is no need to interpolate the original image, as would be the case with quadrature formulas.

Thus, if blur to be eliminated is a fraction of a pixel in size, the use of the Kotelnikov kernel is preferable when processing a blurred image obtained at the resolution limit of the camera. Such a situation is typical, for example, when processing satellite photography of the Earth's surface.

Authors' contribution. All authors equally contributed to the present work.

REFERENCES

1. Sizikov V., Dovgan A. Reconstruction of images smeared uniformly and non-uniformly. *CEUR Workshop Proceedings*. 2019;2344: paper 2. Available from URL: https://ceur-ws.org/Vol-2344/paper2.pdf
2. Sizikov V.S. *Pryamye i obratnye zadachi vosstanovleniya izobrazhenii, spektroskopii i tomografii s MatLab* (*Direct and Inverse Problems of Image Reconstruction, Spectroscopy and Tomography with MATLAB*). St. Petersburg: Lan'; 2011. 410 p. (in Russ.). ISBN 978-5-8114-2754-3
3. Vasilenko G.I., Taratorin A.M. *Vosstanovlenie izobrazhenii* (*Image Recovery*). Moscow: Radio i svyaz'; 1986. 302 p. (in Russ.).
4. Bakushinskii A.B., Goncharskii A.V. *Nekorrektnye zadachi. Chislennye metody i prilozheniya* (*Incorrect Tasks. Numerical Methods and Applications*). Moscow: Moscow University Press; 1989. 199 p. (in Russ.).
5. Bates R., McDonnell M. *Vosstanovlenie i rekonstruktsiya izobrazhenii* (*Image Restoration and Reconstruction*): transl. from Engl. Moscow: Mir; 1989. 336 p. (in Russ.). [Bates R., McDonnell M. *Image Restoration and Reconstruction.* NY: Oxford University Press; 1986. 312 p.]
6. Gonzalez R., Woods R. *Tsifrovaya obrabotka izobrazhenii* (*Digital Image Processing*). Moscow: Tekhnosfera; 2012. 1104 p. (in Russ.). [Gonzalez R., Woods R. *Digital Image Processing.* Prentice Hall; 2008. 954 p.]

СПИСОК ЛИТЕРАТУРЫ

1. Sizikov V., Dovgan A. Reconstruction of images smeared uniformly and non-uniformly. *CEUR Workshop Proceedings*. 2019;2344: paper 2. URL: https://ceur-ws.org/Vol-2344/paper2.pdf
2. Сизиков В.С. *Прямые и обратные задачи восстановления изображений, спектроскопии и томографии с MatLab*. СПб.: Лань; 2011. 410 с. ISBN 978-5-8114-2754-3
3. Василенко Г.И., Тараторин А.М. *Восстановление изображений*. М.: Радио и связь; 1986. 302 с.
4. Бакушинский А.Б., Гончарский А.В. *Некорректные задачи. Численные методы и приложения*. М.: Изд-во МГУ; 1989. 199 с.
5. Бейтс Р., Мак-Доннелл М. *Восстановление и реконструкция изображений*: пер с англ. М.: Мир; 1989. 336 с.
6. Гонсалес Р., Вудс Р. *Цифровая обработка изображений*. М.: Техносфера; 2012. 1104 с.
7. Гончарский А.В., Леонов А.С., Ягола А.Г. Методы решения интегральных уравнений Фредгольма 1-го рода типа свертки. В: *Некоторые вопросы автоматизированной обработки и интерпретации физических экспериментов*. Вып. 1. М.: Изд-во МГУ. 1973. С. 170–191.
8. Медофф Б.П. Реконструкция изображений по ограниченным данным: Теория и применение в компьютерной томографии. В: *Реконструкция изображений*; под ред. Г. Старка. М.: Мир; 1992. С. 384–436.

7. Goncharskii A.V., Leonov A.S., Yagola A.G. Methods for solving Fredholm integral equations of the 1st kind of convolution type. In: *Some Problems of Automated Processing and Interpretation of Physical Experiments*. V. 1. Moscow: Moscow University Press; 1973. P. 170–191 (in Russ.).
8. Medoff B.P. Image Reconstruction from Limited Data: Theory and Application in Computed Tomography. In: G. Stark (Ed.). *Image Reconstruction*. Moscow: Mir; 1992. P. 384–436 (in Russ.).
9. Tikhonov A.N., Goncharskii A.V., Stepanov V.V., Yagola A.G. *Regulyariziruyushchie algoritmy i apriornaya informatsiya* (*Regularizing Algorithms and A Priori Information*). Moscow: Nauka; 1983. 198 p. (in Russ.).
10. Aref'eva M.V., Sysoev A.F. Fast regularizing algorithms for digital image restoration. In: *Vychislitel'nye metody i programmirovanie* (*Numerical Methods and Programming*). 1983. Is. 39. P. 40–55 (in Russ.).
11. Vasilenko G.I. *Teopiya vosstanovleniya signalov: O reduktsii k ideal'nomu priboru v fizike i tekhnike* (*Theory of Signal Recovery: On the Reduction to an Ideal Device in Physics and Technology*). Moscow: Sovetskoe Radio; 1979. 272 p. (in Russ.).
12. Tikhonov A.N., Arsenin V.Ya. *Metody resheniya nekorrektnykh zadach* (*Methods for Solving Ill-Posed Problems*). Moscow: URSS; 2022. 288 p. (in Russ.). ISBN 978-5-9710-9341-1
13. Tikhonov A.N., Goncharskii A.V., Stepanov V.V. Inverse problems of photo image processing. In: Tikhonov A.N., Goncharskii A.V. (Eds.). *Incorrect Problems of Natural Sciences*. Moscow: Moscow University Press; 1987. P. 185–195 (in Russ.).
14. Russ J.C. *The Image Processing Handbook*. Boca Raton: CRC Press; 2007. 852 p.

9. Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г. *Регуляризирующие алгоритмы и априорная информация*. М.: Наука; 1983. 198 с.
10. Арефьева М.В., Сысоев А.Ф. Быстрые регуляризирующие алгоритмы цифрового восстановления изображений. В: *Вычислительные методы и программирование: сборник работ науч.-исслед. центра МГУ*. 1983. Вып. 39. С. 40–55.
11. Василенко Г.И. *Теория восстановления сигналов: О редукции к идеальному прибору в физике и технике*. М.: Сов. радио; 1979. 272 с.
12. Тихонов А.Н., Арсенин В.Я. *Методы решения некорректных задач*. М.: URSS; 2022. 288 с. ISBN 978-5-9710-9341-1
13. Тихонов А.Н., Гончарский А.В, Степанов В.В. Обратные задачи обработки фотоизображений. В: *Некорректные задачи естествознания*; под ред. А.Н. Тихонова, А.В. Гончарского. М.: Изд-во МГУ; 1987. С. 185–195.
14. Russ J.C. *The Image Processing Handbook*. Boca Raton: CRC Press; 2007. 852 p.

About the authors

Victor B. Fedorov, Cand. Sci. (Eng.), Associate Professor, Department of Higher Mathematics, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: feodorov@mirea.ru. Scopus Author ID 57208924592, RSCI SPIN-code 2622-7666, https://orcid.org/0000-0003-1011-5453

Sergey G. Kharlamov, Master Student, Department of Higher Mathematics, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: serhar2000@mail.ru. https://orcid.org/0000-0003-4470-6323

Anatoly I. Starikovskiy, Cand. Sci. (Eng.), Associate Professor, Professor, Department of Radio Electronic Systems and Complexes, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: starikovski@mirea.ru. Scopus Author ID 57208926243, ResearcherID AAH-2239-2020, RSCI SPIN-code 1126-4471, https://orcid.org/0000-0003-4040-3843

Об авторах

Федоров Виктор Борисович, к.т.н., доцент, кафедра высшей математики Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: feodorov@mirea.ru. Scopus Author ID 57208924592, SPIN-код РИНЦ 2622-7666, https://orcid.org/0000-0003-1011-5453

Харламов Сергей Григорьевич, магистрант, кафедра высшей математики Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: serhar2000@mail.ru. https://orcid.org/0000-0003-4470-6323

Стариковский Анатолий Иванович, к.т.н., доцент, профессор кафедры радиоэлектронных систем и комплексов Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: starikovski@mirea.ru. Scopus Author ID 57208926243, ResearcherID AAH-2239-2020, SPIN-код РИНЦ 1126-4471, https://orcid.org/0000-0003-4040-3843

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

UDC 658.51
https://doi.org/10.32362/2500-316X-2023-11-4-105-115

RESEARCH ARTICLE

Organization of an engineering center for industrial import substitution

Dmitry Kh. Mikhailidi [1], Alexander V. Ragutkin [2], Dmitry O. Skobelev [1, 2], Alexey B. Sukhaterin [2, @]

[1] *Environmental Industrial Policy Center, Moscow, 115054 Russia*
[2] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[@] *Corresponding author, e-mail: suhaterin@mirea.ru*

Abstract

Objectives. Following the imposition of sanctions against the Russian Federation, which included a ban on the supply of foreign electronic equipment—including automation systems—to Russian enterprises, the continuing development of science and technology in Russia became a question of ensuring technological sovereignty according to the principle of import substitution. According to plans developed by the Ministry of Industry and Trade of the Russian Federation, the policy of import substitution, including automation systems, will ensure the replacement of imported equipment with domestic counterparts.

Methods. Approaches underlying the joint project of MIREA – Russian Technological University and Environmental Industrial Policy Center to solve the problems of import substitution are described. Various substitution strategies available in the world experience, as well as objective and subjective obstacles to their implementation in Russia, including the insufficiency of domestic regulatory legal acts and previously formed attachments to imported technologies and regulatory frameworks, are considered. Distinctive features of contemporary external relations are adduced to the necessity and urgency of developing technological sovereignty. The main functional requirements for a software and hardware platform for developing modern automated control systems (ACS) for mechanical engineering applications, as well as the required capabilities of an engineering center for solving applied problems of overcoming import dependence, are described. The components of the production of capital goods (engineering) and its role in the product life cycle are shown.

Results. The selection of a pilot engineering object comprising a sectional glass-forming machine, along with a software-hardware complex including elements of industrial electronics and ACS, is justified. The main functional elements of the ACS and their interrelations are shown.

Conclusions. The results confirm the necessity of achieving complete import substitution for the creation of digital products. Prospects for cooperation with interested organizations are shown.

Keywords: product life cycle, technological sovereignty, import substitution, mechanical engineering, software and hardware platform, glass-forming machine, automated control system, reverse engineering

• **Submitted:** 24.04.2023 • **Revised:** 08.06.2023 • **Accepted:** 21.06.2023

For citation: Mikhailidi D.Kh., Ragutkin A.V., Skobelev D.O., Sukhaterin A.B. Organization of an engineering center for industrial import substitution. *Russ. Technol. J.* 2023;11(4):105–115. https://doi.org/10.32362/2500-316X-2023-11-4-105-115

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Организация инжинирингового центра для импортозамещения в промышленности

Д.Х. Михайлиди [1],
А.В. Рагуткин [2],
Д.О. Скобелев [1, 2],
А.Б. Сухатерин [2, @]

[1] *Научно-исследовательский институт «Центр экологической и промышленной политики», Москва, 115054 Россия*
[2] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[@] *Автор для переписки, e-mail: suhaterin@mirea.ru*

Резюме

Цели. После введения санкций против Российской Федерации и запрета поставки иностранной электронной техники, в т.ч. систем автоматизации, российским предприятиям, важнейшей задачей развития науки и техники в России является обеспечение технологического суверенитета. Один из «кирпичиков» в фундаменте решения данной задачи – это импортозамещение. Согласно планам, разработанным Министерством промышленности и торговли Российской Федерации, курс на импортозамещение поможет произвести замену импортного оборудования отечественными аналогами.

Методы. Описываются подходы, положенные в основу совместного проекта РТУ МИРЭА и НИИ «ЦЭПП» по решению задач импортозамещения. Рассмотрены имеющиеся в мировом опыте стратегии замещения, а также объективные и субъективные препятствия для его проведения в России, среди которых недостаточная функциональность нормативно-правовых актов и сформированная привязанность к импортным технологиям и правилам. Показана особенность современных внешних взаимоотношений России как причина необходимости и срочности формирования технологического суверенитета. Описаны основные функциональные требования к программно-аппаратной платформе для построения современных автоматизированных систем управления (АСУ) для машиностроения, а также возможности инжинирингового центра для решения прикладных задач по преодолению импортозависимости. Показаны составные части производства средств производства (машиностроения) и его роль в жизненном цикле продукции.

Результаты. Обоснован выбор пилотного объекта инжиниринга – секционной стеклоформующей машины, предмета разработки – программно-аппаратного комплекса, включающего элементы промышленной электроники и АСУ, показаны основные функциональные элементы АСУ и возникающие между ними связи.

Выводы. Подтверждается, что в создании цифровых продуктов необходимо добиться полного импортозамещения. Представлены перспективы сотрудничества с заинтересованными организациями.

Ключевые слова: жизненный цикл продукции, технологический суверенитет, импортозамещение, машиностроение, программно-аппаратная платформа, стеклоформующая машина, автоматизированная система управления, реверс-инжиниринг

• **Поступила:** 24.04.2023 • **Доработана:** 08.06.2023 • **Принята к опубликованию:** 21.06.2023

Для цитирования: Михайлиди Д.Х., Рагуткин А.В., Скобелев Д.О., Сухатерин А.Б. Организация инжинирингового центра для импортозамещения в промышленности. *Russ. Technol. J.* 2023;11(4):105–115. https://doi.org/10.32362/2500-316X-2023-11-4-105-115

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

From the point of view of classical economics, the expenditure of resources on import substitution cannot be considered highly efficient. This is explained by the fact that the customer of the process will be forced to spend money to re-learn the existing mode of production.

The reasons prompting the Russian Federation to import substitution policy in 2023 were not so much fundamental, stemming from the opportunity to increase domestic economic potential, as situational, caused by sanctions, which created an existential threat to the state and its economy [1]. The prevailing environment of hostile or suspended relations with yesterday's economic partners dictates the need and urgency to localize the full range of products. At the same time, critical types of production activities need to be oriented exclusively to the domestic market, recreating production chains within the country [2].

In 2022, a comprehensive block of urgent measures was introduced in response to the sanctions policy, which, although contradicting the concept of import substitution, made it possible to avoid the risks of a catastrophic fall in production and consumption. Thus, while allowing parallel imports without the permission of rights holders and zeroing import duties merely changed the form of import dependency, they appear to have represented a necessary temporary solution, not only in terms of creating a window for the creation and consumption of domestic products, but also in terms of allowing the characteristics of domestically manufactured goods to be improved as compared with their imported equivalents. This provides a basis on which the imperative to create technological sovereignty can be formulated.

CAPITAL GOODS PRODUCTION AS THE BASIS OF THE TECHNOLOGICAL SOVEREIGNTY OF THE COUNTRY

Under contemporary conditions, a particular country generally specializes in the production of those products where it has an advantage in terms of its possession of factors of production [3]. Comparative cost theory [4, 5] considers that a country cannot have a competitive advantage across the entire spectrum of locally produced goods, which are also available on global markets. Under sanctions, however, the need for domestic production increases regardless of the presence or absence of competitive advantages.

An analysis of the industrial life cycle diagram depicted in Fig. 1 shows that the sector of manufacture of technical means of production or mechanical engineering, which has an impact on the whole chain of transformations (redistribution) of the substance—from the extraction of raw materials to the disposal of the product. Obviously, synchronization of multivector substitution is required—manufacture of the means of production and consumption items. Ultimately, the need for consumption should ensure the effectiveness of import substitution of the technical means of production.

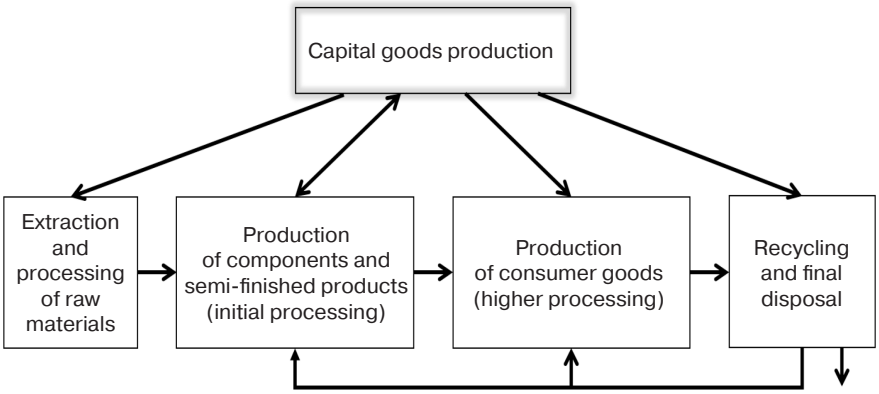


Fig. 1. Diagram of the industrial product lifecycle

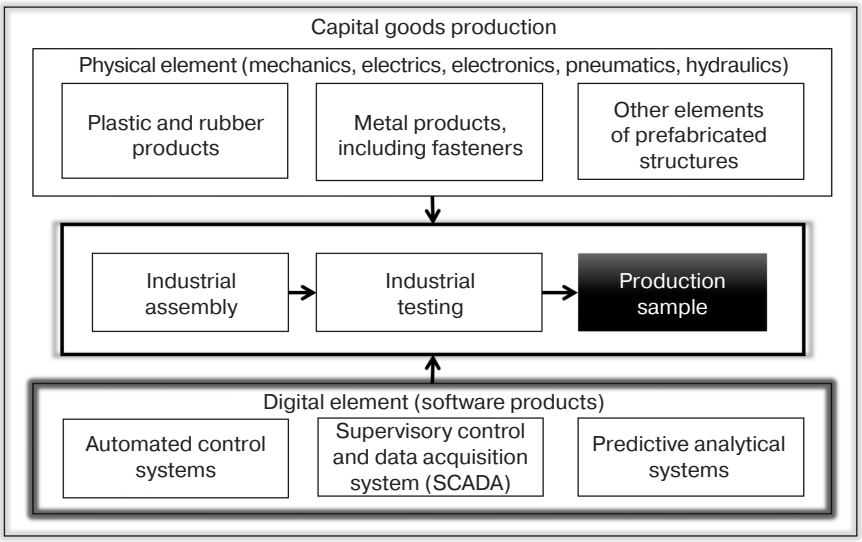


Fig. 2. Interaction of industrial elements in the production of capital goods

The mentioned sector consists of two industrial elements (Fig. 2): physical (machines, machining workstations) and digital (data collection, process control). The digital element is implemented in the physical by means of software and hardware platforms (SHP) to ensure the proper functionality of automatic control systems, relying on actuators (valves, servo drives, etc.). The supervisory control and data acquisition (SCADA) system is used to control complex production equipment, which centralizes the functions of actuators, as well as providing production statistics for automated control systems (ACS) of the technological process and for predictive analytical systems.

As well as constituting a national security goal, the development of mechanical engineering represents a fundamental condition for further growth and development of the country. Even taking into account the reduction already occurring over the past 8 years, the continuing dependence on imports for 50–90% of equipment components represents the main constraint on the development of domestic production and ensuing economic growth.[1,2,3,4,5] This determines import substitution priorities, without which the associated processes are likely to occur haphazardly and lacking proper efficiency [1].

An import substitution phase was experienced by the majority of newly industrialized countries during the period between the 1950s and 1970s [6]. This was concomitant with the economic assumption that the economic situation of commodity-exporting countries (developing countries) will tend to deteriorate if they do not resort to an import-substitution policy during the process of industrialization (Prebisch–Singer hypothesis [7]). While South American countries mainly concentrated their efforts on saturating domestic markets, Southeast Asia tended to be more export-focused. While the export strategy turned out to be more effective [8], this could only be achieved if the national currency is constantly weakened, which tends to lead to a decrease in the added value of the product, high domestic inflation, or restrictions on the growth of living standards [9].

In Russia, import substitution (where it was implemented) was based on an export strategy [10]. In November 2015, the State Council of the Russian Federation adopted basic decisions on this issue. As part of the recommended actions, it was proposed to:

- support the process of import substitution with financial resources, mainly through changes in taxation, as well as with funds from the Industrial Development Fund and the Federal Corporation for the Development of Small and Medium Business, including project financing and the provision of state guarantees;

[1] Order of the Ministry of Industry and Trade of the Russian Federation No. 2486 dated 07.07.2021. https://rulaws.ru/acts/Prikaz-Minpromtorga-Rossii-ot-07.07.2021-N-2486/. Accessed April 04, 2023 (in Russ.).

[2] Order of the Ministry of Industry and Trade of the Russian Federation No. 2913 dated 02.08.2021. https://rulaws.ru/acts/Prikaz-Minpromtorga-Rossii-ot-02.08.2021-N-2913/. Accessed April 04, 2023 (in Russ.).

[3] Order of the Ministry of Industry and Trade of the Russian Federation No. 3273 dated 20.08.2021. https://rulaws.ru/acts/Prikaz-Minpromtorga-Rossii-ot-20.08.2021-N-3273/. Accessed April 04, 2023 (in Russ.).

[4] Order of the Ministry of Industry and Trade of the Russian Federation No. 2882 dated 30.07.2021. https://rulaws.ru/acts/Prikaz-Minpromtorga-Rossii-ot-30.07.2021-N-2882/. Accessed April 04, 2023 (in Russ.).

[5] Order of the Ministry of Industry and Trade of the Russian Federation No. 2881 dated 30.07.2021. https://legalacts.ru/doc/prikaz-minpromtorga-rossii-ot-30072021-n-2881-ob-utverzhdenii/. Accessed April 04, 2023 (in Russ.).

- completely focus the system of public procurement on the purchase of domestic equipment, applying special incentives to stimulate Russian production facilities;
- create a system of centralized management of import substitution and related control, which would allow blocking the purchase of foreign products if Russian analogs are available.[6]

In pursuance of these tasks, several laws and regulations (L&R) were enacted, in particular, RF Government Decree No. 208[7] and RF Government Decree No. 209[8] dated February 18, 2022. While acknowledging their obvious advantages, we would like here to draw attention to some of their disadvantages. For example, the criterion of effectiveness stipulated in RF GD No. 209 is the number of created sets of design documentation, which cannot be said to reflect economic effectiveness. While RF GD No. 208 seems to offer a better mechanism, it lacks mechanisms for the vertical integration and aggregation of development objects according to various attributes; moreover, each component is developed by separate teams, which greatly increases the unit costs of product development as a whole [11].

Local development of software and electronics is supported in Russia. Import substitution programs are additionally implemented within the framework of the national project Small and Medium Entrepreneurship, where the division of efforts occurs in two directions: deployment of serial production, as well as the production of unique equipment and the creation of new technologies facilitating the introduction of know-how and science-intensive capital goods [12].

The purpose of the enacted L&R is to support the deployment of domestic production to create products that replace imported analogs and surpass the latter in their consumer characteristics, as well as leading to the development of new technologies (Fig. 3) [13]. For the Russian economy, this concerns not only products, but also technologies and institutions (rules), whose import has largely determined the continuing dependence on global engineering development centers. This development strategy was imposed not only on industry, but also in education and science, whose development is tied to external assessments and standards: the result was the "implicit sanctions model" [1] implemented long before the beginning of the present confrontational era.

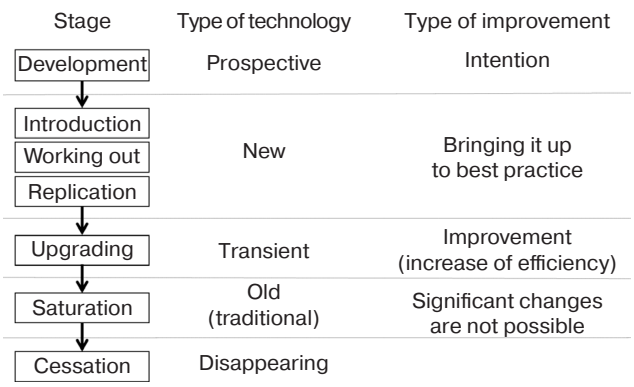


Fig. 3. Evolution of technologies according to O.S. Sukharev [14]

The formed attachments to imported technologies and standards represent one of the main obstacles to their replacement. Since this managerial effect is related to the core of technological content, domestic developers need to create such a core and put it into action in the face of fierce competition with the currently functioning—albeit at the expense of parallel imports—infrastructure [1]. As a rule, manufacturing industries try to exclude from their daily activities the risks associated with the need to improve skills, carry out research and development, and develop other time-consuming innovations. This explains the inertial reaction to the imposition of sanctions, since it is always easier to obtain parallel imports than to force elements of a complex system to localize technology and equipment.

In the context of import substitution, this represents a fundamental limitation. While technological substitution provides an opportunity to create your own product, it will continue to reduce the competitive potential of the individual industry and the economy as a whole if current patterns of economic orientation to the imported product are not altered. Therefore, centers of competence for overcoming import dependence become engineering centers, representing elements of innovative development [15].

Under contemporary conditions, the main focus of import substitution projects should be not be aimed at updating fixed assets, thus merely replicating the already achieved technological level, but instead focus on creating a basis for the rapid development and modernization of the real sector of the economy.

For most areas of manufacture of technical means of production, the creation of automatic and ACS

[6] State Council commissions discussed the idea of creating a unified service for import substitution. https://tass.ru/ekonomika/14848255. Accessed April 04, 2023 (in Russ.).

[7] On the granting of subsidies from the federal budget to the autonomous non-profit organization "Agency for Technological Development" to support projects involving the development of design documentation for components required for industries. Decree of the Government of the Russian Federation No. 208 dated 18.02.2022. http://static.government.ru/media/files/zISr7dzERAaYQY0N2HBwbN4FoBah9M6Y.pdf. Accessed April 04, 2023 (in Russ.).

[8] On providing grants in the form of subsidies from the federal budget for the implementation of projects to create and (or) develop engineering development centers on the basis of educational organizations of higher education and scientific organizations implementing projects related to the development of components. Decree of the Government of the Russian Federation No. 209 dated 18.02.2022. http://government.ru/docs/all/139438/. Accessed April 04, 2023 (in Russ.).

presents a significant challenge. The current uncertainty in maintaining the operability of production equipment in this area is a consequence of its having been characterized by its reliance on foreign microelectronic and software solutions over the past decades. Thus, the insufficient development of the Russian industrial controllers, peripherals and development tools means that problems of localization of mechanical engineering cannot be promptly solved.

The problem of creating domestic SHP solutions, comprising a set of hardware and software used to monitor and control production processes, maintain feedback, and actively influence the course of the process when it deviates from the declared parameters, as well as regulate and optimize the controlled process, is especially acute. Comprising a branched structure with distributed I/O and centralized signal processing, an SHP must ensure the physical interconnection of actuators and sensors of technological equipment. Depending on the type, the signal lines are connected to analog and discrete signal input/output modules, which are in turn connected to a programmable logic controller (PLC) via a fieldbus. To increase fault tolerance, two industrial PLC modules supporting hot redundancy can be used. In the event of a cable break or failure of one of the remote I/O devices, a minimum restoration time of the entire system should be required. Through industrial protocols such as Modbus, Industrial Ethernet, etc. The PLC provides information communication with the central control system, possibly via a fault protection system.

Integrated application development tools (e.g., *Codesys*, *ISAGRaf*, etc.) supporting all 5 IEC 61131-3 programming languages (LD, FBD, IL, ST, SFC)[9]. Thus, the PLC is responsible for receiving information from the sensors in real time through the industrial network, converting it and exchanging with other components of the automation system, as well as controlling actuators. Further data exchange between controllers and operator stations takes place via the backbone network. The operator level includes servers and user automated workstations for monitoring the operational technological and production process and, if necessary, sending commands to change parameters.

In mechanical engineering, SHP is implemented in the form of development and implementation of applied software and hardware complexes (SHC) with different degrees of universalization. The MIREA – Russian Technological University and Environmental Industrial Policy Center began a joint project to establish the functionality of such platform on the example of engineering the ACS of a glass-forming machine (GFM) (Fig. 4). Unlike the physical element of the equipment, where there are objective restrictions to achieving 100% independence (for example, natural rubber, comprising a component of many rubber products, cannot be produced in Russia for climatic reasons), the potential for import substitution of the digital element has almost no limitations and is recognized as a development priority for reasons of production and information security in compliance with the requirements of the best available technologies.

The object of engineering was selected on the basis of scientific experience, close relations with industry enterprises, first of all in the field of compliance with the best available technologies. In cooperation with one of the largest enterprises of the glass industry, an eight-section GFM became available, for which a reverse-engineering program is in the process of being developed (Fig. 5). Our potential joint development of the ACS GFM, which relates to import substitution in the field of mechanical engineering (manufacture of technical means of production), is on the priority list of industrial policy due to mechanical engineering comprising a fundamental condition for the growth and development of the country's economy.

ACS is typically considered in terms of a holistic solution for ensuring the automation of technological processes. The concept of "automated," as opposed to "automatic," emphasizes the need for human participation in individual operations, both in order to maintain control over the process and due to the complexity or inexpediency of automating individual operations. ACS ensures safety of operations by means of highly reliable alarms, interlocks and protections having a minimum response time, exact fulfillment of process regulations, elimination of erroneous actions of operating personnel, maintaining the reliable operation of equipment, and preventing emergency situations.

ACS level includes the development of a control algorithm that collects information from the primary sensors and actuators, processes the received information, automatically regulates (maintaining process parameters at the set value), manually and/or automatically controls electric drives and process equipment, transmits information from controllers and automation systems to the upper level. Industrial networks are used for information communication of all subsystems.

Within the framework of works on import substitution, the main goal consists in the development of ACS (software application part), implementation of the technical means of automation for the existing Russian fleet of sectional SFM based on the Russian element base and use of applied and system products from the register of Russian software[10]. The next step

[9] National Standard of the Russian Federation. *Programmable Controllers*. Part 3. *Programming languages*. https://docs.cntd.ru/document/1200135008. Accessed April 04, 2023 (in Russ.).

[10] https://reestr.digital.gov.ru/. Accessed April 04, 2023 (in Russ.).

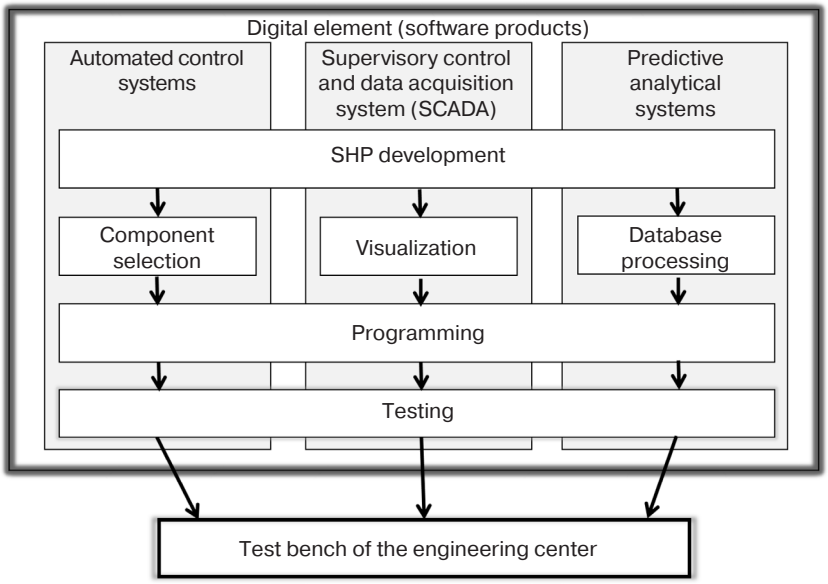


Fig. 4. Engineering center competencies

consists in the adaptation of technological equipment of the machine, including control panels, sensors of media parameters, electromechanical valves, servomotors, stepper motors, pumps, etc., with gradual transition to the use of emerging Russian products. The result of the system development will be the GFC SHC, combining application software, providing algorithms of trouble-free control of actuating mechanisms and control of machine parameters. The software will be implemented on section control PLCs (sectional controllers), devices for setting of synchronous pulse sequences by the central controller, speed and position control of frequency converters through SCADA (Fig. 6).

SCADA is a software tool whose purpose is to automate the control of technological processes, with the adjustment of parameters being carried out in real time. The operator receives comprehensive and reliable information about the objects and performs the necessary actions with the help of efficient means. In industry, the functionality of monitoring, control, archiving of the received data, alerting, and reporting, are in demand. One of the key tools of the software

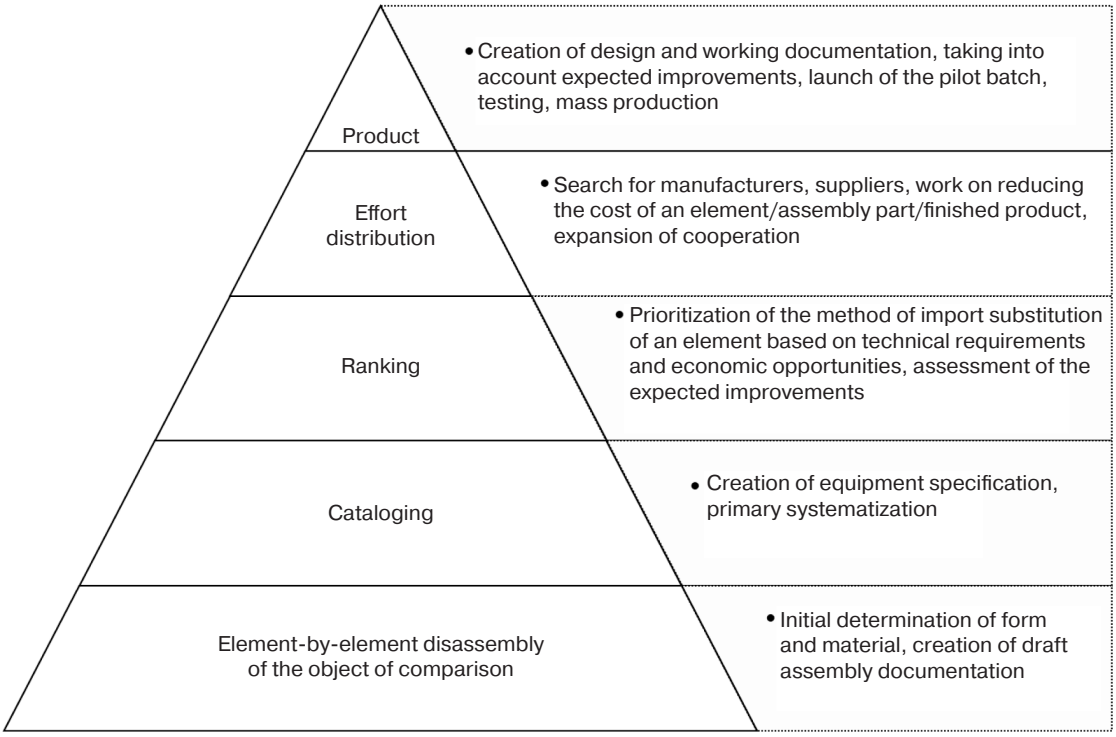


Fig. 5. Hierarchical diagram of reverse engineering

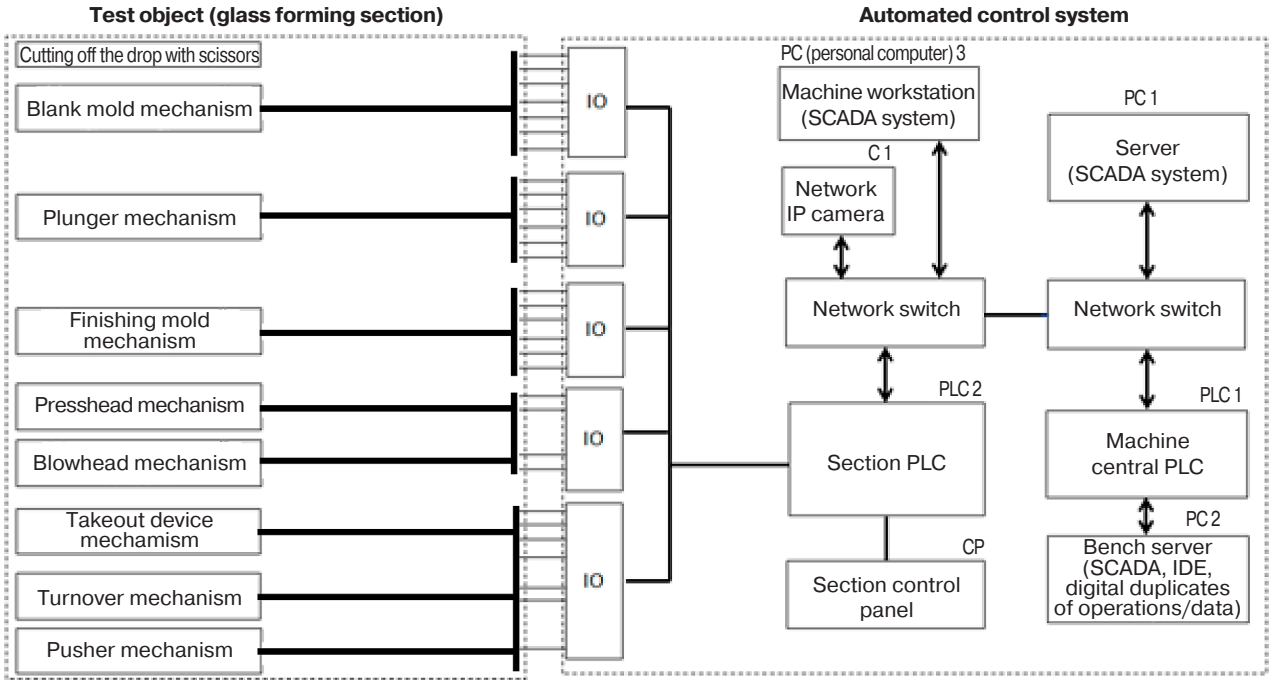


Fig. 6. Functional diagram for the development of the ACS bench. IO—input/output; IDE—integrated development environment

package consists in the hierarchy of access levels. The most common SCADAs on our market were products manufactured by Siemens[11], Aveva[12], GE Digital[13] and other companies. Currently, the functionality of Russian analogs implemented on the platform of Astra Linux[14] system is being improved, including support for object-oriented production models.

In the course of experimental (seminatural) tests of system algorithms carried out on the electromechanical bench, the fastest and cheapest stage of prototyping and testing of ACS algorithms of one section uses the elements of a digital twin to virtually simulate the work of other sections.

In the development of the systems created on the basis of the comparison object, the planned development of versions of the SHC adapted to different versions of the GFM will involve varying degrees of localization of the production of elements of the physical element of the machine. The solution of the tasks requires the combined efforts of interested parties. At the stage of SHC development, financing is provided by own funds. Since the current cost of such systems occupies a significant proportion of the cost of equipment, a substitute product has a significant commercial potential.

Importance is attributed to the possibility of providing practical training to students of Russian education institutes and the personnel of enterprises on the basis of the Engineering Center. To reduce familiarization time, the SCADA interface is designed to be visually similar to that used in the existing GFM.

CONCLUSIONS

Based on the results of research on import dependence overcoming and creating a technological sovereignty, the authors came to the following conclusions:

1. Development of the ACS GFM platform is related to import substitution in the field of mechanical engineering, which is on the priority list of Russian industrial policy.
2. Parallel imports are a temporary mitigation, but not a solution to the problem. Critical components, which include the production of electronic components and software, can and should be locked into the domestic market, recreating production chains within the country.
3. Since technology comprises not only equipment, but also carriers of knowledge, a key role is played by personnel training. Engineering centers thus become elements of innovative development, having the potential to provide additional education for students and industry specialists.
4. Technological sovereignty must be based on technologies and products that are similar or superior to the world's best examples and practices.
5. Developers of technology and equipment must cooperate and coordinate their efforts to maximize

[11] https://www.siemens.com/global/en.html. Accessed April 04, 2023.
[12] https://www.aveva.com/. Accessed April 04, 2023.
[13] https://www.ge.com/digital/. Accessed April 04, 2023.
[14] https://astralinux.ru/. Accessed April 04, 2023 (in Russ.).

the efficiency of both innovation and reverse-engineering. Since we are talking about the manufacture of the means of production rather than mass consumption products, it is inappropriate to develop the same elements (physical and digital) in parallel by different organizations. Combining production orders and their long-term planning will reduce the cost of productive machines, which will affect the economy of mass-market products.

6. Although the development and implementation of ACS of technological processes offer strong prospects for monetization, the initial stage of research in the absence of innovative investors is usually carried out at the expense of self-financing. The existing methods of state support are insufficiently coordinated with each other and characterized by a lack of production planning or the formation of centralized orders, which can affect the cost of the element and product.
7. It is necessary to raise the status of import substitution projects and, if possible, to centralize their implementation and financing, separating them into a separate national project, which establishes the achievements of individual industries and their scaling to the industry as a whole.

Authors' contributions

D.O. Skobelev, A.V. Ragutkin—the research concept, statement of the research problem, and final editing the text of the article.

D.Kh. Mikhailidi (75%), **A.B. Sukhaterin** (25%)—performing routine work on the systematization of the material, analysis of the research results and data preparation, and writing the text of the article.

REFERENCES

1. Sukharev O.S. Import substitution policy: Breaking the limits. *Upravlenets = Manager*. 2023;14(1):33–46 (in Russ.). https://doi.org/10.29141/2218-5003-2023-14-1-3
2. Nikitin G.S., Skobelev D.O. Efficiency of state and corporate investments in the development of the real sector of the economy. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo. Seriya: Sotsial'nye nauki = Vestnik of Lobachevsky State University of Nizhni Novgorod. Series: Social Sciences*. 2022;4(68):32–41 (in Russ.). Available from URL: http://www.unn.ru/pages/e-library/vestnik_soc/18115942_2022_-_4(68)_unicode/4.pdf
3. Ulin B. *Mezhregional'naya i mezhdunarodnaya torgovlya* (*Interregional and International Trade*): transl. from Engl. Moscow: Delo; 2004. 416 p. (in Russ.). [Ohlin B. *Interregional and International Trade*. Cambridge (Massachusetts): Harvard University Press; 1967. 324 p.]
4. Porter M. *Mezhdunarodnaya konkurentsiya. Konkurentnye preimushchestva stran* (*International Competition. Competitive Advantages of Countries*). Moscow: Intellektual'naya Literatura; 2021. 948 p. (in Russ.).
5. Avdokushin E.F. *Mezhdunarodnye ekonomicheskie otnosheniya* (*International Economic Relations*). Moscow: Yurist; 1999. P. 311–315 (in Russ.). Available from URL: https://uchebnik.biz/book/111-mezhdunarodnye-yekonomicheskie-otnosheniya/ (accessed April 04, 2023).
6. Zheng X., Yu H., Yang L. Technology imports, independent innovation, and China's green economic efficiency: an analysis based on spatial and mediating effect. *Environ. Sci. Pollut. Res*. 2022;29(24):36170–36188. https://doi.org/10.1007/s11356-021-17499-y
7. Arezki R., Hadri K., Loungani P., Rao Y. Testing the Prebisch-Singer Hypothesis since 1650: Evidence from Panel Techniques that Allow for Multiple Breaks. *J. Int. Money & Finance*. 2014;42(180):208–223. https://doi.org/10.1016/j.jimonfin.2013.08.012

СПИСОК ЛИТЕРАТУРЫ

1. Сухарев О.С. Государственное управление импортозамещением: преодоление ограничений. *Управленец*. 2023;14(1):33–46. https://doi.org/10.29141/2218-5003-2023-14-1-3
2. Никитин Г.С., Скобелев Д.О. Эффективность государственных и корпоративных инвестиций в развитие реального сектора экономики. *Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки*. 2022;4(68):32–41. URL: http://www.unn.ru/pages/e-library/vestnik_soc/18115942_2022_-_4(68)_unicode/4.pdf
3. Улин Б. *Межрегиональная и международная торговля*: пер. с англ. М.: Дело; 2004. 416 с.
4. Портер М. *Международная конкуренция. Конкурентные преимущества стран.* М.: Интеллектуальная Литература; 2021. 948 с.
5. Авдокушин Е.Ф. *Международные экономические отношения*. М.: Юристъ; 1999. С. 311–315. URL: https://uchebnik.biz/book/111-mezhdunarodnye-yekonomicheskie-otnosheniya/ (дата обращения 04.04.2023).
6. Zheng X., Yu H., Yang L. Technology imports, independent innovation, and China's green economic efficiency: an analysis based on spatial and mediating effect. *Environ. Sci. Pollut. Res*. 2022;29(24):36170–36188. https://doi.org/10.1007/s11356-021-17499-y
7. Arezki R., Hadri K., Loungani P., Rao Y. Testing the Prebisch-Singer Hypothesis since 1650: Evidence from Panel Techniques that Allow for Multiple Breaks. *J. Int. Money & Finance*. 2014;42(180):208–223. https://doi.org/10.1016/j.jimonfin.2013.08.012
8. Нуреев Р.М. *Экономика развития: модели становления и модернизации рыночной экономики*. М.: Норма; 2008. 367 с.
9. Sachs J., Williamson J. External Debt and Macroeconomic Performance in Latin America and East Asia. *Brookings Papers on Economic Activity*. 1985;1985(2):523–573. https://doi.org/10.2307/2534445

8. Nureev R.M. *Ekonomika razvitiya: modeli stanovleniya i modernizatsii rynochnoi ekonomiki* (*Economics of Development: Models of Formation and Modernization of a Market Economy*). Moscow: Norma; 2008. 367 p. (in Russ.).
9. Sachs J., Williamson J. External Debt and Macroeconomic Performance in Latin America and East Asia. *Brookings Papers on Economic Activity*. 1985;1985(2):523–573. https://doi.org/10.2307/2534445
10. Ukhanova R.M., Raiskaya M.V. Innovative Import Substitution as the Main Direction of Technological Modernization in Oil and Gas Enterprises of the Republic of Tatarstan. *Vestnik ekonomiki, prava i sotsiologii* = *The Review of Economics, the Law and Sociology*. 2015;4:134–136 (in Russ.). Available from URL: http://www.vestnykeps.ru/0415/28.pdf
11. Faltsman V. Import Substitution in Energy and Military-Industrial Complexes. *Voprosy Ekonomiki*. 2015;1:116–124 (in Russ.). https://doi.org/10.32609/0042-8736-2015-1-116-124
12. Panshin I.V., Yares O.B. Resursozameschenie and import substitution in solving problems of economic stability and steam-lation region&acute's economic growth. *Sovremennye problemy nauki i obrazovaniya* = *Modern Problems of Science and Education*. 2014;6:511 (in Russ.). Available from URL: https://science-education.ru/en/article/view?id=16092
13. Volosatova A.A., Uchenov A.A., Skobelev D.O. Forming the concept of implementing green economy principles in the Eurasian Economic Union: the role of harmonizing resource efficiency approaches. *Vestnik Evraziiskoi nauki* = *The Eurasian Scientific J*. 2022;14(4):14 (in Russ.). Available from URL: https://esj.today/PDF/23ECVN422.pdf (accessed April 10, 2023).
14. Sukharev O.S. *Ekonomicheskaya teoriya evolyutsii institutov i tekhnologii* (*Economic Theory of the Evolution of Institutions and Technologies*). Moscow: Lenand; 2019. 312 p. (in Russ.).
15. Krasnyuk L.V. Essence and the structural elements of the innovative development of economy. *π-Economy*. 2010;5(107):133–138 (in Russ.). Available from URL: https://cyberleninka.ru/article/n/suschnost-i-strukturnye-elementy-innovatsionnogo-razvitiya-ekonomiki (accessed April 10, 2023).
10. Уханова Р.М., Райская М.В. Инновационное импортозамещение как основное направление технологической модернизации на предприятиях нефтегазового комплекса Республики Татарстан. *Вестник экономики, права и социологии*. 2015;4:134–136. URL: http://www.vestnykeps.ru/0415/28.pdf
11. Фальцман В. Импортозамещение в ТЭК и ОПК. *Вопросы экономики*. 2015;1:116–124. https://doi.org/10.32609/0042-8736-2015-1-116-124
12. Паньшин И.В., Яресь О.Б. Ресурсозамещение и импортозамещение при решении задач обеспечения экономической стабильности и стимулирования экономического роста региона. *Современные проблемы науки и образования*. 2014;6:511. URL: https://science-education.ru/en/article/view?id=16092
13. Волосатова А.А., Ученов А.А., Скобелев Д.О. Формирование концепции внедрения принципов зеленой экономики в евразийском экономическом союзе: роль гармонизации подходов к повышению ресурсной эффективности. *Вестник Евразийской науки*. 2022;14(4):14. URL: https://esj.today/PDF/23ECVN422.pdf (дата обращения 10.04.2023).
14. Сухарев О.С. *Экономическая теория эволюции институтов и технологий*. М.: Ленанд; 2019. 312 с.
15. Краснюк Л.В. Сущность и структурные элементы инновационного развития экономики. *π-Economy*. 2010;5(107):133–138. URL: https://cyberleninka.ru/article/n/suschnost-i-strukturnye-elementy-innovatsionnogo-razvitiya-ekonomiki (дата обращения 10.04.2023).

About the authors

Dmitry Kh. Mikhailidi, Cand. Sci. (Econ.), Researcher, Department of Resource Saving Methodology, Environmental Industrial Policy Center (38, Stremyannyi per., Moscow, 115054 Russia). E-mail: d.Mikhailidi@eipc.center. RSCI SPIN-code 6831-8043, https://orcid.org/0009-0005-6491-0710

Alexander V. Ragutkin, Cand. Sci. (Eng.), Vice-Rector for Innovative Development, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: ragutkin@mirea.ru. Scopus Author ID 56871217700, ResearcherID AAE-4437-2022, RSCI SPIN-code 7531-7376, https://orcid.org/0000-0001-8256-1941

Dmitry O. Skobelev, Dr. Sci. (Econ.), Director, Environmental Industrial Policy Center (38, Stremyannyi per., Moscow, 115054 Russia); Head of the Department of Best Available Technologies and Regulatory Practices, Institute of Management Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: skobelev@mirea.ru. Scopus Author ID 57205144505, RSCI SPIN-code 7830-9773, https://orcid.org/0000-0002-8067-7016

Alexey B. Sukhaterin, Lecturer, Department of Industrial Informatics, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: suhaterin@mirea.ru. https://orcid.org/0009-0004-8997-4342

Об авторах

Михайлиди Дмитрий Христофорович, к.э.н., научный сотрудник, отдел методологии ресурсосбережения, ФГАУ «Научно-исследовательский институт «Центр экологической и промышленной политики» (115054, Россия, Москва, Стремянный переулок, д. 38). E-mail: d.Mikhailidi@eipc.center. SPIN-код РИНЦ 6831-8043, https://orcid.org/0009-0005-6491-0710

Рагуткин Александр Викторович, к.т.н., проректор по вопросам инновационного развития, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: ragutkin@mirea.ru. Scopus Author ID 56871217700, ResearcherID AAE-4437-2022, SPIN-код РИНЦ 7531-7376, https://orcid.org/0000-0001-8256-1941

Скобелев Дмитрий Олегович, д.э.н., директор, ФГАУ «Научно-исследовательский институт «Центр экологической и промышленной политики» (115054, Россия, Москва, Стремянный переулок, д. 38); заведующий кафедрой наилучших доступных технологий и регуляторных практик Института технологий управления ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: skobelev@mirea.ru. Scopus Author ID 57205144505, SPIN-код РИНЦ 7830-9773, https://orcid.org/0000-0002-8067-7016

Сухатерин Алексей Борисович, преподаватель, кафедра промышленной информатики Института искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: suhaterin@mirea.ru. https://orcid.org/0009-0004-8997-4342

Translated from Russian into English by Lyudmila O. Bychkova
Edited for English language and spelling by Thomas A. Beavitt

MIREA – Russian Technological University.
78, Vernadskogo pr., Moscow, 119454 Russian Federation.
Publication date July 31, 2023.
Not for sale.

МИРЭА – Российский технологический университет.
119454, РФ, г. Москва, пр-т Вернадского, д. 78.
Дата опубликования 31.07.2023 г.
Не для продажи.


МИРЭА
Российский
технологический
университет