ISSN 1997-1397 (Print)
ISSN 2313-6022 (Online)

Журнал Сибирского федерального университета

Математика и физика

Journal of Siberian Federal University

Mathematics & Physics

2026 19 (3)

ISSN 1997-1397
(Print)

ISSN 2313-6022
(Online)

2026 19 (3)

ЖУРНАЛ СИБИРСКОГО ФЕДЕРАЛЬНОГО УНИВЕРСИТЕТА
Математика и Физика

JOURNAL OF SIBERIAN FEDERAL UNIVERSITY
Mathematics & Physics

Издание индексируется Scopus (Elsevier), Emerging Sources Citation Index (WoS, Clarivate Analytics), Российским индексом научного цитирования (НЭБ), представлено в международных и российских информационных базах: Ulrich's periodicals directiory, ProQuest, EBSCO (США), Google Scholar, MathNet.ru, КиберЛенинке.

Включено в список Высшей аттестационной комиссии «Рецензируемые научные издания, входящие в международные реферативные базы данных и системы цитирования».

Все статьи представлены в открытом доступе http://journal.sfu-kras.ru/en/series/mathematics_physics.

Журнал Сибирского федерального университета.
Математика и физика.
Journal of Siberian Federal University. Mathematics & Physics.

Учредитель: Федеральное государственное автономное образовательное учреждение высшего образования "Сибирский федеральный университет"(СФУ)
Главный редактор: А.М. Кытманов. Редакторы: В.Е. Зализняк, А.В. Щуплев.
Компьютерная верстка: Г.В. Хрусталева.
№ 3 от 26.06.2026. Индекс: 42327. Тираж: 1000 экз. Свободная цена
Адрес редакции и издателя: 660041 г. Красноярск, пр. Свободный, 82,
стр. 24, оф. 117.
Отпечатано в типографии Издательства БИК СФУ
660041 г. Красноярск, пр. Свободный, 82.
Свидетельство о регистрации СМИ ПИ № ФС 77-28724 от 29.06.2007 г.,
выданное Федеральной службой по надзору в сфере массовых коммуникаций, связи и охраны культурного наследия
http://journal.sfu-kras.ru
Подписано в печать 15.06.26. Формат 84×108/16. Усл.печ. л. 11,8.
Уч.-изд. л. 11,6. Бумага тип. Печать офсетная.
Тираж 1000 экз. Заказ 25230
Возрастная маркировка в соответствии с Федеральным законом № 436-ФЗ:16+

CONTENTS

СОДЕРЖАНИЕ

EDN: BPCWON
УДК 519.24

Fast Numerical Methods for Stochastic Modeling Based on Probabilistic Extensions

Boris S. Dobronets*
Olga A. Popova†
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 10.11.2025, received in revised form 15.12.2025, accepted 21.01.2026

Abstract. The article studies fast algorithms for numerical modeling of problems with random input data. The approaches under consideration are non-intrusive methods based on probabilistic extensions. They rely on existing numerical methods for solving deterministic problems and use them as solvers. Unlike polynomial chaos methods, in some cases it is possible to avoid exponential growth of the number of operations from the number of input parameters.

Keywords: computational probabilistic analysis, probabilistic extensions, uncertainty quantification.

Citation: B.S. Dobronets, O.A. Popova, Fast Numerical Methods for Stochastic Modeling Based on Probabilistic Extensions, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 291–299. EDN: BPCWON.

Introduction

Real computational models usually contain many parameters.They are usually non-deterministic and can be represented either by random constants or random fields. If the input parameters of the model are represented by distribution functions, researchers tend to estimate the output data as distribution functions as well. An example of such a method is the stochastic finite element method for solving differential equations with random coefficients [2].

Methods for solving such problems are usually called Uncertainty Quantification (UQ) [3]. Uncertainty Quantification provides rigorous procedures for describing how the output of a model depends on the uncertainties of its input parameters.

Fig. 1 shows the UQ scheme, where the input parameters $x_1, \ldots, x_n$ of the model $y = f(x_1, \ldots, x_n)$ are represented by their distributions.

Uncertainty quantification methods play a key role in reducing the impact of uncertainties in both optimization and decision-making processes. The main goal of UQ is to construct output distributions of numerical simulations $\boldsymbol{y}$.

Monte Carlo (MC) methods can be attributed to UQ. Despite all its positive qualities, MC has an extremely low convergence rate. To obtain adequate estimates of the statistical characteristics of the modeling results, the original problem must be solved several thousand times.

Reducing the number of operations is one of the goals of stochastic modeling [1]. In the field of fast computing, one way to actually reduce the number of operations is to use polynomial

*BDobronets@yandex.ru https://orcid.org/0000-0002-0167-1637
† https://orcid.org/0000-0002-5739-2741

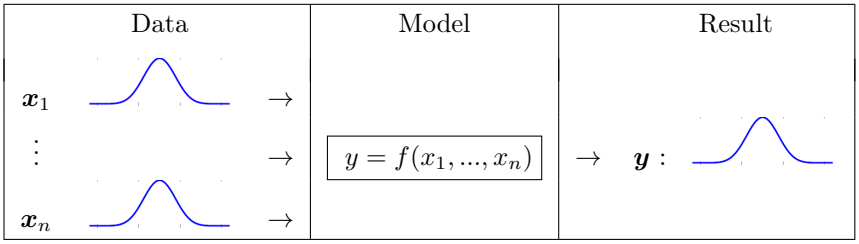


Fig. 1. Uncertainty Quantification

chaos (PC) methods in combination with stochastic collocations. A similar approach has been successfully applied to solving stochastic differential equations [2].

Polynomial chaos expansion seeks to approximate a random function via orthogonal polynomials of random variables. Thus, the Gaussian distribution corresponds to Hermitian polynomials, and the uniform distribution corresponds to the expansion in Legendre polynomials; for gamma distributions, it is convenient to use Laguerre polynomials [1, 3].

Sensitivity analysis quantifies how much of the uncertainty in the model output each uncertain parameter is responsible for. A sensitivity analysis provides a better understanding of the relationship between the parameters and output of a model. This can be useful in a model reduction context. For example, a parameter with a low sensitivity index can essentially be set to any fixed value (within the explored distribution), without affecting the variance of the model much [3].

It should be noted that the MC and polynomial chaos expansion are mainly focused on estimating the mathematical expectation and variance of the modeling results [6]. One of the limitations of using Polynomial chaos methods is the exponential growth of the number of operations from the number of input parameters [6].

The probability density functions of random variables x, y, z will be denoted bold font $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$. Let us denote by $\boldsymbol{R}$ the set of all probability density functions and $\boldsymbol{R}^n$ the space of probability densities of random vectors from $\mathbb{R}^n$.

1. Probabilistic extensions

A competitor of polynomial chaos methods is computational probabilistic analysis (CPA) [5]. The basis of CPA is the calculation of probability extensions, i.e. the construction of distribution laws for functions $f(x_1, \ldots, x_n)$ of random arguments, where the joint probability density function with respect to the random vector $(x_1, \ldots, x_n)$ is known. Unlike PC, CPA works with nonparametric representations of random arguments.

One of the problems considered by CPA is the problem to construct a probability density function of random variables. Let us consider the general case [6], when $(x_1, \ldots, x_n)$ is a system of continuous random variables with a joint probability density function $p(x_1, \ldots, x_n)$ and the random variable z is a function f of random arguments $x_1, \ldots, x_n$

$$z = f(x_1, \ldots, x_n).$$

We say that the random function $\boldsymbol{f} : \boldsymbol{R}^n \to \boldsymbol{R}$ is a probabilistic continuation of the deterministic function $f : \mathbb{R}^n \to \mathbb{R}$ on the set $D \subset \mathbb{R}^n$, if $\boldsymbol{f}(x) = f(x)$ for all arguments $x \in D$ [5].

The random function $\boldsymbol{f} : \boldsymbol{R}^n \to \boldsymbol{R}$ is called the probabilistic extension of the deterministic function $f : \mathbb{R}^n \to \mathbb{R}$ on the set $D \subset \mathbb{R}^n$, if

(i) it is probabilistic continuation of f on D,

(ii) the probability density function $\boldsymbol{f}$ coincides with the probability density function $\boldsymbol{z}$ of the random variable $z = f(x_1, x_2, \ldots, x_n)$, where $(x_1, x_2, \ldots, x_n)$ is a system of continuous random variables with joint probability density functions $\boldsymbol{p}(x_1, x_2, \ldots, x_n)$ [5].

Thus, we can write

$$\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n).$$

In cases where it is necessary to directly indicate the value of $\boldsymbol{f}$ at some point ξ, we will use the notation

$$\boldsymbol{z}(\xi) = \boldsymbol{f}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)(\xi).$$

Theorem 1.1 ([5]). *Let $(x_1, \ldots, x_n)$ be independent continuous random variables, $\boldsymbol{f}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ is a probabilistic extension of the function $f(x_1, x_2, \ldots, x_n)$ and for all real t the function $\boldsymbol{f}(t, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ is a probabilistic extension of the function $f(t, x_2, \ldots, x_n)$.*

Then

$$\boldsymbol{f}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)(\xi) = \int\limits_{supp(\boldsymbol{x}_1)} \boldsymbol{x}_1(t)\boldsymbol{f}(t, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)(\xi)dt \tag{1}$$

Thus, applying the theorem successively to $\boldsymbol{f}(t, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$, we can represent (1) as

$$\boldsymbol{f}(\boldsymbol{x}_1 \ldots \boldsymbol{x}_n)(\xi) = \int_D \boldsymbol{x}_1(t_1) \ldots \boldsymbol{x}_k(t_k)\boldsymbol{f}(t_1, \ldots, t_k, \boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n)(\xi)dt_1dt_2 \ldots dt_k, \tag{2}$$

where $D = \text{supp}(x_1) \times \ldots \times \text{supp}(x_k)$, $t_1, \ldots, t_k$ are deterministic variables.

When calculating (2) using numerical quadratures, the number of operations can be estimated as $\sim Km^k$, where K is the number of operations for calculating the probabilistic extension $\boldsymbol{f}(t_1, \ldots, t_k, \boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n)(\xi)$, m is the number of quadrature nodes in one dimension. Thus, CPA is generally not inferior to polynomial chaos methods in the number of operations.

In a number of problems it is possible to construct algorithms where the number of operations depends linearly on the number of arguments.

Let $f(x_1, \ldots, x_n)$ be a rational function. We can obtained probabilistic extension $\boldsymbol{f}$ of real rational functions f by replacing (i) the real variables x_1, x_2, …, x_n with an probability density functions $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, …, $\boldsymbol{x}_n$ and (ii) the real arithmetic operations with corresponding probabilistic operations. The result $\boldsymbol{f}$ is called a *natural probabilistic extension* [5]. In this case, the number of operations grows linearly from n.

Let us consider the question of calculating a probabilistic extension in the form (2) with as few operations as possible.

It is easy to see that we must strive to represent the construction of a probabilistic extension with the smallest k and the function $\boldsymbol{f}(t_1, \ldots, t_k, \boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_n)$ falls into conditions when it is possible to construct a probabilistic extension for it with non-exponential (polynomial) growth in the number of operations.

2. Karhunen–Loéve expansions

In numerical modeling, input data often consists of random fields. Random fields must be represented in a parameterized form. One way to parameterize a random field is to use Karhunen–Loeve expansions. Next, using the example of solving the Dirichlet problem, we will

consider algorithms that use probability expansions, which reduce computational costs compared to polynomial chaos methods.

Let's consider the Dirichlet problem:

$$Lu = f, \qquad x \in D, \tag{3}$$

$$u(x) = 0, \qquad x \in \partial\overline{D}, \tag{4}$$

D is a bounded open domain in R^2 with piecewise smooth boundary $\partial\overline{\Omega}$,

$$Lu = -\sum_{i=1}^{2} \frac{\partial}{\partial x_i}\left(a_i \frac{\partial}{\partial x_i} u\right).$$

In what follows, to solve (3), it is necessary to parameterize the input coefficients a_i and the right-hand side f. Without loss of generality (to simplify the presentation), we will assume that the coefficients a_i are determined by a set of random constants, and their probability density functions $\boldsymbol{a}_i, i = 1, \ldots, N_a$ are known.

To parameterize the right-hand side of f we will use the Karhunen–Loéve expansions [3, 6]

$$f(x,\omega) \approx f(x,\omega)_N = \sum_{n=1}^{N_f} \sqrt{\lambda_n} F_n(x) f_n(\omega),$$

where $\{\lambda_n, F_n(x)\}_{n=1}^{\infty}$ denote pairs of eigenvalues and functions of a given covariance function, and $f_n(\omega)_{n=1}^{\infty}$ are independent random variables with zero mean and unit variance, represented by their probability density functions $\boldsymbol{f}_n$. Karhunen–Loéve expansions is also known as proper orthogonal decomposition (POD) and principal component analysis (PCA).

Moreover, the error decreases monotonically with an increase in the number of terms in the expansion.

Let it be necessary to find a probabilistic extension of a numerical solution at some point $\boldsymbol{u}^h(x) = \boldsymbol{u}^h(x, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{N_a})$.

The final probabilistic extension can be represented as [8]

$$\boldsymbol{u}^h(x) = \iint \boldsymbol{a}_1(t_1) \ldots \boldsymbol{a}_{N_a}(t_{N_a}) \boldsymbol{u}^h\left(x, t_1, \ldots, t_{N_a}, \boldsymbol{f}_1, \ldots, \boldsymbol{f}_{N_f}\right) dt_1 \ldots dt_{N_a}, \tag{5}$$

where N_a is the number of parameters in the Karhunen–Loéve expansion for coefficients a_i and N_f is the number of parameters in the Karhunen–Loéve expansion for f.

The integral (5) will be calculated using numerical quadratures. The number of calculation operations (5) is proportional to $K\exp(N_a)$, where $\boldsymbol{u}^h\left(x, t_1, ..., t_{N_a}, \boldsymbol{f}_1, ..., \boldsymbol{f}_{N_f}\right)$, due to the independence of $\boldsymbol{f}_n$, can be calculated using natural probabilistic extensions; therefore, the number of operations K depends linearly on N_f.

Thus, the number of direct solutions of the original problem when using the polynomial chaos expansion is proportional to $\sim e^{N_a+N_f}$, and in the case of using the CPA and natural probabilistic extensions it is proportional to the value $\sim e^{N_a}$.

3. Applications of Taylor series expansions

Uncertainty quantification is closely related to sensitivity analysis. It is important to note that sensitivity analysis quantifies how much of the uncertainty in the model output each uncertain parameter is responsible for.

Based on derivatives local methods involve taking the partial derivative of the function $f(x_1, \ldots, x_n)$ to the input parameter x_i:

$$\left.\frac{\partial f(x_1, \ldots, x_n)}{\partial x_i}\right|_{\mathbf{x}_0},$$

where the subscript x_0 indicates that the derivative is taken at a fixed point in the input data space (hence the "local" in the class name). Knowledge of partial derivatives and the use of sensitivity analysis allow us to reduce the number of operations performed using the Taylor series.

Let us consider the case of approximate calculation of the probability extension of a smooth function $f(x_1, \ldots, x_n)$, when the random vector $x = (x_1, \ldots, x_n)$ has independent components.

In the neighborhood of some point $x_0 = (x_{01}, \ldots, x_{0n})$, we approximate the function $f(x_1, \ldots, x_n)$ by a segment of the Taylor series

$$f(x) \approx f(x_0) + \sum_{j=1}^{n} \frac{\partial f(x_0)}{\partial x_j}(x_j - x_{0j}).$$

In this case, the natural probability extension will look like this

$$\boldsymbol{f}(\boldsymbol{x}) \approx f(x_0) + \sum_{j=1}^{n} \frac{\partial f(x_0)}{\partial x_j}(\boldsymbol{x}_j - x_{0j}).$$

Due to the independence of $x = (x_1, \ldots, x_n)$, the number of operations for estimating the probabilistic extension $\boldsymbol{f}(\boldsymbol{x})$ will grow linearly with the dimension n. This approach has much in common with interval sensitivity analysis [7].

Consider the issue of accuracy estimation. Let $\mathsf{E}[x_i] = x_{0i}$ be the mathematical expectation. Then, for the random vector $\Delta_i = x_i - x_{0i}$, we can determine the values of the following characteristics: the mathematical expectation $\mathsf{E}[\Delta_i] = 0$, and the variance $\mathsf{D}[\Delta_i] = \mathsf{E}[\Delta_i^2] = \sigma_i^2$.

Let h be a random vector $h = (\Delta_1, \Delta_2, \ldots, \Delta_n)$. Note that if the second derivatives of f are bounded, the following estimate will hold:

$$|f(x) - f(x_0) + \sum_{j=1}^{n} \frac{\partial f(x_0)}{\partial x_j}\Delta_i| \leqslant C \left\|h\right\|^2$$

where C is a constant, $\left\|h\right\|^2 = \sum\limits_{i=1}^{n} \Delta_i^2$. The mathematical expectation $\left\|h\right\|^2$ can be estimated as $\mathsf{E}\left\|h\right\|^2 = \sum\limits_{i=1}^{n} \sigma_i^2$. Since $\mathsf{E}[\Delta_i^2] = \sigma_i^2$, then the variance

$$\mathsf{D}\Delta_i^2 = \mathsf{E}[(\Delta_i^2 - \sigma_i^2)^2] = \mathsf{E}\Delta_i^4 - 2\sigma_i^2\mathsf{E}[\Delta_i^2] + \sigma_i^4 = \mathsf{E}[\Delta_i^4] - \sigma_i^4.$$

For example, for a normal distribution with zero mean and variance σ_i the variance estimate is $\mathsf{D}\Delta_i^2 = 2\sigma_i^4$ [12].

Let's consider an application of this to solving Cauchy problems for systems of differential equations with random parameters.

$$y' = f(t, y, k), \tag{6}$$

$$y(0) = y0$$

$y \in R^n$ is a solution vector;

$y0 \in R^n$ is a initial data vector;

$k \in R^m$ is a random vector of independent parameters with known probability densities $\boldsymbol{k}$.

Similar problems arise when modeling complex nonlinear reactions in the kinetic region [13]. The solution can be represented as a function of the parameters and initial data

$$y = y(t, k, y0).$$

To construct probability density functions $\boldsymbol{y}$, we approximate the function y by a segment of the series

$$y(t, k) \approx y(t, k_0) + \sum_j \frac{\partial y}{\partial k_j}(k_j - k_{0j}).$$

To find partial derivatives with respect to parameters, we use the original model and obtain a system of differential equations

$$\frac{\partial y'}{\partial k_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial y_i}\frac{\partial y_i}{\partial k_j} + \frac{\partial f}{\partial k_j}, \tag{7}$$

$$\frac{\partial y}{\partial k_j} = 0, \ j = 1, \ldots, m.$$

Simultaneously solving systems (7) and (6), we obtain

$$\boldsymbol{y}(t, \boldsymbol{k}) \approx y(t, k_0) + \sum_j \frac{\partial y}{\partial k_j}(t, k_0)(\boldsymbol{k}_j - k_{0j}).$$

Note that in this case the number of operations for constructing the probabilistic extension $\boldsymbol{y}(t, \boldsymbol{k})$ grows linearly with the dimension of the parameter vector $k \in R^m$.

In some cases it is useful to consider quantile estimates of the solution [14]. Quantile estimates (quantiles) are statistical measures that divide a data set into equal parts. Simply put, a q-quantile is the value below which a q-proportion (or percentage) of all observations lie.

Fig. 2 shows the construction of quantile estimates. If the solution $y(t, k)$ depends monotonically on k, then it is the blue line. In other words, the derivative with respect to k does not vanish

$$0 \notin \frac{\partial y}{\partial k}.$$

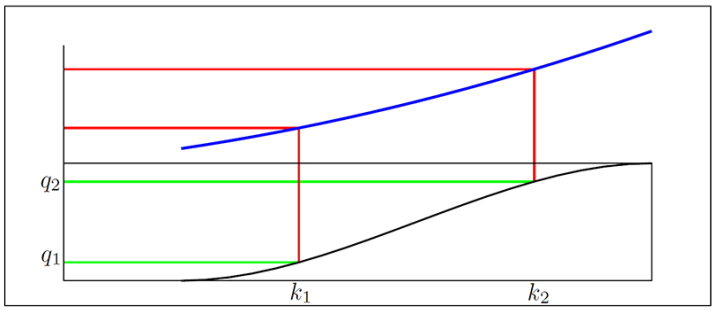


Fig. 2. The choice of values k_1, k_2 for quantiles q_1, q_2

Then, having solved the problem for the selected values of k_1, k_2, we will obtain the estimates of the quantiles as a result.

In Fig. 2 shows the choice of values k_1, k_2 of the parameter k for quantiles q_1, q_2. The blue line shows a solution that depends monotonically on k; the black line is the distribution function of the parameter k; the green lines are the quantities; the red lines are the solution for the given parameters.

The following theorem [7] gives the answer when the solution depends monotonically on k.

Theorem 3.1. *Let's assume that*

$$0 \notin \operatorname{supp}\left(\frac{\partial \boldsymbol{f}}{\partial k_i}(0, \boldsymbol{k})\right),$$

$$0 \notin \operatorname{supp}\left(\frac{\partial \boldsymbol{f}}{\partial y_i}(0, \boldsymbol{k})\right).$$

Then there exists $\bar{t} > 0$ such that for all $t \leqslant \bar{t}$ we have:

$$0 \notin \operatorname{supp}\left(\frac{\partial y}{\partial k_j}(t, \boldsymbol{k})\right).$$

Thus, the theorem shows that if the solution is initially monotone, it will remain monotone for some time.

Let's consider a model example. This system demonstrates the wrapping effect in interval analysis [7].

$$\begin{aligned} y_1' &= y_2, \\ y_2' &= -y_1, \\ y_1(0) &= k_1, \\ y_2(0) &= k_2, \end{aligned}$$

where k_1, k_2 is a random vector of parameters, and $\boldsymbol{k}_1$, $\boldsymbol{k}_2$ are known as probability density functions.

Quantile estimates for example.

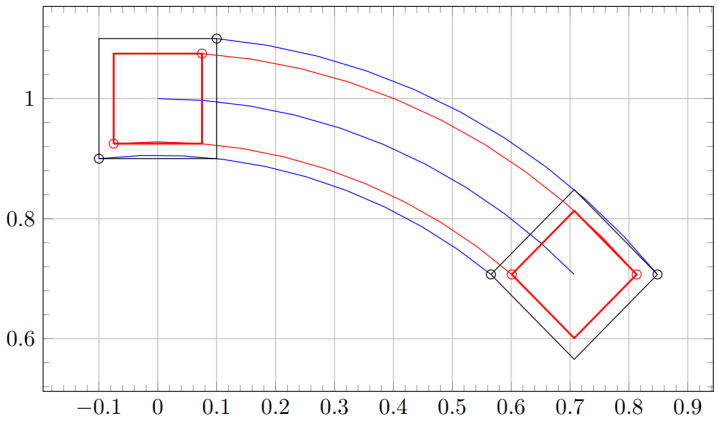


Fig. 3. Quantile estimates

Fig. 3 shows that the quantile estimates consist of the red lines. To construct the quantile estimates, the original system only had to be solved twice. The values of the parameters k are marked with circles.

Conclusion

Thus, it has been shown that using natural probabilistic extensions to solve stochastic modeling problems can, in some cases, significantly reduce the number of operations compared to polynomial chaos methods. To this end, the article considers Karhunen–Loeve expansions and Taylor series. The use of sensitivity analysis allowed us to find values for the model input parameters that yield quantile estimates of the solutions.

References

[1] D.Xiu, Fast Numerical Methods for Stochastic Computations: A Review, *Commun. Comput. Phys.*, **5**(2009) no. 2-4, 242–272.

[2] I.A.Shalimova, K.K.Sabelfeld, Solution to a stochastic Darcy equation by the polynomial chaos expansion, *J. Num. Math.*, **10**(2017), no. 3, 259–271.
DOI: 10.1134/S1995423917030077

[3] R.Smith, Uncertainty Quantification. Theory, Implementation, and Applications, Philadelphia, SIAM, 2014.

[4] S.Tennoe, G.Halnes, G.T.Einevoll, Uncertainpy: A Python Toolbox for Uncertainty Quantification and Sensitivity Analysis in Computational Neuroscience, *Front. Neuroinf.*, **12**(2018). DOI: 10.3389/fninf.2018.00049

[5] B.S.Dobronets, O.A.Popova, Computational probabilistic analysis: models and methods, Krasnoyarsk: SFU, 2020 (in Russian).

[6] M.D.Gunzburger, C.G.Webster, G.Zhang, Stochastic Finite element methods for partial differential equations with random input data, *Acta Numerica*, **23**(2014), 521–650.
DOI: 10.1017/s0962492914000075

[7] B.S.Dobronets, Interval Mathematics, Krasnoyarsk, Krasnoyarsk State University, 2004.

[8] B.S.Dobronets, O.A.Popova, A.M.Merko, Numerical modeling of boundary value problems for differential equations with random coefficients, *Journal of Physics: Conference Series. International Conference "Marchuk Scientific Readings 2021"*, 2021, 012065.
DOI: 10.1088/1742-6596/2099/1/012065

[9] E. de Rocquigny, Modelling Under Risk and Uncertainty An Introduction to Statistical, Phenomenological and Computational Methods, Wiley, 2012.

[10] C.Jiang, Y.Qi, E.Shi, Combination of Karhunen-Loéve and intrusive polynomial chaos for uncertainty quantification of thermomagnetic convection problem with stochastic boundary condition, *Engineering Analysis with Boundary Elements*, **159**(2024), 452–465.
DOI: 10.1016/j.enganabound.2023.12.014

[11] D.Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach. Princeton, NJ, Princeton University Press, 2010.

[12] A.Leon-Garcia, Probability, statistics, and random processes for electrical engineering, University of Toronto, Pearson Education, Inc., 2008.

[13] G.S.Yablonskii, V.I.Bykov, A.N.Gorban , V.I.Elokhin, Kinetic models of catalytic reactions, Amsterdam, Elsevier, 1991.

[14] M.Ichino, The quantile method for symbolic principal component analysis, *Stat. Anal. Data Min.*, **4**(2011), 184–198. DOI: 10.1002/sam.10111

Быстрые численные методы стохастического моделирования, основанные на вероятностных расширениях

Борис С. Добронец
Ольга А. Попова
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. В статье рассматриваются быстрые алгоритмы численного моделирования задач со случайными входными данными. Рассматриваемые подходы представляют собой non-intrusive методы, основанные на вероятностных расширениях. Они основаны на существующих численных методах решения детерминированных задач и используют их в качестве решателей. В отличие от методов полиномиального хаоса, в некоторых случаях удается избежать экспоненциального роста числа операций от числа входных параметров.

Ключевые слова: вычислительный вероятностный анализ, вероятностные расширения, количественная оценка неопределенности, быстрые алгоритмы.

EDN: GMRCGX
УДК 517.956.2

On the Existence of Solution of Some Inverse Problem for the Elliptic Equation

Anna Sh. Lyubanova*
Alexandr V. Velisevich†
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 10.06.2025, received in revised form 05.10.2025, accepted 07.02.2026

Abstract. The sufficient conditions of solution existence are established for the inverse problem of finding a piecewise constant coefficient in the second-order elliptic equation with the Dirichlet boundary condition and the integral overdetermination condition on the boundary of the domain under consideration.

Keywords: inverse problem, elliptic equation, integral overdetermination, existence theorem.

Citation: A.Sh. Lyubanova, A.V. Velisevich, On the Existence of Solution of Some Inverse Problem for the Elliptic Equation, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 300–307. EDN: GMRCGX.

Introduction

In this paper we discuss the existence of the solution to the inverse problem of finding an unknown piecewise constant coefficient in the lower term of the stationary second-order partial differential equation. The problem is considered in the bounded domain $\Omega \subset \mathbf{R}^n$ with a boundary $\partial\Omega \in C^2$.

Problem 1. For given functions $f(x), \beta(x), h(x)$ and constants μ, k_2 find the function $u(x)$ and the constant k_1 such that pair $\{u(x), k(x)\}$ satisfies the equation

$$-\operatorname{div}(\mathcal{M}(x)\nabla u) + m(x)u + k(x)u = f, \tag{1}$$

the boundary condition

$$u|_{\partial\Omega} = \beta(x), \tag{2}$$

and the condition of overdetermination

$$\int_{\partial\Omega} \frac{\partial u}{\partial \overline{N}} h(x)ds = \mu, \tag{3}$$

where the function $k(x)$ takes the form

$$k(x) = \begin{cases} k_1, \ x \in \Omega_1, \\ k_2, \ x \in \Omega_2, \end{cases} \tag{4}$$

*lubanova@mail.ru
†velisevich94@mail.ru

$\Omega_i \subset \Omega$ — a subdomain, $i = 1, 2$, and $\Omega_1 \cap \Omega_2 = \varnothing$, $\overline{\Omega}_1 \cup \overline{\Omega}_2 = \overline{\Omega}$, that is the domain Ω is divided into two non-intersecting parts by a hypersurface $\Gamma \subset \Omega$, $\mathcal{M}(x) = ((m_{ij}(x)))$ — a matrix of functions m_{ij}, $i, j = 1, 2, \ldots, n$, $m(x)$ — scalar function, $\dfrac{\partial}{\partial \overline{N}} = (\mathcal{M}(x)\nabla, \mathbf{n})$, $\mathbf{n}$ is the unit vector of the outward normal to the boundary $\partial\Omega$.

We denote the boundary of the subdomain Ω_i by $\partial\Omega_i$, and the intersection $\partial\Omega \cap \partial\Omega_i$ by S_i, $i = 1, 2$. Γ and S_i will be considered as hypersurfaces of class C^2.

Problem 1 includes two cases:

1) $S_1 = \partial\Omega, S_2 = \varnothing$, that is Ω_2 is an inner subdomain, $\overline{\Omega}_2 \subset \Omega$;

2) $S_2 \neq \varnothing$, that is the subdomain Ω_2 has a common part of the boundary with $\partial\Omega$. In this case we suppose that h is a function is finite in $\partial\Omega$, and supp$h \subset S_1$ is a set of nonzero $(n-1)$-dimensional measure.

A main goal of this paper is to establish the existence of generalized solution of Problem 1. Following the idea [1, 2], based on the method described in [3], the existence of solution is proved by reducing the inverse problem to an operator equation of the second kind due to the corresponding extension of the boundary data $\beta(x)$ and the weight function $h(x)$ from the boundary to the domain.

In [4, 5], the existence and uniqueness of the solution to similar inverse problems for the equation (1) with $k(x) = k = const$ were established in the case of the first and third boundary conditions. The results are also valid for the inverse problems with the boundary conditions of the second kind.

Practical interest in such inverse problems is caused by a lot of applications in the theory of diffusion and filtration [6,7]. The formulation of the boundary value problems for the stationary equations with discontinuous lower coefficient arises in the modelling of physical parameters in the neighborhood of the interfaces of different media, for example, the temperature in the subsurface layer of the ocean, as well as the characteristics of layered semiconductor structures [8]. The coefficient of the equations characterizing the properties of the medium are difficult to determine experimentally. For instance, the properties and structure of a fissured medium depend on the rock occurrence conditions, which are difficult to reproduce in the laboratory with an acceptable accuracy [9].

Many authors considered the coefficient inverse problems for the elliptic equations. The works [7, 10–13] should be noted here. In these works, the unknown coefficients are recovered from information on the values of some integral operator or the trace of the solution on some manifold inside of the domain or on the boundary. The inverse problems with the conditions of integral overdetermination with respect to the flux on the boundary of the domain were not considered.

1. Preliminaries

The following notations are used: $\|\cdot\|_R$, $(\cdot,\cdot)_R$ — the norm and the inner product in $\mathbb{R}^n$; $\|\cdot\|$, $(\cdot,\cdot)$ — the norm and the inner product in $L^2(\Omega)$; $\|\cdot\|_j$, $\langle\cdot,\cdot\rangle_1$ — the norm in $W_2^j(\Omega)$, $j = 1, 2$, and the duality relation between $\overset{\circ}{W}{}_2^1(\Omega)$ and $W_2^{-1}(\Omega)$, respectively.

We introduce the linear operator $M : W_2^1(\Omega) \to W_2^1(\Omega)$ of the form

$$M = -\operatorname{div}(\mathcal{M}(x)\nabla) + m(x)I$$

where I is the identity operator and the notation

$$\langle Mv_1, v_2\rangle_M = \int_\Omega ((\mathcal{M}(x)\nabla v_1, \nabla v_2)_R + m(x)v_1v_2)dx$$

for any $v_1, v_2 \in W_2^1(\Omega)$. The following assumptions of the operator M will be considered to hold.

I. $m_{ij}(x)$, $\partial m_{ij}/\partial x_l$, $i,j,l = 1,2,\ldots,n$, and $m(x)$ are bounded in Ω. The operator M is strongly elliptic, that is, there exist positive constants m_0 and m_1 such that for all $v \in W_2^1(\Omega)$

$$m_0\|v\|_1^2 \leqslant \langle Mv, v\rangle_M \leqslant m_1\|v\|_1^2.$$

II. M is self-adjoint in $\mathring{W}_2^1(\Omega)$, that is $m_{ij}(x) = m_{ji}(x)$ for $i,j = 1,\ldots,n$.

We also introduce functions a, a^σ and b as solutions of the problems

$$Ma = f(x), \qquad a|_{\partial\Omega} = \beta(x); \tag{5}$$

$$Ma^\sigma + \sigma a^\sigma = f, \qquad a^\sigma|_{\partial\Omega} = \beta(x); \tag{6}$$

$$Mb = 0, \qquad b|_{\partial\Omega} = h(x); \tag{7}$$

where $\sigma > 0$ is a real number.

Let us consider an auxiliary inverse problem of finding the pair $\{u_1(x), k_1\}$ consisting of the function $u_1(x)$ and the constant $k_1$which satisfies the equation

$$-\operatorname{div}(\mathcal{M}(x)\nabla u_1) + m(x)u_1 + k_1u_1 = f, \tag{8}$$

the boundary condition

$$u_1|_{\partial\Omega} = \beta(x), \tag{9}$$

and the condition of overdetermination

$$\int_{\partial\Omega} \frac{\partial u_1}{\partial \overline{N}} h(x)ds = \mu. \tag{10}$$

In [4], the sufficient conditions for the existence and uniqueness of the strong solution $\{u_1(x), k_1\}$ to the inverse problem (8)–(10) were gotten due to extension of the boundary data $\beta(x)$ and the weight function $h(x)$ from the boundary $\partial\Omega$ to the domain Ω as the solutions of the boundary value problems (5)–(7). The following theorem gives these conditions.

Theorem 1 ([4]). *Let $\partial\Omega \in C^2$ and the assumptions I and II are fulfilled. Suppose also that*

(i) $f(x) \in L^2(\Omega)$, $\beta(x)$, $h(x) \in W_2^{3/2}(\partial\Omega)$,

(ii) $f(x) \geqslant 0$ *almost everywhere in* Ω; $\beta(x) \geqslant 0$, $h(x) \geqslant 0$ *for almost all* $x \in \partial\Omega$ *and there is a smooth piece* Σ *of the boundary* $\partial\Omega$ *and a constant* $\delta > 0$ *such that* $\beta \geqslant \delta$ *and* $h \geqslant \delta$ *almost everywhere in* Σ, *and the condition*

$$0 \leqslant \mu - \Psi \leqslant \frac{m_0(a,b)^2}{4\|a\|\|b\|},$$

is fulfilled. Here $\Psi = \langle Ma, b\rangle_M - (f,b)$. *Then the problem (8)–(10) has a solution* $\{u_1, k_1\}$, $u_1(x) \in W_2^2(\Omega)$. *And the estimates*

$$0 < a^\sigma \leqslant u_1 \leqslant a, \quad 0 \leqslant k \leqslant \sigma, \quad \|u_1\|_2 \leqslant C\|a\| + \|a\|_2, \tag{11}$$

$$\|u_1\|_1 \leqslant \Big(\frac{\sigma}{m_0}\Big)^{1/2}\|a\| + \|a\|_1, \tag{12}$$

are valid. The constant C depends on $mes\Omega$, m_0, m_1 and the constant $\sigma > 0$ of the form

$$\sigma = \frac{m_0((a,b) - \sqrt{D})}{2\|a\|\|b\|}, \tag{13}$$

$$D \equiv (a,b)^2 - 4(\mu - \Psi)m_0^{-1}\|a\|\|b\|.$$

Moreover, if

$$0 \leqslant \mu - \Psi < \frac{m_0(a,b)^2}{\|a\|\|b\|},$$

then the solution of the problem (8)–(10) is unique.

We define the function $\hat{u}_2(x)$ as a solution of the problem

$$-\operatorname{div}(\mathcal{M}(x)\nabla\hat{u}_2) + m(x)\hat{u}_2 + k_2\hat{u}_2 = (k_1 - k_2)u_1, \text{ for almost all } \Omega_2, \tag{14}$$

$$\hat{u}_2(x)|_{\partial\Omega_2} = 0. \tag{15}$$

The existence and uniqueness theorem for the solution $\hat{u}_2$ of the problem (14)–(15) follows from the results of [14].

Theorem 2. *Let $\partial\Omega \in C^2$, the hypotheses of Theorem 1 are fulfilled and $k_2 \geqslant 0$. Then the problem (14)–(15) has a unique solution $\hat{u}_2 \in \mathring{W}_2^1(\Omega_2) \cap W_2^2(\Omega_2)$ and the estimates*

$$\begin{gathered}\|\hat{u}_2\|_{W_2^2(\Omega_2)} \leqslant c_1\Big(|k_1 - k_2|^2\|a\|^2_{L_2(\Omega_2)} + \|u_2\|^2_{L_2(\Omega_2)}\Big),\\ \int_{\Omega_2}\|\nabla\hat{u}_2\|^2dx \leqslant c_2|k_1 - k_2|^2\|a\|^2_{L_2(\Omega_2)},\end{gathered} \tag{16}$$

are valid. Here c_1 depends on $m_0, m_1, \max\limits_{i,j,l=1,2,\ldots n}\Big\|\dfrac{\partial m_{ij}}{\partial x_l}\Big\|_{L^\infty(\Omega)}$ and $\partial\Omega_2$; c_2 depends on n, $mes\Omega_2$ and m_0. Moreover, if $k_2 = 0$ then the solution $u_2 \geqslant 0$ almost everywhere in Ω_2 and $\|u_2\|_{L^\infty(\Omega_2)} < +\infty$.

2. Existence theorem

The main result of the work is the existence theorem for the generalized solution of Problem 1. By a solution of Problem 1 is meant the pair consisting of the function $u(x)$ and the coefficient $k(x)$ which satisfies the conditions:

1) $u(x) \in W_2^1(\Omega), k(x)$ is of the form (4), where $k_1 \geqslant 0$;

2) the integral identity

$$\langle Mu, v\rangle_M + (k(x)u, v) = (f, v) \tag{17}$$

holds for any $v \in \mathring{W}_2^1(\Omega),\ v|_\Gamma = 0$,;

3) $u(x)$ satisfies the boundary condition (2);

4) the condition of integral overdetermination (3) is fulfilled in terms of the equation

$$\langle Mu, b_1\rangle_M + (k(x)u, b_1) = (f, b_1) + \mu, \tag{18}$$

where function $b_1(x)$ is the extension of $h(x)$ into the domain Ω,

$$Mb_1 = 0 \ \text{ in } \Omega_1, \quad b_1|_{S_1} = h, \quad b_1|_{\Omega\setminus\Omega_1} = 0.$$

Theorem 3. *Let $\partial\Omega \in C^2$, $k_2 \geqslant 0$ and the hypotheses of Theorem 1 guaranteeing the existence of the solution $\{u_1(x), k_1\}$ of the problem (8)–(10) are fulfilled. Suppose also that when the subdomain Ω_2 has a common part of the boundary with $\partial\Omega$, the function h is finite in $\partial\Omega$ and $\Sigma \subset supph \subset S_1$. Then Problem 1 has a solution $\{u(x), k(x)\}$, $u(x) \in W_2^1(\Omega)$, $k(x)$ is of the form (4). Moreover, the estimates*

$$\begin{gathered} 0 \leqslant k(x) \leqslant \max\{\sigma, k_2\}, \\ \|u\|_1 \leqslant K(\|f\| + \|\beta\|_{W_2^{1/2}(\partial\Omega)}), \end{gathered} \tag{19}$$

holds with a constant $K > 0$ which depends on k_2, σ, m_0, m_1 and the domain Ω; the constant $\sigma > 0$ is determined by the formula (13). If $k_2 = 0$, then $u \geqslant 0$ almost everywhere in Ω.

Proof. The solution of Problem 1 will be sought as

$$u(x) = u_1(x) + u_2(x), \tag{20}$$

where $\{u_1(x), k_1(x)\}$ is the solution of the problem (8)–(10) and $u_2(x)$ is the function of the form

$$u_2(x) = \begin{cases} \hat{u}_2(x), x \in \Omega_2, \\ 0, x \in \overline{\Omega} \setminus \Omega_2, \end{cases} \tag{21}$$

the function $\hat{u}_2(x)$ is the solution of the problem (14)–(15).

We construct the solution of Problem 1 according to the following scheme:

1) to solve the inverse provlem (8)–(10), thereby to find the solution $u_1(x)$ and k_1, and hence $k(x)$;

2) to find the function $\hat{u}_2(x)$, as a solution of the problem (14)–(15);

3) to define the function $u_2(x) \in \mathring{W}_2^1(\Omega)$ by (21);

4) to determine a function $u(x)$ of the form (20), $u(x) \in W_2^1(\Omega)$.

By Theorems 1 and 2, there is $0 < k_1 \leqslant \sigma$, $u_1(x) \in W_2^2(\Omega)$, $\hat{u}_2 \in W_2^2(\Omega_2)$. Let us prove that $u_2(x) \in W_2^1(\Omega)$. Indeed, for any $v \in \mathring{C}^\infty(\Omega)$

$$\int_\Omega u_2 v_{x_i} dx = \int_{\Omega_2} \hat{u}_2 v_{x_i} dx = \int_\Gamma \hat{u}_2 v \cos(n, x_i) ds - \int_{\Omega_2} \hat{u}_{2x_i} v dx = -\int_{\Omega_2} \hat{u}_{2x_i} v dx = -\int_\Omega u_{2x_i} v dx,$$

that is, the weak derivative u_{2x_i} exists, $i = 1, \ldots, n$, and

$$u_{2x_i} = \begin{cases} \hat{u}_{2x_i}, \ x \in \Omega_2, \\ 0, \ x \in \Omega_1, \ i = 1, \ldots, n. \end{cases}$$

By (16) and Steklov inequality

$$\|v\|^2 \leqslant c_0 \int_\Omega |\nabla v|^2 dx$$

for any function $v \in \mathring{W}_2^1(\Omega)$, we get

$$\|u_2\|_1^2 = \int_\Omega (|\nabla u_2|^2 + u_2^2) dx \leqslant \max\{1, c_0\} \int_\Omega |\nabla u_2|^2 dx \leqslant c_2 \max\{1, c_0\} |k_1 - k_2| \|u_1\|^2 \leqslant$$

$$\leqslant c_2 \max\{1, c_0\} |k_1 - k_2|^2 \|a\|^2. \tag{22}$$

Let us show that the constructed function $u(x)$ of the form (20) satisfies the integral inequality (17). By virtue of (8) and (9)

$$\langle Mu_1, v\rangle_M + k_1(u_1, v) = (f, v) \tag{23}$$

for any $v \in \mathring{W}_2^1(\Omega)$. Furthermore, in view of (14), (15) and (21) the relation

$$\int_\Omega Mu_2 v dx + (k(x)u_2, v) = \int_{\Omega_2} ((\mathcal{M}(x)\nabla \hat{u}_2, \nabla v) + m(x)\hat{u}_2 v) dx + k_2(\hat{u}_2, v)_{L^2(\Omega_2)} =$$

$$= (k_1 - k_2)(u_1, v)_{L^2(\Omega_2)} = ((k_1 - k(x))u_1, v) \tag{24}$$

is satisfied for any $v \in \mathring{W}_2^1(\Omega), v|_\Gamma = 0$. From (23) and (24) it follows that

$$\langle Mu, v\rangle_M + (k(x)u, v) = \langle Mu_1, v\rangle_M + \langle Mu_2, v\rangle_M + k_1((u_1 + u_2, v))_{L^2(\Omega_1)} + k_2(u_1 + u_2, v)_{L^2(\Omega_2)} =$$

$$= (f, v) - k_1(u_1, v)_{L^2(\Omega_2)} + \int_{\Omega_2} ((\mathcal{M}(x)\nabla \hat{u}_2, \nabla v) + m(x)\hat{u}_2 v) dx + k_2(u_1 + u_2, v)_{L^2(\Omega_2)} =$$

$$= (f, v) - k_1(u_1, v)_{L^2(\Omega_2)} + (k_1 - k_2)(u_1, v)_{L^2(\Omega_2)} + k_2(u_1, v)_{L^2(\Omega_2)} = (f, v).$$

for any $v \in \mathring{W}_2^1(\Omega)$, $v|_\Gamma = 0$.

It remains to check the condition of overdetermination in sense of (18). Multiply (8) by b_1 in terms of inner product of $L_2(\Omega)$ and integrate by parts in the first summand. By (10), this gives

$$\int_\Omega Mu_1 b_1 dx + k_1(u_1, b_1) = -\int_{\partial\Omega} \frac{\partial u_1}{\partial \overline{N}} b_1 dx + \langle Mu_1, b_1\rangle_M + k_1(u_1, b_1)_{L_2(\Omega)} =$$

$$= -\mu + \langle Mu_1, b\rangle_M + k_1(u_1, b_1)_{L_2(\Omega_1)} = (f, b_1).$$

Then

$$\langle Mu, b_1\rangle_M + (k(x)u, b_1) = \langle M(u_1 + u_2), b_1\rangle_M + (k(x)(u_1 + u_2), b_1) =$$

$$= \langle Mu_1, b_1\rangle_M + k_1(u_1, b_1)_{L_2(\Omega_1)} = \mu + (f, b_1),$$

that is, (18) is fulfilled. Thus, the pair $\{u(x), k(x)\}$ is the generalized solution of Problem 1. From (11), (12) and (22) we conclude that for $u(x)$ and $k(x)$ the estimates

$$0 \leqslant k(x) \leqslant \max\{\sigma, k_2\},$$

$$\|u\|_1 \leqslant \|u_1\|_1 + \|u_2\|_1 \leqslant \Big(\frac{\sigma}{m_0}\Big)^{1/2} \|a\| + \|a\|_1 + c_2^{1/2} (\max\{1, c_0\})^{1/2} |k_1 - k_2| \|a\| \leqslant$$

$$\leqslant \|a\|_1 + \Big(\Big(\frac{\sigma}{m_0}\Big)^{1/2} + c_2^{1/2} (\max\{1, c_0\})^{1/2} (\sigma + k_2)\Big) \|a\|$$

are valid. the estimate (19) follows from the last relation and inequality [15]

$$\|a\|_1 \leqslant c(\|f\| + \|\beta\|_{W_2^{1/2}(\partial\Omega)}).$$

The constant $c > 0$ depends on m_0, m_1 and domain Ω.

If $k_2 = 0$, then the right-hand side of the equation (14) is nonnegative. Therefore $\hat{u}_2 \geqslant 0$ and hence $u_2 \geqslant 0$ almost everywhere in Ω in accordance with the comparison theorem for the elliptic equations. By Theorem 1, $u_1 \geqslant 0$ and thus $u(x) \geqslant 0$ almost everywhere in Ω. Theorem is proven. □

This work was supported by the Russian Science Foundation, the Government of the Krasnoyarks Territory and the Krasnoyarsk Regional Science Foundation (Project no. 25-21-20054).

References

[1] A.Sh.Lyubanova, Identification of a constant coefficient in an elliptic equation, *Applicable Analysis*, **87**(2008), 1121–1128. DOI: 10.1080/00036810802189654

[2] A.Sh.Lyubanova, A.Tani, An inverse problem for pseudoparabolic equation of filtration: the existence, uniqueness and regularity, *Applicable Analysis*, **90**(2011), 1557–1571. DOI: 10.1080/00036811.2010.530258

[3] A.I.Prilepko, D.G.Orlovsky, I.A.Vasin, Methods for solving inverse problems in mathematical physics, New York, Marcel Dekker, 2000.

[4] A.Sh.Lyubanova, A.V.Velisevich, Inverse problems for the stationary and pseudoparabolic equations of diffusion, *Applicable Analysis*, **98**(2019), 1997–2010. DOI: 10.1080/00036811.2018.1442001

[5] A.V.Velisevich, On an inverse problem for the stationary equation with a boundary condition of the third kind, *Journal of Siberian Federal University. Math. and Phys.*, **14**(2021), no. 5, 659–666. DOI: 10.17516/1997-1397-2021-14-5-659-666

[6] A.N.Konovalov, Problems of the filtration of incompressible fluid, Novosibirsk, Nauka, 1988 (in Russian).

[7] G.V.Alekseev, E.A.Kalinina, Identification of the lower coefficient for the stationary convection-diffusion-reaction equation, *Sib. Zh. Ind. Math.*, **10**(2007), no. 1, 3–16 (in Russian).

[8] N.T.Levashova, N.N.Nefedov, A.O.Orlov, Time-independent reaction-diffusion equation with discontinuous reactive term, *Computational Mathematics and Mathematical Physics*, **57**(2017), no. 5, 854–866. DOI: 10.7868/S0044466917050064

[9] G.I.Barenblatt, Iu.P.Zheltov, I.N.Kochina, Basic concepts in the theory of seepage of homogeneous liquids in fissured blocks [strata], *J. Appl. Math. Mech.*, **24**(1960), no. 5, 1286–1303.

[10] S.G.Pyatkov, On some inverse problems for the elliptic equations and systems, *Journal of Applied and Industrial Mathematics*, **5**(2011), no. 3, 417–430. DOI: 10.1134/S199047891103015X

[11] A.I.Kozhanov, On the solvability of the inverse problems of parameter recovery in elliptic equations, *Mathematical Notes of NEFU*, **27**(2020), no. 4, 14–29.
DOI: 10.25587/SVFU.2020.57.53.002

[12] A.I.Prilepko, A.B.Kostin, V.V. Soloviev, Inverse source and inverse coefficient problems for elliptic and parabolic equations in Holder and Sobolev spaces, *Journal of Mathematical Sciences*, **237**(2019), no. 4, 576–594. DOI: 10.1007/s10958-019-04184-2

[13] A.Hasanov, V.G.Romanov, An inverse coefficient problem related to elastic plastic torsion of a circular cross-section bar, *Applied Mathematics Letters*, **26**(2013), no. 5, 533–538.

[14] O.A.Ladyzenskaya, N.N.Uralceva, Linear and quasilinear elliptic equations, *Moscow, Nauka*, 1964 (English transl., New York, Academic Press, 1973).

[15] J.-L.Lions, E.Madgenes, Non-homogeneous boundary value problems and applications, Dunod, Paris, 1968.

О существовании решения одной обратной задачи для эллиптического уравнения

Анна Ш. Любанова
Александр В. Велисевич
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. В работе устанавливаются достаточные условия существования решения обратной задачи отыскания кусочно-постоянного коэффициента в линейном эллиптическом уравнении второго порядка с граничным условием первого рода и условием интегрального переопределения на границе исследуемой области.

Ключевые слова: обратная задача, эллиптическое уравнение, интегральное переопределение, теорема существования.

EDN: GYVXFS
УДК 517.55

On Multidimensional Lacunary Hartogs Series with Ostrovsky Lacunae

Arslan P. Mardanov*
Tashkent State Technical University named after Islam Karimov
Tashkent, Uzbekistan
Tahir T. Tuychiev†
Djurabay K. Tishabaev‡
National University of Uzbekistan named after Mirzo Ulugbek
Tashkent, Uzbekistan

Received 10.12.2025, received in revised form 19.01.2026, accepted 21.02.2026

Abstract. This paper is devoted to multidimensional analogues of Ostrovsky's theorem on lacunary series. The work examines the domains of existence of Hartogs lacunary series with Ostrovsky lacunae and series in homogeneous polynomials. Analogues of Ostrovsky's theorem for such series are given and the domains of convergence of these series are described.

Keywords: plurisubharmonic function, singular point, nowhere dense set, power series, lacunary Hartogs series, series in homogeneous polynomials.

Citation: A.P. Mardanov, T.T. Tuychiev, D.K. Tishabaev, On Multidimensional Lacunary Hartogs Series with Ostrovsky Lacunae, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 308–319. EDN: GYVXFS.

Introduction

The work is devoted to the study of the domain of convergence of lacunary series. The main problem discussed in this article is the following question: *what can be said about the features of a power series if the characteristics of the lacunae are known?*

This problem attracted the attention of many mathematicians since the end of the 19th century, such as Hadamard [1], Fabry [2], Faber [3], Mandelbroit [4], G. Pólya [5], Ostrovsky [6] etc. In the twenties of the last century, remarkable results were obtained that allowed the development of this direction to be considered almost complete. Almost all of these results are associated with the name of the outstanding Hungarian mathematician G. Pólya [5]. All classical works on lacunary series are given in detail in the monograph by L. Bieberbach [7]. Multidimensional analogues of lacunary series have been studied in [8–11] and others.

In the first part of the article, in the one-dimensional case, it is clarified under what conditions the disc of convergence of a power series is a natural Weierstrass domain of existence for the sum of the series and under what conditions the domain of existence of a lacunary series will be wider than its disc of convergence. The second part of the work is devoted to a multidimensional analogue of Ostrovsky's theorem on lacunary series. It studies the domain of convergence of multidimensional lacunary Hartogs series. An analogue of Ostrovsky's theorem for Hartogs series is given and the domains of convergence of these series are described. In the last part of the work, an analogue of Ostrovsky's theorem for series in homogeneous polynomials is proved.

*apardayevich@mail.ru https://orcid.org/ 0000-0002-4399-0708
†tahir1955@mail.ru https://orcid.org/ 0000-0001-8092-2432
‡jura63@rambler.ru https://orcid.org/ 0000-0002-7428-3934

1. The domain of existence of one-dimensional lacunary series

Let the following power series be given on the plane

$$f(z) = \sum_{n=0}^{\infty} c_n z^n, \quad \overline{\lim_{n\to\infty}} |c_n|^{\frac{1}{n}} = 1. \tag{1}$$

Then it is clear that, according to the Cauchy–Hadamard formula, the radius of convergence of this series is 1 and this series converges inside the unit disc. Moreover, on the boundary ∂U of the unit disc there lies at least one singular point of the function $f(z)$.

An interesting case is *when each boundary point of the unit disc is singular for $f(z)$, i.e. is the unit disc U a natural Weierstrass domain of existence for the function $f(z)$?*

For a power series

$$f(z) = \sum_{k=0}^{\infty} z^k$$

radius of convergence $R = 1$ and at the boundary $\partial U = \{z : |z| = 1\}$ of the disc of convergence there is only one singular point $z = 1$, i.e. the disc U is not a natural Weierstrass domain of existence for this function. And for the series

$$g(z) = \sum_{k=0}^{\infty} z^{2^k}$$

each boundary point of the disc of convergence is singular (see [12]), and, therefore, the disc of convergence $U = \{z : |z| < 1\}$ is a natural Weierstrass domain of existence for the function $g(z)$.

The second series differs from the first in that it is lacunar, with large intervals of zero terms.

Definition 1. *Series* (1) *is called lacunary if there is a strictly increasing sequence of natural numbers $\{n_k\}$ such that $c_n = 0$ if $n \notin \{n_k\}$ and $c_{n_k} \neq 0$.*

Let the following lacunary series be given on the plane C

$$f(z) = \sum_{k=0}^{\infty} c_{n_k} z^{n_k}, \quad \overline{\lim_{n\to\infty}} |c_{n_k}|^{\frac{1}{n_k}} = 1. \tag{2}$$

The following Hadamard theorem on lacunae holds [1]: *if for the lacunary series* (2) *the condition is satisfied*

$$\frac{n_{k+1} - n_k}{n_k} > \theta_0, \ k = 0, 1, 2, \ldots, \tag{3}$$

θ_0 is some positive number independent of k, then $f(z)$ has a circle $|z| = 1$ by its natural boundary, i.e. $f(z)$ is single-valued and holomorphic in $\{|z| < 1\}$ and each point of the circle $|z| = 1$ is a singular point for the function $f(z)$. Thus, the natural domain of Weierstrass existence is $W_f(z) = \{|z| < 1\}$.

Hadamard's theorem on lacunae was strengthened a little later by Fabry [2] (see also [7]): let the condition be satisfied for series (2)

$$\overline{\lim_{n\to\infty}} \frac{k}{n_k} = 0. \tag{4}$$

Then $f(z)$ has circle $|z| = 1$ as its natural boundary.

Remark. Fabry's theorem on lacunae is a far-reaching generalization of Hadamard's theorem on lacunae, i.e. Fabry condition for lacunae $\overline{\lim\limits_{n\to\infty}}\dfrac{k}{n_k}=0$ is significantly weaker than the Hadamard condition

$$\underline{\lim_{k\to\infty}}\frac{n_{k+1}-n_k}{n_k}>0.$$

◀ Indeed, if the sequence $\{n_k\}$ satisfies the Hadamard condition (3): $\dfrac{n_{k+1}-n_k}{n_k}>\theta_0$, $k=0,1,2,\ldots$, where θ_0 is some positive number independent of k, then the inequality holds

$$n_{k+1}>(1+\theta_0)n_k \tag{5}$$

Using inequality (5) successively k times we obtain the inequality $n_k>(1+\theta_0)^k$ and, therefore, according to Newton's binomial $n_k>(1+\theta_0)^k>ck^2$, where c is some constant. Then

$$0<\frac{k}{n_k}<\frac{k}{ck^2}=\frac{1}{ck}$$

and hence we have $0\leqslant\lim\limits_{k\to\infty}\dfrac{k}{n_k}\leqslant\lim\limits_{k\to\infty}\dfrac{1}{ck}=0$, i.e. on the properties of number sequences $\lim\limits_{k\to\infty}\dfrac{k}{n_k}=0$. ▶

Example 1. Let $n_k=k^2$. For this sequence, we check the fulfillment of the Fabry and Hadamard conditions:

$$\overline{\lim_{k\to\infty}}\frac{k}{n_k}=\overline{\lim_{k\to\infty}}\frac{k}{k^2}=\overline{\lim_{k\to\infty}}\frac{1}{k}=0,$$

But

$$\lim_{k\to\infty}\frac{n_{k+1}-n_k}{n_k}=\lim_{k\to\infty}\frac{(k+1)^2-k^2}{k^2}=\lim_{k\to\infty}\frac{2k+1}{k^2}=0.$$

The Fabry condition is satisfied, but the Hadamard condition is not satisfied. ▶

This example once again proves that the Fabry condition is weaker than the Hadamard condition.

It implies the following question: is it possible to replace condition (4) with a weaker condition, say

$$\overline{\lim_{k\to\infty}}(n_{k+1}-n_k)=\infty \tag{6}$$

or

$$\underline{\lim_{k\to\infty}}\frac{k}{n_k}=0, \tag{7}$$

so that the conclusion about the non-continuity of series (2) through the unit disc still remains valid?

Fabry [2] and Faber [3] constructed examples showing that conditions (6) and (7) no longer lead to the desired result. Fabry for condition (6), and Faber for condition (7) constructed series that satisfy these conditions and have on the circle $|z|=1$ only one singular point. Mandelbroit [4] constructed an example of a series that satisfies condition (6) and has only one singular point in the entire extended plane. Let's take Faber's example.

Example 2. Let a power series be given

$$f(z)=\sum_{k=1}^{\infty}\left(\frac{z+z^2}{2}\right)^{n_k}=\sum_{n=1}^{\infty}c_nz^n. \tag{8}$$

Here $\{n_k\}$ is some increasing sequence of natural numbers. This series converges uniformly inside the lemniscate

$$|z(z+1)| = 2 \tag{9}$$

and under the condition $k = o(n_k)$ this lemniscate, by Fabry's theorem on lacunae, is a natural boundary for it. Lemniscate (9) contains the disc $|z| < 1$ inside itself and touches it at the point $z = 1$.

Let's take $\{n_k\} = \{k!\}$. Then, $\dfrac{n_{k+1}}{n_k} \to \infty$ and the power series

$$f(z) = \sum_{k=0}^{\infty} \left(\frac{z+z^2}{2}\right)^{n_k} = \sum_{k=0}^{\infty} \frac{1}{2^{n_k}} \left[z^{n_k} + n_k z^{n_k+1} + \cdots + z^{2n_k}\right].$$

Since, $2n_k < n_{k+1}$, $k > 1$, then this series is a lacunary series of Ostrovsky (see Definition 2)

$$f(z) = \sum_{n=1}^{\infty} c_n z^n,$$

with convergence radius $R = 1$ with lacunae $c_n = 0, n \in [2n_k + 1, n_{k+1} - 1]$. In addition, it is easy to calculate that

$$\varliminf_{k\to\infty} \frac{k}{n_k} = 0,$$

but $\overline{\lim\limits_{k\to\infty}} \dfrac{k}{n_k} = \dfrac{1}{2}$. ►

This issue was brought to full clarity by Pólya [5], who proved the following statement (Pólya's theorem on lacunae): *if condition* (7) *is satisfied for a lacunary series* (2)*, then this series defines a single-valued holomorphic function with a univalent domain of existence, i.e. its natural domain of existence is* $W_f \subset \mathrm{C}$.

An analogous result holds for power series with Ostrovsky lacunae.

Definition 2. *A power series* (1) *is said to have Ostrovsky lacunae if there exist two strictly increasing sequences* $\{n_k\}$ *and* $\{n'_k\}$ *of natural numbers such that* $c_n = 0$ *if* $n_k < n \leqslant n'_k$ $(k = 0, 1, 2, \dots)$ *and*

$$\varliminf_{k\to\infty} \frac{n_k}{n'_k} = 0. \tag{10}$$

Example 3. If $\{n_k\} = \{k^k\}$ and $\{n'_k\} = \{k^{k+1}\}$, then these sequences satisfy condition (10) and series (1) with lacunae on the intervals $(n_k, n'_k]$, $(k = 2, 3, 4, \dots)$ will be Ostrovsky's lacunary series.

Ostrovsky's lacunary series can always be written in the form

$$f(z) = \sum_{k=0}^{\infty} c_{n_k} z^{n_k}$$

assuming $n_{k+1} = \max\{j > n_k : c_j = 0\} + 1$. Then, $c_n = 0$, for $n_k < n < n_{k+1} (k\ , = 0, 1, 2, \dots)$, $n_{k+1} \geqslant n'_k$ and $\varliminf\limits_{k\to\infty} \dfrac{n_k}{n_{k+1}} = 0$.

Note that Pólya's lacunary series is Ostrovsky's lacunary series, but the converse is not true: Ostrovsky's condition $\varliminf\limits_{k\to\infty} \dfrac{n_k}{n_{k+1}} = 0$ is much weaker than the Pólya condition.

For a power series (1) with Ostrovsky lacunae, the following Ostrovsky theorem holds [6]: *let series* (1) *have Ostrovsky lacunae. Then the lacunary power series* (1) *defines a single-valued holomorphic function with a simply connected domain of existence* $W_f \subset \mathrm{C}$.

In [5], Pólya obtained the following inversion of Pólya's theorem on lacunae: *let the inequality* (2) *be satisfied for the lacunary series*

$$\varliminf_{k\to\infty} \frac{k}{n_k} > 0,$$

Then it is always possible to select the coefficients c_{n_k} *in such a way that series* (2) *gives, upon analytical continuation, a multi-valued analytic function.*

Many scientific works are devoted to the study of singular points of the lacunary series (2). They are presented in detail in the monograph by L. Bieberbach [7].

2. Multidimensional lacunary Hartogs series with Ostrovsky lacunae

It is known that many problems of multidimensional complex analysis are solved using Hartogs series and Jacobi–Hartogs series (see, for example, [13–19], etc.). In this regard, it is of interest to study the domain of convergence of such series. In this work we will study Hartogs series with Ostrovsky lacunae.

The domains of convergence and multidimensional analogues of the Fabry and Pólya theorems for Hartogs series and series in homogeneous polynomials were studied in [8–11] and others. This article is devoted to the study of the domains of existence of Hartogs series with Ostrovsky lacunae.

Consider the Hartogs series

$$f(z,w) = \sum_{n=0}^{\infty} c_n(z)\, w^n \tag{11}$$

with holomorphic coefficients $c_n(z) \in \mathcal{O}(D)$, $n = 0, 1, 2, \ldots$, where $D \subset \mathrm{C}^n$ is some domain. In this case, we will require the natural condition that the radius of convergence $R(z)$ of series (11) is positive for each fixed point $z \in D$.

Let's start with the following result, which is a multidimensional analogue of Ostrovsky's theorem on lacunae.

Theorem 1. *Let series* (11) *satisfy the following conditions:*

1) converges uniformly inside the domain $D \times \{|w| < r\} \subset \mathrm{C}_z^n \times \mathrm{C}_w$, $r > 0$;

2) has Ostrovsky lacunae.

Then series (11) *defines a unique holomorphic function* $f(z,w)$ *in the domain* W^0, *where* W^0 *is the open kernel of the set* $W = \bigcup_{z\in D} W_f(z)$ *and* $W_f(z)$ *is the maximum domain into which* $f(z,w)$ *extends for a fixed* $z \in D$.

To prove this theorem we need the following statement.

Theorem 2 ([16]). *Consider a holomorphic in the domain* $D \times \{|w| < r\} \subset \mathrm{C}_z^n \times \mathrm{C}_w$, $r > 0$, *function* $f(z,w)$ *such that for every fixed* $z \in D$ *the function* $f(z,w)$ *holomorphic in the disk* $|w| < r$ *and univalent in* C_w, *i.e.* $W_f(z) \subset \mathrm{C}_w$.

Let $W = \bigcup_{z\in D} W_f(z)$. *Then*

1) the open kernel W^0 *is a domain (i.e., it is connected), and the set* $P = np_D\left(W \backslash W^0\right)$ *is pluripolar in* D;

2) the function $f(z,w)$ *extends holomorphically into* W^0.

Proof of Theorem 1. From condition 1) of Theorem 1 it follows that

$f(z,w) \in \mathcal{O}(D \times \{|w| < r\}) \subset \mathrm{C}_z^n \times \mathrm{C}_w$, $r > 0$.

We fix an arbitrary point $z \in D$. Then series (11) turns into an ordinary one-dimensional power lacunary series and by condition 2) according to the classical Ostrovsky theorem on lacunae we obtain that for each fixed $z \in D$ the function $f(z,w)$ is holomorphic in the disk $|w| < r$, uniquely continues to C_w, i.e. $W_f(z) \subset \mathrm{C}_w$.
According to Theorem 2, the open kernel W^0 of the set $W = \bigcup_{z\in G} W_f(z)$ is domain and the function $f(z,w)$ extends holomorphically into W^0 and this completes the proof of Theorem 1. □

Using Theorems 1 and 2 we obtain the following multidimensional analogue of Ostrovsky's theorem for a lacunary Hartogs series with a variable radius of convergence (the main result of the work).

Theorem 3. *Let series* (11) *satisfy the following conditions:*
1) $c_n(z) \in \mathcal{O}(D), n = 0, 1, 2, \ldots$
2) for each fixed $z \in D$ *the series converges in the disk* $|w| < R(z)$, $R(z) > r > 0$;
3) has Ostrovsky lacunae, i.e. there are two strictly increasing sequences $\{n_k\}$ *and* $\{n'_k\}$ *of natural numbers such that* $c_n = 0$ *if* $n_k < n \leqslant n'_k$ $(k = 0, 1, 2, \ldots)$ *and*

$$\lim_{k\to\infty} \frac{n_k}{n'_k} = 0.$$

Then there exists a closed nowhere dense set $S \subset D$ *such that series* (11) *defines a unique holomorphic function* $f(z,w)$ *in the domain* $W^0 \backslash (S \times \mathrm{C})$, *where* W^0 *is the open kernel of the set* $W = \bigcup_{z\in D} W_f(z)$.

To prove this theorem we need the following statement.

An analogue of Hartogs' lemma on the upper limit ([19]). *Let a sequence of plurisubharmonic functions* $u_k(z)$ *be given locally uniformly bounded above in the domain* $D \subset \mathrm{C}^n$ *and a real-valued function* $A(z) \in C(D)$ *such that at each fixed point* $z \in D$ *the inequality*

$$\overline{\lim_{k\to\infty}} u_k(z) \leqslant A(z). \tag{12}$$

Then for any compact set $K \subset\subset D$ *and any number* $\varepsilon > 0$ *there is a number* $k_0 \in N$ *such that*

$$u_k(z) \leqslant A(z) + \varepsilon \tag{13}$$

for all $z \in K$ *and* $k > k_0$.

Proof of Theorem 3.
1^0. From the conditions of the theorem it follows that series (11) for each fixed $z^0 \in D$ defines the germ

$$\sum_{n=0}^{\infty} c_n\left(z^0\right) w^n$$

in point $w = 0$, i.e. the radius of convergence of this series $R(z^0) > 0$. Let's consider the sets

$$E_m = \left\{ z \in D : |c_n(z)|^{\frac{1}{n}} \leqslant m,\ n = 1, 2, \ldots \right\}, m = 1, 2, \ldots .$$

By construction and from the holomorphicity of the coefficients $c_n(z)$ in D it follows that these sets are closed, $E_1 \subset E_2 \subset \ldots$ and since $R(z^0) > 0$ for all $z^0 \in D$, then

$$D = \bigcup_{m=1}^{\infty} \mathrm{E}_m.$$

According to Baire's theorem on categories [20], the domain D cannot be represented as a union of nowhere dense sets. Therefore, there exists $m \in N$ such that E_m has an interior point z^0, i.e. there is a ball $B(z^0, \varepsilon)$ such that $B(z^0,\ \varepsilon) \subset E_m$. By the definition of E_m, for all $z \in B(z^0, \varepsilon)$ the inequalities are satisfied

$$\frac{1}{n} \ln |c_n(z)| \leqslant \ln m,\ n = 1,\ 2, \ldots .$$

Let us now denote by D_1 the set of all points $z^0 \in D$ such that for $z^0 \in D_1$ there exists a ball $B(z^0, \varepsilon)$ and a number $M(z^0) \in N$, for which the inequalities $\frac{1}{n} \ln |c_n(z)| \leqslant \ln M(z^0)$, $z \in B(z^0, \varepsilon)\,(n = 1, 2, \ldots)$, i.e. in some neighborhood of the point z^0 all functions $\frac{1}{n} \ln |c_n(z)|$ are bounded from above by some constant $M(z^0)$. Then the set $S = D \backslash D_1$ is closed and nowhere dense in D.

Indeed, if not so, then S contains some neighborhood $U : U \subset S$. Then, by what was proved above, there exists a ball $B \subset U$ such that in this ball the functions $\frac{1}{n} \ln |c_n(z)|$ are uniformly bounded from above, i.e. for some M inequality

$$\frac{1}{n} \ln |c_n(z)| \leqslant \ln M$$

holds for all $z \in B$ and $k = 1, 2, \ldots$. The resulting contradiction proves that the set S is nowhere dense.

2^0. Let now

$$R(z) = \frac{1}{\overline{\lim\limits_{n\to\infty}} |c_n(z)|^{\frac{1}{n}}}$$

be the radius of convergence of series (11). Then

$$-\ln R(z) = \overline{\lim_{n\to\infty}} \frac{1}{n} \ln |c_n(z)|$$

and due to the plurisubharmonicity and local boundedness from above in D_1 of the functions $\frac{1}{n} \ln |c_n(z)|$ we obtain that the function $-\ln R_*(z)$ will be plurisubharmonic in D_1, where $R_*(z) = \underline{\lim\limits_{\xi\to z}} R(\xi)$ is the lower regularization of the radius-function $R(z)$. $R_*(z) \leqslant R(z)$ and outside some pluripolar set $\subset D$ the equality $R_*(z) = R(z)$ [21].

Note that in D_1 the function $R(z)$ is locally uniformly bounded from zero; this fact easily follows from the fact that in D_1 the functions $\frac{1}{n} \ln |c_n(z)|$ are locally uniformly bounded from above. It follows that $R_*(z)$ is also locally bounded away from zero. Let us prove the uniform convergence of series (11) inside the domain $\{(z, w) : z \in D \backslash S, |w| < R_*(z)\}$. To do this, we take an arbitrary compact $K \subset\subset D_1$. From the definition of $R(z)$ for any fixed $z \in D_1$ correct relation

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \ln |c_n(z)| = -\ln R(z) \leqslant -\ln R_*(z).$$

Since $-\ln R_*(z) \in psh(D_1)$, then for an arbitrary domain G satisfying the condition $K \subset G \subset\subset D_1$, there exists monotonically decreasing sequence $v_j(z) \in psh(G) \bigcap C^\infty(G)$ such that $v_j(z) \searrow -\ln R_*(z)$ as $j \to \infty$. Using the notation

$$w_j(z) = \exp\{-v_j(z)\},$$

we get $w_j \in C^\infty(G)$ and $w_j(z) \nearrow R_*(z)$ at $j \to \infty$. Then, according to the analogue of the Hartogs lemma on the upper limit, for any number $j \in N$ and any number $\varepsilon > 0$ there is a number n_0 such that inequality holds

$$\frac{1}{n} \ln |c_n(z)| \leqslant v_j(z) - \ln(1-\varepsilon) = -\ln(1-\varepsilon) w_j(z), \quad n \geqslant n_0, \ z \in K.$$

From here we have

$$|c_n(z)| \leqslant \frac{1}{[(1-\varepsilon) w_j(z)]^n}, \quad n \geqslant n_0, \ z \in K.$$

This proves that series (11) converges absolutely and uniformly on the set

$$\{(z,w) : z \in K, |w| < (1-\varepsilon) w_j(z)\}.$$

Due to the arbitrariness of the compact set $K \subset D \backslash S$ and the number $\varepsilon > 0$ and the fact that $w_j\ (z) \nearrow R_*(z)$ with $j \to \infty$, follows absolute and uniform convergence of series (11) inside the domain

$$\{(z,w) : z \in D \backslash S, |w| < R_*(z)\}.$$

3^0. Let $D_0 \subset\subset D \backslash S = D_1$ arbitrary domain. Then series (11) satisfies the following conditions:

1) series (11) converges uniformly inside the domain $D_0 \times \{|w| < R_*(z)\}$,

2) the coefficients of this series have Ostrovsky lacunae,

that is series (11) satisfies the conditions of Theorem 1.

Then, by Theorem 1, series (11) defines a unique holomorphic function $f(z,w)$ in the domain

$$\left(\bigcup_{z \in D_0 \setminus S} W_f\,(z) \right)^0.$$

From the arbitrariness of the domain D_0 we find that the sum of series (11) defines a unique holomorphic function $f(z,w)$ in the domain $W^0 \backslash (S \times \mathrm{C})$, where W^0 is the open kernel of the set $W = \bigcup_{z \in D} W_f\,(z)$. Theorem 3 is proven. □

Remark. If in Theorem 3 we do not require the condition $R(z) > 0$ for any fixed $z \in D$, then the statement of the theorem will not be true, i.e. Condition 2 of Theorem 3 is necessary (see [18]).

3. An analogue of Ostrovsky's theorem for series in homogeneous polynomials

Consider the formal lacunary series

$$\sum_{k=0}^{\infty} \mathrm{Q}_{n_k}(z) \tag{14}$$

over homogeneous polynomials $\mathrm{Q}_{n_k}(z)$ and we will investigate the domain of convergence of such a series.

In [8] J. Siciak considered the following situation. Let the function of n complex variables $f(z)$ be holomorphic in a neighborhood of zero and expanded into a lacunary series in homogeneous polynomials

$$f(z) = \sum_{k=0}^{\infty} \mathrm{Q}_{n_k}(z). \tag{15}$$

Let $\psi(z) := \overline{\lim\limits_{k\to\infty}} |\mathrm{Q}_{n_k}(z)|^{\frac{1}{n_k}}$ and $h(z) := \psi^*(z)$ be a regularization of the function ψ. Then it is known that the circular domain

$$D := \{z \in \mathrm{C}^n : h(z) < 1\}$$

is the domain of convergence of series (15). For this series, the following multidimensional analogue of Fabry's theorem holds.

Theorem 4 (Siciak [8])**.** *If for series* (15)

$$\overline{\lim_{k\to\infty}} \frac{k}{n_k} = \lim_{k\to\infty} \frac{k}{n_k} = 0,$$

then the domain of convergence D of series (15) *coincides with the natural domain of existence W_f of the function $f(z)$.*

In [9], using the complex theory of pluripotential, a simpler proof of this theorem is given and an analogue of the theorem is considered in more general situations; an analogue of Pólya's theorem is given $\Big($the case when $\underline{\lim\limits_{k\to\infty}} \dfrac{k}{n_k} = 0\Big)$. In the case when $\underline{\lim\limits_{k\to\infty}} \dfrac{k}{n_k} = 0$, series (15) can converge outside the domain D, i.e. D will not be a natural domain of existence of the function f. However, in [9] it was proven that f is a single-valued function in C^n, i.e. $W_f \subset \mathrm{C}^n$ and a simple description of W_f is given in terms of $W_f|_l$, $\ell \in \mathrm{P}^{n-1}$. In this section we will prove an analogue of Pólya's theorem for the lacunary Ostrovsky series in homogeneous polynomials with a variable radius of convergence.

Theorem 5. *Let the series* (15) *and the family of complex lines $L \subset \mathrm{P}^{n-1}$ satisfy the following conditions*

1) L is not pluripolar,

2) *series* (15) *has Ostrovsky lacunae,*

3) for each straight line $\ell \in L$, series (15) *converges in a disc of positive radius $R(\ell) \geqslant r > 0$.*

Then series (15) *defines a unique holomorphic function $f(z)$ in C^n, i.e. the natural domain of existence of W_f of the function f is univalent, $W_f \subset \mathrm{C}^n$. Moreover,*

1) $0 \in W_f$,

2) W_f coincides with the open kernel of the union $\bigcup\limits_{\ell\ni 0} W_{f,\ell}$, i.e.

$$W_f = \left(\bigcup_{\ell\in\mathrm{P}^{n-1}} W_{f,\ell}\right)^0,$$

3) for all complex lines $\ell \in \mathrm{P}^{n-1}$ with the exception of some pluripolar set $\mathrm{P} \subset \mathrm{P}^{n-1}$,

$$\ell \bigcap W_f = W_{f,\ell},\ \ell \notin \mathrm{P}.$$

To prove this theorem, we need the following statement, proven in [9].

Theorem 6. *Let a formal series*

$$\sum_{s=0}^{\infty} \mathrm{Q}_s(z)$$

be given by homogeneous polynomials Q_s and the family $L \subset \mathrm{P}^{n-1}$ of complex lines ℓ. If for each complex line $\ell \in L$ the series

$$\sum_{s=0}^{\infty} \mathrm{Q}_s(z)\ |_l$$

converges in the disc $\ell \cap B(0, r(l))$, $0 < r(l) \leqslant 1$, *then it converges uniformly inside the domain*

$$\Omega = \left\{ z \in \mathrm{C}^n : |z| \cdot \exp V^*\left(\frac{z}{|z|}, E\right) < 1 \right\}, \tag{16}$$

where $E = \bigcup_{l \in L}(l \cap S(0, r(l)))$, $S(0, r(l)) = \partial B(0, r(l))$ *and* $V^*(z, E)$ *is the Green's function of the set* E *(or the Siciak–Zakharyuta extremal function).*

Since the generalized Green's function $V^*(z, E) \equiv +\infty$ if and only if E is a pluripolar set in C^n, then from equality (16) it is clear that Theorem 6 is meaningful only if the set E is non-pluripolar in C^n, which we will also assume. We also exclude the trivial case when $\Omega = \mathrm{C}^n$.

Proof of Theorem 5. From the conditions of Theorem 5 it follows that series (15) satisfies the conditions of Theorem 6. Then, according to Theorem 6, series (15) converges uniformly in a certain neighborhood of zero, and its sum determines the function $f(z)$ holomorphic in this neighborhood

$$f(z) := \sum_{s=0}^{\infty} \mathrm{Q}_s(z). \tag{17}$$

Further on, we use the linear-fractional transformation $('z, z_n) \to ('w, z_n)$, where $w_\nu = \dfrac{z_\nu}{z_n}$ $(\nu = 1, 2, \ldots, n-1)$.

Let us map the domain under consideration into the Hartogs domain (see [9]). Then the straight lines $\ell \ni 0$ will turn into straight lines parallel to the axis z_n, and the neighborhood of zero into some polydisc $'U \times \{|z_n| < r\}, r > 0$. Consequently, the series in homogeneous polynomials (17) turns into a Hartogs series

$$g('z, z_n) = \sum_{k=0}^{\infty} c_k('z) z_n^k, \tag{18}$$

and this series satisfies the conditions of Theorem 1:

Let series (18) *be such that*

1) series (18) *converges uniformly in a polydisc*

$$'U \times \{|z_n| < r\} \subset \mathrm{C}_{'z}^{n-1} \times \mathrm{C}_{z_n}, \quad r > 0,$$

2) series (18) *has Ostrovsky lacunae.*

Then, according to Theorem 1, series (18) defines a single-valued holomorphic function with a univalent domain of existence, i.e.

$$g('z,\, z_n) \in \mathcal{O}\{('z,\, z_n) \in \mathrm{C}_{'z}^{n-1} \times \mathrm{C}_{z_n} : {}'z \in {}'U,\ |z_n| < R_*('z)\}$$

where $R_*('z) = \underline{\lim}_{'w \to 'z} R('w)$ is a lower regularization radius-functions $R('z)$ and $R('z)$ radius of the maximum disc, where series (18) converges, for a fixed $'z \in 'D$.

Next, using Theorem 2, we obtain the proof of Theorem 5. □

The authors express their sincere gratitude to Professor A. Sadullaev for setting the problem and useful discussions.

References

[1] J.Hadamard, Essai sur l'etude des fonctions donnes par leur development de Taylor, *J. de Math. Pures et appl.*, **8**(1892), no. 4, 101–186.

[2] E.Fabry, Sur les points singuliers d'une fonktion donnee par son developpement de Taylor, *Ann. ec. norm. sup. Paris*, **13**(1896), no. 3, 367–399.

[3] G.Faber, Über Potenzreihen mit unendlich vielen verschiedenen Koeffizienten. Sitzgsber . bayr. Akad Wiss. Mat.- nat. Abt., 1906, 581–583.

[4] S.Mandelbrojt, Modern researches on the singularities of functions defined by Taylor series, *Rice Inst. Pamphlet*, **14**(1927), 225–352.

[5] G.Polya, Untersuchungen über Lüchen und Singularitäten von Potenzreihen, *Math. Z.*, **29**(1929), 549–640.

[6] A.Ostrowski, On the representation of analytic functions by power series. *J. Lond. Math. Soc.*, **1**(1926), 236-239.

[7] L.Biberbakh, Analiticheskoe prodoljenie, Moscow, Nauka, 1967 (in Russian).

[8] J.Siciak, Separately analytic functions and envelopes of holomorphy of some lower dimensional subsets of C^N, *Ann. pol. math.*, **22**(1969), no. 1, 145–171.

[9] T.T.Tuychiev, On domains of convergence of multidimensional lacunary series, *Journal of Siberian Federal University. Mathematics & Physics*, **12**(2019), no. 6, 736–746. DOI: 10.17516/1997-1397-2019-12-6-736-746

[10] T.T.Tuychiev, A.P.Mardanov, O mnogomernykh analogakh teorem Fabri i Polia dlya lakunarnykh ryadov Khartogsa, *Uzbekskiy Matematicheskiy zhyrnal*, 2012, no. 3, 126–134 (in Russian).

[11] T.T.Tuychiev, A.P.Mardanov, Ob oblastyakh skhodimosti lakunarnykh ryadov po odnorodnym polinomam, *Uzbekskiy Matematicheskiy zhurnal*, 2014, no. 1, 108–118 (in Russian).

[12] B.V.Shabat, Vvedeniye v kompleksnyy analiz, Moscow, Nauka, 1985, ch. 1 (in Russian).

[13] J.Siciak, Separately analytic functions and envelopes of holomorphy of some lower dimensional subsets of C^N, *Ann. pol. math.*, **22**(1969), no. 1, 145–171.

[14] A.Sadullaev, Plurisubharmonic Functions, Itogi Nauki i Tekhn. Ser. Sovrem. Probl. Mat. Fundam. napravleniya, 8, VINITI, Moscow, 1985, 65–113.

[15] A.Sadullaev, Ye.M.Chirka, O prodolzhenii funktsiy s polyarnymi osobennostyami, *Mat. sb.*, **132**(1987), no. 3, 383–390 (in Russian).

[16] T.T.Tuychiev, Prodolzheniye funktsiy vdol' fiksirovannogo napravleniya, *Sib. mat. zhurn.*, **29**(1988), no. 3, 142–147 (in Russian).

[17] A.S.Sadullaev, T.T.Tuychiev, O prodolzhenii ryadov Khartogsa, dopuskayushikh golomorfnoye prodolzheniye na parallel'nyye secheniya. Uzbekskiy Matematicheskiy zhurnal, 2009, no. 1, 148–157 (in Russian).

[18] T.T.Tuychiev, J.K.Tishabaev, On the continuation of the hartogs series with holomorphic coefficients, *Bulletin of National University of Uzbekistan: Mathematics and Natural Sciences*, **2**(2019), no. 1, Article 5, 69–76. DOI: 10.56017/2181-1318.1021

[19] A.S.Sadullaev, Golomorfnyye funktsii mnogikh peremennykh, Urgench, 2005 (in Russian).

[20] A.N.Kolmogorov, Fomin Elementy teorii funktsiy i funktsianal'noga analiza, Moscow, Nauka, 1976 (in Russian).

[21] A.Sadullaev, Teoriya plyuripotentsiala. Primeneniya Palmarium akademik publishing, Saarbrüchen, Deutschland, 2012 (in Russian).

О многомерных лакунарных рядах Хартогса с лакунами Островского

Арслан П. Марданов
Ташкентский Государственный Технический Университет им. И. Каримова
Ташкент, Узбекистан
Тахир Т. Туйчиев
Джурабай К. Тишабаев
Национальный университет Узбекистана им. М. Улугбека
Ташкент, Узбекистан

Аннотация. Данная работа посвящена многомерным аналогам теоремы Островского о лакунарных рядах. В работе исследуются области существования лакунарных рядов Хартогса с лакунами Островского и рядов по однородным полиномам. Приводятся аналоги теоремы Островского для таких рядов и описываются области сходимости этих рядов.

Ключевые слова: плюрисубгармоническая функция, особая точка, нигде не плотное множество, степенной ряд, лакунарный ряд Хартогса, ряд по однородным многочленам.

EDN: DFABKQ
УДК 517.5

On the One Model of Frank-Kamenetskii

Alexander M. Kytmanov*
Olga V. Khodos†
Siberian Federal University
Krasnoyarsk, Russian Federation
Davlat Kh. Djumabaev ‡
Branch of the Russian State University of Oil and Gas (NRU) named after I.M.Gubkin
Tashkent, Uzbekistan

Received 19.03.2026, received in revised form 09.04.2026, accepted 25.04.2026

Abstract. The work is devoted to the study of the real roots of a system of transcendental equations arising in the Frank-Kamenetskii model. It is shown that the number of real roots is related to the number of real roots of some entire function (resultant). The number of complex roots is investigated.

Keywords: systems of transcendental equations, resultant, simple root.

Citation: A. M. Kytmanov, O. V. Khodos, D. Kh. Djumabaev, On the One Model of Frank-Kamenetskii, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 320–326.
EDN: DFABKQ.

Introduction

Finding the number of real roots of polynomials is a classical problem in algebra. It has been studied using the Hermite method of quadratic forms, the Sturm method, Descartes' sign rule, and the Boudin-Fourier theorem (see, for example, [1]). Further development of these methods for polynomials can be found in [2] and the monograph [3]. For entire functions, the question of the localization of real positive roots was considered in the classical works of N. G. Chebotaryov [4] (pp. 28-56), as well as in the work [5].

For systems of equations, the number of real roots was studied in the papers [6–8]. In the paper [9], the number of real roots was related to the number of real roots of the resultant.

The monographs [10,11] consider algebraic and transcendental systems of equations. Systems of transcendental equations arise, for example, in the study of chemical kinetics equations [12]. One of the problems that arise there is the problem of the number of real positive roots of a system of equations in a reaction polyhedron.

As such, the system of equations arising in one model, corresponding to the autocatalytic scheme of the reaction, namely the Frank–Kamenetskii model, is investigated. It is, in a certain sense, the simplest of nonlinear systems, which has a substantially nonlinear property — the multiplicity of stationary states.

1. Resultant

Let us consider one of the models, the so-called Frank–Kamenetskii model, in a dimensionless form (see [12, Chapter 2])

*AKytmanov@sfu-kras.ru https://orcid.org/0000-0002-7394-1480
†khodos_o@mail.ru
‡djumabaevd@mail.ru https://orcid.org/0009-0003-0563-3974

$$\begin{cases} \dfrac{dy}{d\tau} = \beta f(y)x + s(1-y) = f_1(x,y), \\ \dfrac{dx}{d\tau} = k_1 z - f(y)x - k_2 x z^2 = f_2(x,y), \end{cases} \tag{1}$$

where $f(y) = Dae^{\gamma(1-1/y)}$, $z = 1-x$, x, y are dimensionless variables. All constants are positive.

The stationary states of system (1) are solutions of the stationary system which can be written as

$$\begin{cases} f_1(x,y) = 0, \\ f_2(x,y) = 0, \end{cases} \tag{2}$$

i.e.

$$\begin{cases} \beta Dae^{\gamma(1-1/y)} \cdot x + s(1-y) = 0, \\ k_1(1-x) - Dae^{\gamma(1-1/y)} \cdot x - k_2 x(1-x)^2 = 0. \end{cases} \tag{3}$$

Denoting $Da = b$, $t = \gamma\,(1-1/y)$, and transforming, we obtain the system

$$\begin{cases} \beta b(t-\gamma)e^t \cdot x + st = 0, \\ k_1(1-x) - be^t \cdot x - k_2 x(1-x)^2 = 0. \end{cases} \tag{4}$$

It is obvious that the system (4) has no roots with zero coordinates.

Earlier, in the work [13], the Zeldovich–Semenov model was investigated in a similar way, and then in the work [14], the Aris–Amundson model.

The main idea of the approach is to apply the results related to multidimensional residue theory, and to study power sums of roots and residue integrals (see [10, 11]).

Excluding

$$x = \frac{ste^{-t}}{\beta b(\gamma - t)}$$

from system (4), we obtain the equation

$$\begin{gathered} F(t) = k_2 s^3 t^3 e^{-3t} - 2k_2 s^2 t^2 \beta b(\gamma - t)e^{-2t} + (k_1 + k_2)\beta^2 b^2 st(\gamma - t)^2 e^{-t} + \\ + b^3 st\beta^2(\gamma - t)^2 - k_1\beta^3 b^3(\gamma - t)^3 = 0. \end{gathered} \tag{5}$$

Thus, the resultant of system (2) can be an entire function of the first order of growth $F(t)$. It is difficult to check for multiple zeros directly. However, we can do the following. Instead of $F(t)$, consider the function

$$\begin{gathered} \varphi(t) = e^{-3t} - 2s^{-1}t^{-1}\beta b(\gamma - t) \cdot e^{-2t} + k_2^{-1}(k_1 + k_2)\beta^2 b^2 s^{-2} t^{-2}(\gamma - t)^2 e^{-t} + \\ + k_2^{-1} s^{-2} t^{-2}\beta^2 b^3(\gamma - t)^2 - k_2^{-1} s^{-3} t^{-3} k_1 \beta^3 b^3(\gamma - t)^3. \end{gathered}$$

We will look at the equation $\varphi(t) = 0$ as a third-degree equation in the variable $w = e^{-t}$.

First, we will reduce this equation to a standard form by making a substitution

$$u = w - \frac{2}{3}s^{-1}t^{-1}\beta b(\gamma - t).$$

Then we get

$$\begin{gathered} \varphi(u) = u^3 + s^{-2}t^{-2}\beta^2 b^2(\gamma - t)^2 \left(-\frac{2}{3} + k_2(k_1 + k_2)\right) u - \\ - s^{-3}t^{-3}\beta^3 b^3(\gamma - t)^3 \left(\frac{16}{27} + k_2^{-1}k_1\right) + k_2^{-1}s^{-2}t^{-2}\beta^2 b^3(\gamma - t)^2. \end{gathered}$$

Thus, the function $\varphi(u)$ takes the form

$$\varphi(u) = u^3 + pu + q,$$

where

$$p = s^{-2}t^{-2}\beta^2 b^2(\gamma - t)^2\left(-\frac{2}{3} + k_2(k_1 + k_2)\right), \tag{6}$$

$$q = -s^{-3}t^{-3}\beta^3 b^3(\gamma - t)^3\left(\frac{16}{27} + k_2^{-1}k_1\right) + k_2^{-1}s^{-2}t^{-2}\beta^2 b^3(\gamma - t)^2. \tag{7}$$

Its roots u_j can be found using Cardano's formulas (see, for example, [15, Chapter 9]).

$$u_j = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{8}$$

The roots u_j can also be complex.

Therefore, the roots of the function φ in the variables t will take the form

$$e^{-t} = \frac{2}{3}s^{-1}t^{-1}\beta b(\gamma - t) + \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}},$$

where $p,\ q$ are defined by formulas (6) and (7).

Consider the difference

$$\psi(t) = e^{-t} - \frac{2}{3}s^{-1}t^{-1}\beta b(\gamma - t) - \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} - \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

It is clear that the multiple roots of the function $\varphi(t)$ are the multiple roots of the function $\psi(t)$.

Let us check ψ for multiple zeros. By calculating the derivatives of the function $\psi(t)$, equating them to the function ψ itself, and eliminating the exponent, we obtain an irrational expression involving power functions. By eliminating the irrationalities, we obtain a specific polynomial. By substituting its roots, for example, into the first equation, we find that the exponential function is equal to an irrational power function, which is impossible for almost all parameter values.

Therefore, the next statement is true.

Theorem 1. *For almost all values of the parameters, the function $\varphi(t)$ (and therefore $F(t)$) has no multiple roots.*

2. The number of real roots of the resultant

Next, we use the following statement (see [16]).

Theorem 2. *If the system (2) with real coefficients is such that it has no roots with zero coordinates and all zeros of the resultant $F(t)$ are simple, then the number of real roots of the system (2) coincides with the number of real roots of the resultant $F(t)$.*

Let us return to system (3). By expressing y from the second equation and substituting it into the first equation, we obtain

$$\beta bx e^{\gamma\left(1 - \frac{1}{y(x)}\right)} + s(1 - y(x)) = 0,$$

where

$$y(x) = 1 + (k_1 - k_2 x(1-x))(1-x)\frac{\beta}{s}.$$

Since

$$t = \gamma\,(1 - 1/y)\,,$$

we have

$$t = \frac{(k_1 - k_2 x(1-x))(1-x)\gamma\beta}{s + (k_1 - k_2 x(1-x))(1-x)\beta}.$$

Therefore, we obtain the equation

$$\beta b e^{\frac{(k_1-k_2x(1-x))(1-x)\gamma\beta}{s+(k_1-k_2x(1-x))(1-x)\beta}} + \frac{s}{x}(1 - y(x)) =$$

$$= \beta b e^{\frac{(k_1-k_2x(1-x))(1-x)\gamma\beta}{s+(k_1-k_2x(1-x))(1-x)\beta}} - \frac{\beta}{x}((k_1 - k_2 x(1-x))(1-x)) = 0.$$

That is, the equation

$$b e^{\frac{(k_1-k_2x(1-x))(1-x)\gamma}{s+(k_1-k_2x(1-x))(1-x)}} = \frac{1}{x}((k_1 - k_2 x(1-x))(1-x)) = \frac{k_1}{x} - k_2(1-x)^2. \tag{9}$$

It can be rewritten as follows

$$e^{\gamma} b e^{-\frac{\gamma}{y(x)}} = \frac{1}{x}((k_1 - k_2 x(1-x))(1-x)). \tag{10}$$

Next, we get

$$e^{\gamma} b e^{-\frac{\gamma}{y(x)}} + \frac{s}{x}(1 - y(x)) = 0,$$

or

$$\theta(x) = e^{-\frac{\gamma}{y(x)}} \cdot \frac{x}{y(x) - 1} = e^{-\gamma}\frac{s}{b}. \tag{11}$$

Let us see how many real roots the equation (11) can have. We will find the derivative of this function and set it equal to zero

$$\theta'(x) = e^{-\frac{\gamma}{y(x)}} \left(\frac{\gamma y'(x)}{y^2(x)} \cdot \frac{x}{y(x) - 1} + \frac{y(x) - 1 - x y'(x)}{(y(x) - 1)^2} \right) = 0.$$

Then we get the algebraic equation

$$\gamma x y'(x)(y(x) - 1) + y^2(x)(-y'(x)x + (y(x) - 1)) = 0. \tag{12}$$

We compute

$$y(x) - 1 - x y'(x) = \frac{\beta}{s}\left((k_1 - k_2 x(1-x))(1-x) - x(-k_1 + k_2 x(1-x) - (1-x)(k_2 - 2k_2 x)\right) =$$

$$= \frac{\beta}{s}(k_1 + 2k_2 x^2 - 2k_2 x^3).$$

Equation (12) is an algebraic equation of the ninth degree, so it can have at most 9 real roots. Since the roots of the resultant are simple, the roots of (12) give alternating maxima and minima of the resultant. Therefore, the resultant cannot have more than 10 real roots.

Let us investigate the behavior of the function $\theta(x)$ as $x \to \pm\infty$.

If $x \to +\infty$, then $y(x) \to -\infty$, and $\theta(x) \to -0$. If $x \to -\infty$, then $y(x) \to +\infty$, and $\theta(x) \to +0$.

The singular points of the two-sided infinite gap of the function $\theta(x)$ are the points where $y(x) = 1$ or $y(x) = 0$. These polynomials have degree 3, so each equation has either 1 or 3 simple real roots.

Therefore, the number of singular points of infinite gap for the function $\theta(x)$ is 2, 4, and 6. Moreover, at points where $y(x) = 1$, the function experiences an infinite two-sided gap of different signs, and at points where $y(x) = 0$, the function $\theta(x)$ experiences an infinite one-sided gap.

Therefore, equation (11) has at least one real root if equation $y(x) = 1$ has one real root. And if equation $y(x) = 1$ has 3 real roots, then equation (11) has at least three roots.

We also note that at the point where $y(x) = 0$, equation (11) has no roots on the positive part of the real axis, and the total number of real roots of equation (11) cannot exceed 3.

Let us investigate the equation $y(x) = 1$ in more detail. We have

$$\frac{s}{\beta}(y(x) - 1) = (k_1 - k_2x(1 - x))(1 - x) = -k_2x^3 + 2k_2x^2 - k_2x + k_1 = 0.$$

We find that the roots of the last equation satisfy the cubic equation

$$x^3 - 2x^2 + x - \frac{k_1}{k_2} = 0. \tag{13}$$

By replacing $u = x - \dfrac{2}{3}$ in (13), we obtain the equation

$$u^3 - \frac{1}{3}u + \frac{2}{27} - \frac{k_1}{k_2} = 0.$$

This is an incomplete cubic equation of the form

$$u^3 + pu^2 + q = 0,$$

where $p = -\dfrac{1}{3}$, $q = \dfrac{2}{27} - \dfrac{k_1}{k_2}$.

The number of its real roots is determined by the sign of the discriminant (see, for example, [15, Chapter 9])

$$D = -4p^3 - 27q^2 = \frac{4}{27} - 27\left(\frac{2}{27} - \frac{k_1}{k_2}\right)^2 = \frac{1}{27}\left(4 - \left(2 - \frac{27k_1}{k_2}\right)^2\right) = \frac{4k_1}{k_2} - 27\frac{k_1^2}{k_2^2} =$$

$$= \frac{k_1}{k_2^2}\left(4k_2 - k_1\right).$$

If $D > 0$, i.e. $4k_2 > k_1$, then equation (13) has 3 real roots, and if $D < 0$, i.e. $4k_2 < k_1$, then this equation has one real root.

So we get the statement

Theorem 3. *If $4k_2 > k_1$, then the resultant $F(t)$ (and therefore the system (4)) has at least 3 real roots, and if $4k_2 < k_1$, then the resultant $F(t)$, and therefore the system (4)), has at least one real root for almost all values of the parameters. The total number of real roots of the system (4)) does not exceed 10*

Since the exponential is a periodic function, the number of roots t_k of equation (11) is infinite, and $|t_k| \to \infty$ as $k \to \infty$.

Therefore,

$$x_k = \frac{st_ke^{-t_k}}{\beta b(\gamma - t_k)} \to 0 \quad \text{as} \quad k \to \infty,$$

$$y_k = y(x_k) \to 1 + \frac{k_1\beta}{s} \quad \text{as} \quad k \to \infty.$$

Therefore, we get the statement

Corollary 1. *The resultant $F(t)$ has an infinite set of complex roots t_k, $k \in \mathbb{N}$ and $x_k = x(t_k) \to 0$, and*

$$y_k = y(x_k) \to 1 + \frac{k_1\beta}{s} \quad as \quad k \to \infty.$$

This work is supported by the Krasnoyarsk Mathematical Center and financed by the Ministry of Science and Higher Education of the Russian Federation (Agreement № 075-02-2026-1314).

References

[1] F.R.Gantmakher, Theory of Matrices, New York, Chelsea Pub. Co., 1959.

[2] M.G.Krein, M.A.Naimark, The Method of Symmetric and Hermitian Forms in the Theory of the Separation of the Roots of Algebraic Equation, *Linear Multilin. Algebra*, **10**(1981), no. 4, 265–308.

[3] E.I.Jury, Inners and stability of dynamic systems, New York-London-Sydney-Toronto, Wiley, 1974.

[4] N.G.Chebotarev, Work Collection, Moscow-Leningrad, AN SSSR, Vol. 2, 1949 (in Russian).

[5] A.M.Kytmanov, O.V.Khodos, On localization of the zeros of an entire function of finite order of growth, *Journal Complex Analysis and Operator Theory*, **11**(2017), 393–416. DOI: 10.1007/s11785-016-0606-8

[6] L.A.Ajzenberg, V.A.Bolotov, A.K.Tsikh On the solution of systems of nonlinear algebraic equations using the multidimensional logarithmic residue. On the solvability in radicals, *Sov. Math. Dokl.*, **21**(1980), 645–648.

[7] N.N.Tarkhanov, Calculation of the Poincare index, *Izv. Vyssh. Uchebn. Zaved. Mat.*, (1984), no. 9, 47–50 (in Russian).

[8] A.M.Kytmanov, On the number of real roots of systems of equations, *Soviet Math. (Iz. VUZ)*, **35**(1991), no. 6, 19–22.

[9] A.M.Kytmanov, O.V.Khodos, On the roots of systems of transcendental equations, *Probl. Anal.*, **13**(2024), issue 1, 37–49. DOI: 10.15393/j3.art.2024.14430

[10] V.Bykov, A.Kytmanov, M.Lazman, M.Passare (ed), Elimination Methods in Polynomial Computer Algebra, Springer science+business media, Dordreht, 1998.

[11] A.M.Kytmanov, Algebraic and trascendental systems of equations and transcendental systems of equations, Krasnoyarsk, Siberian Federal University, 2019 (in Russian).

[12] V.I.Bykov, S.B.Tsybenova, Nonlinear models of chemical kinetics, Moscow, KRASAND, 2011 (in Russian).

[13] O.V.Khodos, On Some System of Non-algebraic Equation in $\mathbb{C}^n$, *J. Sib. Fed. Univ. Math. Phys.*, no. 4, **7**(2014), 455–465.

[14] A.M.Kytmanov, O.V.Khodos, On the Aris-Amundson model, *Journal SFU. Mathematics and Physics*, **18**(2025), no. 1, 51–58. edn: FVEGFC

[15] A.G.Kurosh, Higher Algebra, Moscow, Mir, 1972.

[16] A.M.Kytmanov, O.V.Khodos, On the Real Roots of Systems of Transcendental Equations, *J. Sib. Fed. Univ. Math. Phys.*, **17**(2024), no. 3, 328–335.

Об одной модели Франк-Каменецкого

Александр М. Кытманов
Ольга В. Ходос
Сибирский федеральный университет
Красноярск, Российская Федерация
Давлат Х. Джумабаев
Филиал Российского государственного университета нефти и газа (НИУ) им. И. М. Губкина
Ташкент, Узбекистан

Аннотация. Работа посвящена исследованию вещественных корней системы трансцендентных уравнений, возникающих в модели Франк-Каменецкого. Показано, что число вещественных корней связано с числом вещественных корней некоторой целой функции (результанта). Исследовано число комплексных корней.

Ключевые слова: системы трансцендентных уравнений, результант, простой корень.

Journal of Siberian Federal University. Mathematics & Physics 2026, 19(3), 327–334

EDN: LTYPOK
УДК 538.945

Three-level Superconductivity in YBCO Doped with NiO Nanoparticles

Sergey V. Semenov*
Kirensky Institute of Physics
Federal Research Center KSC SB RAS
Krasnoyarsk, Russian Federation
Siberian Federal University
Krasnoyarsk, Russian Federation
Mikhail I. Petrov†
Kirensky Institute of Physics
Federal Research Center KSC SB RAS
Krasnoyarsk, Russian Federation
Timofey D. Balaev‡
Maksim S. Molokeev§
Ivan V. Nemtsev¶
Kirensky Institute of Physics
Federal Research Center KSC SB RAS
Krasnoyarsk, Russian Federation
Siberian Federal University
Krasnoyarsk, Russian Federation
Viktor L. Kirillov‖
Oleg N. Martyanov**
Boreskov Institute of Catalysis SB RAS
Novosibirsk, Russian Federation
Denis M. Gokhfeld††
Kirensky Institute of Physics
Federal Research Center KSC SB RAS
Krasnoyarsk, Russian Federation

Received 10.10.2025, received in revised form 15.11.2025, accepted 21.12.2025

Abstract. We report the existence of three superconducting subsystems in samples of high-temperature superconductor $YBa_2Cu_3O_{7-\delta}$ doped by NiO nanoparticles and then subjected to 18-hour annealing. These subsystems exhibit different superconducting transition temperatures and respond differently to magnetic fields. While two-level superconductivity formed by grain and intergranular boundary subsystems is known in polycrystalline superconductors, our material reveals an additional subsystem formed by surface layers of grains with reduced critical temperature and enhanced sensitivity to magnetic field.

Keywords: high-temperature superconductivity, YBCO, nickel oxide nanoparticles, three-level model, flux pinning.

*svsemenov@iph.krasn.ru
†petrov@kirensky.ru https://orcid.org/0000-0001-7105-4995
‡tima123.balaev@mail.ru
§msmolokeev@mail.ru https://orcid.org/0000-0002-8297-0945
¶ivan_nemtsev@mail.ru https://orcid.org/0000-0002-4222-3099
‖VKirillov@catalysis.ru
**oleg@catalysis.ru https://orcid.org/0000-0001-9999-8680
††gokhfeld@iph.krasn.ru https://orcid.org/0000-0001-9049-9973

Citation: S.V. Semenov, M.I. Petrov, T.D. Balaev, M.S. Molokeev, I.V. Nemtsev, V.L. Kirillov, O.N. Martyanov, D.M. Gokhfeld, Three-level Superconductivity in YBCO Doped with NiO Nanoparticles, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 327–334. EDN: LTYPOK.

Introduction

Structures combining magnetic ordering and superconducting states are of particular interest for condensed matter physics [1]. Superconductors with embedded magnetic inclusions provide an excellent platform for discovering new physical phenomena and magnetic states. However, creating a superconductor with magnetic inclusions is quite challenging. Due to the chemical activity of high-temperature superconductors, adding inclusions usually leads to chemical interactions and degradation of superconductivity. Successful incorporation of a magnetic phase can also suppress the critical temperature, as superconducting correlations in real materials are relatively weak. At the same time, magnetic inclusions in superconductors can serve as particularly effective pinning centers and lead to a significant increase in critical current density [2–4].

Solid-state synthesis is one of the most accessible and popular methods for exploratory synthesis of modified and new superconducting materials [5]. When doping a polycrystalline superconductor with various impurities, the properties of both superconducting grains and intergranular boundaries are modified. Moreover, improvement of grain subsystem properties may be accompanied by a decrease in resistive transition temperature and an increase in resistance of the intergranular boundary subsystem. The contribution of two different subsystems to the resistance and magnetization of a polycrystalline superconductor is described by a two-level model [6–9].

Previously, we used solid-state synthesis to prepare bulk polycrystalline $YBa_2Cu_3O_{7-\delta}$ doped with ultra-small magnetic nanoparticles and investigate the effect of doping on the intragranular critical current density [10, 11]. However, the effect of doping with ultra-small NiO nanoparticles on the formation of superconducting subsystems and their critical temperatures have not been studied.

1. Experiment

The initial superconductor $YBa_2Cu_3O_{7-\delta}$ (YBCO) was prepared using standard solid-state synthesis from Y_2O_3, BaO, and CuO powders. NiO nanoparticles were obtained at the Institute of Catalysis SB RAS [12, 13]. The average size of NiO nanoparticles of 8 nm was determined from micrographs obtained by high-resolution transmission electron microscopy [13].

The required amount of NiO nanoparticles was added to the ground YBCO material. A tenfold amount of alcohol was added to the powder mixture, and the alcohol suspension was stirred. The alcohol was then evaporated with mild heating on a hot plate. The dry residue was thoroughly ground in a mortar and subjected to hydrostatic pressing (100 MPa). The pressed samples were placed in a furnace heated to 920°C and annealed at this temperature for 18 hours. Then the furnace with samples was cooled to 400°C and held at this temperature for 10 hours.

X-ray phase analysis of synthesized materials with nanoparticle content from 0 to 6 wt.% was performed on a Nanoyuan DX-2700BH powder diffractometer. Microstructure of samples was studied using a Hitachi TM4000Plus scanning electron microscope. Resistance measurements were carried out using the standard four-probe method for samples with mass fraction of NiO nanoparticles from $x = 0.001$ to $x = 0.04$. The sample dimensions were $8 \times 0.5 \times 0.5$ mm^3. The samples were cooled in zero external field.

To avoid sample heating during measurements and to obtain low-ohmic contact resistance, clamping gold-plated contacts were used. The transport current through the sample was supplied by a Keithley 2612A source. The voltage drop at potential contacts was measured by a Keithley 2182A nanovoltmeter. Temperature control and stabilization during measurements

were provided by a LakeShore 336 temperature controller. The magnetic field was generated by a FL-1 electromagnet. The transport current I value was 1 mA. The direction of the external magnetic field H was parallel to the direction of I.

2. Results

The peaks on the X-ray patterns correspond to the orthorhombic phase of $YBa_2Cu_3O_{7-\delta}$. NiO peaks are noticeable only in the sample with the highest nanoparticle content $x = 0.06$. The lattice parameters of YBCO determined by the Rietveld method for this sample are $a = 3.82032(11)$ Å, $b = 3.88613(18)$ Å, $c = 11.66301(64)$ Å, $V = 173.152(13)$ Å^3. With increasing NiO content, a slight decrease in the c lattice parameter of YBCO occurs from 11.67753 Å for $x = 0$ to 11.66301 Å for $x = 0.06$. This change indicates partial incorporation of Ni into the YBCO lattice.

According to scanning electron microscopy images, the obtained materials have a polycrystalline microstructure with an average grain size of about 3 μm.

Temperature dependences of resistance measured without applying an external magnetic field for samples with $x = 0.001$, $x = 0.005$, $x = 0.01$, $x = 0.02$, and $x = 0.04$ are shown in Fig. 1. The resistivity ρ values at 95 K (above the onset temperature of superconducting transition) increase with increasing x from 0.001 to 0.02. The $\rho(T)$ dependences show that upon cooling, the transition first occurs in the first subsystem (grains), then the superconducting state is established in the second subsystem (intergranular boundaries). Such $\rho(T)$ dependences are typical for polycrystalline YBCO. However, while in typical polycrystalline YBCO samples the transition width corresponding to resistance vanishing in grains is about 1–3 K, in synthesized samples this transition has a larger width (up to ≈10 K for samples with $x = 0.02$ and $x = 0.04$). The transition width of both first and second subsystems increases with increasing NiO nanoparticle content x. The specific contribution to the total resistance from the second subsystem also increases with increasing x.

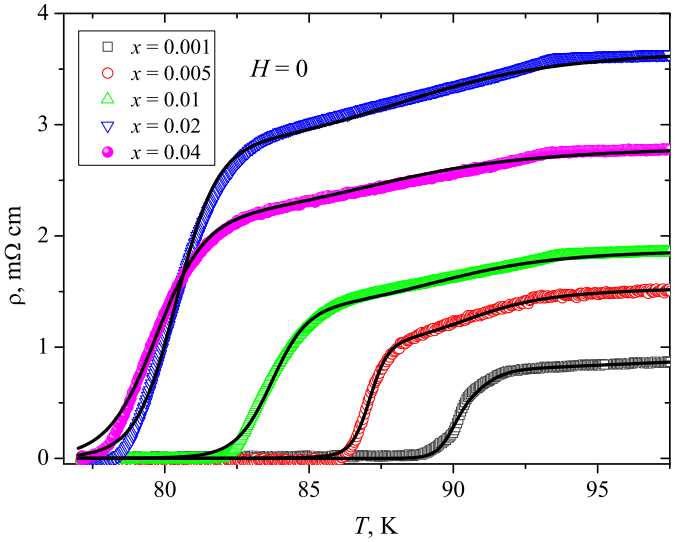


Fig. 1. Temperature dependence of resistivity of YBCO with different NiO content. Lines represent calculation results using formula (3)

Application of an external magnetic field separates the contribution from the second subsystem into two regions with different slopes (Fig. 2). For each region, one can identify its

own transition temperature and transition width. The region existing at lower temperatures is characterized by a strong dependence of transition width on magnetic field. This behavior fully corresponds to the intergranular boundary subsystem of polycrystalline YBCO. For the region at higher temperatures, the transition width also depends on the magnetic field. However, the increase in transition width with increasing magnetic field is weaker than for the intergranular boundary subsystem (see Discussion).

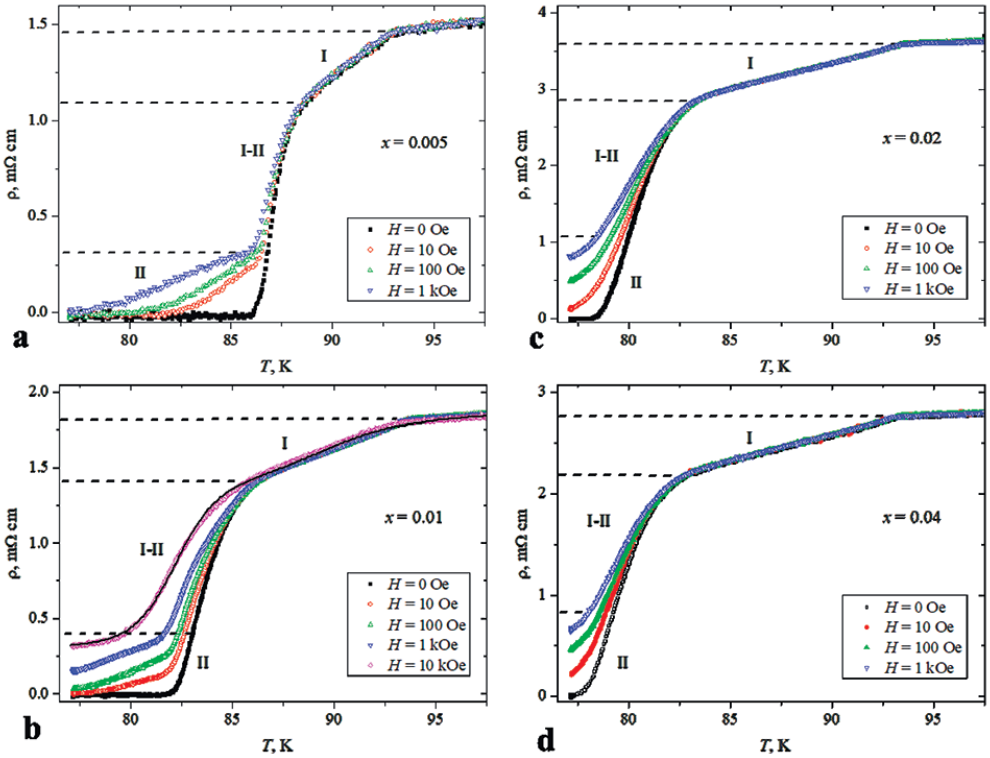


Fig. 2. Temperature dependence of resistivity of YBCO samples with $x = 0.005$ (a), $x = 0.01$ (b), $x = 0.02$ (c), $x = 0.04$ (d) in magnetic fields. Line in Fig. 2b represents calculation results using formula (3)

Hereafter, we denote quantities characterizing the subsystems using indices I for the grain subsystem, II for the intergranular boundary subsystem, and I-II for the subsystem weakly responding to magnetic field.

3. Discussion

The two-level model used for polycrystalline superconductors cannot explain the observed $R(T)$ dependences [6–9]. We propose to consider a three-level model in which the sample resistance is determined by contributions from three subsystems:

$$R(T) = R_{\mathrm{I}}(T) + R_{\mathrm{I\text{-}II}}(T) + R_{\mathrm{II}}(T). \tag{1}$$

The main goal of the description is to determine parameters of each subsystem: transition temperature of each subsystem T_{cI}, T_{cI-II}, T_{cII}, transition widths ΔT_I, ΔT_{I-II}, ΔT_{II} and contributions of subsystems to total resistance p_I, p_{I-II}, p_{II} (ratio of subsystem resistance to

total sample resistance at $T = 95$ K). The $R(T)$ dependences of each subsystem were described using a simple transition model [14] emplo'ying the Boltzmann function:

$$R_N \left[1 + \exp\left(-\frac{T - T_c}{0.125\Delta T}\right)\right]^{-1}. \quad (2)$$

The total resistance in the three-level model is determined as:

$$R(T) = \frac{p_{\rm I} R_N}{1 + \exp\left(-\frac{T - T_{c\rm I}}{0.125\Delta T_{\rm I}}\right)} + \frac{p_{\rm I\text{-}II} R_N}{1 + \exp\left(-\frac{T - T_{c\rm I\text{-}II}}{0.125\Delta T_{\rm I\text{-}II}}\right)} + \frac{p_{\rm II} R_N}{1 + \exp\left(-\frac{T - T_{c\rm II}}{0.125\Delta T_{\rm II}}\right)}. \quad (3)$$

The parameters obtained from describing experimental dependences for samples with different x in zero field (Fig. 1) are given in Tab. 1. Since contributions from the subsystems I-II and II were not separated on these dependences, the parameters for the total contribution from these subsystems are given. The parameters of the sample with $x = 0.001$ practically do not differ from the parameters of ordinary polycrystalline YBCO. With increasing x from 0.001 to 0.04, $T_{c\rm I}$ values decreased by 1.05 times, and $T_{c\rm II}$ values by 1.13 times. Meanwhile, $\Delta T_{\rm I}$ increased by 3.6 times, and $\Delta T_{\rm II}$ by 2.8 times. The contribution to the total sample resistance from the subsystems I-II and II increases with increasing x.

Table 1. Transition parameters in subsystems of samples in zero field

	Subsystem I. Grain cores			Subsystems I-II and II. Grain surfaces and intergranular boundaries		
x	T_c, K	ΔT, K	$\rho_{\rm I}$	T_c, K	ΔT, K	$\rho_{\rm I\text{-}II} + \rho_{\rm II}$
0.001	91.0	3.6	0.34	90.0	2.4	0.66
0.005	90.3	11	0.32	87.1	2.5	0.68
0.01	89.3	16	0.27	83.7	4.9	0.73
0.02	88	20	0.21	80.4	5.6	0.79
0.04	87	20	0.20	79.9	6.7	0.80

Tab. 2 shows parameter values of the sample with $x = 0.005$ at different magnetic fields. The values in the table demonstrate different influence of magnetic field on each of the subsystems. The width and transition temperature of the grain subsystem (I) did not change in magnetic fields up to 1 kOe. Independence of the grain subsystem from magnetic field up to ∼1 kOe is typical for ordinary polycrystalline YBCO samples. In the subsystem I-II, the transition temperature decreases by 1.003 times, and in the intergranular boundary subsystem (II) by 1.08 times with a hundredfold increase in magnetic field. Meanwhile, $\Delta T_{\rm I}$ increased by 1.5 times, and $\Delta T_{\rm II}$ by 3.6 times. Contributions to total sample resistance from different subsystems do not change with increasing H.

Table 2. Transition parameters in subsystems of sample with $x = 0.005$ in magnetic field

	Subsystem I. Grain cores			Subsystem I-II. Grain shells			Subsystem II. Intergranular boundaries		
H, Oe	T_c, K	ΔT, K	$\rho_{\rm I}$	T_c, K	ΔT, K	$\rho_{\rm I\text{-}II}$	T_c, K	ΔT, K	$\rho_{\rm II}$
0	90.3	11	0.32	87.2	2.5	0.51	87.1	2.5	0.17
10	90.3	11	0.32	87.2	2.8	0.51	85	7	0.17
100	90.3	11	0.32	87.1	3.2	0.51	83.5	8	0.17
1000	90.3	11	0.32	86.9	3.7	0.51	81	9	0.17

The obtained data allow us to make assumptions about the nature of superconducting subsystems in the studied samples. The subsystem I, in which superconductivity appears first upon

cooling, is formed by superconducting grains. The broadening of the transition in this subsystem can be explained by a gradient of critical temperature of the material from the grain surface to their center. This gradient arises due to interaction of NiO with YBCO during annealing. Meanwhile, Ni penetrates deep into grains, its concentration is maximum at grain surfaces and decreases toward their center. The critical temperature of YBCO decreases upon Ni incorporation [15–17], and a gradient of critical temperature arises in grains. Thus, in cores of large grains, the superconducting phase exists at $T < 95$ K, while for small grains the critical temperature decreases throughout the entire volume.

The subsystem II is formed by intergranular boundaries; the change in resistance of this subsystem in different fields and temperatures is completely identical to the behavior of a polycrystalline superconductor during the transition of its intergranular boundaries to the superconducting state.

The subsystem I-II qualitatively resembles the intergranular boundary subsystem, since the width of resistive transition in this subsystem increases with increasing magnetic field. However, the influence of magnetic field on this subsystem is weaker than on intergranular boundaries. It can be assumed that this subsystem is formed by material on grain surfaces with suppressed superconducting properties due to Ni incorporation. This "poisoned" grain surface quickly transitions to the superconducting state upon cooling, but upon heating contributes significantly to the total sample resistance.

The schematic presentation of three subsystems in polycrystalline superconductor is shown in Fig. 3.

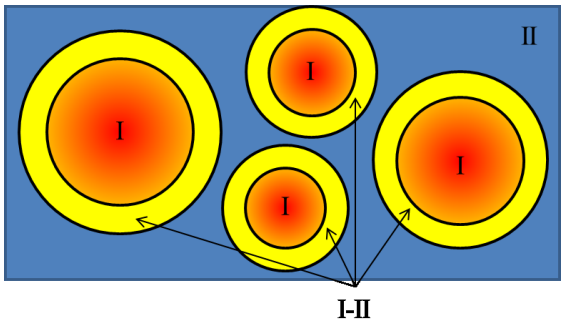


Fig. 3. Schematic of three-level model. The subsystem I is the core of grains. The subsystem I-II is the shell of grains. The subsystem II is the intergrain boundaries

Conclusions

Thus, we have demonstrated the realization of three-level superconductivity in YBCO samples doped with NiO nanoparticles. Subsystems with different resistive transition temperatures are formed by grain cores, grain shells, and intergranular boundaries.

The appearance of superconducting subsystems with special properties is possible when creating superconductors under high pressures (hydrides, nickelates, etc.) and may complicate the analysis of experimental data. Our work provides an example of accounting for contributions from different superconducting subsystems.

The measurements were carried out on the equipment of Krasnoyarsk Regional Center of Research Equipment of Federal Research Center "Krasnoyarsk Science Center SB RAS".

This study was supported by grant 24-22-00053 provided by the Russian Science Foundation, https://rscf.ru/project/24-22-00053//.

References

[1] D.Makarov, O.M.Volkov, A.Kakay, O.V.Pylypovskyi, B.Budinska, O.V.Dobrovolskiy, New Dimension in Magnetism and Superconductivity: 3D and Curvilinear Nanoarchitectures, *Adv. Mater.*, **34**(2022), 2101758. DOI: 10.1002/adma.202101758.

[2] Y.Zhao, C.H.Cheng, J.S.Wang, Flux pinning by NiO-induced nano-pinning centres in melt-textured YBCO superconductor, *Supercond. Sci. Technol.*, **18**(2005), S43–S46. DOI: 10.1088/0953-2048/18/2/010.

[3] A.N.Maksimova, I.A.Rudnev, I.A.Kashurnikov, A.N.Moroz, Effect of an Array of Submicron Magnetic Dots on Magnetization, Critical Current, and the Structure of Vortex Configurations in HTS, *Phys. Solid State.*, **65**(2023), 517. DOI: 10.21883/PSS.2023.04.55989.500.

[4] R.Kandari, M.Dahiya, M.Faraz, N.Khare, Effect of Magnetic NiO Nanoparticles on Pinning Properties of $YBa_2Cu_3O_{7-x}$ Superconductor, *J. Supercond. Nov. Magn.*, 38(2025), 37. DOI: 10.1007/s10948-024-06885-5.

[5] D.M.Gokhfeld, S.V.Semenov, I.V.Nemtsev, M.I.Petrov, Superconducting Tm123 granules on non-superconducting Pr123 seeds, *Phys. Solid State.*, **67**(2025), 1105–1110.

[6] L.Ji, M.S.Rzchowski, N.Anand, M.Tinkham, Magnetic-field-dependent surface resistance and two-level critical-state model for granular superconductors, *Phys. Rev. B.*, **47**(1993), 470–483. DOI: 10.1103/PhysRevB.47.470.

[7] V.V.Derevyanko, T.V.Sukhareva, V.A.Finkel', Effect of the Temperature, External Magnetic Field, and Transport Current on Electrical Properties, Vortex Structure Evolution Processes, and Phase Transitions in Subsystems of Superconducting Grains and "Weak Links" of Granular Two-Level High-Temperature Superconductor $YBa_2Cu_3O_{7-\delta}$, *Phys. Solid State.*, **60**(2018), 470–480. DOI: 10.1134/S1063783418030083.

[8] S.V.Semenov, A.D.Balaev, D.A.Balaev, Dissipation in granular high-temperature superconductors: New approach to describing the magnetoresistance hysteresis and the resistive transition in external magnetic fields, *J. Appl. Phys.*, **125**(2019), 033903. DOI: 10.1063/1.5066602.

[9] D.M.Gokhfeld, On Estimating the Critical Current Density in Polycrystalline Superconductors Synthesized by Solid-State Method, *J. Supercond. Nov. Magn.*, **36**(2023), 1089–1092. DOI: 10.1007/s10948-023-06575-8.

[10] S.V.Semenov, D.M.Gokhfeld, et al., Effect of ε-Fe_2O_3 on the superconducting properties of YBCO, *Phys. Solid State.*, **66**(2024), 2135–2140. DOI: journals.ioffe.ru/articles/60207.

[11] D.M.Gokhfeld, S.V.Semenov, et al., Effect of decorating NiO nanoparticles on superconducting properties of YBCO, arxiv cond-mat, 2025. DOI: 10.48550/arXiv.2510.14322.

[12] D.A.Balaev, A.A.Dubrovskiy, A.A.Krasikov, et al., Magnetic properties of NiO nano particles: Contributions of the antiferromagnetic and ferromagnetic subsystems in different magnetic field ranges up to 250 kOe, *Phys. Solid State.*, **59**(2017), 1547–1552. DOI: 10.1134/S1063783417080029.

[13] S.I.Popkov, A.A.Krasikov, A.A.Dubrovskiy, et al., Size effects in the formation of an uncompensated ferromagnetic moment in NiO nanoparticles, *J. Appl. Phys.*, **126**(2019), 103904. DOI: 10.1063/1.5109054.

[14] D.Gokhfeld, Description of Resistive Transition for Type-II Superconductors, *J. Supercond. Nov. Magn.*, **38**(2025), 32. DOI: 10.1007/s10948-024-06894-4.

[15] L.Shlyk, G.Krabbes, G.Fuchs, K.Nenkov, B.Schupp, Effects of annealing treatments on the superconducting properties of melt-processed YBCO doped with Ni, *Supercond. Sci. Technol.*, **18**(2004), S10. DOI: 10.1088/0953-2048/18/2/003.

[16] Y.X.Zhou, S.Scruggs, K.Salama, Effects of ionic doping on superconducting properties of melt textured $YBa_2(Cu_{1-x}M_x)_3O_{7-\delta}$ (M = Co, Ni, Zn or Ga) large grains, *Supercond. Sci. Technol.*, **19**(2006), S556. DOI: 10.1088/0953-2048/19/7/S26.

[17] H.Shakeripour, S.S.Hosseini, S.S.Ghotb, B.Hadi-Sichani, S.Pourasad, Magnetic doping effects on the superconductivity of $Y_{1-x}M_xBa_2Cu_3O_{7-\delta}$ (M = Fe, Co, Ni), *Ceram. Int.*, **47**(2021), 10635–10642. DOI: 10.1016/J.CERAMINT.2020.12.176.

Трехуровневая сверхпроводимость в YBCO, допированном наночастицами NiO

Сергей В. Семёнов
Институт физики им. Л. В. Киренского ФИЦ КНЦ СО РАН
Красноярск, Российская Федерация
Сибирский федеральный университет
Красноярск, Российская Федерация
Михаил И. Петров
Институт физики им. Л. В. Киренского ФИЦ КНЦ СО РАН
Красноярск, Российская Федерация
Тимофей Д. Балаев
Максим С. Молокеев
Иван В. Немцев
Институт физики им. Л. В. Киренского ФИЦ КНЦ СО РАН
Красноярск, Российская Федерация
Сибирский федеральный университет
Красноярск, Российская Федерация
Виктор Л. Кириллов
Олег Н. Мартьянов
Институт катализа им. Борескова СО РАН
Новосибирск, Российская Федерация
Денис М. Гохфельд
Институт физики им. Л. В. Киренского ФИЦ КНЦ СО РАН
Красноярск, Российская Федерация

Аннотация. Обнаружено существование трех сверхпроводящих подсистем в образцах высокотемпературного сверхпроводника $YBa_2Cu_3O_{7-\delta}$ с добавлением наночастиц NiO, подвергнутых восемнадцатичасовому отжигу. Эти подсистемы имеют разную температуру сверхпроводящего перехода и демонстрируют разный отклик на магнитное поле. В поликристаллических сверхпроводниках известна реализация двухуровневой сверхпроводимости, образованной подсистемами гранул и межгранульных границ. В полученном материале возникает новая подсистема, образованная поверхностными слоями гранул с пониженной критической температурой и повышенной чувствительностью к магнитному полю.

Ключевые слова: высокотемпературная сверхпроводимость, YBCO, наночастицы оксида никеля, трехуровневая модель, пиннинг магнитного потока.

Journal of Siberian Federal University. Mathematics & Physics 2026, 19(3), 335–346

EDN: KHDDDY
УДК 510.23

General Metalanguage for Formal Proofs

Nikolay A. Mishko*
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 10.06.2025, received in revised form 05.10.2025, accepted 07.02.2026

Abstract. Big-step semantics is presented for the language that is powerful enough to encode arbitrary formal systems with computable inference rules and metatheorems about them but simple enough to implement and to reason about and potentially suitable for self-verification.

Keywords: formal system, big-step semantics, proof verification.

Citation: N.A. Mishko, General Metalanguage for Formal Proofs, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 335–346. EDN: KHDDDY.

Introduction

Most modern proof checkers are based on exploiting dependent type theory and the Curry–Howard correspondence: propositions correspond to types, proofs to terms of a given type (that is, computer programs), proof checking to type checking of the constructed term. Many of them allow the user to introduce additional axioms, such as the axiom of choice or function extensionality.

However, these systems use a considerable number of built-in axioms and inference rules by construction (for example, the type theory used in Lean 3 consists of 52 inference rules [1], not counting computational ones and rules about algorithmic equality), which significantly limits the set of additional axioms that do not trivialize the system. For instance, 'subsingletion elimination' in Lean implies axiom K that is known [2] to be inconsistent with the univalence axiom, thus making Lean incompatible with homotopy type theory without extra tricks.

Moreover, the possibility of the introduction of additional reduction rules in such systems is associated with a number of problems; this makes it impossible to express, for example, cubical type theory without modifying the kernel of the system. It is also problematic to express theories (such as variants of modal type theory) based on the use of split contexts $\Delta \mid \Gamma \vdash x : A$ without explicit support for them in the kernel.

Metamath [3] contains only one built-in inference rule — the substitution rule, — the usage of which is limited only by the free variable distinctness conditions specified by the author of the axiomatic system; which makes it possible (potentially) to formulate an arbitrary axiomatic system admitting this rule. However, Metamath doesn't contain its own metaprogramming tools, and the theorem-as-string representation significantly complicates the use of Metamath as a compilation target language; therefore, for example, within the Hilbert-style calculus there is no possibility to express the deduction theorem. In addition, such a representation makes formula manipulation difficult inside Metamath too: that is, for instance, the strings `ph -> ps -> ch` and `( ph -> ( ps -> ch ) )` are different, which leads to obvious problems.

*berlinfall@inbox.ru

LCF-style systems [4] use strongly typed general-purpose programming language — Standard ML or derived — as a metalanguage. This solves the problems described above for Metamath but complicates the kernel. Standard ML doesn't support metaprogramming (in relation to itself), which significantly limits its expressiveness, especially in the context of proof verification.

Languages of the LISP family [5] support metaprogramming perhaps in the most elegant way; but — being dynamic languages — they lack reliable encapsulation tools essential for the implementation of LCF-like systems. It is proposed to solve the encapsulation problem by introducing formulae and theorems as separate data types alongside numbers, strings, and so on. Next, we will construct a theoretical model for such a language.

1. Formulae and operations on them

Let us describe a fairly general syntax suitable for describing standard axiomatic theories. We fix some infinite set of names Σ, consisting, for example, of all possible sequences of Latin letters, possibly with a numeric subscript. So, for first-order logic, we may set

$$\Sigma = \{x_1, x_2, x_3, \ldots, \forall, \exists, =, \neg, \supset, \wedge, \vee\}.$$

We define the set of formulae Ω inductively:

$$\frac{x \in \Sigma}{\mathrm{Fv}(x) \in \Omega}, \quad \frac{n \in \mathbb{N}_{>0}}{\mathrm{Bv}(n) \in \Omega}, \quad \frac{f \in \Sigma \qquad e_1, \ldots, e_n \in \Omega}{f(e_1, \ldots, e_n) \in \Omega}, \quad \frac{Q \in \Sigma \qquad e \in \Omega}{Qe \in \Omega}.$$

The formula $\mathrm{Fv}(x)$ is called *a free variable,* and $\mathrm{Bv}(n)$ *a bound variable,* where the number n denotes the nesting level of the quantifier referred to by the bound variable. Thus, for example, in $\forall\forall\,\mathrm{Bv}(1)$ the bound variable refers to the inner quantifier, and in $\forall\forall\,\mathrm{Bv}(2)$, accordingly, to the outer one. This approach is called a locally nameless representation [6, 7].

Formulae of the form $f(e_1, \ldots, e_n)$ are used both to represent logical connectives (like $\supset(p, q)$ for the implication $p \supset q$ or $\vee(p, q)$ for the disjunction $p \vee q$) and non-logical ones — function symbols, — for example, $\circ(f, g)$ for the composition of morphisms $f \circ g$ in some category; as well as for logical and non-logical constants: for example, $\top()$ to denote logical truth or $0()$ to denote zero. Finally, similarly, Qe formulae are used for logical quantifiers, e.g., $\forall{=}(\mathrm{Bv}(1), \mathrm{Bv}(1))$ for $\forall x\,.\,x = x$ or $\exists\neg(=(\mathrm{Bv}(1), 0()))$ for $\exists y\,.\,y \neq 0$; and for quantifiers in a broad sense: $\lambda(\mathrm{Bv}(1))$ for λ-abstraction $\lambda x\,.\,x$, $\varepsilon(\forall(\neg(\in(\mathrm{Bv}(1), \mathrm{Bv}(2)))))$ for ε-operator $\varepsilon x\,.\,\forall y\,.\,y \notin x$, or $\int(f(\mathrm{Bv}(1), 0()))$ for the integral $\int f(x, 0)\,\mathrm{d}x$.

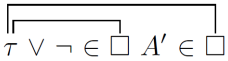
For example, the formula

$$\forall x\,.\,x = x_1 \vee x \neq x_2 \supset \exists y\,.\,x = y$$

will be written as follows:

$$\forall\supset(\vee(=(\mathrm{Bv}(1), \mathrm{Fv}(x_1)), \neg(=(\mathrm{Bv}(1), \mathrm{Fv}(x_2)))), \exists{=}(\mathrm{Bv}(2), \mathrm{Bv}(1))).$$

By and large, this notation is equivalent to that used in Bourbaki's 'Theory of Sets' [8]. Indeed, the formula



corresponds to the formula

$$\tau\vee(\neg(\in(\mathrm{Bv}(1), \mathrm{Fv}(A'))), \in(\mathrm{Bv}(1), \mathrm{Fv}(A''))),$$

where a graphical link is used instead of a numerical index to denote the quantifier referred to by a given bound variable.

Let us formulate the induction principle for Ω.

Theorem. *Let $P(e)$ be a predicate defined on $e \in \Omega$ such that:*

1. $\forall x \in \Sigma \,.\, P(\mathrm{Fv}(x))$.
2. $\forall n \in \mathbb{N}_{>0} \,.\, P(\mathrm{Bv}(n))$.
3. $\forall f \in \Sigma \; \forall e_1, \ldots, e_n \in \Omega \,.\, P(e_1) \wedge \ldots \wedge P(e_n) \Rightarrow P(f(e_1, \ldots, e_n))$.
4. $\forall Q \in \Sigma \; \forall e \in \Omega \,.\, P(e) \Rightarrow P(Qe)$.

Then $\forall e \in \Omega \,.\, P(e)$.

Free variable $x \in \Sigma$ *capturing* $x^\flat = x_1^\flat : \Omega \to \Omega$ and (its inverse) bound variable *freeing* $t^\sharp = t_1^\sharp : \Omega \to \Omega$ operations are implemented recursively:

$$x_k^\flat(\mathrm{Fv}(y)) = \begin{cases} \mathrm{Bv}(k), & \text{if } x = y, \\ \mathrm{Fv}(y), & \text{otherwise,} \end{cases} \qquad t_k^\sharp(\mathrm{Fv}(x)) = \mathrm{Fv}(x),$$

$$x_k^\flat(\mathrm{Bv}(n)) = \mathrm{Bv}(n), \qquad t_k^\sharp(\mathrm{Bv}(k')) = \begin{cases} t, & \text{if } k = k', \\ \mathrm{Bv}(k'), & \text{otherwise,} \end{cases}$$

$$x_k^\flat(f(e_1, \ldots, e_m)) = f(x_k^\flat(e_1), \ldots, x_k^\flat(e_m)), \qquad t_k^\sharp(f(e_1, \ldots, e_m)) = f(t_k^\sharp(e_1), \ldots, t_k^\sharp(e_m)),$$

$$x_k^\flat(Qe) = Q x_{k+1}^\flat(e), \qquad t_k^\sharp(Qe) = Q t_{k+1}^\sharp(e).$$

Let us define a set of free variables $\mathbf{Fv} : \Omega \to \mathcal{P}(\Sigma)$:

$$\begin{aligned} \mathbf{Fv}(\mathrm{Fv}(x)) &= \{x\}, \\ \mathbf{Fv}(\mathrm{Bv}(k)) &= \varnothing, \\ \mathbf{Fv}(f(e_1, \ldots, e_m)) &= \mathbf{Fv}(e_1) \cup \ldots \cup \mathbf{Fv}(e_m), \\ \mathbf{Fv}(Qe) &= \mathbf{Fv}(e). \end{aligned}$$

Lemma 1. $\mathbf{Fv}(x^\flat(e)) = \mathbf{Fv}(e) \setminus \{x\}$ *and* $\mathbf{Fv}(e) \subseteq \mathbf{Fv}(t^\sharp(e)) \subseteq \mathbf{Fv}(e) \cup \mathbf{Fv}(t)$.

Lemma 2. *Let $x \in \Sigma$. Then $x \notin \mathbf{Fv}(e) \Leftrightarrow x^\flat(\mathrm{Fv}(x)^\sharp(e)) = e$.*

Proof. We see that $\mathbf{Fv}(x^\flat(\mathrm{Fv}(x)^\sharp(e))) = \mathbf{Fv}(\mathrm{Fv}(x)^\sharp(e)) \setminus \{x\} \subseteq (\mathbf{Fv}(e) \cup \{x\}) \setminus \{x\} = \mathbf{Fv}(e) \setminus \{x\}$, so if $x \in \mathbf{Fv}(e)$ and $x^\flat(\mathrm{Fv}(x)^\sharp(e)) = e$, then $x \in \mathbf{Fv}(e) = \mathbf{Fv}(x^\flat(\mathrm{Fv}(x)^\sharp(e))) \subseteq \mathbf{Fv}(e) \setminus \{x\}$, but this is impossible.

Conversely, it suffices to consider the predicate $P(e) = \Big(\forall k \,.\, x \notin \mathbf{Fv}(e) \Rightarrow x_k^\flat(\mathrm{Fv}(x)_k^\sharp(e)) = e\Big)$ and prove by induction that $\forall e \in \Omega \,.\, P(e)$. □

Note that not all formulae in Ω are equally meaningful: indeed, a free variable in $\forall\, \mathrm{Bv}(2)$ refers to the non-existent quantifier. Following the terminology from [6], formulae that do not contain such 'dangling' (or 'loose') variables are called *locally closed*. We express this property using the recursive predicate $\mathrm{lc}(e) = \mathrm{lc}_1(e)$:

$$\begin{aligned} \mathrm{lc}_k(\mathrm{Fv}(x)) &= \top, \\ \mathrm{lc}_k(\mathrm{Bv}(n)) &= n < k, \\ \mathrm{lc}_k(f(e_1, \ldots, e_n)) &= \mathrm{lc}_k(e_1) \wedge \ldots \wedge \mathrm{lc}_k(e_n), \\ \mathrm{lc}_k(Qe) &= \mathrm{lc}_{k+1}(e). \end{aligned}$$

We also denote $\Omega_{\mathrm{lc}}(k) = \{e \in \Omega \mid \mathrm{lc}_k(e)\}$ and $\Omega_{\mathrm{lc}} = \Omega_{\mathrm{lc}}(1)$. So there is a chain of inclusions $\Omega_{\mathrm{lc}} = \Omega_{\mathrm{lc}}(1) \subset \Omega_{\mathrm{lc}}(2) \subset \ldots \subset \Omega_{\mathrm{lc}}(k) \subset \Omega_{\mathrm{lc}}(k+1) \subset \ldots \subset \Omega$. Moreover, by the definition of local closedness, $Q\,(\Omega_{\mathrm{lc}}(k+1)) \subset \Omega_{\mathrm{lc}}(k)$.

Lemma 3. *If* $\mathrm{lc}(t)$*, then* $t^\sharp_k\,(\Omega_{\mathrm{lc}}(k+1)) \subset \Omega_{\mathrm{lc}}(k)$ *and* $x^\flat_k\,(\Omega_{\mathrm{lc}}(k)) \subset \Omega_{\mathrm{lc}}(k+1)$.

Proof. It suffices to consider predicates

$$P_1(e) = \Big(\forall k\,.\,\mathrm{lc}_k(e) \Rightarrow \mathrm{lc}_{k+1}(x^\flat_k(e))\Big)\,,\ P_2(e) = \Big(\forall k\,.\,\mathrm{lc}_{k+1}(e) \Rightarrow \mathrm{lc}_k(t^\sharp_k(e))\Big)\,,$$

and prove by induction (in an obvious way) propositions $\forall e\,.\,P_1(e)$ and $\forall e\,.\,P_2(e)$. □

Thus, we have chains of mappings:

$$\Omega_{\mathrm{lc}} = \Omega_{\mathrm{lc}}(1) \xrightarrow{x^\flat_1} \Omega_{\mathrm{lc}}(2) \xrightarrow{x^\flat_2} \cdots \xrightarrow{x^\flat_{k-1}} \Omega_{\mathrm{lc}}(k) \xrightarrow{x^\flat_k} \Omega_{\mathrm{lc}}(k+1) \xrightarrow{x^\flat_{k+1}} \cdots\,,$$
$$\Omega_{\mathrm{lc}} = \Omega_{\mathrm{lc}}(1) \xleftarrow{t^\sharp_1} \Omega_{\mathrm{lc}}(2) \xleftarrow{t^\sharp_2} \cdots \xleftarrow{t^\sharp_{k-1}} \Omega_{\mathrm{lc}}(k) \xleftarrow{t^\sharp_k} \Omega_{\mathrm{lc}}(k+1) \xleftarrow{t^\sharp_{k+1}} \cdots\,.$$

Lemma. *Let* $x_1, \ldots, x_{k-1} \in \Sigma$*. If* $\mathrm{lc}_k(e)$*, then* $\mathrm{lc}\,\Big((\mathrm{Fv}(x_1)^\sharp_1 \circ \ldots \circ \mathrm{Fv}(x_{k-1})^\sharp_{k-1})(e)\Big)$.

Lemma 7. *Let* $L \subset \Sigma$ *be a finite set of names and* $P(e)$ *be a predicate defined on* $e \in \Omega_{\mathrm{lc}}$ *such that:*

1. $\forall x \in \Sigma\,.\,P(\mathrm{Fv}(x))$.
2. $\forall f \in \Sigma\ \forall e_1, \ldots, e_n \in \Omega_{\mathrm{lc}}\,.\,P(e_1) \wedge \ldots \wedge P(e_n) \Rightarrow P(f(e_1, \ldots, e_n))$.
3. $\forall Q \in \Sigma\ \forall x \in \Sigma \setminus L\ \forall e \in \Omega_{\mathrm{lc}}\,.\,P(e) \Rightarrow P(Qx^\flat(e))$.

Then $\forall e \in \Omega_{\mathrm{lc}}\,.\,P(e)$.

Proof. We define a family of predicates:

$$\widetilde{P}_k(e) = \Big[\forall x_1, \ldots, x_{k-1} \in \Sigma \setminus \mathbf{Fv}(e)\,.\,\mathrm{lc}_k(e) \Rightarrow P\,\Big((\mathrm{Fv}(x_1)^\sharp_1 \circ \ldots \circ \mathrm{Fv}(x_{k-1})^\sharp_{k-1})(e)\Big)\Big]\,.$$

We see that $\widetilde{P}_1(e) = \mathrm{lc}(e) \Rightarrow P(e)$, therefore $(\forall e \in \Omega_{\mathrm{lc}}\,.\,P(e)) \Leftrightarrow (\forall e \in \Omega\,.\,\widetilde{P}_1(e))$. We denote $\widetilde{P}(e) = \forall k \in \mathbb{N}_{>0}\,.\,\widetilde{P}_k(e)$. The proposition $\forall e \in \Omega\,.\,\widetilde{P}(e)$ is easy to prove by induction. □

Theorem 1. *Let* $P(e)$ *be a predicate defined on* $e \in \Omega_{\mathrm{lc}}$ *such that:*

1. $\forall x \in \Sigma\,.\,P(\mathrm{Fv}(x))$.
2. $\forall f \in \Sigma\ \forall e_1, \ldots, e_n \in \Omega_{\mathrm{lc}}\,.\,P(e_1) \wedge \ldots \wedge P(e_n) \Rightarrow P(f(e_1, \ldots, e_n))$.
3. $\forall Q, x \in \Sigma\ \forall e \in \Omega_{\mathrm{lc}}\,.\,P(e) \Rightarrow P(Qx^\flat(e))$.

Then $\forall e \in \Omega_{\mathrm{lc}}\,.\,P(e)$.

Proof. It suffices to apply Lemma 7 for $L = \varnothing$. □

2. Syntax and semantics of calculus

2.1. Syntax

We define two sets of expressions: 'external' syntactic $\mathcal{S}$ (which the user operates with) and 'internal' semantic $\mathcal{M}$ (which the user cannot — what is important — operate with directly). 'External' syntax consists of ordinary S-expressions:

$$\frac{x \in \Sigma}{x \in \mathcal{S}}, \quad \frac{s_1 \in \mathcal{S} \qquad s_2 \in \mathcal{S}}{s_1 \,.\, s_2 \in \mathcal{S}}.$$

'Internal' syntax consists of atoms, cons cells, closures, formulae, theorems, and special values intended to encode axioms and inference rules:

$$\frac{x \in \Sigma}{\overline{x} \in \mathcal{M}}, \quad \frac{\alpha \in \mathcal{M} \qquad \beta \in \mathcal{M}}{\alpha \,.\, \beta \in \mathcal{M}}, \quad \frac{\Gamma \in \overrightarrow{\mathcal{M}^{\Sigma}} \qquad x_1 \in \Sigma \qquad \cdots \qquad x_n \in \Sigma \qquad t \in \mathcal{S}}{\overline{\Gamma \,.\, x_1, \ldots, x_n \vdash t} \in \mathcal{M}},$$

$$\frac{e \in \Omega}{\overline{\mathrm{Prop}}(e) \in \mathcal{M}}, \quad \frac{e \in \Omega}{\overline{\mathrm{The}}(e) \in \mathcal{M}}, \quad \frac{\Gamma \in \overrightarrow{\mathcal{M}^{\Sigma}} \qquad x_1 \in \Sigma \qquad \cdots \qquad x_n \in \Sigma \qquad t \in \mathcal{S}}{\overline{\mathrm{Assume}}_{\Gamma.x_1,\ldots,x_n}(t) \in \mathcal{M}},$$

where the set of (computational) contexts $\overrightarrow{\mathcal{M}^{\Sigma}}$ is defined mutually recursively with $\mathcal{M}$:

$$\frac{}{\varepsilon \in \overrightarrow{\mathcal{M}^{\Sigma}}}, \quad \frac{\Gamma \in \overrightarrow{\mathcal{M}^{\Sigma}} \qquad \sigma \in \mathcal{M} \qquad x \in \Sigma}{\Gamma, \sigma^x \in \overrightarrow{\mathcal{M}^{\Sigma}}}.$$

For brevity, instead of $\varepsilon, \sigma_1^{x_1}, \ldots, \sigma_n^{x_n}$ we write simply $\sigma_1^{x_1}, \ldots, \sigma_n^{x_n}$.

$\overline{\mathrm{The}}(e)$ shall be interpreted as a 'proof certificate' [9] of a proposition e: a term evaluated to $\overline{\mathrm{The}}(e)$ is interpreted as a proof, while $\overline{\mathrm{The}}(e)$ itself does not contain any information about its own derivation.

For a name $x \in \Sigma$ and context $\Gamma \in \overrightarrow{\mathcal{M}^{\Sigma}}$ by $\sigma = \Gamma(x) \in \mathcal{M}$ we mean σ such that σ^x is the first occurrence of x from the right in the context, if one exists. Accordingly, we write $x \in \Gamma$ if $\Gamma(x)$ makes sense.

2.2. Syntactic sugar

We assume that the cons cell constructor is right-associative. Let us introduce some standard abbreviations:

$$\begin{aligned}
\lambda x_1, \ldots, x_n \,.\, t &= \mathrm{lambda} \,.\, x_1 \,.\, (\ldots) \,.\, x_n \,.\, t \,.\, \mathrm{nil}, \\
f(x_1, \ldots, x_n) &= f \,.\, x_1 \,.\, (\ldots) \,.\, x_n \,.\, \mathrm{nil}, \\
(\mathbf{let}\ x = t\ \mathbf{in}\ s) &= (\lambda x \,.\, s)(t), \\
\lambda\langle x_1, \ldots, x_n\rangle \,.\, t &= (\lambda f \,.\, \lambda x \,.\, f(\mathrm{car}(x), \mathrm{car}(\mathrm{cdr}(x)), \ldots))(\lambda x_1, \ldots, x_n \,.\, t), \\
(\mathbf{let}\ \langle x_1, \ldots, x_n\rangle = x\ \mathbf{in}\ t) &= (\lambda x_1, \ldots, x_n \,.\, t)(\mathrm{car}(x), \mathrm{car}(\mathrm{cdr}(x)), \ldots).
\end{aligned}$$

In addition, for formulae we write $f\langle x_1, \ldots, x_n\rangle = \mathrm{fu}(f, x_1, \ldots, x_n)$ and $Q_x t = \mathrm{qu}(Q, x, t)$.

An ordered n-tuple of objects $e_1, \ldots, e_n \in \mathcal{S}$ is defined recursively:

$$\langle\rangle = \mathrm{nil}, \langle e_1, e_2, \ldots, e_n\rangle = \mathrm{cons}(e_1, \langle e_2, \ldots, e_n\rangle),$$

as well as the n-tuple of objects $\sigma_1, \ldots, \sigma_n \in \mathcal{M}$:

$$\langle\rangle = \overline{\text{nil}}, \langle\sigma_1, \sigma_2, \ldots, \sigma_n\rangle = \sigma_1 \,.\, \langle\sigma_2, \ldots, \sigma_n\rangle.$$

We define syntax for conditional expressions:

$$(\textbf{case}\ p\ \textbf{of}\ q_1 \to x_1, \ldots, q_n \to x_n) = \text{cond}(p, q_1 \,.\, x_1, \ldots, q_n \,.\, x_n).$$

We will encode inductive types using cons cells. Specifically, let t be an inductive type:

$$t \triangleq c_1(x_1, \ldots, x_{N(1)}) \mid \ldots \mid c_n(x_1, \ldots, x_{N(n)}),$$

then the i-th constructor is encoded as follows:

$$c_i = \lambda x_1, \ldots, x_{N(i)} \,.\, \text{cons}(c_i^{\#}, \langle x_1, \ldots, x_{N(i)}\rangle).$$

A function (possibly partially) defined on constructors

$$f(c_{i_1}(x_1, \ldots, x_{N(i_1)})) = f_{i_1},\ \ldots,\ \ f(c_{i_k}(x_1, \ldots, x_{N(i_k)})) = f_{i_k},$$

is represented as:

$$f = \lambda c \,.\, (\textbf{case}\ \text{car}(c)\ \textbf{of}\ c_{i_1}^{\#} \to \lambda\langle x_1, \ldots, x_{N(i_1)}\rangle \,.\, f_{i_1},\ \ldots,\ c_{i_k}^{\#} \to \lambda\langle x_1, \ldots, x_{N(i_k)}\rangle \,.\, f_{i_k})(\text{cdr}(c));$$

moreover, the recursive definition is reduced to the non-recursive one in the standard way, using the fixed-point combinator.

2.3. Semantics

Finally, we define the big-step semantics. The first group of rules describes the subset of language corresponding to the standard λ-calculus. For simplicity of description, a variable $x \in \Sigma$ outside the context is evaluated to an atom $\overline{x} \in \mathcal{M}$. By $\overrightarrow{x}$ we mean $x_1, \ldots, x_n$, by $\overrightarrow{P(x)}$ we mean $P(x_1), \ldots, P(x_n)$, and by $\overrightarrow{P(x, y)}$, respectively, $P(x_1, y_1), \ldots, P(x_n, y_n)$.

$$\frac{x \in \Gamma}{\Gamma \vdash x \Longrightarrow \Gamma(x)},\quad \frac{x \notin \Gamma}{\Gamma \vdash x \Longrightarrow \overline{x}},\quad \frac{\overrightarrow{x \in \Sigma} \qquad t \in \mathcal{S}}{\Gamma \vdash \lambda\overrightarrow{x} \,.\, t \Longrightarrow \overline{\Gamma \,.\, \overrightarrow{x} \vdash t}},$$
$$\frac{\Gamma \vdash f \Longrightarrow \overline{\Delta \,.\, \overrightarrow{y} \vdash t} \qquad \overrightarrow{\Gamma \vdash x \Longrightarrow \xi} \qquad \Delta, \overrightarrow{\xi^y} \vdash t \Longrightarrow \tau}{\Gamma \vdash f(\overrightarrow{x}) \Longrightarrow \tau}.$$

The second group consists of rules describing operations on formulae and theorems; in terms of LISP languages, these are *special forms.*

The first set of rules concerns the constructors and destructors of cons cells:

$$\frac{\Gamma \vdash x_1 \Longrightarrow \alpha \qquad \Gamma \vdash x_2 \Longrightarrow \beta}{\Gamma \vdash \text{cons}(x_1, x_2) \Longrightarrow \alpha \,.\, \beta},\quad \frac{\Gamma \vdash x \Longrightarrow \alpha \,.\, \beta}{\Gamma \vdash \text{car}(x) \Longrightarrow \alpha},\quad \frac{\Gamma \vdash x \Longrightarrow \alpha \,.\, \beta}{\Gamma \vdash \text{cdr}(x) \Longrightarrow \beta},$$

$$\frac{\Gamma \vdash x \Longrightarrow \overline{\text{nil}}}{\Gamma \vdash \text{car}(x) \Longrightarrow \overline{\text{nil}}},\quad \frac{\Gamma \vdash x \Longrightarrow \overline{\text{nil}}}{\Gamma \vdash \text{cdr}(x) \Longrightarrow \overline{\text{nil}}},$$

as well as the corresponding predicate functions:

$$\frac{\Gamma \vdash x \Longrightarrow \overline{y} \qquad \Gamma \vdash x_1 \Longrightarrow \sigma_1}{\Gamma \vdash \mathrm{elim}_{\mathrm{atom}}(x, x_1, x_2) \Longrightarrow \sigma_1}, \quad \frac{\Gamma \vdash x \Longrightarrow \sigma \qquad \Gamma \vdash x_2 \Longrightarrow \sigma_2 \qquad \sigma \notin \overline{\Sigma}}{\Gamma \vdash \mathrm{elim}_{\mathrm{atom}}(x, x_1, x_2) \Longrightarrow \sigma_2},$$

$$\frac{\Gamma \vdash x \Longrightarrow \alpha \,.\, \beta \qquad \Gamma \vdash x_1 \Longrightarrow \sigma_1}{\Gamma \vdash \mathrm{elim}_{\mathrm{cc}}(x, x_1, x_2) \Longrightarrow \sigma_1}, \quad \frac{\Gamma \vdash x \Longrightarrow \sigma \qquad \Gamma \vdash x_2 \Longrightarrow \sigma_2 \qquad \sigma \notin \mathcal{M} \,.\, \mathcal{M}}{\Gamma \vdash \mathrm{elim}_{\mathrm{cc}}(x, x_1, x_2) \Longrightarrow \sigma_2}.$$

The conditional expression $\mathrm{cond}(t, \overrightarrow{x \,.\, s})$ is reduced to the first from the left s_i, for which the corresponding x_i reduces to t, if such exists. Otherwise, the conditional expression does not reduce.

$$\frac{\Gamma \vdash t \Longrightarrow \overline{y_i} \qquad \overrightarrow{\Gamma \vdash x \Longrightarrow y} \qquad \Gamma \vdash s_i \Longrightarrow \sigma_i \qquad \forall j < i \,.\, y_j \neq y_i}{\Gamma \vdash \mathrm{cond}(t, \overrightarrow{x \,.\, s}) \Longrightarrow \sigma_i}.$$

The value $\overline{\mathrm{Assume}}_{\Delta.\overrightarrow{y}}(t)$ reduces as closure $\overline{\Delta \,.\, \overrightarrow{y} \vdash t}$, but produces a value of the form $\overline{\mathrm{The}}(e)$. It acts as *a constructor* for theorems.

$$\frac{\Gamma \vdash f \Longrightarrow \overline{\mathrm{Assume}}_{\Delta.\overrightarrow{y}}(t) \qquad \overrightarrow{\Gamma \vdash x \Longrightarrow \xi} \qquad \Delta, \overrightarrow{\xi^{y}} \vdash t \Longrightarrow \overline{\mathrm{Prop}}(e)}{\Gamma \vdash f(\overrightarrow{x}) \Longrightarrow \overline{\mathrm{The}}(e)}.$$

The destructor of a theorem, respectively, extracts its proposition:

$$\frac{\Gamma \vdash x \Longrightarrow \overline{\mathrm{The}}(e)}{\Gamma \vdash \mathrm{prop}(x) \Longrightarrow \overline{\mathrm{Prop}}(e)}.$$

Next, we have a set of constructors for formulae Ω:

$$\frac{\Gamma \vdash x \Longrightarrow \overline{y}}{\Gamma \vdash \mathrm{fv}(x) \Longrightarrow \overline{\mathrm{Prop}}(\mathrm{Fv}(y))}, \quad \frac{\Gamma \vdash x \Longrightarrow \overline{g} \qquad \overrightarrow{\Gamma \vdash y \Longrightarrow \overline{\mathrm{Prop}}(e)}}{\Gamma \vdash \mathrm{fu}(x, \overrightarrow{y}) \Longrightarrow \overline{\mathrm{Prop}}(g(\overrightarrow{e}))},$$

$$\frac{\Gamma \vdash x_1 \Longrightarrow \overline{Q} \qquad \Gamma \vdash x_2 \Longrightarrow \overline{y} \qquad \Gamma \vdash x_3 \Longrightarrow \overline{\mathrm{Prop}}(e)}{\Gamma \vdash \mathrm{qu}(x_1, x_2, x_3) \Longrightarrow \overline{\mathrm{Prop}}(Qy^{\flat}(e))}.$$

And also destructors:

$$\frac{\Gamma \vdash x \Longrightarrow \overline{\mathrm{Prop}}(\mathrm{Fv}(x))}{\Gamma \vdash \mathrm{car}(x) \Longrightarrow \overline{x}}, \quad \frac{\Gamma \vdash x \Longrightarrow \overline{\mathrm{Prop}}(g(e_1, \ldots, e_n))}{\Gamma \vdash \mathrm{car}(x) \Longrightarrow \overline{g}}, \quad \frac{\Gamma \vdash x \Longrightarrow \overline{\mathrm{Prop}}(Qe)}{\Gamma \vdash \mathrm{car}(x) \Longrightarrow \overline{Q}},$$

$$\frac{\Gamma \vdash x \Longrightarrow \overline{\mathrm{Prop}}(g(e_1, \ldots, e_n))}{\Gamma \vdash \mathrm{cdr}(x) \Longrightarrow \langle \overline{\mathrm{Prop}}(e_1), \ldots, \overline{\mathrm{Prop}}(e_n) \rangle},$$

$$\frac{\Gamma \vdash x_1 \Longrightarrow \overline{Q} \qquad \Gamma \vdash x_2 \Longrightarrow \overline{\mathrm{Prop}}(t) \qquad \Gamma \vdash x_3 \Longrightarrow \overline{\mathrm{Prop}}(Qe)}{\Gamma \vdash \mathrm{free}(x_1, x_2, x_3) \Longrightarrow \overline{\mathrm{Prop}}(t^{\sharp}(e))},$$

and the corresponding predicates:

$$\frac{\Gamma \vdash x_1 \Longrightarrow \sigma_1 \qquad \Gamma \vdash x \Longrightarrow \overline{\mathrm{Prop}}(\mathrm{Fv}(y))}{\Gamma \vdash \mathrm{elim}_\Omega(x_1,x_2,x_3,x_4,x) \Longrightarrow \sigma_1}, \quad \frac{\Gamma \vdash x_2 \Longrightarrow \sigma_2 \qquad \Gamma \vdash x \Longrightarrow \overline{\mathrm{Prop}}(g(e_1,\ldots,e_n))}{\Gamma \vdash \mathrm{elim}_\Omega(x_1,x_2,x_3,x_4,x) \Longrightarrow \sigma_2},$$

$$\frac{\Gamma \vdash x_3 \Longrightarrow \sigma_3 \qquad \Gamma \vdash x \Longrightarrow \overline{\mathrm{Prop}}(Qe)}{\Gamma \vdash \mathrm{elim}_\Omega(x_1,x_2,x_3,x_4,x) \Longrightarrow \sigma_3}, \quad \frac{\Gamma \vdash x_4 \Longrightarrow \sigma_4 \qquad \Gamma \vdash x \Longrightarrow \sigma \qquad \sigma \notin \overline{\mathrm{Prop}}(\Omega)}{\Gamma \vdash \mathrm{elim}_\Omega(x_1,x_2,x_3,x_4,x) \Longrightarrow \sigma_4}.$$

Finally, for each formula $e \in \Omega$ we fix a name $\mathrm{fresh}_e \in \Sigma \setminus \mathbf{Fv}(e)$, and define the corresponding function fresh_Ω:

$$\frac{\Gamma \vdash x \Longrightarrow \overline{\mathrm{Prop}}(e)}{\Gamma \vdash \mathrm{fresh}_\Omega(x) \Longrightarrow \overline{\mathrm{fresh}_e}}.$$

For $s \in \mathcal{S}$ we write $\Gamma \nvdash s$, if $\nexists \sigma \,.\, \Gamma \vdash s \Longrightarrow \sigma$.

For example, using the introduced abbreviations, in the semantics used, the term from the previous section is written as follows:

$$\forall_x \supset \langle \vee \langle = \langle \mathrm{fv}(x), \mathrm{fv}(x_1) \rangle, \neg \langle = \langle \mathrm{fv}(x), \mathrm{fv}(x_2) \rangle \rangle \rangle, \exists_y = \langle \mathrm{fv}(x), \mathrm{fv}(y) \rangle \rangle.$$

It is easy to prove that the following analogues of logical rules hold.

Lemma (exchange). *Let* $x_1 \neq x_2$. *If* $\Gamma_1, {\xi_1}^{x_1}, {\xi_2}^{x_2}, \Gamma_2 \vdash t \Longrightarrow \tau$, *then* $\Gamma_1, {\xi_2}^{x_2}, {\xi_1}^{x_1}, \Gamma_2 \vdash t \Longrightarrow \tau$.

Lemma (contraction). *If* $\Gamma_1, {\xi_1}^{x}, {\xi_2}^{x}, \Gamma_2 \vdash t \Longrightarrow \tau$, *then* $\Gamma_1, {\xi_2}^{x}, \Gamma_2 \vdash t \Longrightarrow \tau$.

Let us extend the predicate lc to the elements of $\mathcal{M}$ and $\overrightarrow{\mathcal{M}^\Sigma}$:

$$\begin{array}{ccc} \mathrm{lc}(\overline{x}) = \top, & \mathrm{lc}(\overline{\mathrm{Prop}}(e)) = \mathrm{lc}(e), & \mathrm{lc}(\varepsilon) = \top, \\ \mathrm{lc}(\alpha \,.\, \beta) = \mathrm{lc}(\alpha) \wedge \mathrm{lc}(\beta), & \mathrm{lc}(\overline{\mathrm{The}}(e)) = \mathrm{lc}(e), & \mathrm{lc}(\Gamma, \sigma^x) = \mathrm{lc}(\Gamma) \wedge \mathrm{lc}(\sigma). \\ \mathrm{lc}(\overline{\Gamma \,.\, \overrightarrow{x} \vdash t}) = \mathrm{lc}(\Gamma), & \mathrm{lc}(\overline{\mathrm{Assume}}_{\Gamma \,.\, \overrightarrow{x}}(t)) = \mathrm{lc}(\Gamma), & \end{array}$$

Lemma. *If* $\mathrm{lc}(e)$, *then* $\mathrm{lc}(Qx^\flat(e))$. *Conversely, if* $\mathrm{lc}(t)$ *and* $\mathrm{lc}(Qe)$, *then* $\mathrm{lc}(t^\sharp(e))$.

Proof. Indeed, if $\mathrm{lc}(e)$, then, by Lemma 3, $\mathrm{lc}_2(x^\flat(e))$, therefore $\mathrm{lc}(Qx^\flat(e))$. If $\mathrm{lc}(Qe)$, then $\mathrm{lc}_2(e)$, therefore, again by Lemma 3, $\mathrm{lc}_1(t^\sharp(e)) = \mathrm{lc}(t^\sharp(e))$. □

Theorem 2. *Let* $\mathrm{lc}(\Gamma)$. *If* $\Gamma \vdash s \Longrightarrow \sigma$ *for some* $s \in \mathcal{S}$ *and* $\sigma \in \Omega$, *then* $\mathrm{lc}(\sigma)$.

Proof. On the inductive set $\mathcal{S} \Longrightarrow \mathcal{M} = \{\Gamma \vdash s \Longrightarrow \sigma \mid \Gamma \in \overrightarrow{\mathcal{M}^\Sigma} \wedge s \in \mathcal{S} \wedge \sigma \in \mathcal{M}\}$ we define the predicate $P(\Gamma \vdash s \Longrightarrow \sigma) = (\mathrm{lc}(\Gamma) \Rightarrow \mathrm{lc}(\sigma))$. The assertion $\forall \tau \,.\, P(\tau)$ is then proved by induction with obvious reasoning. □

Consequence. *Let* $\mathrm{lc}(\Gamma)$. *If* $\Gamma \vdash s \Longrightarrow \overline{\mathrm{Prop}}(e)$ *for some* $s \in \mathcal{S}$ *and* $e \in \Omega$, *then* $\mathrm{lc}(e)$.

Consequence. *Let* $\mathrm{lc}(\Gamma)$. *If* $\Gamma \vdash s \Longrightarrow \overline{\mathrm{The}}(e)$ *for some* $s \in \mathcal{S}$ *and* $e \in \Omega$, *then* $\mathrm{lc}(e)$.

We write $\Gamma \Vdash e$ if $\exists s \,.\, \Gamma \vdash s \Longrightarrow \overline{\mathrm{The}}(e)$. By *the theory* of the context Γ we understand the set of formulae derivable in it $\mathrm{Th}(\Gamma) = \{e \in \Omega \mid \Gamma \Vdash e\}$. In particular, from Theorem 2 it follows that $\Gamma \nVdash e$ if $\neg \mathrm{lc}(e)$; or, in other words, $\mathrm{Th}(\Gamma) \subseteq \Omega_{\mathrm{lc}}$.

2.4. General constructions

Next, we define the operations common to all axiomatic theories. First, the logical connectives $\mathrm{andalso}(p,q) = (\textbf{case}\ p\ \textbf{of}\ \mathrm{true} \to q, \mathrm{false} \to \mathrm{false})$ and $\mathrm{orelse}(p,q) = (\textbf{case}\ p\ \textbf{of}\ \mathrm{true} \to \mathrm{true}, \mathrm{false} \to q)$, the equality for names $\mathrm{dec}^{=}_{\Sigma}(x_1,x_2) = (\textbf{case}\ x_1\ \textbf{of}\ x_2 \to \mathrm{true}, x_1 \to \mathrm{false})$, predicates for formulae:

$$\begin{aligned}\mathrm{isfv}(x) &= \mathrm{elim}_{\Omega}(\mathrm{true},\mathrm{false},\mathrm{false},\mathrm{false},x),\\ \mathrm{isfu}(x) &= \mathrm{elim}_{\Omega}(\mathrm{false},\mathrm{true},\mathrm{false},\mathrm{false},x),\\ \mathrm{isqu}(x) &= \mathrm{elim}_{\Omega}(\mathrm{false},\mathrm{false},\mathrm{true},\mathrm{false},x),\end{aligned}$$

and a set of standard predicates for tuples:

$$\begin{aligned}\mathrm{isnil}(x) &= \mathrm{elim}_{\mathrm{atom}}(x, \mathrm{dec}^{=}_{\Sigma}(x,\mathrm{nil}), \mathrm{false}),\\ \mathrm{iscc}(x) &= \mathrm{elim}_{\mathrm{cc}}(x,\mathrm{true},\mathrm{false}),\\ \mathrm{exists}(f,xs) &= \textbf{case}\ \mathrm{true}\ \textbf{of}\\ &\quad \mathrm{isnil}(xs) \to \mathrm{false},\\ &\quad \mathrm{iscc}(xs) \to \mathrm{orelse}(f(\mathrm{car}(xs)), \mathrm{exists}(f,\mathrm{cdr}(xs))),\\ \mathrm{forall}^2(f,xs,ys) &= \textbf{case}\ \mathrm{true}\ \textbf{of}\\ &\quad \mathrm{andalso}(\mathrm{isnil}(xs),\mathrm{isnil}(ys)) \to \mathrm{true},\\ &\quad \mathrm{andalso}(\mathrm{isnil}(xs),\mathrm{iscc}(ys)) \to \mathrm{false},\\ &\quad \mathrm{andalso}(\mathrm{iscc}(xs),\mathrm{isnil}(ys)) \to \mathrm{false},\\ &\quad \mathrm{andalso}(\mathrm{iscc}(xs),\mathrm{iscc}(ys)) \to \textbf{let}\ \theta = f(\mathrm{car}(xs),\mathrm{car}(ys))\ \textbf{in}\\ &\qquad\qquad \mathrm{andalso}(\theta, \mathrm{forall}^2(f,\mathrm{cdr}(xs),\mathrm{cdr}(ys))).\end{aligned}$$

Next, we recursively express equality for the formulae:

$$\begin{aligned}\mathrm{dec}^{=}_{\Omega}(t_1,t_2) = \mathrm{andalso}&(\mathrm{dec}^{=}_{\Sigma}(\mathrm{car}(t_1),\mathrm{car}(t_2)),\\ &\textbf{case}\ \mathrm{true}\ \textbf{of}\\ &\mathrm{andalso}(\mathrm{isfv}(t_1),\mathrm{isfv}(t_2)) \to \mathrm{true},\\ &\mathrm{andalso}(\mathrm{isfu}(t_1),\mathrm{isfu}(t_2)) \to \mathrm{forall}^2(\mathrm{dec}^{=}_{\Omega},\mathrm{cdr}(t_1),\mathrm{cdr}(t_2)),\\ &\mathrm{andalso}(\mathrm{isqu}(t_1),\mathrm{isqu}(t_2)) \to \textbf{let}\ x = \mathrm{fv}(\mathrm{fresh}_{\Omega}(\varepsilon^{\#}\langle t_1,t_2\rangle))\ \textbf{in}\\ &\qquad\qquad \mathrm{dec}^{=}_{\Omega}(\mathrm{free}(\mathrm{car}(t_1),x,t_1),\mathrm{free}(\mathrm{car}(t_2),x,t_2)),\\ &\mathrm{true} \to \mathrm{false}),\end{aligned}$$

and the predicate that determines whether a given variable occurs free in a given formula:

$$\begin{aligned}\mathrm{occur}_{\Omega}(x,t) = &\ \textbf{case}\ \mathrm{true}\ \textbf{of}\\ &\mathrm{isfv}(t) \to \mathrm{dec}^{=}_{\Sigma}(x,\mathrm{car}(t)),\\ &\mathrm{isfu}(t) \to \mathrm{exists}(\mathrm{occur}_{\Omega}(x,-),\mathrm{cdr}(t)),\\ &\mathrm{isqu}(t) \to \textbf{let}\ y = \mathrm{fv}(\mathrm{fresh}_{\Omega}(\varepsilon^{\#}\langle \mathrm{fv}(x),t\rangle))\ \textbf{in}\ \mathrm{occur}_{\Omega}(x,\mathrm{free}(\mathrm{car}(t),y,t)),\end{aligned}$$

where $\varepsilon^{\#} \in \Sigma$ is an arbitrary name.

Finally, *the unpacking* operation $\mathrm{of}(g,e) = \textbf{case}\ \mathrm{isfu}(e)\ \textbf{of}\ \mathrm{true} \to \textbf{case}\ \mathrm{car}(e)\ \textbf{of}\ g \to \mathrm{cdr}(e)$ for a formula of the form $\overline{\mathrm{Prop}}(g(e_1,\ldots,e_n))$ extracts the tuple of arguments $\langle \overline{\mathrm{Prop}}(e_1),\ldots,\overline{\mathrm{Prop}}(e_n)\rangle$.

3. Propositional logic and first-order logic

Let's consider how to formulate propositional logic within the constructed language. First, we introduce the following abbreviations:

$$\begin{aligned}
\overline{\text{mp}} &= \overline{\text{Assume}}_{\varepsilon.t_1,t_2}(\textbf{let } \langle p,q\rangle = \text{of}(\supset, \text{prop}(t_1)) \textbf{ in case } \text{dec}^{=}_{\Omega}(p, \text{prop}(t_2)) \textbf{ of } \text{true} \to q),\\
\overline{\text{veq}} &= \overline{\text{Assume}}_{\varepsilon.p,q}(\supset\langle p, \supset\langle q, p\rangle\rangle),\\
\overline{\text{cr}} &= \overline{\text{Assume}}_{\varepsilon.p,q,r}(\supset\langle\supset\langle p, \supset\langle q, r\rangle\rangle, \supset\langle\supset\langle p, q\rangle, \supset\langle p, r\rangle\rangle\rangle),\\
\overline{\text{elim}_{\neg\neg}} &= \overline{\text{Assume}}_{\varepsilon.p}(\supset\langle\neg\neg p, p\rangle),
\end{aligned}$$

where $\neg p = \supset\langle p, \bot\langle\rangle\rangle$. Let $\Gamma_{\text{PL}} = (\overline{\text{mp}}^{\text{mp}}, \overline{\text{veq}}^{\text{veq}}, \overline{\text{cr}}^{\text{cr}})$ and $\Gamma_{\text{CL}} = \left(\Gamma_{\text{PL}}, \overline{\text{elim}_{\neg\neg}}^{\text{elim}_{\neg\neg}}\right)$.

As noted above, $\overline{\text{The}}(e)$ expressions do not contain any information about their own derivation. The deduction theorem is proved by induction on the length of the derivation, so it is not possible to prove it directly (within the system). However, we can encode the derivation trees [10] separately:

$$\text{deriv} \triangleq \text{assume}(p) \mid \text{theorem}(t) \mid \text{apply}(d_1, d_2).$$

Next, we prove auxiliary theorems. We define:

$$\begin{aligned}
\text{idthm}(p) = {} & \textbf{let } h_1 = \text{veq}(p, \supset\langle p, p\rangle) \textbf{ in}\\
& \textbf{let } h_2 = \text{cr}(p, \supset\langle p, p\rangle, p) \textbf{ in}\\
& \textbf{let } h_3 = \text{mp}(h_2, h_1) \textbf{ in}\\
& \textbf{let } h_4 = \text{veq}(p, p) \textbf{ in}\\
& \text{mp}(h_3, h_4),\\
\text{prefix}(p, t) = {} & \text{mp}(\text{veq}(\text{prop}(t), p), t).
\end{aligned}$$

For simplicity (if necessary), we assume that Γ does not contain names from the definition of big-step semantics, as well as names from Γ_{PL}, Γ_{CL}, and others.

Lemma. *If* $\Gamma_{\text{PL}}, \Gamma \vdash p \Longrightarrow \overline{\text{Prop}}(\phi)$ *and* $\Gamma_{\text{PL}}, \Gamma \vdash t \Longrightarrow \overline{\text{The}}(\psi)$*, then* $\Gamma_{\text{PL}}, \Gamma \vdash \text{idthm}(p) \Longrightarrow \overline{\text{The}}(\supset(\phi, \phi))$ *and* $\Gamma_{\text{PL}}, \Gamma \vdash \text{prefix}(p, t) \Longrightarrow \overline{\text{The}}(\supset(\phi, \psi))$.

For any derivation tree d, its goal $\text{goal}(d)$ is defined as

$$\begin{aligned}
\text{goal}(\text{assume}(p)) &= p,\\
\text{goal}(\text{theorem}(t)) &= \text{prop}(t),\\
\text{goal}(\text{apply}(d_1, d_2)) &= \textbf{let } \langle p, q\rangle = \text{of}(\supset, \text{goal}(d_1)) \textbf{ in } q.
\end{aligned}$$

For the derivation without nodes of the form $\text{assume}(q)$, one can construct a proof of its goal:

$$\begin{aligned}
\text{verify}(\text{theorem}(t)) &= t,\\
\text{verify}(\text{apply}(d_1, d_2)) &= \text{mp}(\text{verify}(d_1), \text{verify}(d_2));
\end{aligned}$$

that is, if $\Gamma_{\text{PL}}, \Gamma \vdash \text{goal}(d) \Longrightarrow \overline{\text{Prop}}(\phi)$, then $\Gamma_{\text{PL}}, \Gamma \vdash \text{verify}(d) \Longrightarrow \overline{\text{The}}(\phi)$ (provided that all necessary terms are reducible).

The definition of $\mathrm{discharge}_p$ corresponds to the proof of the deduction theorem:

$$\begin{aligned}
\mathrm{discharge}_p(\mathrm{assume}(q)) &= \mathbf{case}\ \mathrm{dec}^{=}_{\Omega}(p,q)\ \mathbf{of}\\
&\qquad \mathrm{true} \to \mathrm{theorem}(\mathrm{idthm}(p)),\\
&\qquad \mathrm{false} \to \mathrm{apply}(\mathrm{theorem}(\mathrm{veq}(q,p)), \mathrm{assume}(q)),\\
\mathrm{discharge}_p(\mathrm{theorem}(t)) &= \mathrm{theorem}(\mathrm{prefix}(p,t)),\\
\mathrm{discharge}_p(\mathrm{apply}(d_1,d_2)) &= \mathbf{let}\ \langle q,r\rangle = \mathrm{of}(\supset, \mathrm{goal}(d_1))\ \mathbf{in}\\
&\qquad \mathrm{apply}(\mathrm{apply}(\mathrm{theorem}(\mathrm{cr}(p,q,r)), \mathrm{discharge}_p(d_1)), \mathrm{discharge}_p(d_2)).
\end{aligned}$$

Let's also consider the rules for (classical) first-order logic:

$$\begin{aligned}
\overline{\mathrm{intro}_\forall} &= \overline{\mathrm{Assume}}_{\varepsilon.x,t}(\forall_x \mathrm{prop}(t)),\\
\overline{\mathrm{elim}_\forall} &= \overline{\mathrm{Assume}}_{\varepsilon.t,p}(\supset\langle p, \mathrm{free}(\forall,t,p)\rangle),\\
\overline{\mathrm{exch}_\forall} &= \overline{\mathrm{Assume}}_{\varepsilon.x,p,q}(\mathbf{case}\ \mathrm{occur}_\Omega(x,p)\ \mathbf{of}\ \mathrm{false} \to \supset\langle \forall_x \supset\langle p,q\rangle, \supset\langle p, \forall_x q\rangle\rangle),
\end{aligned}$$

and define the corresponding context $\Gamma_{\mathrm{FOL}} = \Big(\Gamma_{\mathrm{CL}}, \overline{\mathrm{intro}_\forall}^{\mathrm{intro}_\forall}, \overline{\mathrm{elim}_\forall}^{\mathrm{elim}_\forall}, \overline{\mathrm{exch}_\forall}^{\mathrm{exch}_\forall}\Big)$.

The deduction theorem extends naturally to first-order logic; to do this, we define deriv and related functions further:

$$\begin{aligned}
\mathrm{deriv} &\triangleq \ldots \mid \mathrm{intro}(x,d),\\
\mathrm{goal}(\mathrm{intro}(x,d)) &= \forall_x \mathrm{goal}(d),\\
\mathrm{verify}(\mathrm{intro}(x,d)) &= \mathrm{intro}_\forall(x, \mathrm{verify}(d)),\\
\mathrm{discharge}_p(\mathrm{intro}(x,d)) &= \mathrm{apply}(\mathrm{theorem}(\mathrm{exch}_\forall(x,p,\mathrm{goal}(d))), \mathrm{intro}(x, \mathrm{discharge}_p(d))).
\end{aligned}$$

Example. *Let* $\phi, \psi \in \Omega_{\mathrm{lc}}(2)$. *Then* $\Gamma_{\mathrm{FOL}}, \Gamma \Vdash \supset(\forall\supset(\phi,\psi), \supset(\forall\phi, \forall\psi))$.

Proof. By our assumptions, we can choose a name $x \notin \mathbf{Fv}(\phi) \cup \mathbf{Fv}(\psi)$ and $p, q \in \mathcal{S}$, such that $\Gamma_{\mathrm{FOL}}, \Gamma \vdash p \Longrightarrow \mathrm{Fv}(x)^\sharp(\phi)$ and $\Gamma_{\mathrm{FOL}}, \Gamma \vdash q \Longrightarrow \mathrm{Fv}(x)^\sharp(\psi)$. Then, in particular, the following holds:

$$\Gamma_{\mathrm{FOL}}, \Gamma \vdash \supset\langle \forall_x \supset\langle p,q\rangle, \supset\langle \forall_x p, \forall_x q\rangle\rangle \Longrightarrow \overline{\mathrm{Prop}}(\supset(\forall\supset(\phi,\psi), \supset(\forall\phi,\forall\psi))).$$

We denote $h_t = \mathrm{apply}(\mathrm{theorem}(\mathrm{elim}_\forall(\mathrm{fv}(x), \forall_x t)), \mathrm{assume}(\forall_x t))$. Then:

$$\begin{aligned}
\Gamma_{\mathrm{FOL}}, \Gamma \vdash \mathrm{goal}(h_p) &\Longrightarrow \overline{\mathrm{Prop}}(\mathrm{Fv}(x)^\sharp(\phi)),\\
\Gamma_{\mathrm{FOL}}, \Gamma \vdash \mathrm{goal}(h_{\supset\langle p,q\rangle}) &\Longrightarrow \overline{\mathrm{Prop}}(\mathrm{Fv}(x)^\sharp(\supset(\phi,\psi))),\\
\Gamma_{\mathrm{FOL}}, \Gamma \vdash \mathrm{goal}(\mathrm{apply}(h_{\supset\langle p,q\rangle}, h_p)) &\Longrightarrow \overline{\mathrm{Prop}}(\mathrm{Fv}(x)^\sharp(\psi)),\\
\Gamma_{\mathrm{FOL}}, \Gamma \vdash \mathrm{goal}(\mathrm{intro}(x, \mathrm{apply}(h_{\supset\langle p,q\rangle}, h_p))) &\Longrightarrow \overline{\mathrm{Prop}}(\forall\psi).
\end{aligned}$$

We write $h = \mathrm{intro}(x, \mathrm{apply}(h_{\supset\langle p,q\rangle}, h_p))$. Finally,

$$\begin{aligned}
\Gamma_{\mathrm{FOL}}, \Gamma \vdash \mathrm{goal}(\mathrm{discharge}_{\forall_x p}(h)) &\Longrightarrow \overline{\mathrm{Prop}}(\supset(\forall\phi,\forall\psi)),\\
\Gamma_{\mathrm{FOL}}, \Gamma \vdash \mathrm{goal}(\mathrm{discharge}_{\forall_x \supset\langle p,q\rangle}(\mathrm{discharge}_{\forall_x p}(h))) &\Longrightarrow \overline{\mathrm{Prop}}(\supset(\forall\supset(\phi,\psi), \supset(\forall\phi,\forall\psi))),\\
\Gamma_{\mathrm{FOL}}, \Gamma \vdash \mathrm{verify}(\mathrm{discharge}_{\forall_x \supset\langle p,q\rangle}(\mathrm{discharge}_{\forall_x p}(h))) &\Longrightarrow \overline{\mathrm{The}}(\supset(\forall\supset(\phi,\psi), \supset(\forall\phi,\forall\psi))),
\end{aligned}$$

which is what was required. □

Alternatively, the deduction theorem can be added as an inference rule as follows:

$$\overline{\text{intro}_{\supset}} = \overline{\text{Assume}}_{\overline{\text{Assume}}_{\varepsilon.q(q)^h.p,\phi}(\supset\langle p,\text{prop}(\phi(h(p)))\rangle)}.$$

This work is supported by the Krasnoyarsk Mathematical Center and financed by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-02-2025-1790).

References

[1] M.Carneiro, The Type Theory of Lean, GitHub : [online] (accessed: 23.07.2025); URL: `https://github.com/digama0/lean-type-theory/releases`.

[2] Homotopy Type Theory: Univalent Foundations of Mathematics, The Univalent Foundations Program; Princeton: Institute for Advanced Study, 2013.

[3] N.Megill, D.A.Wheeler, Metamath: A Computer Language for Mathematical Proofs, Morrisville, Lulu Press, 2019.

[4] L.Paulson, Logic and computation: interactive proof with Cambridge LCF, New York, Cambridge Univerity Press, 1987.

[5] J.McCarthy, Recursive functions of symbolic expressions and their computation by machine, Part I, Communications of the ACM, 1960, Vol. 3, Iss. 4, 184–195.

[6] A.Charguéraud, The Locally Nameless Representation, *Journal of Automated Reasoning*, **49**(2011), 363–408.

[7] A.M.Pitts, Locally Nameless Sets, *Proceedings* of the ACM on Programming Languages, **7**(2023), no. POPL, 488–514. DOI: 10.1145/3571210

[8] N.Bourbaki, Théorie des ensembles, Heidelberg, Springer Berlin, 2006.

[9] J.Sterling, Implementing Inference Rules in Standard ML, CMU School of Computer Science: [online] (accessed: 28.08.2025); URL: `https://www.cs.cmu.edu/~fp/courses/15317-f17/recitations/recitation7.pdf`.

[10] P.Scott, Write your own Theorem Prover, GitHub: [online] (accessed: 28.08.2025); URL: `https://github.com/bollu/minitt/blob/c1dd42e89ee58baba3a8a41a9873fc63f0efcbd9/write-your-own-theorem-prover.pdf`.

Общий метаязык для формальных доказательств

Николай А. Мишко
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. В работе представлена операционная семантика языка, достаточного мощного для выражения произвольных формальных систем с вычислимыми правилами вывода и метатеорем об этих системах, при этом достаточно простого для реализации и теоретического анализа, а также потенциально пригодного для самоверификации.

Ключевые слова: формальная система, операционная семантика, верификация доказательств.

EDN: RDSRHT
УДК 546.55+54.182+544.77.023.55+538.958

Chemical Synthesis of Gold Nanoparticles: Mechanisms of Formation and Morphology Control

Amina Sh. Umarova*
Reshetnev Siberian State University of Science and Technology
Krasnoyarsk, Russian Federation
Anna V. Lukyanenko†
Andrey V. Tsarenko‡
Aleksandr S. Fedorov§
Aleksandr S. Aleksandrovsky¶
Mikhail N. Volochaev‖
Kirensky Institute of Physics
Federal Research Center KSC SB RAS
Krasnoyarsk, Russian Federation

Received 19.10.2025, received in revised form 02.12.2025, accepted 27.01.2026

Abstract. In this work, gold nanoparticles were synthesized by two methods: the Turkevich–Frens method and seed-mediated growth method. The features of each method and their effect on the size, morphology, and optical properties of the obtained particles are described. The dependence of the plasmon resonance wavelength on the shape of nanoparticles has been analysed using optical spectroscopy. It has also been shown that changes in synthesis conditions affect the spectral characteristics of colloidal gold.

Keywords: gold nanoparticles, Turkevich–Frens method, seed-mediated growth method, localized surface plasmon resonance.

Citation: A.Sh. Umarova, A.V. Lukyanenko, A.V. Tsarenko, A.S. Fedorov, A.S. Aleksandrovsky, M.N. Volochaev, Chemical Synthesis of Gold Nanoparticles: Mechanisms of Formation and Morphology Control, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 347–356. EDN: RDSRHT.

Introduction

In recent decades, metallic nanoparticles have attracted great attention from researchers due to their unique optical, physical and chemical properties. Gold nanoparticles (AuNPs) are of particular interest due to their applications in fields such as biomedicine [1], photonics and optoelectronics [2], and even in the space industry [3]. The study of colloidal gold (CG) as a scientific object began in the middle of the 19th century, when it was first observed that gold solutions have a characteristic colour depending on particle size. Later, this phenomenon was explained by the optical properties of gold nanoparticles, specifically through the excitation of localized surface plasmon resonance (LSPR). LSPR represents the collective oscillation of free electrons on the surface of metallic nanoparticles under incident light [4]. As a result of this phenomenon, the intensity of the electromagnetic field around the nanoparticle increases and a resonance peak of absorption or scattering appears in its optical spectrum. The properties of

*amina-umarova-2016@mail.ru https://orcid.org/0009-0004-3704-9446
†lav@iph.krasn.ru https://orcid.org/0000-0001-7496-242X
‡tsarienko.2001@mail.ru
§qchem99@yandex.ru https://orcid.org/0000-0002-7911-3301
¶aleksandrovsky@kirensky.ru https://orcid.org/0000-0003-1821-6718
‖volochaev91@mail.ru https://orcid.org/0000-0001-6245-4153

LSPR are influenced by several factors, including particle size and shape. By changing these parameters, it is possible to tune the resonance peak in the visible to near-infrared range [5]. The above mentioned characteristics of LSPR allow it to be used in various scientific researches, such as the creation of photodetectors and biosensors, enhanced photocatalysis, etc. CG can be synthesized by two methods: the dispersion method and the condensation method [6]. The dispersion method based on the destruction of the crystal lattice of the bulk metallic gold, for instance, by electrical discharge [7]. The advantage of this method is the absence of residual chemical compounds in the resulting colloid. The main disadvantage of this method is the heterogeneity of the particle size, which complicates the control of their optical characteristics. That's why the condensation method is more commonly used. It is based on the reduction of gold ions (e.g. from chloroauric acid $HAuCl_4$) by chemical reducing agents (e.g. sodium borohydride $NaBH_4$ or trisodium citrate $C_6H_5Na_3O_7$), or through thermal or photochemical reduction [8]. The opportunity to control the size, shape, and optical properties of the particles required for LSPR, makes the condensation method preferable for this work. In this work, gold reduction was carried out using one of the most widely applied techniques — the Turkevich–Frens method, which involves the reduction of gold ions from a boiling aqueous solution of $HAuCl_4$ using $C_6H_5Na_3O_7$ as a reducing agent [9]. As a result, gold nanoparticles with sizes in the range of 20–30 nm were obtained. The final particle size depends strongly on the amount of reducing agent: increasing the volume of citrate leads to the formation of smaller particles due to the faster reduction of gold ions [10]. However at larger particle sizes the Turkevich–Frens method produces polydisperse systems, making size control more challenging. That's why for better particle size control we used the seed-mediated growth method. This method involves a two-step synthesis: in the first stage, small seed nanoparticles are produced; in the second stage, these seeds are added to a solution containing $HAuCl_4$ and a reducing agent [11]. Controlled deposition of gold on the seed surfaces leads to particle growth and allows the formation of nanoparticles of the desired size. The main advantage of this method is the opportunity to produce monodisperse nanoparticles, which is important for precise control of their optical characteristics. Moreover, by adjusting the synthesis conditions, it is possible to vary the shape of the nanoparticles, for example, by synthesizing spherical and anisotropic structures. For reliable characterization of nanoparticles, methods capable of determining both geometric and spectral characteristics were employed. The transmission electron microscopy (TEM) technique was used for morphological analysis, while spectrophotometry was applied to study LSPR. In this work, the characteristics of gold nanoparticles synthesized by the Turkevich–Frens and seed-mediated growth methods are presented. The morphology of nanoparticles was studied by using TEM, which was allowed for the determination of size and shape of the particles. Furthermore, the influence of synthesis conditions on the manifestation of LSPR was also considered: an increase in particle size led to a red-shift of the plasmonic peak to the long-wavelength range, but a change in shape, particularly from spherical to anisotropic, led to the appearance of several resonance modes, which also affects the wavelength shift to the red range. The novelty of the present work consists in a comparative experimental analysis of gold nanoparticles synthesized by the Turkevich—Frens method and the seed-mediated growth method. Special attention is paid to the correlation between synthesis parameters, particle morphology determined by TEM, and optical properties associated with LSPR. In addition, the applicability limits of the dipole approximation for quasi-spherical and anisotropic gold nanoparticles are experimentally demonstrated by direct comparison of measured absorption spectra with simulated absorption spectra.

1. Materials and methods

Materials. Chloroauric acid (99.9%, $HAuCl_4$), trisodium citrate (99.0%, $C_6H_5Na_3O_7$), ascorbic acid (>= 99.0%, $C_6H_8O_6$, AA), sodium borohydride (98%, $NaBH_4$), cetyltrimethylammonium bromide (>= 99%, $C_{19}H_{42}BrN$, CTAB), cetyltrimethylammonium chloride (25 wt.%

aqueous solution, $C_{19}H_{42}ClN$, CTAC).

Equipment. Magnetic stirrer with heating Labdevices-20D (Labdevices/China), mini-centrifuge BioSan Microspin 12 (BioSan SIA/Latvia), drying oven ULAB UT-4620 (ULAB/China), original syringe pump based on Arduino (Krasnoyarsk, Russia), transmission electron microscope Hitachi HT7700 (Hitachi High-Technologies Corporation/Japan), spectrophotometer Shimadzu UV-3600 (Shimadzu Corporation/Japan), quartz cuvette Hellma Analytics (Hellma GmbH I& Co. KG/Germany).

Synthesis of nanoparticles. *Turkevich–Frens Method.* To investigate the effect of synthesis temperature, samples 1, 2, and 3 were prepared. As an example, the synthesis of sample 1 is described below: A 50 mL aqueous solution of $HAuCl_4$ (100 mM) was heated to boiling, after that 3 mL of freshly prepared $C_6H_5Na_3O_7$ (38.8 mM) was added. The resulting reaction mixture was boiled for 20 minutes on magnetic stirrer, and then cooled to room temperature. The final solution was centrifuged at 6000 rpm for 25 minutes. After that the pH was adjusted to neutral (pH 7). As a result, a ruby-red CG solution was obtained, characteristic of particles about 20 nm in size. For a more detailed analyse of the influence of the temperature and reagent ratios, an additional series of the synthesis were performed. The detailed syntheses are summarised in Tab. 1.

Table 1. Synthesis conditions for nanoparticles obtained by the Turkevich-Frens method

Sample	Volume of $HAuCl_4$ (ml)	Volume of $C_6H_5Na_3O_7$ (ml)	Comment
1	50	3	T=100°C, UV filter
2	50	5	T=95°C, UV filter
3	50	3	T=92°C, without UV filter

In all experiments, the reduction process was initiated by adding $C_6H_5Na_3O_7$ to the aqueous $HAuCl_4$ solution at its boiling point. The temperatures listed in Tab. 1 correspond to the temperature at which citrate was added. The concentration of the reducing agent was varied independently. Therefore, the temperature and citrate concentration were not varied simultaneously during the synthesis.

Seed–mediated growth method. This method involves the preparation of gold seeds that serve as nucleation centers for controlled particle growth. The seeds are added to a growth solution containing $HAuCl_4$ and a reducing agent.

(I) Preparation of Au clusters. In the first stage gold clusters were synthesised. In a 20 ml glass vial, 5 ml of aqueous solution $HAuCl_4$ (0.5 mM) and 5 ml of CTAB solution (200 mM), which is used as a stabilizing agent, were mixed. Then, 0.6 ml of freshly prepared $NaBH_4$ solution (10 mM) was added as a reducing agent. The resulting solution was kept in a drying oven at 27 °C for 3 hours to completely decompose the excess $NaBH_4$.

(II) Preparation of Au seeds. At the second stage the previously synthesised gold clusters were used to grow seed nanoparticles. In a 20 ml glass vial, 4 ml of CTAC solution (200 mM), 3 ml of AA solution (100 mM) and 0.1 ml of the prepared gold clusters were mixed. Under continuous stirring 4 ml of aqueous solution $HAuCl_4$ (0.5 mM) was added. The final colloidal solution was centrifuged at 14500 rpm for 30 minutes to isolate the gold nanoparticles as a precipitate, which was subsequently used for pH adjustment.

(III) Growth of gold nanoparticles on seeds. At this stage samples 4, 5, 6, 7 and 8 were synthesised. As an example, the synthesis of sample 4 is described below: in a 30 ml glass vial placed on a magnetic stirrer, 10 ml of CTAC solution (100 mM), 0.675 ml of AA solution (10 mM) and 50 μl of the seed nanoparticles were added. Using a syringe pump 10 ml of aqueous solution HAuCl4 (0.5 mM) was injected into the reaction mixture. After injection, the mixture was left to react at 24°C for 10 minutes. The final colloidal solution was centrifuged three times

at 8000 rpm for 10–15 minutes each time. Then, pH level was adjusted to neutral (pH 7). As a result, a violet-blue CG solution was obtained, characteristic of particles about 80 nm in size. To synthesize samples 5, 6, 7 and 8, the addition method and feed rate of $HAuCl_4$ (υ) were varied to study their effect on the morphological and optical properties of the nanoparticles. For samples 6 and 7 $HAuCl_4$ was added manually, while for samples 5 and 8 $HAuCl_4$ was added using a syringe pump. The detailed syntheses are summarised in Tab. 2.

Table 2. Synthesis conditions for nanoparticles obtained by seed-mediated growth method

Sample	Volume of CTAC+AA (ml)	Volume of seed nanoparticles (μl)	Volume of $HAuCl_4$ (ml)	Comment
4	10.675	50	10	Mechanical addition $HAuCl_4$ υ = 11 ml/h)
5	10.675	25	10	Mechanical addition $HAuCl_4$ (υ = 8.6 ml/h)
6	10.675	100	2	Mechanical addition $HAuCl_4$ (υ = 2.2 ml/h)
7	10.675	400	8	Mechanical addition $HAuCl_4$ (υ = 5.6 ml/h)
8	10.675	160	6.2	Mechanical addition $HAuCl_4$ (υ = 8.3 ml/h)

2. Results and discussion

The presented TEM images (Fig. 1a-c) show gold nanoparticles synthesized using the Turkevich–Frens method. Analysis of samples obtained using the Turkevich–Frens method indicates a slight deviation from a spherical shape, which indicates the influence of synthesis parameters on the morphology of particles. The final particle size and morphology are determined

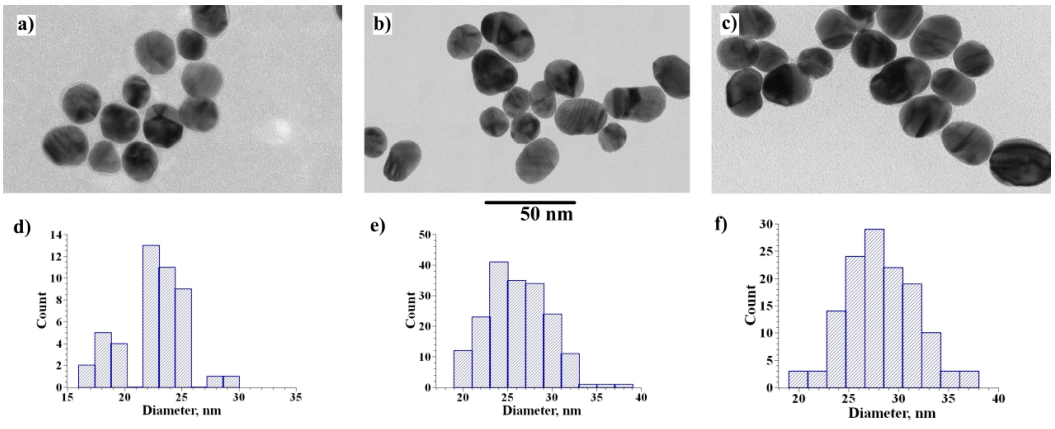


Fig. 1. TEM images of gold nanoparticles synthesised by the Turkevich-Frens method: a) Sample 1; b) Sample 2; c) Sample 3. Particle size distribution histograms: d) Sample 1; e) Sample 2; f) Sample 3

by the combined effects of the temperature of the aqueous $HAuCl_4$, the concentration of the

reducing agent, and the kinetics of reagent addition. This behaviour can be attributed to the combined effects of synthesis conditions on the nucleation and growth processes. Faster reduction of gold ions and the formation of more nucleation centers, promoted by higher temperatures of the aqueous $HAuCl_4$ and higher concentrations of the reducing agent, lead to the formation of a larger number of smaller nanoparticles. Conversely, slower nucleation under lower temperatures or lower reducing agent concentrations favours the growth of fewer, larger particles [12]. To quantitatively assess particle sizes, size distribution histograms were constructed (Fig. 1d-f). For each sample, the particle size distributions were obtained by analysing 10–15 TEM images, with more than 100 individual nanoparticles measured per sample. Sample 1 (Fig. 1d) is characterized by particle sizes ranging from 18 to 26 nm, with the histogram maximum shifted toward smaller sizes, which means that a large number of small gold nanoparticles were formed. Sample 2 (Fig. 1e) demonstrates a broader size distribution (20–32 nm) and a shift of the histogram maximum toward larger particles. Sample 3 (Fig. 1f) shows a normal size distribution with a peak around 27–28 nm, suggesting a high degree of monodispersity in the system.

To further study the effect of synthesis conditions on the shape and size of nanoparticles, we will consider samples obtained by the seed-mediated growth method.

As can be seen from the obtained TEM images (Fig. 2a-e), the morphological characteristics of gold nanoparticles are significantly influenced by the synthesis parameters. In particular, an increase of the volume of the added seed nanoparticles leads to a larger number of growth centers at the beginning of the synthesis, which in turn promotes the formation of smaller nanoparticles [13]. The particle size distribution histograms shown in Figs. 2f and 2g demonstrate the dependence of the average nanoparticle size on the amount of added seeds nanoparticles. However, an increase of the volume of the aqueous solution $HAuCl_4$ provides a bigger number of gold ions per a growth center, which promotes the formation of larger nanoparticles. These dependencies correspond to the model of nanoparticle growth under controlled synthesis conditions [14].

Moreover, a slower addition rate of the $HAuCl_4$ solution leads to the formation of larger particles, as demonstrated in Fig. 2a. This can be explained by the fact that a low feeding rate of $HAuCl_4$ solution maintains a stable concentration of gold ions in the solution, reducing the probability of local supersaturation. Uniform addition of $HAuCl_4$ using a syringe pump promotes isotropic particle growth, and manual addition may cause temporary excesses of gold ions, resulting in non-uniform growth in Samples 6 and 7. The histogram shown in Fig. 2i exhibits a broad size distribution, which is probably due to the non-uniform rate of $HAuCl_4$ solution. To clearly demonstrate the observed trends, the sizes of the obtained gold nanoparticles are summarized in Tab. 3.

Table 3. Characteristics of all samples

Methods								
Turkevich–Frens method (Quasi–spherical samples)			Seed–mediated growth method (Spherical samples)			Seed–mediated growth method (Anisotropic samples)		
Sample	Size, nm	LSPR wavelength (nm)	Sample	Size, nm	LSPR wavelength (nm)	Sample	Size, nm	LSPR wavelength (nm)
1	22.0 ± 0.3	522	4	76.6 ± 3.2	546	6	35.7 ± 1.7	544
2	23.9 ± 0.8	528	5	94.2 ± 3.2	560	7	39.2 ± 1.9	552
3	27.1 ± 0.2	532				8	61.6 ± 6.4	574

The optical spectra can be used to analyse the effect of the morphology of gold nanoparticles on their optical properties.

Analysis of the experimental spectra (Fig. 3) reveals a redshift in the wavelength of maximum absorption with increasing particle size, as Tab. 3 shows. In the size range of 20–40 nm,

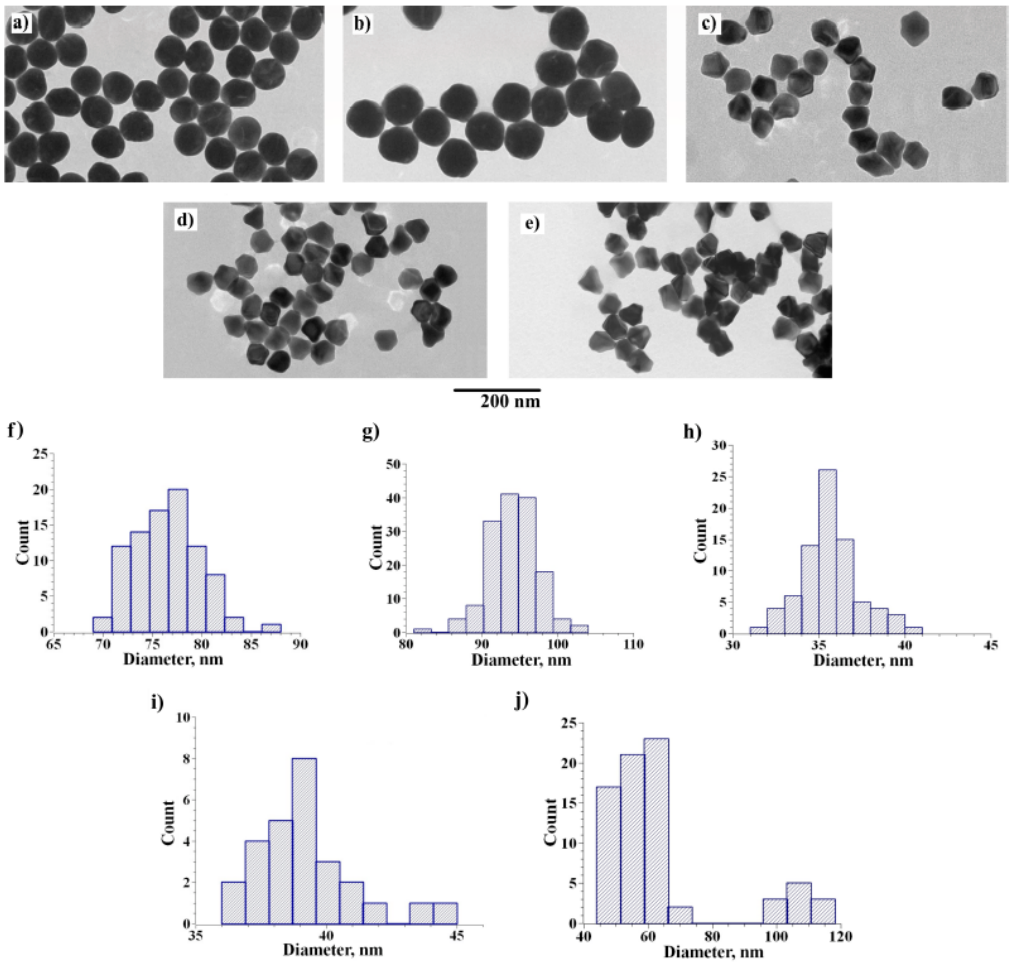


Fig. 2. TEM images of gold nanoparticles synthesised by the seed-mediated growth method: a) Sample 4; b) Sample 5; c) Sample 6; d) Sample 7; e) Sample 8. Particle size distribution histograms: f) Sample 4; g) Sample 5; h) Sample 6; i) Sample 7; j) Sample 8

an approximately linear relationship is observed between particle diameter and the LSPR peak position, which agrees with Mie theory [6] in the dipole approximation. In this approach, equation (1) was obtained for the polarizability of a nanoparticle α with a radius R in an external electromagnetic field with a wavelength λ under the condition $R \ll \lambda$, when the quasi-static approximation is applicable [15, 16]:

$$\alpha = 4\pi R^3 \varepsilon_{\text{out}} \frac{(\varepsilon_{\text{in}} - \varepsilon_{\text{out}})}{(\varepsilon_{\text{in}} + \chi\varepsilon_{\text{out}})}, \tag{1}$$

where R is the particle radius, ϵ_{in} is the complex frequency-dependent ω dielectric permittivity of the particle, ϵ_{out} is the permittivity of the surrounding medium; χ is a geometric factor (for spheres $\chi = 2$).

Analysis of this equation shows that the polarizability maximum, which determines the LSPR wavelength, occurs under the resonance condition $\epsilon_{\text{in}}{=}{-}\chi\epsilon_{\text{out}}$, with polarizability proportional to the particle (R^3) indicating that larger particles have higher polarizability.

The optical properties of spherical and anisotropic nanoparticles follow different trends. Samples 4 and 5 consisting of perfectly spherical nanoparticles, and samples 1, 2, and 3 consisting

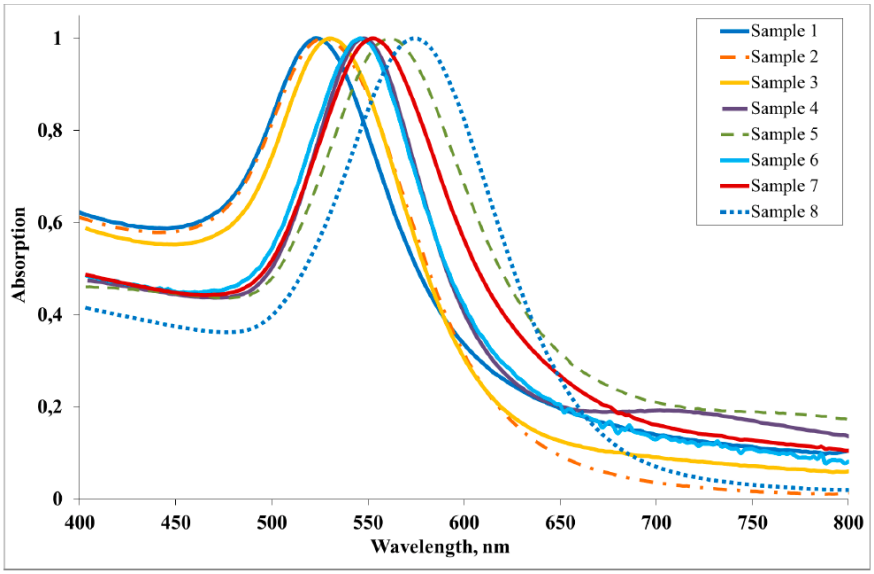


Fig. 3. Experimental absorption spectra of gold nanoparticles

of quasi-spherical nanoparticles with slight ellipsoidal deformations, maintain high symmetry, allowing the dipole approximation to calculate LSPR. However, samples 6, 7, and 8 exhibit more anisotropic shapes, shifting the absorption peak corresponding to the LSPR wavelength. These shifts are probably due to the non-spherical shape, splitting the resonance modes of spherical particles into multiple modes for complex-shaped nanoparticles, resulting in redshifts requiring consideration of higher-order modes [17].

Theoretical calculations of these modes and LSPR frequencies for non-spherical particles can be performed using numerical methods such as FDTD (finite-difference time domain) or FEM (finite-element method). However, they were not performed in this work due to computational complexity. Nevertheless, to assess the effect of anisotropic shape on LSPR wavelength shifts, the optical spectra of all samples were simulated using Equation (1) within the dipole approximation, assuming all particles are spherical (Fig. 4).

Analysis of the simulated absorption spectra for spherical particles shows a linear relationship between particle size and LSPR wavelength. Comparison with experimental spectra reveals deviations in the ordering of absorption maxima. Therefore, it can be concluded that nanoparticle anisotropy affects optical properties. Specifically, discrepancies are observed between expected and experimental peak positions: according to the theoretical model, the spectrum of sample 7 should correspond to the experimental spectrum of sample 4, and the theoretical spectrum of sample 8 corresponds to the experimental spectrum of sample 7. Similarly, the theoretical spectrum of sample 4 corresponds to the experimental position of sample 5, and that of sample 5 to sample 8. Tab. 4 presents a comparison of theoretical and experimental LSPR peak wavelengths for all samples.

The analysis of Tab. 4 demonstrates that samples 4 and 5 exhibit complete agreement between theoretical and experimental values, validating the dipole approximation. Small deviations for quasi-spherical samples 1, 2, and 3 may be attributed to slight geometric asymmetry. The largest discrepancies are observed for anisotropic samples 6, 7, and 8. In these cases, the discrepancy reaches 20–30 nm, which indicates significant contribution of particle geometry.

These results confirm that the morphology of gold nanoparticles directly influences LSPR spectral characteristics [18]. At the same time, it should be emphasized that the agreement

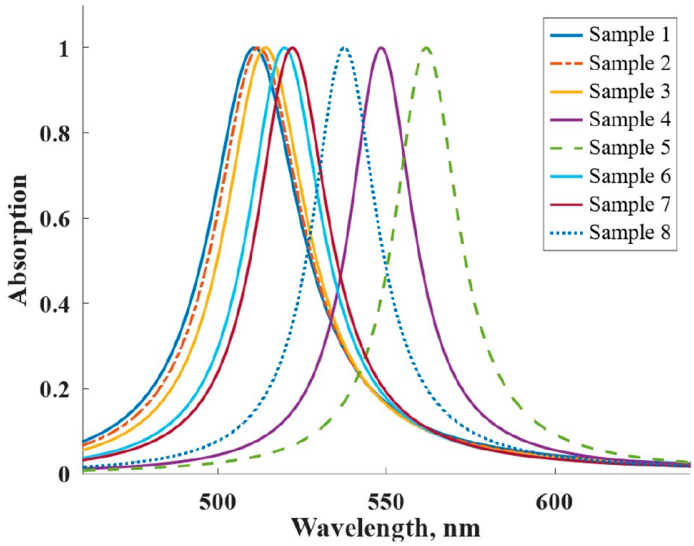


Fig. 4. Simulated optical spectra of gold nanoparticles

Table 4. Theoretical and experimental LSPR peak wavelengths

Sample	Shape of the sample	Theoretical LSPR peak wavelength (nm)	Experimental LSPR peak wavelength (nm)
1	Quasi–spherical with ellipsoidal deformations	511	522
2		512	528
3		514	532
4	Spherical	546	546
5		562	560
6	Anisotropic	521	544
7		524	552
8		540	574

between experimental absorption spectra and simulated absorption spectra performed within the dipole approximation is determined not only by particle shape, but also by the average particle size and the width of the size distribution. In particular, larger nanoparticles with narrow size distributions demonstrate better correspondence with the theoretical model, while increased polydispersity leads to spectral shifts.

Conclusion

In this study, two main methods for synthesizing colloidal gold were investigated and implemented: the Turkevich-Frens method and the seed-mediated growth method. The experiments confirmed that variations in synthesis conditions, such as temperature, reagent ratios, and the mode of $HAuCl_4$ solution addition, have a significant impact on particle size and final morphology, and consequently on the position of the LSPR peak. The Turkevich–Frens method demonstrated high efficiency for producing nanoparticles in the 20–30 nm size range. The seed-

mediated growth method enabled a wider size range of gold nanoparticles and allowed control over their morphology through a two-stage synthesis process. Comparison of the experimental optical spectra with theoretical models based on Mie theory showed good agreement for the LSPR peak positions of spherical particles. At the same time, non-spherical particles exhibited noticeable deviations between experimental and simulated spectra. Overall, this study highlights the importance of tuning synthesis parameters to control the optical properties of gold nanoparticles and emphasizes the necessity of considering particle morphology for accurate interpretation of their spectral characteristics.

The study has been supported by the Russian Science Foundation, Agreement no. 23-12-20007 and the Government of the Krasnoyarsk Territory and the Krasnoyarsk Territorial Foundation for Support of Scientific and R&D Activities, Agreement no. 256.

The electron microscopy studies were carried out on the equipment of Krasnoyarsk Territorial Center for Collective Use, Krasnoyarsk Scientific Center, Siberian Branch of Russian Academy of Sciences.

References

[1] L.A.Dykman, N.G.Khlebtsov, Gold nanoparticles in biology and medicine: recent advances and prospects, *Acta Naturae*, **3**(2011), no. 2 (in Russian).

[2] L.Nie, et al., Applications of gold nanoparticles in optical biosensors, *Journal of Biomedical Nanotechnology*, **10**(2014), no. 10. DOI: 10.1166/jbn.2014.1987

[3] M.V.Matveeva, et al., Perspectives of the use of nanomaterials in space projects, In: Novoe slovo v nauke i praktike: gipotezy i aprobatsiya rezultatov issledovanii, Новосибирск, 2016 (in Russian).

[4] I.V.Kavetskaya, et al., Optical properties of the gold nanoparticle, *Kondensirovannye sredy i mezhfaznye granitsy*, **11**(2009), no. 1 (in Russian).

[5] J.M.Park, et al., Recent advances in hollow gold nanostructures for biomedical applications, *Frontiers in Chemistry*, **9**(2021). DOI: 10.3389/fchem.2021.699284

[6] G.Mie, Über die Optik trüber Medien, speziell kolloidaler Metallösungen, *Annalen der Physik*, **330**(1908), no. 3.

[7] P.M.Bychkovskii, et al., Gold nanoparticles: synthesis, properties, biomedical applications, *Rossiyskii Bioterapevticheskii Zhurnal*, **10**(2011), no. 3 (in Russian).

[8] L.A.Dykman, N.G.Khlebtsov, Methods for chemical synthesis of colloidal gold, *Russ. Chem. Rev.*, **88**(2019), no. 3, 229–247. DOI: /10.1070/RCR4843

[9] N.G.Khlebtsov, L.A.Dykman, B.N.Khlebtsov, Synthesis and plasmonic tuning of gold and gold–silver nanoparticles, *Russ. Chem. Rev.*, **91**(2022), no. 10. DOI: 10.57634/RCR5058

[10] M.O.Besenhard, et al., New insight into the effect of mass transfer on the synthesis of silver and gold nanoparticles, *Cryst. Eng. Com.*, **20**(2018), no. 44. DOI: 10.1039/C8CE01014E

[11] Y.Zheng, et al., Successive, Seed-Mediated Growth for the Synthesis of Single-Crystal Gold Nanospheres with Uniform Diameters Controlled in the Range of 5–150 nm, *Particle & Particle Systems Characterization*, **31**(2014), no. 2. DOI: 10.1002/ppsc.201300256

[12] S.A.Tovstun, V.F.Razumov, Theoretical analysis of methods for the colloidal syn-thesis of monodisperse nanoparticles, *High Energy Chem.*, **44**(2010), no. 3, 196–203.

[13] N.R.Jana, L.Gearheart, C.J.Murphy, Evidence for Seed-Mediated Nucleation in the Chemical Reduction of Gold Salts to Gold Nanoparticles, *Chemistry of Materials*, **13**(2001), no. 7. DOI: 10.1021/cm000662n

[14] M.Chen, et al., Synthesis and optical properties of size-controlled gold nanoparticles, *Powder Technology*, **311**(2017). DOI: 10.1016/j.powtec.2017.01.087

[15] U.Kreibig, M.Vollmer, Optical Properties of Metal Clusters, Vol. 25, 2013.

[16] K M.Mayer, J.H.Hafner, Localized Surface Plasmon Resonance Sensors, *Chemical Reviews*, **111**(2011), no. 6. DOI: 10.1021/cr100313v

[17] Y.M.Lee, S.E.Kim, J.E.Park, Strong coupling in plasmonic metal nanoparticles, *Nano Convergence*, **10**(2023), no. 1. DOI: 10.1186/s40580-023-00383-5

[18] M.Shabaninezhad, G.Ramakrishna, Theoretical investigation of size, shape and aspect ratio effect on the LSPR sensitivity of hollow-gold nanoshells, *The Journal of Chemical Physics*, **150**(2019), no. 14. DOI: 10.1063/1.5090885

Химический синтез золотых наночастиц: механизмы образования и контроль морфологии

Амина Ш. Умарова
Сибирский государственный университет науки и технологий им. академика М. Ф. Решетнева
Красноярск, Российская Федерация
Анна В. Лукьяненко
Андрей В. Царенко
Александр С. Федоров
Александр С. Александровский
Михаил Н. Волочаев
Институт физики им. Л. В. Киренского
Федеральный исследовательский центр КНЦ СО РАН
Красноярск, Российская Федерация

Аннотация. В данной работе были синтезированы наночастицы золота двумя методами: методом Туркевича–Френса и методом роста на затравках. Описаны особенности каждого метода, их влияние на размер, морфологию и оптические свойства полученных частиц. С помощью оптической спектроскопии проанализирована зависимость длины волны плазмонного резонанса от формы наночастиц. Также показано, что изменение условий синтеза влияет на спектральные характеристики коллоидного золота.

Ключевые слова: золотые наночастицы, метод Туркевича–Френса, метод роста на затравках, локализованный поверхностно-плазмонный резонанс.

Journal of Siberian Federal University. Mathematics & Physics 2026, 19(3), 357–368

EDN: PQRZCV
УДК 621.396.669

Modification of Spatiotemporal Interference Suppression Algorithms to Improve Their Computational Efficiency and Accuracy of Navigation Parameter Measurements

Valery N. Tyapkin*
Dmitry D. Dmitriev†
Siberian Federal University
Krasnoyarsk, Russian Federation
Pavel V. Shtro
JSC NPP Radiosvyaz
Krasnoyarsk, Russian Federation
Igor V. Tyapkin‡
Evgenii D. Mikhov§
Vasilii N. Ratushniak¶
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 15.11.2025, received in revised form 19.01.2026, accepted 27.02.2026

Abstract. This article discusses an algorithm for optimal spatiotemporal processing of navigation parameters in the presence of interference. The problem of estimating radio navigation parameters in GNSS receivers equipped with an antenna array is addressed. An optimal adaptive algorithm for determining navigation parameters using an antenna array based on GNSS signals in an unknown interference environment is synthesized. The main approaches to improving the computational efficiency of adaptive spatial-temporal processing algorithms in GNSS receivers are presented. It is shown that recursive updating of the R matrix using the QR factorization based on Givens rotations significantly improves computational efficiency. Examples of applying the CORDIC algorithm to recursive estimation of the R matrix and antenna array weights are given.

Keywords: antenna array, spatiotemporal interference suppression algorithms, adaptive algorithms, global navigation satellite systems, GNSS receivers, unmanned aerial vehicles, QR decomposition, recurrent matrix estimation, computational efficiency, accuracy of measurement of navigation parameters.

Citation: V.N. Tyapkin, D.D. Dmitriev, P.V. Shtro, I.V. Tyapkin, E.D. Mikhov, V.N. Ratushniak, Modification of Spatiotemporal Interference Suppression Algorithms to Improve their Computational Efficiency and Accuracy of Navigation Parameter Measurements, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 357–368. EDN: PQRZCV.

*vtyapkin@sfu-kras.ru
†ddmitriev@sfu-kras.ru
‡ityapkin@sfu-kras.ru
§emikhov@sfu-kras.ru
¶vratushniyk@sfu-kras.ru

Introduction

Unmanned aerial vehicles (UAVs) are increasingly being used by government and commercial organizations, as well as private individuals. These applications include terrain mapping, disaster monitoring, search and rescue operations, cargo delivery, reconnaissance, patrolling remote areas and extended sites, traffic monitoring, and much more [1, 2]. The equipment on board the UAV must meet strict requirements for weight, dimensions, and power consumption. This necessitates the use of GNSS receivers as the foundation of UAV navigation systems. They offer low cost, weight, dimensions, and power consumption, as well as acceptable accuracy. Typically, a standard GNSS receiver provides a coordinate measurement error of less than 5 m. However, low noise immunity hinders their widespread use on UAVs.

The need to improve the interference immunity of onboard GNSS receivers is driven by the following factors [3]:

- the increase in natural and artificial radio interference. Interference has become a standard countermeasure against UAV navigation equipment [4];

- the potential for catastrophic consequences associated with the loss of a UAV or mission disruption;

- the availability of electronic countermeasures for individuals;

- economic feasibility - the cost of protection against interference is 10 times less than potential losses;

- regulatory requirements - interference immunity requirements also apply to commercial systems.

Taking into account the above, the implementation of multi-level protection against radio interference (hardware and algorithmic) should be mandatory when designing onboard navigation systems for UAVs of any class.

1. Synthesis of an algorithm for optimal spatiotemporal processing of navigation parameters against a background of interference

A general approach to solving the problem of optimal filtering of navigation signals against a background of interference is considered in [5–7]. The problem of optimal filtering of navigation parameters is generally formulated as follows:

A) The navigation parameters are expressed by a state vector, the optimal estimate of which must be obtained. In this case, the vector formation model is described by a stochastic differential equation [8]:

$$\frac{d\boldsymbol{\lambda}}{dt} = \mathbf{F}(t)\boldsymbol{\lambda}(t) + \mathbf{G}(t)\boldsymbol{\xi}(t),\ \lambda(0) = \lambda_0, \tag{1}$$

where $\mathbf{F}(t), \mathbf{G}(t)$ are known functions of time; $\boldsymbol{\xi}(t)$ is white Gaussian noise with two-sided spectral density; λ_0 is a random number distributed according to the Gaussian law with zero mathematical expectation and variance.

Under the assumption of mutual independence of noise, the posterior probability density of the state vector satisfies the Stratonovich equation and has a Gaussian approximation.

B) The observed process, i.e. the additive mixture of received navigation signals and inter-

ference, is represented as

$$\mathbf{Y}(t) = \mathbf{S}_{signals}(t, \boldsymbol{\lambda}(t)) + \mathbf{n}(t)),\ t \geqslant t_0, \tag{2}$$

where $\mathbf{S}_{signals}$ is the vector of useful signals; $\boldsymbol{\lambda}(t)$ is the vector of estimated parameters; $\mathbf{n}(t)$ is the vector of white Gaussian noise, which is also characterized by zero mathematical expectation and correlation function.

In most cases, the normal distribution law for received signals is physically justified. This is due to the fact that interference is normalized in the relatively narrowband receiving paths of a GNSS receiver. Based on this, the probability density function of the interference signal will be described by the expression:

$$P_{noise}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-y^2}{2\sigma^2}\right), \tag{3}$$

where σ^2 is the variance of the noise

Let's assume that the input of a multichannel M-element phased array antenna is exposed to a useful signal and interference. The set of signals received from the outputs of the M-element antenna is described by time functions $y_1(t), y_2(t), \ldots, y_M(t)$ and forms a column vector $\mathbf{y}(t) = [y_1(t), y_2(t), \ldots, y_M(t)]^T$. Single-channel reception ($M = 1$) can be considered a special case of multichannel reception.

Sampling of received signals, including the interference component, can be performed directly at the radio frequency. The key feature of this approach is that the sampling frequency must be very high. This is due to the fact that the interval between Δt readings should be approximately half the period of the high-frequency oscillation with a frequency f_0, i.e., $\Delta t \approx 1/(2f_0)$. The result is random discrete readings of radio interference, the mathematical representation of which as a column vector is most convenient.

In the case of multichannel reception, the vector of received signals can be written as follows:

$$\dot{\mathbf{Y}} = \begin{vmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_k & \cdots & \mathbf{y}_N \end{vmatrix}, \tag{4}$$

where $\mathbf{y}_k$ are the values of the received oscillations on the antenna elements at discrete moments of time $t_k = k\Delta t$, $k = \overline{1, N}$

The delay of the signal envelope and interference across the antenna array aperture can be neglected. Then the signals received by the antenna array will correspond to a multidimensional distribution density represented by the following formula:

$$p_{noise}(\mathbf{Y}_t) = \frac{1}{(2 \cdot \pi)^m |\boldsymbol{\Phi}|} exp(-\mathbf{Y}^H \dot{\boldsymbol{\Phi}}^{-1} \dot{\mathbf{Y}}/2), \tag{5}$$

where

$$\dot{\mathbf{Y}} = ||\dot{Y}_n||;\quad n = \overline{1, m};\quad \boldsymbol{\Phi} = ||\Phi_{ik}|| = M||\dot{Y}_i \dot{Y}_k^*/2||. \tag{6}$$

Here and below, the superscript H denotes the complex conjugate transposed matrix, and the superscript (*) denotes the complex conjugate matrix.

B) It is necessary to ensure optimization of the selected criterion. For example, use the maximum likelihood ratio criterion with a quadratic loss function and a Gaussian approximation of the posterior probability density function. This is valid if the estimation errors do not exceed the linear region of the discriminatory characteristics of the optimal discriminators.

Thus, the problem of estimating radio navigation parameters in general can be formulated as follows: using observations (2) and a priori information about the statistical characteristics of the processes, it is necessary to form the best estimate of the navigation parameters.

Let us consider a flat antenna array with M omnidirectional antenna elements (Fig. 1). Let us define a coordinate system $0XYZ$, the origin of which is combined with one of the antenna elements, and the $0XY$ plane coincides with the plane of the antenna array.

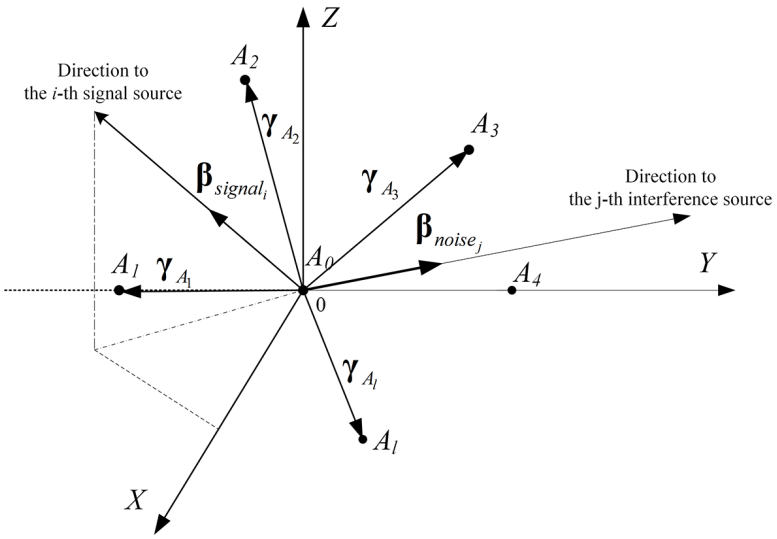


Fig. 1. Geometric interpretation of the arrangement of antenna array elements and received signals

The position of each l-th antenna element in the $OXYZ$ coordinate system is characterized by vector $\boldsymbol{\gamma}_{Al}$. The oscillations of useful signals and interference received by the antenna array have a plane wavefront. The directions to signal sources are specified by unit vectors $\boldsymbol{\beta}_{signal_i}$, and the directions to interference sources are specified by unit vectors $\boldsymbol{\beta}_{noise_j}$.

At the output of the l-th antenna element there is a continuous signal [8], described by the equation:

$$y_l(t) = \sum_{i=1}^{n} \sqrt{P_{signal_i}} s_{signal_i,l}(t, \lambda_{i,l}) + \sum_{j=1}^{p} \sqrt{P_{noise_j}} s_{noise_j,l}(t) + n_l(t), \;\; l = \overline{1,m}\,, \qquad (7)$$

where $s_{signal_i,l}(t, \lambda_{i,l})$ is the unit power signal of the i-th navigation satellite at the output of the l-th antenna element; P_{signal_i} is the signal power of the i-th navigation satellite; $s_{noise_j,l}$ is the j-th interference signal of unit power at the output of the l-th antenna element; P_{noise_j} is the power of the j-th interference signal; $n_l(t)$ is white Gaussian noise with correlation matrix $M[n_l(t)n_l(t+\tau)] = \delta(\tau)N_0/2$, $M[n_m(t)n_l(t_1)] = 0$, $m \neq l$ for any t and t_l; $\lambda_{i,l}$ is the vector of parameters that vary time and are subject to estimation.

The synthesis of an optimal adaptive algorithm for determining navigation parameters from GNSS signals using an antenna array in an unknown interference environment requires an estimate of the joint a posteriori probability density $p(\boldsymbol{\lambda}, \boldsymbol{\Phi}|\dot{\mathbf{Y}})$, where $\boldsymbol{\lambda}$ is the vector of the sought informative parameters, $\boldsymbol{\Phi}$ is the interference correlation matrix, $\mathbf{Y}$ is the observation vector at the output of the antenna array.

The estimate of the required informative parameters λ can be obtained from the relation:

$$p(\boldsymbol{\lambda}/\mathbf{Y}) = \int p(\boldsymbol{\lambda}, \boldsymbol{\Phi}/\mathbf{Y}) p(\boldsymbol{\Phi}/\mathbf{Y}) d\boldsymbol{\Phi}. \tag{8}$$

The interference correlation matrix $\boldsymbol{\Phi}$ can be calculated with sufficiently high accuracy, therefore expression (7) can be represented as:

$$\begin{cases} \hat{\boldsymbol{\Phi}} = \int p(\hat{\boldsymbol{\Phi}}/\mathbf{Y}) d\boldsymbol{\Phi}, \\ p(\boldsymbol{\lambda}/\mathbf{Y}) = p(\boldsymbol{\lambda}, \hat{\boldsymbol{\Phi}}/\mathbf{Y}). \end{cases} \tag{9}$$

Representing the optimal adaptive algorithm in an unknown interference environment in the form (9) allows us to divide it into two stages: 1) adaptation stage (estimation of the correlation matrix); 2) stage of estimating informative navigation parameters.

Naturally, the accuracy of algorithm (9) will depend on the accuracy of the estimate of the correlation matrix of noise $\hat{\boldsymbol{\Phi}}$. Therefore, when implementing algorithm (9), it is necessary to carry out a separate study of the influence of the error in estimating the correlation matrix of noise $\hat{\boldsymbol{\Phi}}$ on the accuracy characteristics of the estimates of the informative parameters $\boldsymbol{\lambda}$.

Using observation equation (7), we obtain an optimal algorithm for determining the parameters $\boldsymbol{\lambda}$ of the $S_{signal}(\boldsymbol{\lambda})$ signal based on multichannel observations $\mathbf{Y}$ over a certain time interval $t = \overline{1, T}$. The value of the interval T is not particularly significant. For convenience, it can be taken to correspond to 1 ms — the basic interval of GLONASS and GPS.

Taking into account the independence of noise in the received observation (9) at different points in time, the observation functional of the entire set of observations $\mathbf{Y}_k^T$ on the interval will be equal to [9]:

$$p(\mathbf{Y}_k^T|\boldsymbol{\lambda}) = \prod_{k=1}^{T} p(\mathbf{Y}_k|\boldsymbol{\lambda}). \tag{10}$$

The $\hat{\boldsymbol{\lambda}} = \max_{\lambda}{}^{-1} p(\mathbf{Y}_k^T|\boldsymbol{\lambda})$ estimate will be determined:

$$\begin{aligned} \hat{\boldsymbol{\lambda}} = \max_{\lambda}{}^{-1} Re\{\ln l(\boldsymbol{\lambda})\} = \max_{\lambda}{}^{-1} Re\{\zeta(\boldsymbol{\lambda})\} = \max_{\lambda}{}^{-1} Re\{\mathbf{H}\dot{\dot{\mathbf{S}}}_{signal}(\boldsymbol{\lambda})\}^H \dot{\boldsymbol{\Phi}}^{-1} \dot{\mathbf{Y}} = \\ = \max_{\lambda}{}^{-1} Re\{\mathbf{H}^* \boldsymbol{\Phi}^{-1} \mathbf{Y}_k \dot{\dot{\mathbf{S}}}^H_{signal,k}(\boldsymbol{\lambda}) \,. \end{aligned} \tag{11}$$

From the analysis of formula (11) it follows that the solution to the problem of signal processing during navigational-temporal determinations is based on the digital implementation in real time of matrix operations such as matrix inversion or the solution of systems of linear algebraic or differential equations (Kalman filter).

Another problem that arises when processing navigation signals is the effect of multipath. In this case, the signal from the navigation satellite arrives at the GNSS receiver not directly, but after additional reflection from surrounding objects (buildings, walls, ground, water, etc.). Multipath signals create correlated interference (they are coherent with the useful signal, but are delayed and come from other directions), distort the shape of the correlation peak (lead to errors in measuring pseudorange and carrier phase), and reduce the signal-to-noise ratio. To solve this problem, spatiotemporal signal processing in an antenna array is used. It combines spatial (antenna array) and temporal (delay line or filter chains on each antenna element of the array) processing into a single adaptive system. This allows: spatial interference selection by

creating deep notches in the antenna's radiation pattern; effective suppression of broadband and narrowband interference operating simultaneously; optimization of the reception of navigation signals operating along a direct propagation path and suppression of multipath signals.

Multipath suppression allows: to restore the sharp shape of the correlation peak; to reduce the magnitude of pseudorange measurement errors, which is especially critical for high-precision applications; to improve the accuracy and reliability of the GNSS receiver in challenging conditions (urban, mountainous, forested, etc.).

Thus, spatiotemporal processing is a necessary technology for modern high-precision and interference-immune GNSS receivers [9–11]. The main disadvantage of this technology is the extremely high computational complexity (inversion of a matrix of size $(M \times N) \times (M \times N)$ in real time) and the cost associated with the use of multichannel antenna arrays and powerful processors.

2. Main directions for improving the computational efficiency of adaptive algorithms for space-time processing in navigation receivers

The transition to implementing adaptive signal processing algorithms on FPGAs requires computations based on fixed-point representations. This makes solving spatiotemporal processing problems using conventional computational methods virtually impossible. The main possible method to solve this problem is the development and use of algorithms based on the calculation of not the sample covariance matrix of total noise $\mathbf{\Phi}$, but the functional of this matrix.

One of the radical approaches to solving the problem of spatiotemporal processing is the use of QR matrix decomposition algorithms. These algorithms are based on the possibility of representing an arbitrary matrix $\mathbf{A}$ as the product of an orthogonal matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$, i.e., $\mathbf{A} = \mathbf{Q} \times \mathbf{R}$ [12]. The QR factorization ensures numerical stability and reduces the computational complexity of algorithms compared to direct methods (e.g., matrix inversion). Furthermore, it allows updating antenna array weights without explicit matrix inversion. In this case, at each step, the $\mathbf{Q}$ and $\mathbf{R}$ matrices are updated (via Givens or Householder rotations), and the weight coefficients are calculated via back substitution.

When solving the algorithm synthesis problem using QR decomposition, a matrix of input signal samples (observations) of the adaptive filter, transformed using QR decomposition, is used instead of a correlation matrix. This matrix has dimensions $(M \times N)$. This reduces the dynamic range of the numbers involved in the calculations compared to the numbers involved when solving the same problem using an algorithm estimating the inverse correlation matrix of noise. This fact directly demonstrates the improved stability of the QR decomposition algorithm. Let's consider the main stages of this approach.

2.1. QR decomposition of the data matrix.

In the case of multichannel reception, the received signal vector has the form (4). In this case, the total number of samples will be $m = N \times M$, where N is the number of time samples and M is the number of antenna array elements. The dimension of the input data vector will be $(N \times M)$.

The components of the decomposition $\mathbf{Y} = \mathbf{Q} \times \mathbf{R}$ will have dimensions:

- $\mathbf{Q}$ is a unitary matrix $\mathbf{Q}^H \times \mathbf{Q} = \mathbf{I}$ of dimensions $N \times M$;
- $\mathbf{R}$ is an upper triangular matrix of dimensions $M \times M$.

2.2. Relationship with the covariance matrix:

Let us represent the covariance matrix $\mathbf{\Phi}$ using the matrices $\mathbf{Q}$ and $\mathbf{R}$:

$$\mathbf{\Phi} = \mathbf{Y}^H \times \mathbf{Y} = (\mathbf{Q} \times \mathbf{R})^H \times (\mathbf{Q} \times \mathbf{R}) = \mathbf{R}^H \times \mathbf{Q}^H \times \mathbf{Q} \times \mathbf{R} = \mathbf{R}^H \times \mathbf{R} \tag{12}$$

From (12) it follows that the estimation of the covariance matrix is reduced to the estimation of the upper triangular matrix $\mathbf{R}$ and requires fewer calculations:

$$\mathbf{\Phi} \approx \frac{1}{N}\mathbf{Y}^H\mathbf{Y} = \frac{1}{\mathbf{N}}\mathbf{R}^H \times \mathbf{R} \tag{13}$$

2.3. Recursive updating of the matrix R

Recursive updating of the matrix $\mathbf{R}$ eliminates the need to recalculate the entire QR decomposition when new data samples $\mathbf{y}_{(n+1)}$ arrive. This is achieved by using a modified QR decomposition update via Givens rotations or Householder reflections. Let's consider the basic steps for updating the matrix $\mathbf{R}_{(n+1)}$ from the previous $\mathbf{R}_{(n)}$ without completely recalculating the QR decomposition:

Step 1. Assume that at step n there is an upper triangular matrix $\mathbf{R}(n)$ of dimension $M \times M$, where M is the number of antenna array elements. In the absence of a priori data, the identity matrix multiplied by a loading coefficient equal to $K_{load} = 0.1 \ldots 0.3$ can be used as an initial estimate of matrix $\mathbf{R}$.

Step 2. When a new data vector $\mathbf{y}_{(n+1)}$ is received in the form of a column of size M, an extended matrix $\mathbf{R}_{ext}$ is formed:

$$\mathbf{R}_{ext} = \sqrt{\lambda_{forg}} \begin{vmatrix} r_{11(n)} & r_{12(n)} & r_{13(n)} & \cdots & r_{1M(n)} \\ 0 & r_{22(n)} & r_{23(n)} & \cdots & r_{2M(n)} \\ 0 & 0 & r_{33(n)} & \cdots & r_{3M(n)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & r_{MM(n)} \\ y_{1(n+1)} & y_{2(n+1)} & y_{3(n+1)} & \cdots & y_{M(n+1)} \end{vmatrix} \tag{14}$$

where λ_{forg} is the forgetting parameter, the subscript (n) means the current step, (n + 1) means the next step of adaptation.

The forgetting parameter can take values $0 < \lambda_{forg} \leqslant 1$, but is usually equal to $0.95 \ldots 0.999$. The closer λ_{forg} to 1, the greater the «memory» of the algorithm and the slower the adaptation; the smaller λ_{forg}, the faster old data is forgotten and the faster the adaptation. The resulting matrix has dimensions $(M + 1) \times M$.

By applying the Givens rotation sequence, the elements of the vector $\mathbf{y}(n+1)$ are set to zero, starting from the first element to the last element (usually M rotations).

$$\begin{vmatrix} \mathbf{R}_{M(n+1)} \\ 0_M^T \end{vmatrix} = \sqrt{\lambda_{forg}} \begin{vmatrix} r_{11(n+1)} & r_{12(n+1)} & r_{13(n+1)} & \cdots & r_{1M(n+1)} \\ 0 & r_{22(n+1)} & r_{23(n+1)} & \cdots & r_{2M(n+1)} \\ 0 & 0 & r_{33(n+1)} & \cdots & r_{3M(n+1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & r_{MM(n+1)} \\ 0 & 0 & 0 & \cdots & 0 \end{vmatrix} \tag{15}$$

Let's consider the proposed algorithm in more detail. Suppose we need to zero out the element $\mathbf{r}_{ij}$ in the augmented matrix $\mathbf{R}_{ext}$ (here i is the row number, and j is the column number of the matrix). This is accomplished by using the Givens rotation $\mathbf{G}(i,j,\theta)$, which acts on rows i and j. The Givens rotation is determined by the angle θ. It does not need to be calculated, it is sufficient to find the values of $\cos(\theta)$ (denoted c) and $\sin(\theta)$ (denoted s). Denoting $a = r_{jj}$ — the element on the matrix diagonal in the target row j and $b = r_{ij}$ — the matrix element to be zeroed in row i, we obtain the rotation coefficients $c = a/\sqrt{a^2+b^2}$; $s = -b/\sqrt{a^2+b^2}$. The rotation $\mathbf{G}(i,j,\theta)$ is applied to the entire matrix $\mathbf{R}_{ext}$, but since it only affects rows j and i, the entire matrix does not need to be recalculated, and only these two rows are updated. To do this, in rows i and j in each column k of the matrix $\mathbf{R}_{ext}$, the calculation of new values of its elements r'_{ik} and r'_{jk} is performed in accordance with the formulas $r'_{jk} = c \cdot r_{jk} - s \cdot r_{ik}$. Returning to the stated problem of zeroing the element r_{ij} in the extended matrix $\mathbf{R}_{ext}$, we obtain: for the diagonal element $r_{ij} - r'_{jj} = c \cdot a - s \cdot b = \dfrac{a}{\sqrt{a^2+b^2}} \cdot a - \dfrac{(-b)}{\sqrt{a^2+b^2}} \cdot b = \sqrt{a^2+b^2}$; for the element r_{ij}, equated to zero - $r'_{ij} = s \cdot a + c \cdot b = \dfrac{(-b)}{\sqrt{a^2+b^2}} \cdot a + \dfrac{a}{\sqrt{a^2+b^2}} \cdot b = 0$

Thus, the algorithm for converting the matrix $\mathbf{R}_{ext}$ to upper triangular form can be represented as a step-by-step process of zeroing the elements of the new row from left to right. For this purpose, the Givens rotation method is used between the new row and the corresponding rows of the original matrix $\mathbf{R}$.

Let's consider an example of using Givens rotations to zero out the elements of a new row of the matrix $\mathbf{R}_{ext}$ In the first step, we set the element $y_1(n+1)$ in (14), located at position $(+1, 1)$, to zero.

Let's calculate the coefficients c and s: $c = r_{11}/\sqrt{r_{11}^2 + y_{1(n+1)}^2}$, $s = -y_{1(n+1)}/\sqrt{r_{11}^2 + y_{1(n+1)}^2}$. The desired value of the element r'_{11} will be equal to $\sqrt{r_{11}^2 + y_{1(n+1)}^2}$ and the element $y_1(n+1) - 0$. As a result, we obtain the transformed matrix $\mathbf{R'}_{ext}$ (the indices (n) and (n + 1) of the matrix elements are not shown):

$$\mathbf{R'}_{ext} = \sqrt{\lambda_{forg}} \begin{vmatrix} r'_{11} & r'_{12} & r'_{13} & \dots & r'_{1M} \\ 0 & r_{22} & r_{23} & \dots & r_{2M} \\ 0 & 0 & r_{33} & \dots & r_{3M} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & r_{MM} \\ 0 & y'_2 & y'_3 & \dots & y'_M \end{vmatrix} \tag{16}$$

The second step is to set the second element of the new row y'_2 of the matrix $\mathbf{R}'_{ext}$ equal to zero. We calculate the coefficients c and s: $c = r_{22}/\sqrt{r_{22}^2 + y'^2_2}$, $s = -y'_2/\sqrt{r_{22}^2 + y'^2_2}$.

After processing all columns, we obtain a matrix described by formula (15).

The advantages of this approach include high computational efficiency, as only the values of one row need be set to zero, while the original zeros below the matrix diagonal are preserved.

2.4. Application of the CORDIC algorithm for recursive evaluation of the matrix R

The CORDIC algorithm allows one to calculate the key elements of the matrix $\mathbf{R}$ obtained using the Givens rotation (cos and sin) using a sequence of simple and fast shift and addition

operations. The Givens rotation matrix $G(i, j, \theta)$ in the plane looks like this:

$$\mathbf{G}(i, j, \theta) = \begin{vmatrix} c & s \\ -s & c \end{vmatrix} \tag{17}$$

Applying it to the elements of the extended matrix $\mathbf{R}_{ext}$ (14), we obtain the desired values: $\begin{vmatrix} c & s \\ -s & c \end{vmatrix} \times \begin{vmatrix} r_{11} \\ y_{1(n+1)} \end{vmatrix} = \begin{vmatrix} c \cdot r_{11} + s \cdot y_{1(n+1)} \\ -s \cdot y_{1(n+1)} + c \cdot r_{11} \end{vmatrix} = \begin{vmatrix} r'_{11} \\ 0 \end{vmatrix}$. Multiplying a vector by this matrix requires 4 multiplication operations and 2 additions. The CORDIC algorithm replaces multiplication operations with a sequence of microrotations, each performed through shift and addition operations. The angle θ is represented as the sum of pre-computed angles:

$$\theta \approx \sum_{i=0}^{n-1} \nu_i \alpha_i, \tag{18}$$

where $\alpha_i = \text{acrtan}(2^i)$, $\nu_i \in \{-1, 1\}$.

At each step i, the rotation by angle αi is performed as follows $\begin{cases} r'_{11} = r_{11} - \nu_i \cdot y_{1(m+1)} \cdot 2^{-i}, \\ 0 = y_{1(m+1)} + \nu_i \cdot r_{11} \cdot 2^{-i} \end{cases}$ where 2^{-i} is the bit shift, which in real hardware does not consume computing resources. After n iterations, the vector is multiplied by the scaling coefficient $K_{sc} = \prod_{i=0}^{n-1} \sqrt{1 + 2^{-2i}}$ which can be pre-calculated and taken into account at the end of the calculations.

2.5. Diagonal loading

Diagonal loading allows to increase the robustness of signal processing algorithms by improving the conditionality of the covariance matrix, especially when the number of samples is small or when there are collinear signals. The solution to this problem is possible by modifying the matrix by adding a small random value to its diagonal elements.

$$\mathbf{R}_{mod} = \mathbf{R} + K_{load} \cdot \sigma^2 \cdot \mathbf{I}, \tag{19}$$

where K_{load} is the load coefficient; σ^2 is the uncorrelated unit noise; $\mathbf{I}$ is the identity matrix K_{load} is usually taken to be $0,1 \ldots 0,3$ of the mean value of the diagonal elements.

2.6. Calculating antenna array weighting coefficients

The antenna array weighting coefficients $\mathbf{w}_{(n)}$ are calculated by the back substitution method through the solution of the Wiener–Hopf equation in matrix form:

$$\mathbf{\Phi}_{(n)} \times \mathbf{w}_{(n)} = \mathbf{H}_{(n)}, \tag{20}$$

where $\mathbf{\Phi}(n)$ is a covariance matrix of size $M \times M$ (M is the number of antenna array elements) with the Emyrt property; $\mathbf{w}(n)$is the vector of antenna array weight coefficients; $\mathbf{H}_{(n)}$ is the vector of the direction to the useful signal.

To calculate the normalized weight vector, it is necessary to solve the matrix equation (20) for $\mathbf{w}(n)$.

$$\mathbf{w}(n) = \frac{\mathbf{\Phi}_{(n)}^{-1} \times \mathbf{H}_{(n)}}{\mathbf{H}_{(n)}^{H} \mathbf{\Phi}_{(n)}^{-1} \times \mathbf{H}_{(n)}}. \tag{21}$$

Substituting equation (12) into (21), we obtain

$$w(n) = \frac{\left(\mathbf{R}_{(n)}^H \mathbf{R}_{(n)}\right)^{-1} \times \mathbf{H}_{(n)}}{\mathbf{H}_{(n)}^H \left(\mathbf{R}_{(n)}^H \mathbf{R}_{(n)}\right)^{-1} \times \mathbf{H}_{(n)}}. \tag{22}$$

We introduce an auxiliary vector V such that $\mathbf{R}^H \times \mathbf{V} = \mathbf{H}$, where $\mathbf{R}^H$ is a lower triangular matrix, and solve this system with respect to $\mathbf{V}$. Considering that $\mathbf{R}^H$ is a lower triangular matrix, the system is solved from top to bottom by the direct substitution method.

$$\begin{cases} \nu_1 = h_1 / r_{11}^H \\ \nu_2 = (h_2 - r_{12}^H \nu_1) / r_{22}^H \\ \nu_3 = (h_3 - r_{13}^H \nu_1 - r_{23}^H \nu_2) / r_{33}^H \\ \nu_4 = (h_4 - r_{14}^H \nu_1 - r_{24}^H \nu_2 - r_{34}^H \nu_3) / r_{44}^H \\ \quad \cdots \\ \nu_m = (h_m - r_{1m}^H \nu_1 - r_{2m}^H \nu_2 - \cdots - r_{(m-1)m}^H \nu_{(m-1)}) / r_{mm}^H \end{cases} \tag{23}$$

From here $\mathbf{R} \times \mathbf{w} = \mathbf{V}$ and the system is transformed to a form with an upper triangular matrix $\mathbf{R}$:

$$\begin{vmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1M} \\ 0 & r_{22} & r_{23} & \cdots & r_{2M} \\ 0 & 0 & r_{33} & \cdots & r_{3M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & r_{MM} \end{vmatrix} \times \begin{vmatrix} w_1 \\ w_2 \\ w_3 \\ \cdots \\ w_M \end{vmatrix} = \begin{vmatrix} v_1 \\ v_2 \\ v_3 \\ \cdots \\ v_M \end{vmatrix} \tag{24}$$

This equation is solved by the backward substitution method, starting with the last equation:

$$\begin{cases} w_m = v_m / r_{mm} \\ w_{(m-1)} = (v_{(m-1)} - r_{(m-1)(m)} w_{(m)}) / r_{(m-1)(m-1)} \\ \quad \cdots \\ w_3 = (v_3 - r_{34} w_4 - r_{35} w_5 - r_{36} w_6 \cdots - r_{3m} w_m) / r_{33} \\ w_2 = (v_2 - r_{23} w_3 - r_{24} w_4 - r_{25} w_5 \cdots - r_{2m} w_m) / r_{22} \\ w_1 = (v_1 - r_{12} w_2 - r_{13} w_3 - r_{14} w_4 \cdots - r_{1m} w_m) / r_{11} \end{cases} \tag{25}$$

Thus, the sequential reduction of the matrix $\mathbf{R}$ first to a lower triangular and then to an upper triangular form allows us to significantly simplify the finding of the values of the weight coefficients of the antenna array.

Conclusion

Thus, the paper presents the main approaches to improving the computational efficiency of adaptive spatial-temporal processing algorithms in radio navigation systems. The advantage of using QR decomposition algorithms to calculate the sample covariance matrix of the total interference F is as follows: high numerical stability, since unitary transformations (rotations) do not increase the error rate; recursivity, which enables efficient updating of weighting coefficients when new data arrives; parallel execution of the algorithm is possible, since Givens rotations can

be performed in parallel; no degeneracy problem, since the **R** matrix is always well-conditioned due to the use of diagonal loading.

Thus, determining the array weights using the upper-triangular matrix **R** of the QR transform is a modern, numerically stable method that enables updating the adaptive array weights in real time. This approach has wide application in real-world digital signal processing systems, such as adaptive antenna arrays and noise suppression systems.

The study was supported by a grant from the Russian Science Foundation no. 25–19–20070, https://rscf.ru/project/25–19–20070/, and a grant from the Krasnoyarsk Regional Science Foundation

References

[1] B.Alzahrani, O.S.Oubbati, A.Barnawi, M.Atiquzzaman, D.Alghazzawi, UAV assistance paradigm: State-f-the-art in applications and challenges, *Journal of Network and Computer Applications*, **166**(2020), no. 1. DOI:10.1016/j.jnca.2020.102706

[2] X.Li, A.V.Savkin, Networked Unmanned Aerial Vehicles for Surveillance and Monitoring: A Survey, *Future Internet*, **13**(2021), no. 7, 174. DOI: 10.3390/fi13070174

[3] A.S.Raskaliyev, Interference immunity of GNSS signals, *Bulletin of KazATC*, **97**(2016), no. 2.

[4] S.I.Babusenko, V.V.Kiryushkin, A.V.Zhuravlev,UBLOX Global Navigation Satellite System Modules Noise Immunity Experimental Evaluation, *Radioengineering*, **85**(2021), no. 6, 77-85 (in Russian). DOI 10.18127/j00338486-202106-13

[5] GLONASS. Principles of Design and Operation, Ed. 4nd,revised and additional, Edited by A.I.Perov, V.N.Kharisov, Moscow, Radiotekhnika, 2010.

[6] A.I.Perov, Methods and algorithms for optimal signal reception in consumer equipment of satellite radio navigation systems,Moscow,Radiotekhnika,2012.

[7] Statistical Theory of Radio Engineering Systems,Ed. 2nd,revised and additional, Edited by A.I.Perov, Moscow, Radiotekhnika, 2022.

[8] P.V.Shtro, V.N.Tyapkin, P.V.Luferchik, et al., Assessment of the Accuracy Characteristics of Navigation Receivers with Interference Compensator in Complex Interference Environments, *Radioengineering*, **89**(2025), no. 2, 65–72. DOI 10.18127/j00338486-202502-09.

[9] V.N.Tyapkin, D.D.Dmitriev, P.V.Shtro, et al., Model of Receiving Channels of an Adaptive Antenna Array to Assess theImpact of Differences in their Characteristics on the Efficiency of Interference Suppression, *J. Sib. Fed. Univ. Math. Phys.*, 18(2025), no. 1, 81–90. EDN: MLYWGB

[10] V.N.Tyapkin, Minimum Variance Distortionless Response Adaptive Beamforming, Science Intensive Technologies, **17**(2016), no. 8, 39–43 (in Russian).

[11] V.N.Tyapkin, V.N.Ratushnyak, D.D.Dmitriev, A.B.Gladyshev, Increase of Noise Suppression Efficiency Due to Correction of Frequency Characteristics of Receiving Channels in the Navigation Equipment of Consumers, *Spacecrafts & Technologies*, **2**(2018), no. 1, 30–37 (in Russian). DOI: 10.26732/2618-7957-2018-1-30-37

[12] V.I.Dzhigan, Adaptive signal filtering: theory and algorithms, Moscow, Technosphere, 2013.

Модификация пространственно-временных алгоритмов подавления помех в целях улучшения их вычислительной эффективности и точности измерения навигационных параметров

Валерий Н. Тяпкин
Дмитрий Д. Дмитриев
Сибирский федеральный университет
Красноярск, Российская Федерация
Павел В. Штро
АО «НПП «Радиосвязь»
Российская Федерация
Игорь В. Тяпкин
Евгений Д. Михов
Василий Н. Ратушняк
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. В статье рассмотрен алгоритм оптимальной пространственно-временной обработки навигационных параметров на фоне помех. Раскрыта задача оценки радионавигационных параметров в условиях помех в ГНСС-приемниках, оснащенных антенной решеткой. Проведен синтез оптимального адаптивного алгоритма определения навигационных параметров с антенной решеткой по сигналам ГНСС в условиях неизвестной помеховой обстановки. Приведены основные пути улучшения вычислительной эффективности адаптивных алгоритмов пространственно-временной обработки в ГНСС-приемниках. Показано, что рекуррентное обновление матрицы R с использованием QR-разложения на основе вращений Гивенса позволяет существенно улучшить эффективность вычислений. Приведены примеры применения алгоритма CORDIC для рекуррентной оценки матрицы R и весовых коэффициентов антенной решетки.

Ключевые слова: антенная решетка, пространственно-временные алгоритмы подавления помех, адаптивные алгоритмы, глобальные навигационные спутниковые системы, ГНСС-приемники, беспилотные летательные аппараты, QR-разложение, рекуррентная оценка матриц, вычислительная эффективность, точность измерения навигационных параметров.

EDN: TOLSVD
УДК 519.872

Transient Asymptotic Analysis of Queueing System with Non-stationary Poisson Arrivals and Server Unit Switches

Yana A. Tyulenina*
Ekaterina A. Fedorova†
National Research Tomsk State University
Tomsk, Russian Federation

Received 18.12.2025, received in revised form 21.01.2026, accepted 27.02.2026

Abstract. The paper presents a study of a queuing system with two service units, an unlimited number of servers, and a non-stationary arrival Poisson process as a mathematical model of a hybrid power system combining two energy sources. The transient asymptotic method is proposed for the deriving the time-dependent formulas for the probability distribution of the number of customers in the system under the limit condition of a high arrival rate. Gaussian form of the asymptotic probability distribution of the number of customers in the system is proved. The results of the numerical analysis are presented.

Keywords: mathematical modelling, queueing theory, transient asymptotic analysis, non-stationary Poisson arrival process.

Citation: Y.A. Tyulenina, E.A. Fedorova, Transient Asymptotic Analysis of Queueing System with Non-stationary Poisson Arrivals and Server Unit Switches, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 369–380. EDN: TOLSVD.

Introduction

Renewable energy is becoming increasingly important in the modern world. The escalating global demand for energy together with climate changes and the depletion of fossil fuel reserves are reasons of the development of renewable energy sources. However, its instability (particularly for solar and wind power) presents significant challenges to grid reliability. In addition, renewable energy plants have less capacity per unit comparing fuel ones. In this way, it is necessary to construct hybrid energy supply systems combining both renewable and traditional energy sources. Also, hybrid energy systems have several advantages, such as independence from municipal power plants in the event of failures or accidents [1, 2].

The number of theoretical studies devoted to the analysis of hybrid energy systems remains relatively small [3–7,9]. Most of the existing works focus on an economic assessment of investment projects related to the implementation of renewable energy sources [7], or on some technical parameters calculation of the designed energy systems based on mathematical physics methods and electrical engineering [2–4]. Usually, authors make decisions based on averaged values of data, without taking into account the random nature of the processes under study. To explain this issue let us note that the operation of traditional energy supply systems is usually considered as deterministic. But it is known that the volumes of electricity production by a renewable source are stochastic dependent on climatic conditions. Additionally, sometimes electricity consumption can be unpredictable [4, 5]. Thus, not taking stochastic factors into account leads to significant deviations in the estimates of the performance characteristics of power plants.

Stochastic modelling, such as queuing theory, is rarely used in energy field. Examples are presented in papers [8–11], where queuing theory perspectives are described for energy systems

*yanccch@yandex.ru https://orcid.org/0009-0003-7704-7237
†ekat_fedorova@mail.ru https://orcid.org/0000-0001-8933-5322

study. It can be noted that authors consider the simplest queueing models (such as a Markovian single-server queue and a queueing network) and use known formulas. Hybrid energy systems have not been studied in these terms yet.

In this paper, we apply queueing theory approach for modelling hybrid energy systems. We propose a queueing system (QS) with two service units defining two energy sources. The service unit defining a renewable energy source is unstable, it has random operating times. The other one is stable. The arrival process in QS describes electricity consumption requests, the number of servers is unlimited, and presents the number of turning on devices in electricity grid. The asymptotic analysis approach [13, 14] is developed for the model under study.

The rest of the paper is organized as follows. Section 1 provides the detailed description of the developed mathematical model in terms of queueing theory. Section 2 is devoted to the asymptotic analysis method under the limit condition of a high arrival rate. Section 3 presents numerical examples.

1. Mathematical model

Let us consider a mathematical model of a hybrid energy system in the form of the following queueing system. There are two service units with unlimited numbers of servers, which model two energy sources. One service unit is stable, the second one has a stochastic nature. It turns on and off during random times distributed exponentially with parameters γ_1 and γ_0 respectively. Customers arrive in the system according non-stationary Poisson arrival process with intensity $\lambda(t)$ and define consumers demand for electricity. Only one energy source can be used. In this way, the service is prioritized for the second (unstable) unit, if it is enabled. The service time (electricity consumption time) is distributed exponentially with parameters μ_1 and μ_2 for the first and the second service units respectively. The proposed model is illustrated schematically in Fig. 1.

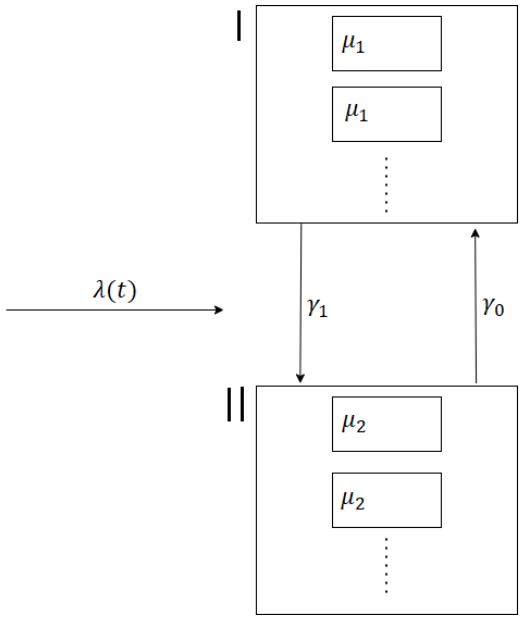


Fig. 1. Mathematical model

Note that the considered model is a particular case of infinite queueing systems in a random

environment with two states and a variable service parameter (an example is studied in [14]).

Let us introduce $i(t)$ — the number of customers in the system at time t, $k(t) = 1, 2$ — the state of the service units at time t, where state "1" means that the unstable unit is disabled, so the service be in the first unit; state "2" means that the second unstable service unit is working.

We denote the probability that the "k"-th service unit is enabled and there are i customers in the system at time t by probabilities $P(k, i, t)$, where $k(t) = 1, 2$, $i(t) = 0, 1, \dots$.

The stochastic process $\{k(t), i(t)\}$ is a two-dimensional continuous-time Markov process. The following system of Kolmogorov differential equations for $P(k, i, t)$ can be composed.

$$\begin{cases} \dfrac{\partial P(1,i,t)}{\partial t} = -P(1,i,t)(\lambda(t) + i\mu_1 + \gamma_1) + \\ \quad +P(2,i,t)\gamma_0 + P(1,i-1,t)\lambda(t) + P(1,i+1,t)(i+1)\mu_1, \\ \dfrac{\partial P(2,i,t)}{\partial t} = -P(2,i,t)(\lambda(t) + i\mu_2 + \gamma_0) + \\ \quad +P(1,i,t)\gamma_1 + P(2,i-1,t)\lambda(t) + P(2,i+1,t)(i+1)\mu_2, \end{cases} \tag{1}$$

where $P(k, -1, t) \equiv 0$.

Let us introduce the partial characteristic functions in the following form

$$H_k(u,t) = \sum_{i=0}^{\infty} e^{jui} P(k,i,t), \quad j = \sqrt{-1}.$$

System of equations (1) is rewritten for the partial characteristic functions as follows

$$\begin{cases} \dfrac{\partial H_1(u,t)}{\partial t} = H_1(u,t)(\lambda(t)(e^{ju} - 1) - \gamma_1) + j\mu_1 \dfrac{\partial H_1(u,t)}{\partial u}(e^{-ju} - 1) + \gamma_0 H_2(u,t), \\ \dfrac{\partial H_2(u,t)}{\partial t} = H_2(u,t)(\lambda(t)(e^{ju} - 1) - \gamma_0) + j\mu_2 \dfrac{\partial H_2(u,t)}{\partial u}(e^{-ju} - 1) + \gamma_1 H_1(u,t). \end{cases} \tag{2}$$

By summing equations (2), an additional equation can be derived.

$$\frac{\partial H_1(u,t)}{\partial t} + \frac{\partial H_2(u,t)}{\partial t} = \lambda(t)(H_1(u,t) + H_2(u,t))(e^{ju} - 1) - \\ - \left(j\mu_1 \frac{\partial H_1(u,t)}{\partial u} + j\mu_2 \frac{\partial H_2(u,t)}{\partial u} \right)(e^{-ju} - 1) \tag{3}$$

System (2)–(3) is hardly solved directly, so the asymptotic method is proposed.

2. Asymptotic analysis

The method of asymptotic analysis in the queuing theory is an approach for solving equations determining any probability characteristics under the given limit condition, which is specified for different models and practical aims (i.e. asymptotic conditions of long service time, heavy load, long delay in an orbit, etc.) [13, 14].

In the paper, we will consider the system under a limit condition of a high arrival rate as the most important case in real energy systems. Here, we will develop the asymptotic method proposed in [13].

First, we introduce an infinitely large parameter $N \to \infty$ and present the arrival intensity as $\lambda(t) = \lambda_1(t)N$. Traditionally, the asymptotic analysis has two steps.

2.1. First-order asymptotics

Let us perform the following asymptotic substitutions

$$N = \frac{1}{\epsilon}, \quad \lambda(t) = N\lambda_1(t) = \frac{1}{\epsilon}\lambda_1(t), \quad \gamma_m = \frac{1}{\epsilon}\delta_m, \tag{4}$$
$$u = \epsilon w, \ H_k(u,t) = F_k(w,t,\epsilon),$$

where $\epsilon \to 0$ is an infinitesimal parameter.

Let us note that the introduced notations imply a transition to the study of the characteristic function of stochastic process $\epsilon \cdot i(t)$.

By substituting the notations into system (2)–(3), the following asymptotic equations can be written

$$\begin{cases} \dfrac{\partial F_1(w,t,\epsilon)}{\partial t} = F_1(w,t,\epsilon)\left(\dfrac{\lambda_1(t)}{\epsilon}(e^{j\epsilon w}-1) - \dfrac{\delta_1}{\epsilon}\right) - \\ \qquad -j\mu_1\dfrac{\partial F_1(w,t,\epsilon)}{\partial \epsilon w}(e^{-j\epsilon w}-1) + \dfrac{\delta_0}{\epsilon}F_2(w,t,\epsilon), \\ \dfrac{\partial F_2(w,t,\epsilon)}{\partial t} = F_2(w,t,\epsilon)\left(\dfrac{\lambda_1(t)}{\epsilon}(e^{j\epsilon w}-1) - \dfrac{\delta_0}{\epsilon}\right) - \\ \qquad -j\mu_2\dfrac{\partial F_2(w,t,\epsilon)}{\partial \epsilon w}(e^{-j\epsilon w}-1) + \dfrac{\delta_1}{\epsilon}F_1(w,t,\epsilon) \end{cases} \tag{5}$$

and the additional equation

$$\frac{\partial F_1(w,t,\epsilon)}{\partial t} + \frac{\partial F_2(w,t,\epsilon)}{\partial t} = (F_1(w,t,\epsilon)+F_2(w,t,\epsilon))\frac{\lambda_1(t)}{\epsilon}(e^{jw\epsilon}-1) - \\ - \left(j\mu_1\frac{\partial F_1(w,t,\epsilon)}{\partial w\epsilon} + j\mu_2\frac{\partial F_2(w,t,\epsilon)}{\partial w\epsilon}\right)(e^{-jw\epsilon}-1). \tag{6}$$

Multiplying by ϵ and applying Taylor series, we obtain

$$\begin{cases} \epsilon\dfrac{\partial F_1(w,t,\epsilon)}{\partial t} = F_1(w,t,\epsilon)(\lambda_1(t)j\epsilon w - \delta_1) + j\mu_1\dfrac{\partial F_1(w,t,\epsilon)}{\partial w}j\epsilon w + \delta_0 F_2(w,t,\epsilon) + O(\epsilon^2), \\ \epsilon\dfrac{\partial F_2(w,t,\epsilon)}{\partial t} = F_2(w,t,\epsilon)(\lambda_1(t)j\epsilon w - \delta_0) + j\mu_2\dfrac{\partial F_2(w,t,\epsilon)}{\partial w}j\epsilon w + \delta_1 F_1(w,t,\epsilon) \end{cases} \tag{7}$$

and

$$\frac{\partial F_1(w,t,\epsilon)}{\partial t} + \frac{\partial F_2(w,t,\epsilon)}{\partial t} = (F_1(w,t,\epsilon)+F_2(w,t,\epsilon))\lambda_1(t)\left(jw + \epsilon\frac{(jw)^2}{2}\right) - \\ - \left(j\mu_1\frac{\partial F_1(w,t,\epsilon)}{\partial w} + j\mu_2\frac{\partial F_2(w,t,\epsilon)}{\partial w}\right)\left(-jw + \epsilon\frac{(jw)^2}{2}\right) + O(\epsilon^2). \tag{8}$$

We denote $F_k(w,t) = \lim\limits_{\epsilon\to 0} F_k(w,t,\epsilon)$. We suppose that $\lim\limits_{\epsilon\to 0}\epsilon\dfrac{\partial F_k(w,t,\epsilon)}{\partial t} = 0$. Then we have the following equations under $\epsilon \to 0$

$$\begin{cases} -\delta_1 F_1(w,t) + \delta_0 F_2(w,t) = 0, \\ -\delta_0 F_2(w,t) + \delta_1 F_1(w,t) = 0 \end{cases} \tag{9}$$

and

$$\frac{\partial F_1(w,t)}{\partial t} + \frac{\partial F_2(w,t)}{\partial t} = jw\lambda_1(t)(F_1(w,t)+F_2(w,t)) + \\ + jw\left(j\mu_1\frac{\partial F_1(w,t)}{\partial w} + j\mu_2\frac{\partial F_2(w,t)}{\partial w}\right). \tag{10}$$

From equalities (9), we conclude that the solution have the form

$$F_k(w,t) = R_k \Phi(w,t), \tag{11}$$

where R_k is defined as

$$\begin{cases} R_1 = \dfrac{\delta_0}{\delta_0+\delta_1} = \dfrac{\gamma_0}{\gamma_0+\gamma_1}, \\ R_2 = \dfrac{\delta_1}{\delta_0+\delta_1} = \dfrac{\gamma_1}{\gamma_0+\gamma_1}. \end{cases} \tag{12}$$

By substituting expression (11) into equation (10), we obtain

$$\frac{\partial \Phi(w,t)}{\partial t} = jw\lambda_1(t)\Phi(w,t) + jw(j\mu_1 R_1 + j\mu_2 R_2)\frac{\partial \Phi(w,t)}{\partial w}. \tag{13}$$

Under the steady state (if it exists), the solution of equation (13) has the form

$$\Phi(w) = e^{jwa}, \quad a = \frac{\alpha_1}{\mu_1 R_1 + \mu_2 R_2},$$

where $\alpha_1 = \lim\limits_{t\to\infty} \lambda_1(t)$.

Let us find the solution of equation (13) in the form

$$\Phi(w,t) = e^{jwax(t)}. \tag{14}$$

From equation (13), we can write

$$ax'(t) = \lambda_1(t) - ax(t)(\mu_1 R_1 + \mu_2 R_2). \tag{15}$$

Equation (15) is a non-homogeneous differential equation. Its solution can be written as

$$x(t) = e^{-(\mu_1 R_1+\mu_2 R_2)t}\left(x(0) + \frac{1}{a}\int_0^t \lambda_1(\tau)e^{(\mu_1 R_1+\mu_2 R_2)\tau}d\tau\right), \tag{16}$$

where $x(0) = x_0$ is an initial condition.

By substituting (16) to (14) and taking into account the normalization condition, we have derived the solution

$$\Phi(w,t) = \exp\left\{jw \cdot e^{-(\mu_1 R_1+\mu_2 R_2)t}\left(ax(0) + \int_0^t \lambda_1(\tau)e^{(\mu_1 R_1+\mu_2 R_2)\tau}d\tau\right)\right\}.$$

Returning to the substitutions (4), the first-order asymptotic characteristic function of the number of customers in the system is written as follows

$$H(u,t) = \exp\{ju \cdot m(t)\},$$

that determines the transient asymptotic mean of the considered process as

$$m(t) = \frac{\alpha \cdot x(0) \cdot e^{-(\mu_1 R_1+\mu_2 R_2)t}}{\mu_1 R_1 + \mu_2 R_2} + e^{-(\mu_1 R_1+\mu_2 R_2)t}\int_0^t \lambda(\tau)e^{(\mu_1 R_1+\mu_2 R_2)\tau}d\tau \tag{17}$$

where $\alpha = \lim\limits_{t\to\infty} \lambda(t)$.

Note that an additional condition for the method applying is the existence of limit $\alpha = \lim\limits_{t\to\infty} \lambda(t)$.

2.2. Second-order asymptotics

Let us suppose that

$$H_k(u,t) = H_k^{(2)}(u,t)e^{juNax(t)}.$$

By substituting into equations (2), we obtain

$$\begin{cases} \dfrac{\partial H_1^{(2)}(u,t)}{\partial t} + H_1^{(2)}(u,t)juNax'(t) = H_1^{(2)}(u,t)\left(\lambda(t)(e^{ju}-1)-\gamma_1\right) - \\ -j\mu_1\dfrac{\partial H_1^{(2)}(u,t)}{\partial u}(e^{-ju}-1) - j\mu_1 H_1^{(2)}(u,t)jNax(t)(e^{-ju}-1) + \gamma_0 H_2^{(2)}(u,t), \\ \dfrac{\partial H_2^{(2)}(u,t)}{\partial t} + H_2^{(2)}(u,t)juNax'(t) = H_2^{(2)}(u,t)(\lambda(t)(e^{ju}-1)-\gamma_0) - \\ -j\mu_2\dfrac{\partial H_2^{(2)}(u,t)}{\partial u}(e^{-ju}-1) - j\mu_2 H_2^{(2)}(u,t)jNax(t)(e^{-ju}-1) + \gamma_1 H_1^{(2)}(u,t). \end{cases} \tag{18}$$

By summing up equations (18), we can derive an additional equation

$$\begin{aligned} &\frac{\partial H_1^{(2)}(u,t)}{\partial t} + \frac{\partial H_2^{(2)}(u,t)}{\partial t} + juNax'(t)(H_1^{(2)}(u,t)+H_2^{(2)}(u,t)) = \\ &= H_1^{(2)}(u,t)(N\lambda_1(t)(e^{ju}-1)-\gamma_1) - j\mu_1\frac{\partial H_1^{(2)}(u,t)}{\partial u}(e^{-ju}-1) + \\ &\quad +\mu_1 H_1^{(2)}(u,t)Nax(t)(e^{-ju}-1) + \gamma_0 H_2^{(2)}(u,t) + \\ &+H_2^{(2)}(u,t)(N\lambda_1(t)(e^{ju}-1)-\gamma_0) - j\mu_2\frac{\partial H_2^{(2)}(u,t)}{\partial u}(e^{-ju}-1) + \\ &\quad +\mu_2 H_2^{(2)}(u,t)Nax(t)(e^{-ju}-1) + \gamma_1 H_1^{(2)}(u,t). \end{aligned} \tag{19}$$

For the second-order asymptotics, we will use the following asymptotic substitutions

$$N = \frac{1}{\epsilon^2},\ \ N\lambda_1(t) = \frac{1}{\epsilon^2}\lambda_1(t),\ \ \gamma_k = N\delta_k = \frac{1}{\epsilon^2}\delta_k,\ \ u = \epsilon w,\ H_k^{(2)}(u,t) = F_k^{(2)}(w,t,\epsilon), \tag{20}$$

where $\epsilon \to 0$ is an infinitesimal parameter.

From system (18)–(19), the following asymptotic equations can be written after some transformations

$$\begin{cases} \epsilon\dfrac{\partial F_1^{(2)}(w,\epsilon,t)}{\partial t} + F_1^{(2)}(w,\epsilon,t)jwax'(t) = F_1^{(2)}(w,\epsilon,t)\left(\lambda_1(t)(jw+\epsilon\dfrac{(jw)^2}{2}) - \dfrac{1}{\epsilon}\delta_1\right) - \\ -\epsilon j\mu_1\dfrac{\partial F_1^{(2)}(w,\epsilon,t)}{\partial w}\left(-jw+\epsilon\dfrac{(jw)^2}{2}\right) - j\mu_1 F_1^{(2)}(w,\epsilon,t)jax(t)\left(-jw+\epsilon\dfrac{(jw)^2}{2}\right) + \\ +\dfrac{1}{\epsilon}\delta_0 F_2^{(2)}(w,\epsilon,t) + O(\epsilon^2), \\ \epsilon\dfrac{\partial F_2^{(2)}(w,\epsilon,t)}{\partial t} + F_2^{(2)}(w,\epsilon,t)jwax'(t) = F_2^{(2)}(w,\epsilon,t)\left(\lambda_1(t)(jw+\epsilon\dfrac{(jw)^2}{2}) - \dfrac{1}{\epsilon}\delta_0\right) - \\ -\epsilon j\mu_2\dfrac{\partial F_2^{(2)}(w,\epsilon,t)}{\partial w}\left(-jw+\epsilon\dfrac{(jw)^2}{2}\right) - j\mu_2 F_2^{(2)}(w,\epsilon,t)jax(t)\left(-jw+\epsilon\dfrac{(jw)^2}{2}\right) + \\ +\dfrac{1}{\epsilon}\delta_1 F_1^{(2)}(w,\epsilon,t) + O(\epsilon^2). \end{cases} \tag{21}$$

From equation (19), we have

$$\frac{\partial F_1^{(2)}(w,\epsilon,t)}{\partial t}+\frac{\partial F_2^{(2)}(w,\epsilon,t)}{\partial t}=\frac{1}{\epsilon}jw(F_1^{(2)}(w,\epsilon,t)+F_2^{(2)}(w,\epsilon,t))\lambda_1(t)\left(1+\epsilon\frac{jw}{2}-ax'(t)\right)-$$
$$-\left(j\mu_1\frac{\partial F_1^{(2)}(w,\epsilon,t)}{\partial w}+j\mu_1F_1^{(2)}(w,\epsilon,t)j\frac{1}{\epsilon}ax(t)+j\mu_2\frac{\partial F_2^{(2)}(w,\epsilon,t)}{\partial w}+j\mu_2F_2^{(2)}(w,\epsilon,t)j\frac{1}{\epsilon}ax(t)\right)\times$$
$$\times\left(-jw+\epsilon\frac{(jw)^2}{2}\right)+O(\epsilon^2). \qquad (22)$$

Let us find the solution in the following form

$$F_k^{(2)}(w,t,\epsilon)=\Phi^{(2)}(w,t)(R_k+jw\epsilon\phi_k(t))+O(\epsilon^2) \qquad (23)$$

By substituting expression (23) into equations (21) and making some transformations, we obtain that

$$\begin{cases} R_1(-ax'(t)+\lambda_1(t)-\mu_1ax(t))-\delta_1\phi_1(t)+\delta_0\phi_2(t)=0,\\ R_2(-ax'(t)+\lambda_1(t)-\mu_2ax(t))+\delta_1\phi_1(t)-\delta_0\phi_2(t)=0. \end{cases} \qquad (24)$$

Making some transformations, we get:

$$\delta_1\phi_1(t)-\delta_0\phi_2(t)=ax(t)R_1((\mu_1R_1+\mu_2R_2)-\mu_1).$$

Taking into account expression (15), equations (24) are linearly dependent. So we introduce an additional expression to uniquely define the function $\phi_k(t)$:

$$\phi_1(t)+\phi_2(t)=0. \qquad (25)$$

Then $\phi_k(t)$ are expressed as

$$\begin{cases} \phi_2(t)=-\phi_1(t),\\ \phi_1(t)=ax(t)\dfrac{\delta_0}{(\delta_0+\delta_1)^2}((\mu_1R_1+\mu_2R_2)-\mu_1). \end{cases} \qquad (26)$$

For function $\Phi^{(2)}(w,t)$ derivation, let us consider equation (22). By substituting expression (15) and perform some transformations, we obtain the following equation under $\epsilon\to 0$:

$$\frac{\partial\Phi^{(2)}(w,t)}{\partial t}=jw\frac{\partial\Phi^{(2)}(w,t)}{\partial w}(j\mu_1R_1+j\mu_2R_2)+\frac{(jw)^2}{2}\Phi^{(2)}(w,t)(\lambda_1(t)+2(\phi_1(t)+$$
$$+\phi_2(t))(\lambda_1(t)-ax'(t))-2ax(t)(\mu_1\phi_1(t)+\mu_2\phi_2(t))+ax(t)(\mu_1R_1+\mu_2R_2)), \qquad (27)$$

It is obvious that the solution has the form $\Phi^{(2)}(w,t)=\exp\left\{\dfrac{(jw)^2}{2}y(t)\right\}$ where $y(t)$ is defined by the following equation:

$$y'(t)=-2y(t)(\mu_1R_1+\mu_2R_2)+b(t). \qquad (28)$$

Denote

$$b(t)=\lambda_1(t)+ax(t)(\mu_1R_1+\mu_2R_2)+2(\phi_1(t)+\phi_2(t))(\lambda_1(t)-ax'(t))-2ax(t)(\mu_1\phi_1(t)+\mu_2\phi_2(t)).$$

Taking into account expressions (12), (15) and (25), we simplify $b(t)$ as follows

$$b(t) = \frac{1}{N}\lambda(t) + \frac{1}{N}\alpha x(t) - 2\frac{1}{N}\left(\frac{\alpha}{(\mu_1 R_1 + \mu_2 R_2)}x(t)\right)^2 \frac{\gamma_0(\mu_1 - \mu_2)}{(\gamma_0 + \gamma_1)^2}((\mu_1 R_1 + \mu_2 R_2) - \mu_1).$$

Thus, the solution of equation (28) has the form

$$y(t) = e^{-2(\mu_1 R_1 + \mu_2 R_2)t}\left(y(0) + \int_0^t e^{2(\mu_1 R_1 + \mu_2 R_2)\tau} b(\tau)d\tau\right). \tag{29}$$

Returning to substitutions (20), the asymptotic characteristic function has Gaussian form

$$H(u,t) = \exp\left\{juNax(t) + \frac{(ju)^2}{2}Ny(t)\right\},$$

with the time-dependent asymptotic mean $m = Nax(t)$ and variance $D = Ny(t)$ defined by expressions (18) and (29), respectively.

3. Numerical analysis

To illustrate the results application, let us consider the following numerical example. We suppose that an object is supplied by two sources of electricity — a municipal grid and an individual solar installation. The average operating time of solar a solar power system is 8 hours per day in summer. The model parameters of service units switching have the following values:

$$\gamma_1 = 8/24,\ \gamma_0 = 16/24.$$

We consider a non-stationary arrival Poisson Process of electricity requests with the following intensity:

$$\lambda(t) = \alpha + (t-2)^2 \exp\left\{-\frac{t}{\beta}\right\},$$

where $\alpha = 10$, $\beta = 6$. The form of the function $\lambda(t)$ is illustrated in Fig. 2. This form of the arrival Process describes a situation of a sharp increase in the load on the power grid at some moment ($t = 18$ hours in the example).

Let the service time parameters and an initial value (typical load) have the following values:

$$\mu_1 = 5,\ \mu_2 = 4,\ \ x_0 = 2.$$

Applying the modelling results, we can demonstrate the dependence of the main characteristics on time. Fig. 3 shows the calculated transient asymptotic mean $m(t)$ and variance $s(t)$ of the number of customers in the system. From Fig. 3, it can be seen that the system returns to the steady state at $t \geqslant 40$.

Let us calculate the transient asymptotic distributions for the given parameter values. Figs. 4–5 show the asymptotic probability distributions the number of customers in the system at various time points (Fig. 4 — before maximum load, Fig. 5 — after maximum load).

In this way, we can observe the dynamic behaviour of the system and predict the necessary performance characteristic.

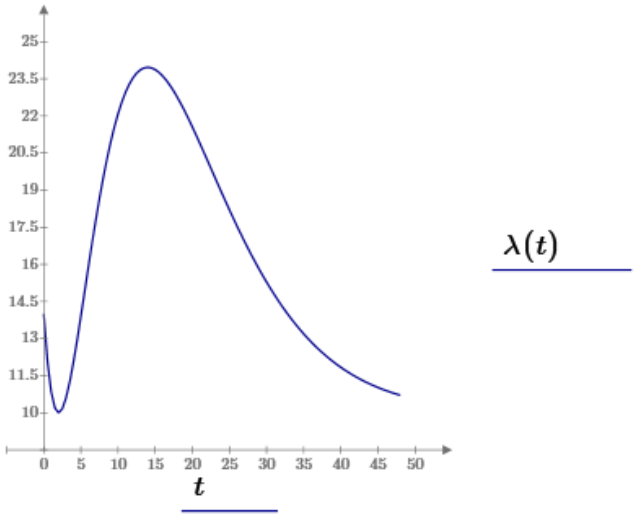


Fig. 2. Example of arrival intensity function

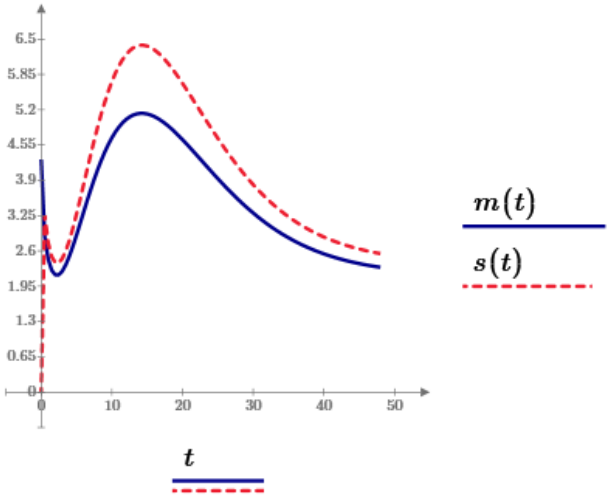


Fig. 3. Transient asymptotic mean $m(t)$ and variance $s(t)$

Conclusion

This paper analyzes a mathematical model of a hybrid energy system represented as the queueing system with switching between two service units, the unlimited number of servers and non-stationary Poisson arrivals. To find the probability distribution of the number of customers in the system, an asymptotic analysis method was used under the condition of a high arrival rate.

In the future, it is planned to use the obtained analytical results to calculate the operational characteristics of hybrid energy systems and assess economic efficiency, taking into account various configurations and geographical features based on technical and climatic data. Additionally, more complex models can be proposed, such as resource queueing models.

The study is supported by Russian Science Foundation according to the research project no. 24-71-00022, https://rscf.ru/project/24-71-00022/.

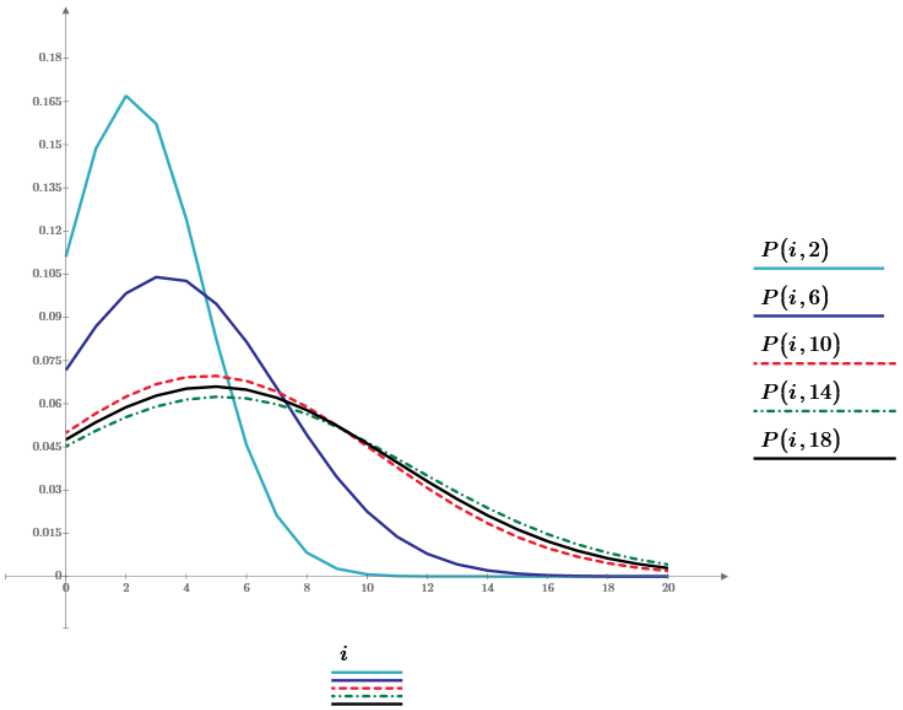


Fig. 4. Asymptotic probability distribution the number of customers in the system for $t \leqslant 18$

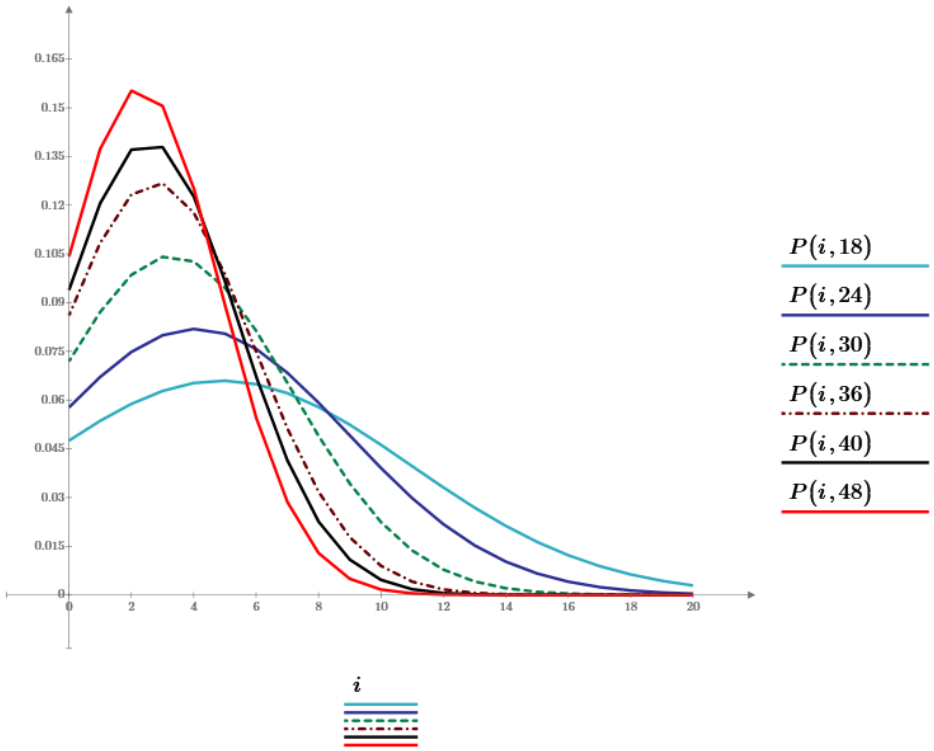


Fig. 5. Asymptotic probability distribution the number of customers in the system for $t \geqslant 18$

References

[1] K.S.Krishna, K.S.Kumar, A review on hybrid renewable energy systems, *Renewable and Sustainable Energy Reviews*, **52**(2015), 907-916.

[2] A.Herbst, F.Toro, F.Reitze, E.Jochem, Introduction to Energy Systems Modelling, *Swiss Journal of Economics and Statistics*, **148**(2012), no. 2, 111–135. DOI: 10.1007/BF03399363

[3] M.King, Li.Dacheng, M.Dooner, G.Saikat, J.N.Roy, C.Chakraborty, W.Jihong, Mathematical Modelling of a System for Solar PV Efficiency Improvement Using Compressed Air for Panel Cleaning and Cooling, *Energies*, (2021), no. 14. DOI: 10.3390/en14144072

[4] G.Perveen, M.Rizwan, N.Goel, Comparison of intelligent modelling techniques for forecasting solar energy and its application in solar PV based energy system, *IET Energy Systems Integration*, (2019), no. 11, 34–51. DOI: 10.1049/iet-esi.2018.0011

[5] A.Kumar, Y.Sood, N.M.Ram, Hybrid Energy System Simulation and Modelling Incorporating Wind and Solar Power, 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2024.

[6] S.Tyagi, N.Goyal, A.Kumar, M.Ram, Stochastic hybrid energy system modelling with component failure and repair, *International Journal of System Assurance Engineering and Management*, **13**(2021), 842–852. DOI: 10.1007/s13198-021-01129-4

[7] S.Varshney, C.Shekhar, Optimal management strategies of renewable energy systems with hyperexponential service provisioning: an economic investigation, *Frontiers. Energy Research*, (2023), no. 11, 1–22. DOI: 10.3389/fenrg.2023.1329899

[8] I.Ferretti, M.Camparada, L.E.Zavanella, Queuing Theory-Based Design Methods for the Definition of Power Requirements in Manufacturing Systems, *Energies*, (2022), no. 15, 7621. DOI: 10.3390/en15207621

[9] M.Watermeyer, M.Thomas, O.Grothe, F.Musgens, A hybrid model for day-ahead electricity price forecasting: Combining fundamental and stochastic modelling, Engineering, Economics, 2023, 1–38.

[10] L.Zavanella, S.Zanoni, I.Ferretti, L.Mazzoldi, Energy Demand in Production Systems: A Queuing Theory Perspective, *International Journal of Production Economics*, **170**(2015), 393–400. DOI: 10.1016/j.ijpe.2015.06.019

[11] S.Varshney, K.P.Panda, M.Gupta, Shah M., et al., Stochastic modeling of electric vehicle infrastructure using queueing-theoretical approach, *Results in Engineering*, **2025**(25), Art. no. 104149. DOI: 10.1016/j.rineng.2025.104149

[12] G.Kou , S.Yüksel, H.Dincer, M.Acar, S.Eti, Queuing Theory-Driven Renewable Energy Investment Decisions through Artificial Expert Choices with Dynamic Hybrid Multi-Facet Fuzzy Modelling, *Science*, (2025), no. 28, 113711. DOI: 10.1016/j.isci.2025.113711

[13] V.A.Vavilov, A.A.Nazarov, Asymptotic Analysis of the $M(T)|M(T)|\infty$ System, Information Technologies and Mathematical Modelling (ITMM-2024): Proceedings of the XXIII International Conference named after A.F. Terpugov, October 20–26, 2024. Tomsk: Publishing House of Tomsk State University, 2024, 184–189 (In Russian).

[14] S.P.Moiseeva, I.A.Turenova, A.A.Imomov, Asymptotic Analysis of Multi-Arrival Heterogenous Resource Queueing System $MMPP/GI(2)/\infty$ in a Random Markovian Environment, *J. Sib. Fed. Univ. Math. Phys.*, **18**(2025), no. 6, 770–781. EDN: HGXLZE

Асимптотический анализ системы массового обслуживания с нестационарным пуассоновским потоком и переключением в переходном режиме

Яна А. Тюленина
Екатерина А. Фёдорова
Национальный исследовательский Томский государственный университет
Томск, Российская Федерация

Аннотация. В статье представлено исследование системы массового обслуживания с двумя блоками приборов и переключением с неограниченным числом приборов и нестационарным пуассоновским потоком как математической модели гибридной энергосистемы, использующей два источника электроэнергии. Предложен оригинальный метод асимптотического анализа переходных процессов для получения нестационарного распределения вероятностей состояний системы в предельном условии высокой интенсивности входящего потока. Доказано, что асимптотическое распределение числа заявок в системе имеет форму нормального распределения, получены выражения для асимптотического математического ожидания и дисперсии. Представлены результаты численного анализа.

Ключевые слова: математическое моделирование, теория массового обслуживания, асимптотический анализ в переходном режиме, нестационарный пуассоновский входящий поток.

EDN: SLDKXC
УДК 532.517.4

Approximate Solution for the Self-similar Decay of the Far Swirling Turbulent Wake

Alexey V. Shmidt*
Institute of Computational Modelling SB RAS
Krasnoyarsk, Russia

Received 08.01.2026, received in revised form 11.02.2026, accepted 17.03.2026

Abstract. Approximate solution of a boundary-value problem for a model of the far swirling turbulent wake past a self-propelled body is constructed using asymptotic expansion of the solution in a neighbourhood of the singular point. A good agreement between constructed solution and numerical solution is obtained.

Keywords: far swirling turbulent wake, boundary-value problem, semi-empirical model of turbulence, approximate solution, asymptotic expansion.

Citation: A.V. Shmidt, Approximate Solution for the Self-similar Decay of the far Swirling Turbulent Wake, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 381–390. EDN: SLDKXC.

Introduction

Swirling turbulent wakes can be formed in various types of technological devices and in the flows past bodies. Such flows were investigated in a number of publications (see, e.g., [1–18] and references therein). The results of detailed laboratory measurements in swirling turbulent wakes with moderate excess momentum and angular momentum were presented [6, 9].

The particular case is a wake past the self-propelled body with an engine mounted in its trailing edge. The engine thrust compensates for the hydrodynamic drag force that results in momentumless wake. The engine may twist the flow. Swirl must be balanced by rotation of the body part in the opposite direction to ensure that the body does not rotate around its axis. In this case, the angular momentum in the wake with respect to its axis is equal to zero.

An approximate solution to a model of the far momentumless turbulent wake without swirl was obtained by matching asymptotic expansions at the axis of wake [19–21]. The values of similarity exponents were determined during the matching procedure. In the present work, previously proposed technique to construct approximate solution of a semi-empirical model of the far swirling turbulent wake past the self-propelled body with zero momentum and angular momentum is used.

1. Problem statement

A self-propelled body of revolution moves uniformly and rectilinearly in unbounded homogeneous incompressible fluid so that the stream-wise components of excess momentum and angular

*schmidt@icm.krasn.ru

momentum in a wake past the body are equal to zero (see, Fig. 1). A trailing part of the body rotates to compensate for the swirl in wake. U_0 is the undisturbed flow velocity, U is the longitudinal velocity component of averaged motion. Coordinate axis Ox is directed downstream and coincides with the axis of symmetry of the body. The coordinate origin is placed at its trailing edge. The radial coordinate is denoted by r. Moreover, it is assumed that flow is steady in the far fields of wake.

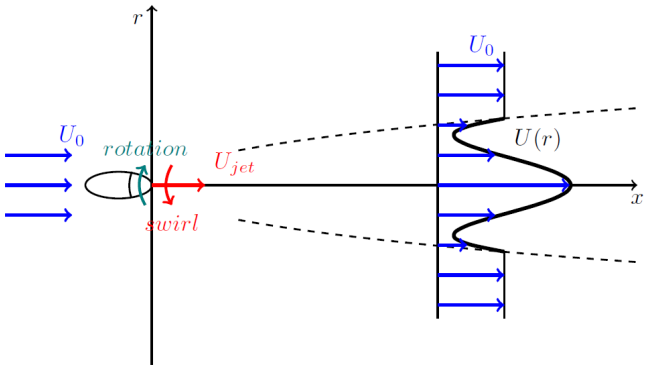


Fig. 1. Schematic diagram for the flow past a self-propelled body

According to the results of detailed numerical simulation [14], the flow in a swirling turbulent wake becomes almost isotropic and shear-free at large distances from the self-propelled body. Therefore, the influence of averaged longitudinal and tangential velocity components on the flow pattern in the far wake regions can be neglected.

2. Similarity reduction

The following semi-empirical model of turbulence is used to describe flow in the far swirling turbulent wake past a self-propelled body [14]:

$$U_0\frac{\partial U_1}{\partial x}=\frac{1}{r}\frac{\partial}{\partial r}\left(C_\mu r\frac{e^2}{\varepsilon}\frac{\partial U_1}{\partial r}\right), \tag{1}$$

$$U_0\frac{\partial W}{\partial x}=\frac{1}{r^2}\frac{\partial}{\partial r}\left(C_\mu r^3\frac{e^2}{\varepsilon}\frac{\partial W/r}{\partial r}\right), \tag{2}$$

$$U_0\frac{\partial e}{\partial x}=\frac{1}{r}\frac{\partial}{\partial r}\left(C_e r\frac{e^2}{\varepsilon}\frac{\partial e}{\partial r}\right)-\varepsilon, \tag{3}$$

$$U_0\frac{\partial \varepsilon}{\partial x}=\frac{1}{r}\frac{\partial}{\partial r}\left(C_\varepsilon r\frac{e^2}{\varepsilon}\frac{\partial \varepsilon}{\partial r}\right)-C_{\varepsilon 2}\frac{\varepsilon^2}{e}. \tag{4}$$

Here $U_1=U-U_0$ is the defect of the mean longitudinal velocity component, W is the mean tangential velocity component, e is the kinetic energy of turbulence, and ε is the kinetic energy dissipation rate. In what follows the undisturbed flow velocity U_0 is taken to be unity. The empirical constants in (1)-(4) are as follows

$$C_\mu=0.25,\ C_e=0.147,\ C_\varepsilon=0.113,\ C_{\varepsilon 2}=1.92.$$

Model (1)–(4) was constructed [14] as a simplification of more complicated mathematical model [9, 12]. It includes the averaged equations of motion, continuity, transfer of normal

Reynolds stresses, and the kinetic energy dissipation rate in the approximation of a thin shear layer.

Integral relations

$$J = \int_0^\infty rU_1 dr = 0, \tag{5}$$

$$M = \int_0^\infty r^2 W dr = 0, \tag{6}$$

follow from equations (1), (2). They are valid under the following conditions

$$\frac{e^2}{\varepsilon}\frac{\partial U_1}{\partial r} \to 0, \quad \frac{e^2}{\varepsilon}\frac{\partial (W/r)}{\partial r} \to 0 \tag{7}$$

at $r \to \infty$.

Since the flow in the far swirling turbulent wake past a self-propelled body is close to the self-similar one [14] group-theoretical analysis [22] to construct similarity reduction of model (1)–(4) is used. The Lie algebra basis of (1)–(4) consists of the following operators

$$X_1 = \frac{\partial}{\partial x}, \quad X_2 = \frac{\partial}{\partial U_1} \quad X_3 = U_1\frac{\partial}{\partial U_1}, \quad X_4 = W\frac{\partial}{\partial W}, \quad X_5 = r\frac{\partial}{\partial W}, \quad X_6 = x\frac{\partial}{\partial x} - 2e\frac{\partial}{\partial e} - 3\varepsilon\frac{\partial}{\partial \varepsilon},$$

$$X_7 = r\frac{\partial}{\partial r} + 2e\frac{\partial}{\partial e} + 2\varepsilon\frac{\partial}{\partial \varepsilon}.$$

Using the linear combination of operators X_3, X_4, X_6 and X_7, the following representation for the solution of model (1)–(4) is obtained

$$U_1 = x^\beta U_2(t), \; W = x^\gamma W_1(t), \quad e = x^{2\alpha-2}K(t), \quad \varepsilon = x^{2\alpha-3}E(t), \quad t = r/x^\alpha, \tag{8}$$

where t is the self-similar variable, α, β, and γ are arbitrary constants appearing in the linear combination of operators X_3, X_4, X_6, and X_7.

Using (8), one can reduce model (1)–(4) to the system of ordinary differential equations

$$C_\mu \frac{K^2 U_2''}{E} + \left(C_\mu \frac{K}{E}\left(2K' - \frac{KE'}{E} + \frac{K}{t}\right) + \alpha t\right) U_2' - \beta U_2 = 0, \tag{9}$$

$$C_\mu \frac{K^2 W_1''}{E} + \left(C_\mu \frac{K}{E}\left(2K' - \frac{KE'}{E} + \frac{K}{t}\right) + \alpha t\right) W_1' - \left(\frac{C_\mu}{t^2}\left(\frac{tK^2}{E}\right)' + \gamma\right) W_1 = 0, \tag{10}$$

$$C_e \frac{K^2 K''}{E} + 2C_e \frac{KK'^2}{E} - \left(C_e \frac{K^2}{E}\left(\frac{E'}{E} - \frac{1}{t}\right) + \alpha t\right) K' - 2(\alpha - 1)K - E = 0, \tag{11}$$

$$C_\varepsilon \frac{K^2 E''}{E} - C_\varepsilon \frac{K^2 E'^2}{E^2} + \left(C_\varepsilon \frac{K}{E}\left(2K' + \frac{K}{t}\right) + \alpha t\right) E' - C_{\varepsilon_2}\frac{E^2}{K} - (2\alpha - 3)E = 0. \tag{12}$$

Solution of reduced system (9)-(11) has to satisfy the boundary conditions

$$U_2'(0) = W_1(0) = K'(0) = E'(0) = 0, \tag{13}$$

$$U_2(a) = W_1(a) = K(a) = E(a) = 0. \tag{14}$$

Boundary conditions (13) take into account the flow symmetry with respect to the Ox axis. Boundary conditions (14) follow from the requirement that flow is undisturbed outside the turbulent wake. The value of a in (14) is related to the turbulent wake semi-width which is found from the condition

$$e(a, x) = 0.5e(0, x). \tag{15}$$

Coefficients in equations (9)–(12) have singularities at the boundary point. It complicates the construction of a solution of boundary-value problem (9)–(14). Self-similar solutions to the $k-\varepsilon$ turbulence model for plane and radial jets were constructed [23], using a transformation that stretches the similarity variable and simplifies the reduced system of ordinary differential equations. The first terms of asymptotic expansion of a solution at the flow boundary and modified shooting method were used to obtain a self-similar solution to the $k-\varepsilon$ model for the far turbulent wake past a towed body [24].

It should be noted that self-similar solution (8) is of the second kind [25]. To obtain values of the similarity exponents, the corresponding nonlinear eigenvalue problem for a model of the far axisymmetric momentumless turbulent wake without swirl (equations (1), (3), and (4)) was solved numerically [26]. The value $\alpha = 0.27$ was obtained. This nonlinear eigenvalue problem was also solved analytically [27]. The similarity exponent α is expressed in the form

$$\alpha = \frac{2C_{\varepsilon_2} - 3}{4(C_{\varepsilon_2} - 1)}. \tag{16}$$

After substitution of $C_{\varepsilon_2} = 1.92$ in (16) one can obtain $\alpha = 21/92 \approx 0.2283$.

3. Approximate solution

To construct approximate solution of boundary-value problem (9)–(14) previously propose approach is used [19–21]. It is based on matching of asymptotic expansions. Values of the similarity exponents α, β, and γ can be found during the matching procedure.

There is the following asymptotic expansion of a solution of system (9)–(12) in a neighbourhood of the singular point $t = a$

$$U_2^a = \sum_{i=1}^{16} u_i \left|t-a\right|^{i/2} + o\left(\left|t-a\right|^8\right), \tag{17}$$

$$W_1^a = \sum_{i=1}^{16} w_i \left|t-a\right|^{i/2} + o\left(\left|t-a\right|^8\right), \tag{18}$$

$$K^a = \sum_{i=1}^{16} k_i \left|t-a\right|^{i/2} + o\left(\left|t-a\right|^8\right), \tag{19}$$

$$E^a = \sum_{i=1}^{16} e_i \left|t-a\right|^{i/2} + o\left(\left|t-a\right|^8\right), \tag{20}$$

where u_1, w_1, k_1, and e_1 are arbitrary constants. The remaining coefficients depend on a, α, β, γ, u_1, w_1, k_1 and e_1. They are not presented here because they have quite cumbersome representations.

It is not difficult to check that there exist power series expansions for U_2, W_1, K and E near $t = 0$ subject to boundary conditions (13). They are

$$U_2^0 = U_{20} + \sum_{i=1}^{5} \alpha_i t^{2i} + o(t^{10}), \tag{21}$$

$$W_1^0 = \sum_{i=1}^{6} \beta_i t^{2i-1} + o(t^{11}), \tag{22}$$

$$K^0 = K_0 + \sum_{i=1}^{5} \gamma_i t^{2i} + o(t^{10}), \tag{23}$$

$$E^0 = E_0 + \sum_{i=1}^{5} \kappa_i t^{2i} + o(t^{10}), \tag{24}$$

where U_{20}, β_1, K_0 and E_0 are arbitrary constants. Coefficients α_i, γ_i, κ_i $(i = 1, \ldots, 5)$ and β_i $(i = 2, \ldots, 6)$ in (21)–(24) are some expressions of α, β, γ, U_{20}, β_1, K_0, and E_0. These expressions are not presented here because they are quite cumbersome.

Since equations (9), (10) admit the scaling transformations on U_2 and W_1 respectively, the values of U_{20} and β_1 can be chosen arbitrarily. It is assumed that $U_{20} = -1$ and $\beta_1 = -1$. The approximate solution is constructed on the basis of expansions (17)-(20). It is assumed that coefficients u_i $(i = 8, \ldots, 16)$, w_i $(i = 8, \ldots, 16)$, k_i $(i = 8, \ldots, 16)$ and e_i $(i = 8, \ldots, 16)$ are arbitrary, and their values are determined during the matching procedure which is performed at the point $t = 0$. Equating derivatives of like order of the corresponding truncated expansions (17)–(20) and (21)–(24)

$$\left.\frac{d^i}{dt^i} U_2^0\right|_{t=0} = \left.\frac{d^i}{dt^i} U_2^a\right|_{t=0}, \ i = 0..10, \tag{25}$$

$$\left.\frac{d^i}{dt^i} W_1^0\right|_{t=0} = \left.\frac{d^i}{dt^i} W_1^a\right|_{t=0}, \ i = 0..10, \tag{26}$$

$$\left.\frac{d^i}{dt^i} K^0\right|_{t=0} = \left.\frac{d^i}{dt^i} K^a\right|_{t=0}, \ i = 0..11, \tag{27}$$

$$\left.\frac{d^i}{dt^i} E^0\right|_{t=0} = \left.\frac{d^i}{dt^i} E^a\right|_{t=0}, \ i = 0..10, \tag{28}$$

one can obtain a system of algebraic equations with unknown quantities α, β, γ, K_0, E_0, u_i $(i = 1, 8, \ldots, 16)$, w_i $(i = 1, 8, \ldots, 16)$, k_i $(i = 1, 8, \ldots, 16)$, and e_i $(i = 1, 8, \ldots, 16)$. The value of a is determined using (15). The numerical solution of system of algebraic equations (15), (25)–(28) is

$$a = 1.4565,\ \alpha = 0.2312,\ \beta = -2.3068,\ \gamma = -3.2759,\ K_0 = 1.4875,\ E_0 = 1.7113,\ u_1 = 0.5566,$$

$$u_8 = -159.1227,\ u_9 = 391.7203,\ u_{10} = -577.9103,\ u_{11} = 568.2446,\ u_{12} = -384.2439,$$

$$u_{13} = 177.7355,\ u_{14} = -54.0127,\ u_{15} = 9.7570,\ u_{16} = -0.7968,\ w_1 = 0.3778,\ w_8 = -123.4300,$$

$$w_9 = 266.4117,\ w_{10} = -364.0304,\ w_{11} = 335.5019,\ w_{12} = -212.2378,\ w_{13} = 91.2582,$$

$$w_{14} = -25.5212,\ w_{15} = 4.1835,\ w_{16} = -0.3038,\ k_1 = 0.1245,\ k_8 = -14.7676,\ k_9 = -54.2749,$$

$$k_{10} = 152.0286,\ k_{11} = -195.8013,\ k_{12} = 155.2477,\ k_{13} = -80.0900,\ k_{14} = 26.4075,\ k_{15} = -5.0879,$$

$$k_{16} = 0.4379,\ e_1 = 0.0527,\ e_8 = 93.8003,\ e_9 = -285.8002,\ e_{10} = 451.4314,\ e_{11} = -462.4311,$$

$$e_{12} = 324.3951,\ e_{13} = -155.8908,\ e_{14} = 49.3333,\ e_{15} = -9.2983,\ e_{16} = 0.7932.$$

The obtained values K_0 and E_0 are used to solve boundary-value problem (9)–(14) with fixed $\alpha = 21/92$ by the shooting method. As a result of numerical calculations, the following values were found: $K_0 = 1.5110$, $E_0 = 1.7413$, $\beta = -2.369$, and $\gamma = -3.383$. Comparison of the constructed solution in the form of truncated Puiseux series [28, 29] and numerical solution shows that maximum discrepancy between them does not exceed 2% (see Fig. 2).

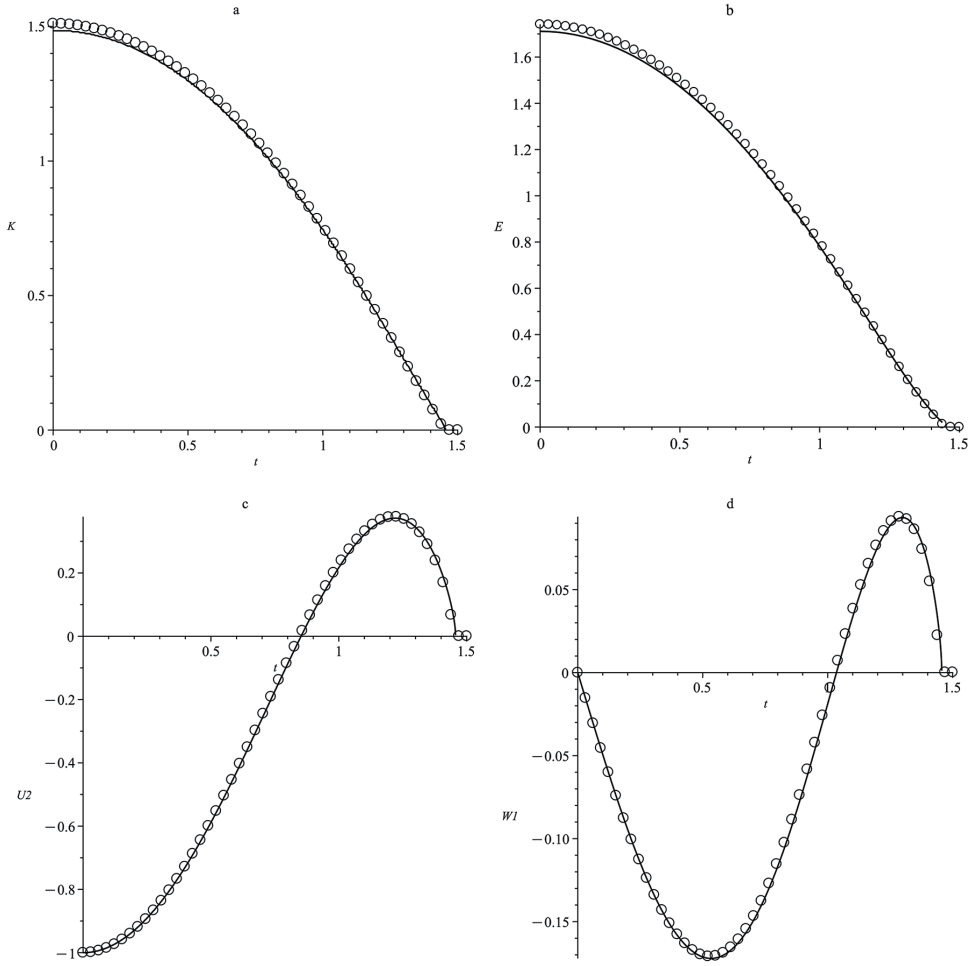


Fig. 2. Profiles of approximate and numerical solutions: a — kinetic energy of turbulence; b — dissipation rate of kinetic energy; c — defect of the longitudinal averaged velocity component; d — tangential averaged velocity component; solid lines — approximate solution, open circles — numerical solution

Moreover, the matching procedure is performed for $\alpha = 21/92$ (the equation for $i = 11$ is omitted in (27)). The following results are obtained

$$a = 1.4403,\ \beta = -2.5153,\ \gamma = -3.4477,\ K_0 = 1.5121,\ E_0 = 1.7421,\ u_1 = 0.6546,$$

$$u_8 = -159.1227,\ u_9 = -248.9156,\ u_{10} = 694.5620,\ u_{11} = -1004.8357,\ u_{12} = 893.1924,$$

$$u_{13} = -508.7194,\ u_{14} = 182.2401,\ u_{15} = -37.5929,\ u_{16} = 3.4189,\ w_1 = 0.4155,\ w_8 = -49.5955,$$

$$w_9 = 77.5824,\ w_{10} = -86.8760,\ w_{11} = 68.4323,\ w_{12} = -36.8537,\ w_{13} = 13.0603,$$

$$w_{14} = -2.7786,\ w_{15} = 0.2818,\ w_{16} = -0.0040,\ k_1 = 0.1813,\ k_8 = -6.1228,\ k_9 = -35.7628,$$

$$k_{10} = 93.3509,\ k_{11} = -118.4767,\ k_{12} = 93.8299,\ k_{13} = -48.6601,\ k_{14} = 16.1910,\ k_{15} = -3.1562,$$

$$k_{16} = 0.2754,\ e_1 = 0.0902,\ e_8 = 48.9870,\ e_9 = -163.6724,\ e_{10} = 266.5966,\ e_{11} = -279.7097,$$

$$e_{12} = 201.1470,\ e_{13} = -99.2699,\ e_{14} = 32.3033,\ e_{15} = -6.2643,\ e_{16} = 0.5498.$$

A good agreement between approximate solution (17)–(20) and numerical solution is obtained in this case (see Fig. 3).

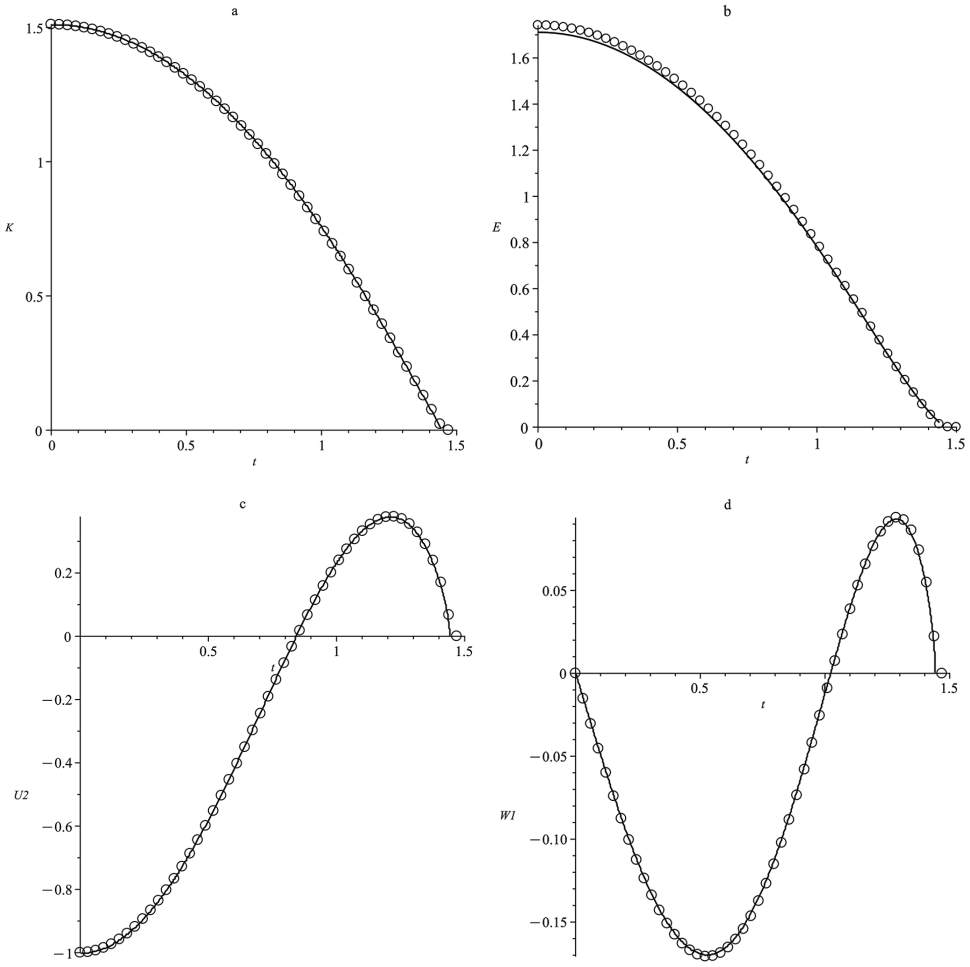


Fig. 3. Profiles of approximate and numerical solutions ($\alpha = 21/92$): a — kinetic energy of turbulence; b — dissipation rate of kinetic energy; c — defect of the longitudinal averaged velocity component; d — tangential averaged velocity component; solid lines — approximate solution, open circles — numerical solution

The established laws of similarity decay in the far swirling turbulent wake past a self-propelled body are in consistence with those presented in [14, 15, 27].

The proposed approach for constructing approximate solutions can be applied to other boundary-value problems unrelated to turbulence.

The author is grateful to Professor O. V. Kaptsov and Professor G. G. Chernykh for their attention to the work.

This work is supported by the Krasnoyarsk Mathematical Center and financed by the Ministry of Science and Higher Education of the Russian Federation in the framework of the establish-

ment and development of regional Centers for Mathematics Research and Education (Agreement no. 075-02-2026-735).

References

[1] A.J.Reynolds, Similarity in swirling wakes and jets, *J. Fluid Mech.* **15**(1962), no. 2, 241–243.

[2] J.A.Schetz, Injection and mixing in turbulent flow, *Progr. in Astronautics and Aeronautics*, **68**(1980).

[3] B.S.Hyun, V.C.Patel, Measurements in the flow around a marine propeller at the stern of an axisymmetric body. Part 1. Circumferentially-averaged flow, *Exp. Fluids*, **11**(1991), 33–44.

[4] B.S.Hyun, V.C.Patel, Measurements in the flow around a marine propeller at the stern of an axisymmetric body. Part 2. Phase-averaged flow, *Exp. Fluids*, **11**(1991), 105–117.

[5] A.G.Gumilevskii, Similarity and decay laws of wakes with zero momentum and angular momentum, *Fluid Dyn.*, **28**(1993), no. 5, 619–623.

[6] V.A.Kostomakha, N.V.Lesnova, Turbulent swirling wake behind a sphere with complete or partial drag compensation, *J. Appl. Mech. Tech. Phys.*, **36**(1995), no. 2, 226–233. DOI: 10.1007/BF02369653

[7] G.G.Chernykh, A.G.Demenkov, V.A.Kostomakha, Numerical model of a swirling momentumless turbulent wake, *Russ. J. Numer. Anal. Math. Model.*, **13**(1998), no. 4, 279–288. DOI: 10.1515/rnam.1998.13.4.279

[8] J.Piquet, Turbulent flows. Models and physics, Berlin, Heidelberg, Springer-Verlag, 1999.

[9] N.V.Gavrilov, A.G.Demenkov, V.A.Kostomakha, G.G.Chernykh, Experimental and numerical modeling of the turbulent wake of a self-propelled body, *J. Appl. Mech. Tech. Phys.*, **41**(2000), no. 4, 619–627. DOI: 10.1007/BF02466859

[10] A.I.Sirviente, V.Patel, Wake of a self-propelled body. Part 2. Momentumless wake with swirl, *AIAA J.*, **38**(2000), no. 4, 620–627. DOI: 10.2514/2.1033

[11] O.F.Vasil'ev, A.G.Demenkov, G.G.Chernykh, V.A.Kostomakha, Numerical simulation of a twisted turbulent wake beyond a self-propelled body, *Doklady Physics*, **46**(2001), no. 1. 52–55. DOI:10.1134/1.1348589

[12] G.G.Chernykh, A.G.Demenkov, V.A.Kostomakha, Swirling turbulent wake behind a self-propelled body, *Int. J. Comp. Fluid Dyn.*, **19**(2005), no. 5, 399–408. DOI: 10.1080/10618560500228024

[13] M.B. de Stadler, S.Sarkar, Simulation of a propelled wake with moderate excess momentum in a stratified fluid, *J. Fluid Mech.*, **692**(2012), no. 5, 28–52. DOI: 10.1017/jfm.2011.489

[14] A.G.Demenkov, G.G.Chernykh, Numerical modeling of the swirling turbulent wake decay past a self-propelled body, *Thermophys. Aeromech.*, **23**(2016), no. 5, 667–675. DOI: 10.1134/S086986431605005X

[15] A.G.Demenkov, G.G.Chernykh, Numerical modelling of swirling momentumless turbulent wake dynamics, *Computat. Techn.*, **23**(2018), no. 5, 37–48 (in Russian). DOI: 10.25743/ICT.2018.23.5.004

[16] A.V.Shmidt, Similarity in the far swirling momentumless turbulent wake, *J. SFU. Math. & Phys.*, **13**(2020), no. 1, 79–86. DOI: 10.17516/1997-1397-2020-13-1-79-86.

[17] G.G.Chernykh, A.G.Demenkov, O.V.Kaptsov, A.V.Shmidt, On mathematical modeling of swirling turbulent wakes with varied total excess momentum and angular momentum, *J. of Eng. Thermophys.*, **29**(2020), no. 2, 222–233. DOI: 10.1134/S1810232820020046

[18] G.G.Chernykh, A.G.Demenkov, O.V. Kaptsov, A.V. Shmidt, Dynamics of passive scalar in swirling turbulent far wakes, *J. of Eng. Thermophys.*, **31**(2022), no. 3, 512–521. DOI:10.1134/S1810232822030122

[19] A.V.Shmidt, Approximate solution for a flow in the far flat momentumless turbulent wake, *J. Appl. Mech. and Tech. Phys.*, **64**(2023), no. 6, 987–992. DOI: 10.1134/S002189442306007X

[20] A.V.Shmidt, Approximate solution to a model of the far momentumless axisymmetric turbulent wake, *J. SFU Math. & Phys.*, **17**(2024), no. 2, 229–237. EDN: LOHSKC

[21] A.V.Shmidt, Approximate solution of a boundary-value problem for a model of the far momentumless turbulent wake, *Mathematical models and integration methods: Seminar Proc. Appl. Mech. Phys.*, (2024), 125–134. DOI: 10.1515/9783111546667-008

[22] L.V.Ovsyannikov, Group analysis of differential equations, New York, Academic Press, 1982.

[23] J.B.Cazalbou, P.R.Spalart, P.Bradshaw, On the behavior of two-equation models at the edge of a turbulent region, *Phys. Fluids*, **6**(1994), no. 5, 1797–1804. DOI: 10.1063/1.868241

[24] O.V.Kaptsov, Yu.V.Shanko, A family of self-similar solutions for one model of a far turbulent wake, *Proc. Int. Conf. Comp. and Inf. Tech. in Sci., Pavlodar (Kazakhstan)*, **1**(2006), 576–579.

[25] G.I.Barenblatt, Similarity, self-similarity and intermediate asymptotics, *New York: Plenum*, 2003.

[26] S.Hassid, Similarity and decay laws of momentumless wakes, *Phys. Fluids*, **23**(1980), no. 2, 404–405.

[27] V.Maderich, S.Konstantinov, Asymptotic and numerical analysis of momentumless turbulent wakes, *Fluid Dyn. Res.*, **42**(2010). DOI: 10.1088/0169-5983/42/4/045503

[28] V.A.Puiseux, Recherches sur les fonctions algebriques, *J. Math. Pures Appl*, **15**(1850), 365–480.

[29] V.A.Puiseux, Nouvelles recherches sur les fonctions algebriques, *J. Math. Pures Appl*, **16**(1851), 228–240.

Приближенное решение, описывающее автомодельное вырождение дальнего закрученного турбулентного следа

Алексей В. Шмидт
Институт вычислительного моделирования СО РАН
Красноярск, Российская Федерация

Аннотация. Приближенное решение краевой задачи для модели дальнего закрученного турбулентного следа за самодвижущимся телом строится на основе асимптотического разложения решения в окрестности особой точки. Получено хорошее согласие сконструированного и численного решения.

Ключевые слова: дальний закрученный турбулентный след, краевая задача, полуэмпирическая модель турбулентности, приближенное решение, асимптотическое разложение.

EDN: UBDPFK
УДК 512.6

New Explicit Formula for Generalized Bernoulli Polynomials

Faiza Azzouni*
Farid Bencherif†
LA3C Laboratory, Faculty of Mathematics
USTHB, Algiers, Algeria
Laala Khaldi‡
LIM Laboratory, Department of Mathematics
University of Bouira
10000 Bouira, Algeria

Received 05.01.2025, received in revised form 05.02.2026, accepted 07.03.2026

Abstract. Let ℓ be a positive integer. In this paper, we prove a new explicit formula for the generalized Bernoulli polynomials of order ℓ. This formula generalizes many known identities for the Bernoulli polynomials. Our main tools are linear operators.

Keywords: Bernoulli polynomial, Bernoulli number, binomial coefficient, Stirling number of the second kind.

Citation: F. Azzouni, F. Bencherif, L. Khaldi, New Explicit Formula for Generalized Bernoulli Polynomials, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 391–398.
EDN: UBDPFK.

1. Introduction and main result

The generalized Bernoulli polynomials $B_n^{(\ell)}(x)$ of order ℓ ($\ell \in \mathbb{C}$) are defined by the generating function

$$\left(\frac{t}{e^t-1}\right)^{\ell} e^{xt} = \sum_{n=0}^{\infty} B_n^{(\ell)}(x)\frac{t^n}{n!}. \tag{1}$$

It is a straightforward to see that for $\ell = 1$, we obtain the classical Bernoulli polynomials $B_n(x)$. By setting $x = 0$ in Equation (1), we obtain $B_n^{(\ell)}$, the generalized Bernoulli numbers of order ℓ. Specifically, for $x = 0$ and $\ell = 1$ in Identity (1), we recover the classical Bernoulli numbers B_n (see, eg., Comtet [7, p. 48] and Roman [23, p. 94]).

These numbers and polynomials have many applications in various fields of Mathematics, such as analysis and number theory [18]. Many explicit formulas for classical, remarkable, and generalized numbers and polynomials can be found in [1, 2, 10, 14–16, 24, 26]. In the case with $\ell = -k$, k being a positive integer, we easily deduce from (1) the following representation for $B_n^{(-k)}(x)$:

$$B_n^{(-k)}(x) = \binom{n+k}{k}^{-1} S(n+k, k, x), \tag{2}$$

*f.azzouni@univ-boumerdes.dz
†fbencherif@usthb.dz
‡l.khaldi@univ-bouira.dz

where $S(n,k,x)$ denotes the generalized Stirling numbers of the second kind [6, p. 152, Relation (3.9)], defined by the generating function

$$\sum_{n=0}^{\infty} S(n,k,x)\frac{t^n}{n!} = \frac{1}{k!}(e^t-1)^k e^{tx},$$

and explicitly by

$$S(n,k,x) = \frac{1}{k!}\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}(x+j)^n.$$

For other values of ℓ, it is less obvious to find an explicit formula for $B_n^{(\ell)}(x)$ or simply for $B_n^{(\ell)}$. The following well-known explicit formula for B_n

$$B_n = \sum_{k=0}^{n}(-1)^k\binom{n+1}{k+1}\binom{n+k}{n}^{-1}S(n+k,k), \quad (n\geqslant 0), \tag{3}$$

where $S(n,k) = S(n,k,0)$ denotes the classical Stirling numbers of the second kind, is an old result, it was proved by many authors, among whom we can cite Jordan, who proved this identity in 1950 [13, p. 219], Gould, who gave a proof in 1972 [10, p. 48, Formula (11)], and it was also obtained by Shirai and Sato in 2001 [24, p. 140], Jeong, Kim and Son in 2005 [12, p.59], by Muthumalai [19, p. 61, Theorem 3.1] in 2013, as well as by Guo and Qi [8, Formula (6)] in 2015, by Qi and Chapman [20, p. 91, Relation (1.3)] in 2016, by Qi et al [21, p. 12, Theorem 2.2] in 2020, and also by Li and Chu in 2024 [17, p. 28187, Relation (4.3)]. Recall that Formula (3) has had multiple generalizations.

In 1985, Todorov [26, p. 665, Formula (3)] has gave the following generalization of (3)

$$B_n^{(\ell)} = \sum_{k=0}^{n}(-1)^k\binom{n+k}{k}^{-1}\binom{n+\ell}{n-k}\binom{\ell+k-1}{k}S(n+k,k),\ n\geqslant 0 \text{ and } \ell\in\mathbb{C}. \tag{4}$$

In 1988, Srivastava and Todorov [25] proved the following explicit formula:

$$B_n^{(\ell)}(x) = \sum_{k=0}^{n}\binom{n}{k}\binom{\ell+k-1}{k}\frac{k!}{(2k)!}\sum_{j=0}^{k}(-1)^j\binom{k}{j}j^{2k}(x+j)^{n-k}\times$$
$$\times F[k-n,k-\ell;2k+1;j/(x+j)],\ (\ell\in\mathbb{C}), \tag{5}$$

where $F[a,b;c;z]$ denotes the Gaussian hypergeometric function, defined by

$$F[a,b;c;z] = 1+\frac{ab}{c}\frac{z}{1!}+\frac{a(a+1)b(b+1)}{c(c+1)}\frac{z^2}{2!}+\cdots$$

In the applications of this formula, they find Todorov's formula [25, p. 513, Formula (15)].

In 2001, Shirai and Sato [24, p. 141, Corollary 7, Relation (1)] proved the following formula which, is not a special case of (4):

$$B_n = \frac{n}{n-1}\sum_{k=1}^{n+1}(-1)^k\binom{n+2}{k+1}\binom{n+k-1}{n-1}^{-1}S(n+k,k),\ (n\geqslant 2). \tag{6}$$

In 2005 [12, p. 60], Jeong, Kim and Son proved the following generalization of (3),

$$B_n^{(\ell)} = \sum_{k=0}^{n}(-1)^k\binom{n+1}{k+1}\binom{n+k\ell}{n}^{-1}S(n+k\ell,k\ell),\ n\geqslant 0 \text{ and } \ell\in\mathbb{N}. \tag{7}$$

Note that Jeong, Kim and Son formula (7) is also not a special case of (4).

In 2017, Boutiche et al. [4] exploited the demonstration of the formula (5) to prove the following explicit relation:

$$B_n^{(\ell)}(x) = \sum_{k=0}^{n} (-1)^k \binom{n+k}{k}^{-1} \binom{n+\ell}{n-k} \binom{\ell+k-1}{k} S(n+k,k,x),\ n \geqslant 0,\ \ell \in \mathbb{C}. \tag{8}$$

Formula (8) is, in fact, a transfer of Todorov's formula (4) for generalized Bernoulli numbers to generalized Bernoulli polynomials.

In 2017, Bencherif et al. [1] improved Formula (8) by proving the following identity:

$$B_n^{(\ell)}(x) = \sum_{k=0}^{m} (-1)^k \binom{n+k}{k}^{-1} \binom{m+\ell}{m-k} \binom{\ell+k-1}{k} S(n+k,k,x),\ m \geqslant n \geqslant 0,\ \ell \in \mathbb{C}. \tag{9}$$

Note again that Jeong, Kim and Son's formula (7) is not a special case of (9) either.

In 2022, Bencherif et al. [2, Corollary 2] generalized (9) by proving the following new identity, where $\lambda \in \mathbb{C},\ \alpha \in \mathbb{C},\ m, n,\ \ell \in \mathbb{N}$, and $m \geqslant n$:

$$B_n^{(\alpha\ell)}(x) = \sum_{k=0}^{m} (-1)^k \binom{\alpha+m}{m-k} \binom{\alpha+k-1}{k} \binom{n+k\ell}{k\ell}^{-1} S(n+k\ell, k\ell, x+\lambda\ell(\alpha+k)), \tag{10}$$

from which they deduced the following representation of $B_n^{(\ell)}(x)$ [2, p. 4, Formula (10)]

$$B_n^{(\ell)}(x) = \sum_{k=0}^{m} (-1)^k \binom{m+1}{k+1} \binom{n+k\ell}{n}^{-1} S(n+k\ell, k\ell, x),\ m \geqslant n \geqslant 0,\ \ell \in \mathbb{N}. \tag{11}$$

Letting $x = 0$ and $m = n$ in (11), we obtain (7). Formula (10) is effectively a generalization of the Jeong, Kim and Son formula (7).

In 2024, Li and Chu [17, p. 28187, Relation (4.4)] provided further proof of Shirai's formula (6), but it does not seem that this formula has been generalized to date. None of the above generalizations can be used to obtain it. The motivation for this work was to find a generalization of Formula (6). Our main result is the following :

Theorem 1.1. *For every integers m, n, r, ℓ with $m \geqslant n \geqslant r+1 \geqslant 1$, and $\ell \geqslant 0$, we have*

$$B_n^{(\ell)}(x) = \frac{n}{n-r} \sum_{k=0}^{m+r} (-1)^k \binom{n+k}{k} \binom{m+r+1}{k+1} \binom{n+k\ell}{k\ell}^{-1} \binom{n+k-r}{n-r}^{-1} \times S(n+k\ell, k\ell, x). \tag{12}$$

Letting $r = 1, \ell = 1$, and $m = n$ in (12), we get

$$B_n(x) = \frac{n}{n-1} \sum_{k=0}^{n+1} (-1)^k \binom{n+2}{k+1} \binom{n+k-1}{n-1}^{-1} S(n+k,k,x),\ (n \geqslant 2). \tag{13}$$

Identity (13) is a transfer of Formula (6) for the Bernoulli numbers B_n to the Bernoulli polynomials $B_n(x)$. We have therefore generalized Shirai's formula.

Theorem 1.1 also generalizes Formula (11). Indeed, for $r = 0$ in (12), we get (11).

Theorem 1.1 also provides new explicit formula for B_n. From (12) we can deduce the new representation of B_n given by the following Corollary:

Corollary 1.1. *For any integer $n \geqslant 2$, we have*

$$B_n = \frac{n}{n-1}\sum_{k=1}^{n+2}(-1)^{n+k-1}\binom{n+2}{k}\binom{n+k-2}{n-1}^{-1}S(n+k,k). \tag{14}$$

Proof. For $r=1$, $\ell=1$, $m=n$, and $x=1$ in (12), we obtain

$$B_n(1) = \frac{n}{n-1}\sum_{k=0}^{n+1}(-1)^{k}\binom{n+2}{k+1}\binom{n+k-1}{n-1}^{-1}S(n+k,k,1). \tag{15}$$

Noting that

$$B_n(1) = (-1)^n B_n,$$

and [6, p. 153, Formula (3.16)]

$$S(n+k,k,1) = S(n+k+1,k+1),$$

we can easily deduce (14) from (15). □

2. Lemma

The proof of Theorem 1.1 is mainly based on the following lemma:

Lemma 1.2. *For all integers m, n, r with $m \geqslant 0$, $n \geqslant 0$, and $r \geqslant 0$, we have*

$$\begin{aligned}&\binom{n-1}{r} - \sum_{k=0}^{m+r}(-1)^k\binom{n+k}{r}\binom{m+r+1}{k+1}x^{k+1} = \\ &= \sum_{k=0}^{r}(-1)^{r-k}\binom{n+m+k}{k}\binom{m+r+1}{r-k}(1-x)^{m+1+k}.\end{aligned} \tag{16}$$

Proof. Let

$$P_{n,m,r}(x) := \sum_{k=0}^{r}(-1)^{r-k}\binom{n+m+k}{k}\binom{m+r+1}{r-k}(1-x)^{m+1+k}. \tag{17}$$

It is clear that proving Relation (16) is equivalent to proving that

$$P_{n,m,r}(x) = \sum_{j=0}^{m+r+1}(-1)^j\binom{n+j-1}{r}\binom{m+r+1}{j}x^j. \tag{18}$$

By expanding $(1-x)^{m+1+k}$ using the binomial formula, we obtain:

$$P_{n,m,r}(x) = \sum_{j=0}^{m+r+1}(-1)^j A(n,m,r,j)x^j, \tag{19}$$

where

$$A(n,m,r,j) := \sum_{k=0}^{r}(-1)^{r-k}\binom{n+m+k}{k}\binom{m+r+1}{r-k}\binom{m+1+k}{j}. \tag{20}$$

Noting that

$$\binom{m+r+1}{r-k}\binom{m+1+k}{j} = \binom{m+r+1}{j}\binom{m+r+1-j}{r-k}.$$

Now, $A(n,m,r,j)$ can be written as

$$A(n,m,r,j)=\binom{m+r+1}{j}\sum_{k=0}^{r}(-1)^{r-k}\binom{n+m+k}{k}\binom{m+r+1-j}{r-k}.$$

Noticing that

$$(-1)^{r-k}\binom{n+m+k}{k}=(-1)^{r}\binom{-n-m-1}{k},$$

we deduce that

$$A(n,m,r,j)=\binom{m+r+1}{j}(-1)^{r}\sum_{k=0}^{r}\binom{-n-m-1}{k}\binom{m+r+1-j}{r-k}.$$

Thanks to Vandermonde convolution formula from Gould's book [11, p. 22, Relation (3.1)], we have

$$\sum_{k=0}^{r}\binom{-n-m-1}{k}\binom{m+r+1-j}{r-k}=\binom{r-n-j}{r}.$$

Therefore

$$A(n,m,r,j)=\binom{m+r+1}{j}(-1)^{r}\binom{r-n-j}{r}.$$

This simplifies to

$$A(n,m,r,j)=\binom{m+r+1}{j}\binom{n+j-1}{r}.$$

Substituting the expression for $A(n,m,r,j)$ into (19), we obtain

$$P_{n,m,r}(x)=\sum_{j=0}^{m+r+1}(-1)^{j}\binom{n+j-1}{r}\binom{m+r+1}{j}x^{j}.$$

Relation (18) is then proven. The proof of the lemma is complete. □

3. Proof of Theorem 1.1

Let D denote the derivation operator of $\mathbb{C}[x]$, and $\mathbb{C}[[D]]$ the commutative $\mathbb{C}$-algebra of composition operators of $\mathbb{C}[x]$. The order of a composition operator Ω is, by definition, the order of the unique formal series $S(t)\in\mathbb{C}[[t]]$ such that $\Omega=S(D)$ (see [3, Chap. VI], [22, p. 198]). Consider the automorphism $J=\dfrac{e^{D}-1}{D}$. From Lemma 1.2, for $\ell\in\mathbb{C}$, we have

$$\binom{n-1}{r}J^{-\ell}=\sum_{k=0}^{m+r}(-1)^{k}\binom{n+k}{r}\binom{m+r+1}{k+1}J^{k\ell}+J^{-\ell}P_{n,m,r}\left(J^{\ell}\right). \tag{21}$$

It is easy to see that the order of the composition operator $J^{-\ell}P_{n,m,r}\left(J^{\ell}\right)$ is greater than or equal to $m+1$. Indeed, according to the expression (17), $J^{-\ell}P_{n,m,r}(J^{\ell})$ can be written as a linear combination of the operators $J^{-\ell}(1-J^{\ell})^{m+1+k}$, where $k\in\{0,1,\ldots,r\}$, and each of these operators has an order greater than or equal to $m+1$. It follows that

$$J^{-\ell}P_{n,m,r}(J^{\ell})(x^{n})=0,\quad \text{for all integers } n\leqslant m.$$

Thus, by applying each of the members of (21) to x^n, for $m \geqslant n \geqslant r+1 \geqslant 1$, we obtain

$$\binom{n-1}{r} J^{-\ell}(x^n) = \sum_{k=0}^{m+r} (-1)^k \binom{n+k}{r} \binom{m+r+1}{k+1} J^{k\ell}(x^n).$$

Now, we have $J^{-\ell}(x^n) = B_n^{(\ell)}(x)$ and $J^{k\ell}(x^n) = B_n^{(-k\ell)}(x)$. It follows that

$$\binom{n-1}{r} B_n^{(\ell)}(x) = \sum_{k=0}^{m+r} (-1)^k \binom{n+k}{r} \binom{m+r+1}{k+1} B_n^{(-k\ell)}(x), \ (\ell \in \mathbb{C}). \tag{22}$$

From (2), we have

$$B_n^{(-k\ell)}(x) = \binom{n+k\ell}{k\ell}^{-1} S(n+k\ell, k\ell, x), \ (\ell \in \mathbb{N}).$$

Moreover, we have

$$\binom{n-1}{r} = \frac{n-r}{n} \binom{n}{r}.$$

We deduce that

$$B_n^{(\ell)}(x) = \frac{n}{n-r} \sum_{k=0}^{m+r} (-1)^k \binom{n+k}{r} \binom{n}{r}^{-1} \binom{m+r+1}{k+1} \binom{n+k\ell}{k\ell}^{-1} S(n+k\ell, k\ell, x).$$

Note that

$$\binom{n+k}{r} \binom{n}{r}^{-1} = \binom{n+k}{k} \binom{n+k-r}{n-r}^{-1},$$

Next, we observe that

$$B_n^{(\ell)}(x) = \frac{n}{n-r} \sum_{k=0}^{m+r} (-1)^k \binom{n+k}{k} \binom{m+r+1}{k+1} \binom{n+k\ell}{k\ell}^{-1} \binom{n+k-r}{n-r}^{-1} S(n+k\ell, k\ell, x),$$

where $m \geqslant n \geqslant r+1 \geqslant 1$ and $\ell \in \mathbb{N}$. This proves the desired result.

Note that Formula (22) for $r = 0$ and $m = n$, simplifies to

$$B_n^{(\ell)}(x) = \sum_{k=0}^{n} (-1)^k \binom{n+1}{k+1} B_n^{(-k\ell)}(x), \ (\ell \in \mathbb{C}). \tag{23}$$

For $x = 0$, (23) simplifies to

$$B_n^{(\ell)} = \sum_{k=0}^{n} (-1)^k \binom{n+1}{k+1} B_n^{(-k\ell)}, \ (\ell \in \mathbb{C}). \tag{24}$$

Then we obtain the identity (24) which was proved by Gould [9, p. 449, Relation (2.5)] and also by Carlitz [5] in 1959.

References

[1] F.Bencherif, B.Benzaghou, S.Zerroukhat, Une identité pour des polynômes d'Appell, *C. R. Math. Acad. Sci. Paris*, **355**(2017), 1201–1204.

[2] F.Bencherif, N.Mokhtari, S.Zerroukhat, A new identity for Appell polnomials, *J. Integer Seq.*, **25**(2022), 1–8.

[3] N.Bourbaki, Functions of a Real Variable, Elements of Mathematics, Springer-Verlag, Berlin, 2004.

[4] M.A.Boutiche, M.Rahmani, H.M.Srivastava, Explicit formulas associated with some families of generalized Bernoulli and Euler polynomials, *Mediterr. J. Math.*, **14**(2017), 1–6. DOI: 10.1007/s00009-017-0891-0

[5] L.Carlitz, Note on Nörlund's polynomial $B_n^{(z)}$, *Proc. Amer. Math. Soc.*, **11**(1960), 452–455.

[6] L.Carlitz, Weighted Stirling numbers of the first and second kind–I, *Fibonacci Quart.*, **18**(1980), 147–162.

[7] L.Comtet, Advanced Combinatorics, D. Reidel Pub. Co., Dordrecht-Holland/Boston-U.S.A., 1974.

[8] B.-N.Guo, F.Qi, An explicit formula for Bernoulli numbers in terms of Stirling numbers of the second kind, *J. Anal. Number Theory*, **3**(2015), no. 1, 27–30. DOI: 10.12785/jant/030105

[9] H.W.Gould, Stirling number representation problems, *Proc. Amer. Math. Soc.*, **11**(1960), 447–451.

[10] H.W.Gould, Explicit formulas for Bernoulli numbers, *Amer. Math. Monthly*, **79**(1972), 44–51.

[11] H.W.Gould, *Combinatorial identities*, revised edition, Morgantown, West-Virginia, 1972.

[12] S.Jeong, M.-S Kim, J.-W.Son, On explicit formulae for Bernoulli numbers and their counterparts in positive characteristic, *J. Number Theory*, **113**(2005), 53–68.

[13] C.Jordan, *Calculus of Finite Differences*, Chelsea Publ. Co., New York, 1950.

[14] L.Khaldi, F.Bencherif, A.Derbal, A note on explicit formulas for Bernoulli polynomials, *Zh. Sib. Fed. Univ. Mat. Fiz.*, **15**(2022), 224–233. DOI: 10.17516/1997-1397-2022-15-2-226-235

[15] L.Khaldi, F.Bencherif, and M. Mihoubi, Explicit formulas for Euler polynomials and Bernoulli numbers, *Notes on Number Theory and Discrete Mathematics*, **27**(2021), 80–89. DOI:10.7546/nntdm.2021.27.4.80-89

[16] T.Komatsu, B.K.Patel, C.Pita-Ruiz, Several formulas for Bernoulli numbers and polynomials, *Adv. Math. Commun.*, **17**(2023), 522–535. DOI: 10.3934/amc.2021006

[17] N.Li, W.Chu, Explicit formulae for Bernoulli numbers, *AIMS Mathematics*, **9**(2024), no. 10, 28170–28194. DOI: 10.3934/math.20241366

[18] B.Mazur, Bernoulli numbers and the unity of mathematics, 2008.
Available at https://wstein.org/wiki/attachments/2008(2f)480a/Bernoulli.pdf.

[19] R.K.Muthumalai, A note on Bernoulli numbers, *Notes on Number Theory and Discrete Mathematics*, *19*(2013), no. 1, 59–65.

[20] F.Qi, R.J.Chapman, Two closed forms for the Bernoulli polynomials, *J. Number Theory*, **159**(2016), 89–100. DOI: 10.1016/j.jnt.2015.07.021

[21] F.Qi, D.-W.Niu, D.Lim, Y.-H.Yao, Special values of the Bell polynomials of the second kind for some sequences and functions, *J. Math. Anal. Appl.*, **491**(2020), no. 2, Article 124382. DOI: 10.1016/j.jmaa.2020.124382

[22] A.M.Robert, A Course in p-adic Analysis, Springer-Verlag, New York, 2000.

[23] S.Roman, *The Umbral Calculus*, Academic Press, 1984.

[24] S.Shirai, K.-I.Sato, Some identities involving Bernoulli and Stirling numbers, *J. Number Theory*, **90**(2001), no. 1, 130–142.

[25] H.M.Srivastava, P.G.Todorov, An explicit formula for the generalized Bernoulli polynomials, *J. Math. Anal. Appl.*, **130**(1988), 509–513.

[26] P.G.Todorov, Une formule simple explicite des nombres de Bernoulli généralisés, *C. R. Acad. Sci. Paris, Ser. I*, **301**(1985), 665–666.

Новая явная формула для обобщенных полиномов Бернулли

Фаиза Аззуни
Фарид Беншериф
Лаборатория LA3C, Математический факультет
USTHB, Алжир, Алжир
Лаала Халди
Лаборатория LIM, кафедра математики
Университет Буира
10000 Буира, Алжир

Аннотация. Пусть ℓ — положительное целое число. В этой статье мы доказываем новую явную формулу для обобщенных многочленов Бернулли порядка ℓ. Эта формула обобщает многие известные тождества для многочленов Бернулли. Нашими основными инструментами являются линейные операторы.

Ключевые слова: полином Бернулли, число Бернулли, биномиальный коэффициент, число Стирлинга второго рода.

Journal of Siberian Federal University. Mathematics & Physics 2026, 19(3), 399–405

EDN: TWHOYE
УДК 519.6

Accurate Conservation of Invariant Quantities for Keplerian Orbit Integration

Oleg A. Zolotov*
Nikita V. Morozov†
Victor E. Zalizniak‡
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 10.12.2025, received in revised form 18.01.2026, accepted 20.02.2026

Abstract. A correction algorithm has been developed to preserve required motion invariants at each step of numerical integration of gravitational N-body problem, independent of the underlying numerical integration method. The effectiveness of the proposed algorithm is demonstrated using the Kepler problem as an example. Specifically, it is shown that the algorithm extends the applicability range of classical integrators of lower order.

Keywords: numerical solution of equations of motion, symplectic scheme, invariants of motion, Kepler problem.

Citation: O.A. Zolotov, N.V. Morozov, V.E. Zalizniak, Accurate Conservation of Invariant Quantities for Keplerian Orbit Integration, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 399–405. EDN: TWHOYE.

Introduction

The Kepler problem is the simplest case in celestial mechanics and astrodynamics. In the relative reference frame, there are seven conserved quantities: total energy E, three components (L_x, L_y, L_z) of the angular momentum vector $\mathbf{L}$ and three components (P_x, P_y, P_z) of the Laplace-Runge-Lenz vector $\mathbf{P}$. The listed quantities are closely related to orbital elements of the Kepler problem. Specifically, energy determines the semi-major axis, the magnitude of the Laplace vector determines the eccentricity, and the components of the angular momentum vector allow us to determine the argument of periapsis and the longitude of the ascending node [1]. It seems evident that adhering to conservation laws throughout the process of numerical integration would significantly improve the accuracy of the computed orbital elements [1].

Classical numerical integration methods do not preserve the invariants of motion. For example, when integrating a conservative system using an explicit Runge-Kutta scheme, a non-physical drift in total energy occurs [2]. For that reason, symplectic methods are widely used in celestial mechanics. Algorithms of this class are stable due to their preservation of the geometric structure of the phase space, which, in turn, limits the error in total energy [2]. It should be noted, however, that this property does not allow us to strictly conserve the energy integral from a theoretical standpoint. There are known applications of higher-order implicit numerical schemes [3], but all of them prove to be computationally expensive. Therefore, it is reasonable to consider introducing corrections to simpler numerical integration schemes that would strictly conserve invariants of motion.

*ozolot_@mail.ru
†tappmeist.718@gmail.com
‡vzalizniak@yandex.ru

We note that the introduction of corrections to preserve invariants has been applied in the works of Nacozy [4], Fukushima [5], and Liu & Liao [6]. However, these methods conserve only the energy integral.

In this work, we propose the direct method employing two scaling coefficients. It is independent of the chosen numerical integration scheme and strictly conserves the motion invariants for any system of bodies interacting via Newtonian potential.

1. Description of the proposed method

In the Hamiltonian formulation of the motion problem for multiple gravitating bodies, each of the N bodies is described by a position vector $\mathbf{r}_n$, momentum $\mathbf{p}_n$, and mass m_n (each body is treated as a point particle). The system evolves according to the equations of motion:

$$\frac{d\mathbf{p}_n}{dt} = -\frac{\partial H(\mathbf{r},\mathbf{p})}{\partial \mathbf{r}_n}, \quad \frac{d\mathbf{r}_n}{dt} = \frac{\partial H(\mathbf{r},\mathbf{p})}{\partial \mathbf{p}_n},$$

where

$$\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N), \quad \mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_N),$$

and the Hamiltonian $H(\mathbf{r},\mathbf{p})$, which does not explicitly depend on time, represents the total energy E of the system. Numerical integration of the problem involves time discretization. Let E_0 and L_0 denote the total energy and angular momentum of the system at the initial moment, respectively. For given $(\mathbf{r}^{k-1}, \mathbf{p}^{k-1})$, an arbitrary numerical integration algorithm produces $(\mathbf{r}^k, \mathbf{p}^k)$. The total energy $E(\mathbf{r}^k, \mathbf{p}^k)$ and angular momentum $L(\mathbf{r}^k, \mathbf{p}^k)$ are not conserved, i.e.,

$$E(\mathbf{r}^k, \mathbf{p}^k) \neq E_0 \quad \text{and} \quad L(\mathbf{r}^k, \mathbf{p}^k) \neq L_0.$$

For each integration step, we seek scaling corrections

$$(\mathbf{r}^k, \mathbf{p}^k) \to (\beta \mathbf{r}^k, \alpha \mathbf{p}^k),$$

which satisfy the condition of the minimum sum of squares of relative deviations of energy and angular momentum:

$$\frac{\left(\sum_i \mathbf{p}_i^2 - \sum_k G\frac{m_i m_j}{\|\mathbf{r}_i - \mathbf{r}_j\|} - E_0\right)^2}{E_0^2} + \frac{\left(\sum_i \mathbf{r}_i \times \mathbf{p}_i - \mathbf{L}_0\right)^2}{\mathbf{L}_0^2} \to \min.$$

Minimization of it using the method of Lagrange multipliers under the constraints of conserving linear momentum and the system's center-of-mass leads to the following two cases:

$$\begin{cases} \alpha^2 \dfrac{K}{E_0} + \dfrac{1}{\beta}\dfrac{U}{E_0} - 1 = 0 \\ \alpha\beta \dfrac{K}{E_0} - \dfrac{\mathbf{L}\mathbf{L}_0}{\mathbf{L}_0^2} = 0 \end{cases} \tag{1}$$

$$\begin{cases} 2\alpha^2 \dfrac{K}{E_0} + \dfrac{1}{\beta}\dfrac{U}{E_0} = 0 \\ (\alpha - 1)^2 + (\beta - 1)^2 \to \min \end{cases} \tag{2}$$

Note that in the case 1 both energy and angular momentum are conserved.

$$\alpha = \frac{\mathbf{L}}{\mathbf{L}_0}\frac{U}{2K}\left(-1 \pm \sqrt{1 + 4\left(\frac{\mathbf{L}\mathbf{L}_0^2}{\mathbf{L}^2}\right)^2 \frac{KE_0}{U^2}}\right),$$

$$\beta = -\frac{U}{2E_0}\left(-1 \mp \sqrt{1 + 4\left(\frac{\mathbf{L}\mathbf{L}_0^2}{\mathbf{L}^2}\right)^2 \frac{KE_0}{U^2}}\right).$$

From the pair of corrections, we select those that exert minimal influence on the trajectory and velocity of the body, i.e.,

$$(\alpha - 1)^2 + (\beta - 1)^2 \to \min.$$

When $\left(\frac{LL_0}{L^2}\right)^2 \frac{KE_0}{U^2} \leqslant -\frac{1}{4}$, system 1 has no solution, and conservation laws cannot be fulfilled. Solving system 2 leads to a fifth-degree equation in α, whose numerical solution at each integration step requires significant computational effort. Given that the corrections to the position vector must have minimal impact on the trajectory ($\beta \approx 1$), the approximation $\beta \approx \frac{1}{\beta}$ can be made. Applying this approximation, system 2 yields the following expression for α:

$$\alpha^3 + \frac{U}{2K}\left(1 + \frac{1}{2}\frac{U}{2K}\right)\alpha - \frac{1}{2}\left(\frac{U}{2K}\right)^2 = 0, \tag{3}$$

with

$$\frac{1}{\beta} = -2\alpha^2\frac{K}{U}.$$

Equation 3 will have three solutions, at least one of which is always real. In case of ambiguity in selecting the appropriate one, we use the condition from the previous case (in practice, we have not encountered instances where all three solutions are real):

$$(\alpha - 1)^2 + \left(\frac{1}{\beta} - 1\right)^2 \to \min.$$

It is important to note that these corrections permit a temporary departure of the position vectors and system momenta beyond the physically permissible phase volume. However, numerical experiments have demonstrated that in the vast majority of cases, this departure persists for no more than a few integration steps.

2. Numerical experiments

We analyze the influence of eccentricity on the efficiency of the algorithm and compare it with the methods proposed in [5, 6], and [7]. For simplicity, consider the Kepler problem in the form

$$H = \frac{|\mathbf{p}|^2}{2} - \frac{1}{|\mathbf{r}|},$$

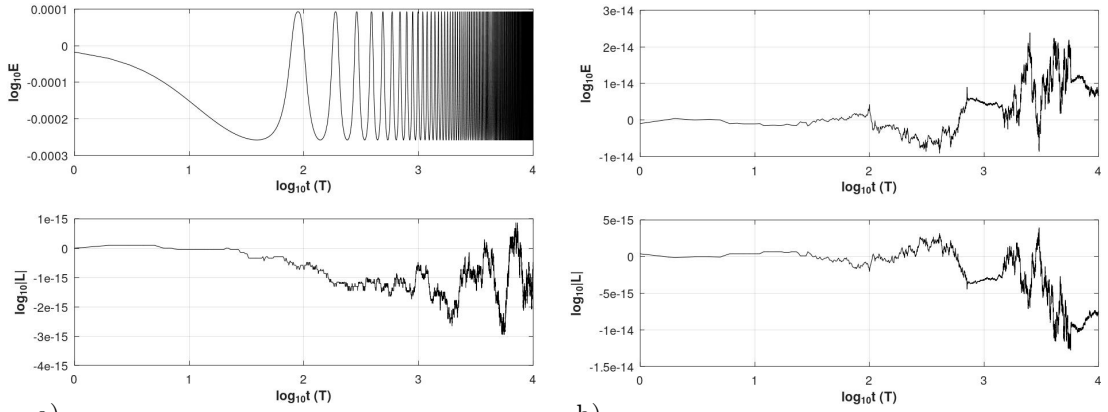
with fixed orbital elements:

$$a = 2, \quad i = 20^\circ, \quad \Omega = 50^\circ, \quad \omega = 30^\circ, \quad \text{and} \quad M = 40^\circ.$$

A second-order symplectic integration scheme with a fixed step size $h = T/100$ is used, where T is the orbital period. During the integration period, the body completes 10 full revolutions. The figures below illustrate the time evolution of the total energy E and the magnitude of angular momentum $|\mathbf{L}|$ during the numerical integration of the Keplerian orbit with eccentricity $e = 0.3$. From these orbital elements and eccentricity we deduce initial conditions (position and momentum of the body):

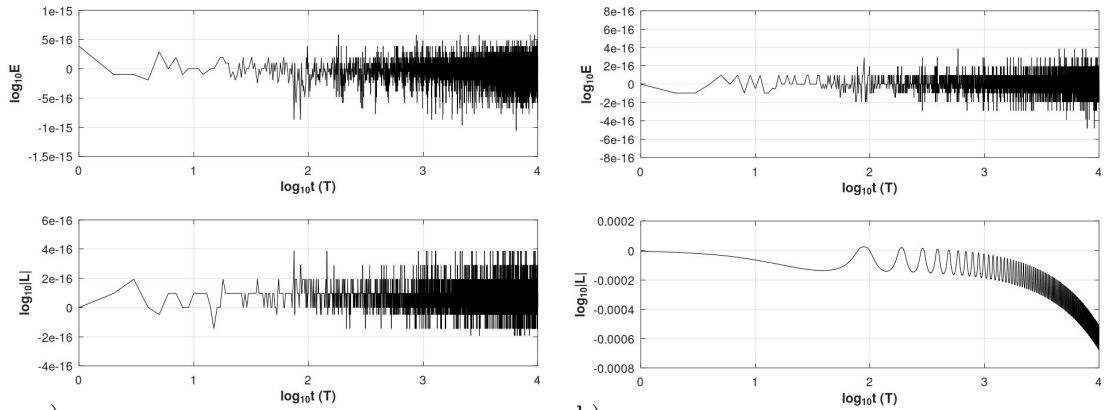
$$\mathbf{r} = (-1.3423, 0.7747, 0.5555), \quad \mathbf{p} = (-0.5928, -0.6023, 0.0244).$$

When applying any of the listed corrections, the accuracy of preserving the selected motion invariants is practically near a machine epsilon. The methods [5, 6], and [7] conserve only the energy integral. Although the symplectic algorithm preserves system's phase-space structure (thereby conserving angular momentum), the property is lost after applying aforementioned corrections. The results of the numerical experiments are listed in Figs. 1–4. These figures show decimal logarithms of the relative deviations of energy and angular momentum from their initial values.



a) b)

Fig. 1. Relative deviations of total energy and angular momentum magnitude for the Kepler problem ($e = 0.3$) during numerical integration; a) 2nd-order symplectic scheme with no correction; b) 6th-order Runge–Kutta–Feldberg scheme with adaptive step size and no correction



a) b)

Fig. 2. Relative deviations of total energy and angular momentum magnitude for the Kepler problem ($e = 0.3$) during numerical integration; 2nd-order symplectic scheme; a) with our correction scheme; b) with corrections proposed in [7].

We now consider the same problem with an increased eccentricity value $e = 0.98$. From this eccentricity and aforementioned orbital elements we deduce the following initial conditions:

$$\mathbf{r} = (-0.7998, -2.0338, -0.2528), \quad \mathbf{p} = (-0.1135, -0.6192, -0.1132).$$

It is well-known that integrating orbits with high eccentricity in regions of strong gravitational attraction requires a sufficiently small timestep. For such cases, high-order predictor-corrector

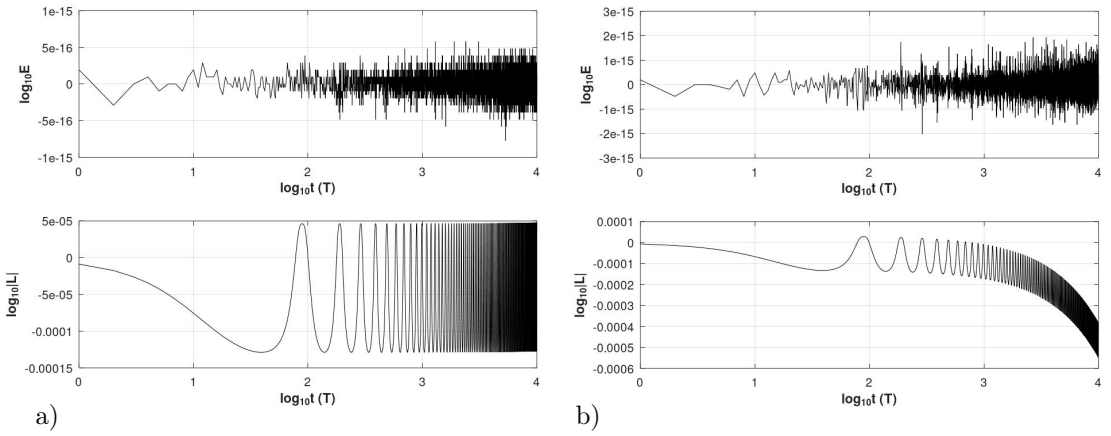


Fig. 3. Relative deviations of total energy and angular momentum magnitude for the Kepler problem ($e = 0.3$) during numerical integration; 2nd-order symplectic scheme; a) with corrections proposed in [6]; b) with corrections proposed in [5]

schemes (e.g., Runge–Kutta–Feldberg or Everhart) with adaptive step sizes are particularly reasonable. After applying the proposed corrections to a second-order fixed-step symplectic algorithm with $h = T/1000$, stable system behavior was achieved in the interval $[0, 100T]$. Below is a comparison with the 6th order Runge-Kutta-Feldberg method, where the step size varies within $[h/8, 4h]$.

Despite the extreme orbital precession, the shape of the orbit remains unchanged (within machine epsilon precision) because of the strict conservation of motion invariants. In contrast, the corrections proposed in [5, 6], and [7] do not achieve this effect, causing the system to eventually become unstable. This instability most likely arises from the accumulation of errors in the angular momentum vector.

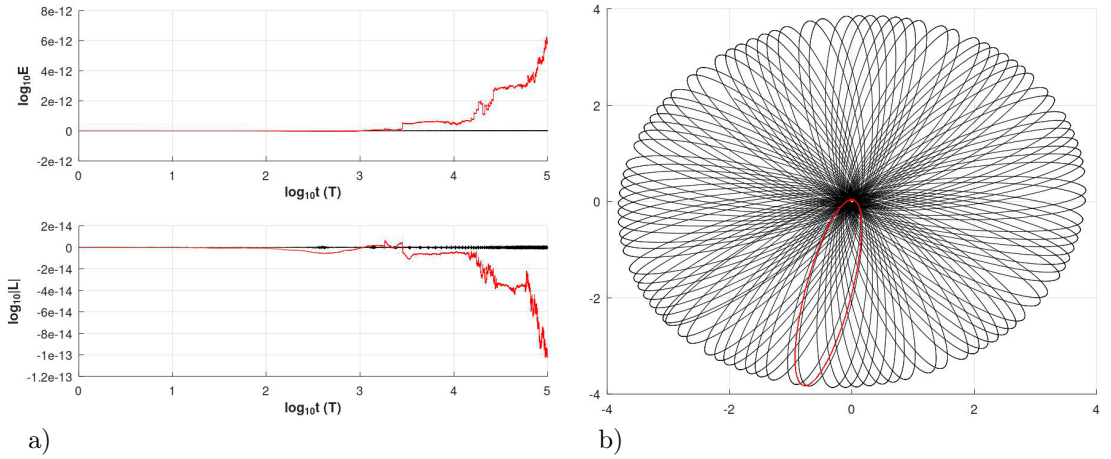


Fig. 4. Relative deviations of total energy and angular momentum magnitude for the Kepler problem ($e = 0.98$) during numerical integration using 6th-order Runge–Kutta–Feldberg scheme with adaptive step size with no correction (red) and using 2nd-order symplectic scheme with corrections proposed in this work (black); a) total energy E and the magnitude of angular momentum $|\mathbf{L}|$; b) Keplerian orbit

3. Algorithmic orbital precession

To better match the true orbit, we must deal with the orbital precession that is inherent in symplectic methods. We let $\mathbf{P}$ and $\mathbf{P}_0$ be current (numerical) and initial values of the Laplace vector, respectively. First we normalize them:

$$\mathbf{A} = \frac{\mathbf{P}}{\|\mathbf{P}\|}, \quad \mathbf{A}_0 = \frac{\mathbf{P}_0}{\|\mathbf{P}_0\|}.$$

We introduce rotation $\hat{\mathbf{R}}$ after each integration step that keeps the direction of Laplace vector:

$$\hat{\mathbf{R}}\mathbf{A} = \hat{\mathbf{R}}\mathbf{p} \times \hat{\mathbf{R}}\mathbf{L} - U\hat{\mathbf{R}}\mathbf{r} = \mathbf{A}_0,$$

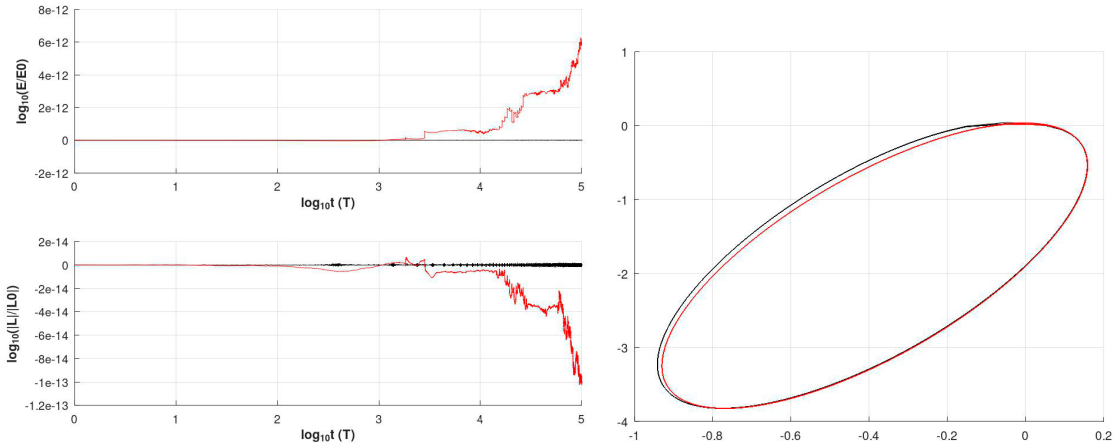
where U is the potential energy of the system. Then we get:

$$\hat{\mathbf{R}}\mathbf{A} = \mathbf{A} + (\sin\varphi)\mathbf{n} \times \mathbf{A} + (1 - \cos\varphi)\mathbf{n} \times (\mathbf{n} \times \mathbf{A}),$$

where

$$\mathbf{n} = \frac{\mathbf{A} \times \mathbf{A}_0}{\|\mathbf{A} \times \mathbf{A}_0\|}, \quad \varphi = (\widehat{\mathbf{A}, \mathbf{A}_0}).$$

We then apply this rotation to both position and momentum vectors before scaling correction. As a result, we completely eliminate orbital precession with this rotation (see Fig. 5). Note that both the energy and the angular momentum values are not affected by this transformation because we applied it beforehand



a) b)

Fig. 5. Relative deviations of total energy and angular momentum magnitude for the Kepler problem ($e = 0.98$) during numerical integration using 6th-order Runge–Kutta–Feldberg scheme with adaptive step size with no correction (red) and using 2nd-order symplectic scheme with both rotation and scale corrections proposed in this work (black); a) total energy E and the magnitude of angular momentum $|\mathbf{L}|$; b) Keplerian orbit (100 revolutions)

Conclusions

In this work, we have proposed a correction algorithm for the numerical integration of N-body orbits. By introducing two scaling coefficients, applied to the momentum and position vectors at each integration step, the method enforces strict conservation of energy and angular momentum. Specifically for the Kepler problem, preliminary rotation allows one to achieve consistency of the Runge–Laplace vector. The proposed direct correction is independent of the underlying numerical integration scheme and requires minimal computational overhead.

References

[1] D.Z.Ma, X.Wu, S.Y.Zhong, Extending Nacozy's approach to correct all orbital elements for each of multiple bodies, *The Astrophysical Journal*, **687**(2008), no. 2, 1294.

[2] H.Yoshida, Construction of higher order symplectic integrators, *Physics letters A*, **150**(1990), no. 5-7, 262–268.

[3] E.Everhart, An efficient integrator that uses Gauss-Radau spacings, *International Astronomical Union Colloquium. – Cambridge University Press*, **83**(1985), 185–202.

[4] P.E.Nacozy, The use of integrals in numerical integrations of the N-body problem, *International Astronomical Union Colloquium. – Cambridge University Press*, **10**(1971), 40–51.

[5] T.Fukushima, Efficient orbit integration by scaling for Kepler energy consistency, *The Astronomical Journal* **126**(2003), no. 2, 1097. DOI: 10.1086/376476

[6] L.Liu, X.H.Liao, *Acta Astron. Sinica*, **28**(1987), 225.

[7] O.A.Zolotov, V.E.Zalizniak, Accurate energy conservation in molecular dynamics simulation, *Nanosystems: physics, chemistry, mathematics*, **4**(2013), no. 5, 657–669.

Точное сохранение инвариантов движения в численном решении задачи Кеплера

Олег А. Золотов
Никита В. Морозов
Виктор Е. Зализняк
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. Разработан алгоритм введения поправок, сохраняющих необходимые инварианты движения на каждом шаге численных расчетов гравитационной задачи многих тел, не зависящий от базового метода численного интегрирования. На примере задачи Кеплера показана эффективность предложенного алгоритма. В частности, показано, что предложенный алгоритм расширяет область применимости классических интеграторов.

Ключевые слова: численное решение уравнений движения, симплектический алгоритм, инварианты движения, задача Кеплера.

EDN: XDSERQ
УДК 517.55

New Properties of the Laplace Transform in the Classical Domains of Second Type

Gulmirza Kh. Khudayberganov*
National University of Uzbekistan named after M. Ulugbek
Tashkent, Uzbekistan
Shokhrukh Sh. Rajabov†
Tashkent State Transport University
Tashkent, Uzbekistan

Received 10.06.2025, received in revised form 05.10.2025, accepted 07.02.2026

Abstract. In this article, matrix analogues of the general properties of the Laplace transform for the classical domains of second type are established. In particular, the theorem on the holomorphicity of the image function, the uniqueness theorem for original function, and an explicit inverse Laplace transform formula are derived in the matrix setting. Furthermore, new properties specific to the classical domains of second type are obtained and rigorously proved using direct matrix differential operators.

Keywords: classical domains, Hermitian matrix, holomorphic function, Laplace transform, matrix image function, matrix differential operators.

Citation: G.Kh. Khudayberganov, Sh.Sh. Rajabov, New Properties of the Laplace Transform in the Classical Domains of Second Type, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 406–416. EDN: XDSERQ.

Introduction

The Laplace transform is named after the great mathematician and astronomer Pierre-Simon Marquis de Laplace (1749–1827), who used a similar transformation in his work on probability theory [1], and as a result, the integral form of the Laplace transform naturally developed. The theory of the Laplace transform was further developed in the 19th and early 20th centuries by Matthias Lerch, Oliver Heaviside, and Thomas Bromwich. The scientist who brought the theory of the Laplace transform into a complete theory of the now widespread operational calculus is Gustav Doetsch. The concept of the Laplace transform was first used in his work published in 1937 [2]. He was the first to apply the Laplace transform to solving engineering problems in his scientific work. After that the Laplace transform relations for functions of one variable were studied, the question of constructing Laplace transform relations for functions of several arguments naturally arose. The solution to this problem was found in 1981 by the Russian mathematician L.G. Smyshlyayeva in her monograph [3].

The Laplace transform is used to study the properties of dynamical systems and to facilitate the solution of differential equations [4, 5]. For example, the Laplace transform reduces the solution of differential equations to the solution of algebraic equations, and then obtains solutions of equal strength [3, 6, 7]. After the emergence of matrix analysis, direct Laplace transforms were applied to matrix domains and their matrix analogues were obtained [4, 8, 9]. Until now, the Laplace transforms have been used by leading scientists from major scientific schools of the world for hypergeometric functions with matrix arguments [10], Laguerre matrix polynomials

*gkhudaiberg@mail.ru https://orcid.org/0000-0002-4837-8331
†sh.sh.rajabov@gmail.com https://orcid.org/0000-0002-3400-8364

with application to the numerical inversion of matrix functions [11], for the Bernstein operational matrix [12] and lately for functions with several matrix arguments [13]; for functions with symmetric matrix arguments the Laplace transform analogues were obtained [14–18].

1. Fundamental concepts of the Laplace transform in the classical domains of second type

The French mathematician E. Cartan showed in 1935 that there are six types of classes of unbounded symmetric domains [19]. Domains belonging to the first four of these classes are called classical domains [20].

Definition 1. *If the group of automorphisms of a domain $D \subset \mathbb{C}^n$ is transitive, that is, if for $z_1,\ z_2 \in D$ there exists an automorphism satisfying the condition $\varphi \in Aut\,(D)$ for an arbitrary $\varphi\,(z_1) = z_2$, then the domain $D \subset \mathbb{C}^n$ is called* ***a homogeneous domain.***

Definition 2. *If for an arbitrary point $\zeta \in D$ in a homogeneous domain $D \subset \mathbb{C}^n$ there exists an automorphism $\varphi \in Aut\,(D)$ satisfying the following conditions:*
1) the equality $\varphi\,(\zeta) = \zeta$ is valid for the only point $z \in D$;
2) $\varphi \circ \varphi = e$, where $e \in Aut\,(D)$ is the identity mapping;
then $D \subset \mathbb{C}^n$ is called ***a symmetric domain****.*

Definition 3. *The domain $D \subset \mathbb{C}^n$ is called* ***an irreducible domain*** *if it is not a direct product of bounded symmetric domains of lower dimension.*

Definition 4. *If the automorphism group of a bounded domain $D \subset \mathbb{C}^n$ is transitive and forms a Lie group, then the domain is called* ***a classical domain****.*

Below are the classical domains classified by E. Cartan [19, 20]:

$$\Re_I\,(m,k) = \left\{Z \in \mathbb{C}\,[m,k] :\ I^{(m)} - ZZ^* > 0\right\},$$

$$\Re_{II}\,(m) = \left\{Z \in \mathbb{C}\,[m,m] :\ I^{(m)} - Z\bar{Z} > 0,\ \forall Z' = Z\right\},$$

$$\Re_{III}\,(m) = \left\{Z \in \mathbb{C}\,[m,m] :\ I^{(m)} + Z\bar{Z} > 0,\ \forall Z' = -Z\right\},$$

$$\Re_{IV}\,(n) = \left\{Z \in \mathbb{C}^n :\ |\langle z,z\rangle|^2 - 2|z|^2 + 1 > 0,\ |\langle z,z\rangle| < 1\right\},$$

where $I^{(m)}$ is the m-order unit matrix, Z^* is the complex conjugate of the transposed matrix ($H > 0$ for a Hermitian matrix indicates that it is a positive definite matrix, i.e., all its eigenvalues are positive: $\det|\lambda I - H| = 0 \Rightarrow \forall \lambda_i > 0$) [20, 21].

Each of these classical domains is a homogeneous, symmetric, irreducible, convex complete circular domain centered at a point O (O is the zero matrix of m-order). These domains do not have a biholomorphic equivalence relation with each other, so a complex analysis is constructed for each of them separately.

We conduct this research work in **the classical domains of second type**:

$$\Re_{II}\,(m) = \left\{Z \in \mathbb{C}[m,m] :\ I^{(m)} - Z\bar{Z} > 0,\ \forall Z' = Z\right\}.$$

Let $f : S_m \to S_m$ $(A \in S_m \subset \mathbb{R}\,[m \times m]\,|\,f(A) = f\,(UAU') \in S_m)$ be a symmetric matrix-function with matrix argument (here, S_m is the class of real symmetric matrices, $UU' = I$ are orthogonal matrices) [10].

Definition 5. *A function $f(A)$ that satisfies the following conditions is called* ***a matrix original****:*

I. *$f(A) \equiv 0$ for $A < 0$ (here the relation $A < 0$ is understood as each element of the matrix A being less than 0);*

II. *A matrix-function $f(A)$ with a real matrix argument is continuous or piecewise continuous (i.e. $\forall X_0 \in \Upsilon : \lim\limits_{X \to X_0} f(X) = f(X_0)$, or it has a finite number $n_0 \leq m$ of type I discontinuities) in the matrix right half-plane: $\Upsilon = \{A = (a_{ij}) \in \mathbb{R}[m \times m] : \forall a_{ij} \geqslant 0\}$;*

III. *$\forall A \in S_m$ $\exists M > 0$ and $\alpha \geqslant 0$ such that $|f(A)| \leqslant M \cdot e^{\alpha A}$ (here, each element of $|f(A)|$ is understood to be equal to the modulus of the element of the symmetric matrix-function $f(A)$).*

Now, based on Definition 5 above, we give the definition of a matrix image.

Definition 6. *The* ***matrix image*** *of a matrix original in the classical domains of second type, $Z = X + iY$ $(Z(m))$ is defined as*

$$F(Z) = \mathcal{L}_Z\{f(A)\} = \int\limits_{A>0} e^{-Sp(ZA)} f(A)\, dA\,. \tag{1}$$

Here, for $A = (a_{ij})$ we have $dA = \prod\limits_{i \leqslant j} da_{ij}$, $Sp(ZA) = \sum\limits_{i \geqslant j} z_{ij} \cdot a_{ij}$ is the trace of the matrix [10, 21, 22].

Remark 1. In the integral (1) a complex parametric matrix Z is defined as $Z = (\eta_{ij} z_{ij})$, and the coefficients η_{ij} (invariant coefficients) of the matrix elements are defined as follows [22]:

$$\eta_{ij} = \begin{cases} \frac{1}{2}, & i \neq j. \\ 1, & i = j. \end{cases} \tag{2}$$

Definition 7. *The transformation from a matrix original to a matrix image by formula (1) is called the* ***Laplace transform for the classical domains of second type*** *$\Re_{II}(m)$.*

Remark 2. The matrix original $f(A)$ and the matrix image $F(Z)$ is defined as $F(Z) \stackrel{\cdot}{\to} f(A)$ or $f(A) \stackrel{\cdot}{\leftarrow} F(Z)$. Here, the direction of the sign "$\stackrel{\cdot}{\to}$" is always from the matrix image to the matrix original [14, 18]. Also, the notation $\mathcal{L}_Z\{f(A)\} = F(Z)$ is also used [10, 22].

2. Analogs of general properties of the Laplace transform

Theorem 1 (Theorem on the holomorphicity of image)**.** *The matrix image $\mathcal{L}_Z\{f(A)\} = F(Z)$ of a matrix original $f(A)$ is a holomorphic function of the matrix variable Z from the matrix right half-plane: $\Upsilon_{\Re_{II}} = \{Z(m) : \operatorname{Re} Z = X > X_0 > 0\}$.*

Proof. Suppose that $F(Z)$ is the image function at a point $\forall Z = \{Z \in \Re_{II}(m) : \operatorname{Re} Z = X > X_0 > 0\}$, corresponding to the matrix original $f(A)$, determined via the Laplace transform (1). That is, $\mathcal{L}_Z\{f(A)\} = F(Z)$. For the matrix-exponential $e^{Sp(-ZA)}$ the following estimate holds: for a fixed $Z \in \Re_{II}(m)$ and $\forall A \in S_m \subset \mathbb{R}[m \times m]$ there exist real numbers $C > 0$ and $\frac{\alpha}{2} \geqslant 0$ such that for $\sigma := \min\limits_i \sigma_i > \frac{\alpha}{2}$ (here σ_i are the eigenvalues of the matrix $\lambda_i(Z) - Z$) we have

$$\left| e^{-Sp(ZA)} \right| \leqslant C \cdot e^{-\left(\sigma - \frac{\alpha}{2}\right) \cdot Sp(A)}. \tag{3}$$

On the other hand, according to Definition 5 of the matrix original $f(A)$ by condition *III* $\forall A \in S_m \subset \mathbb{R}[m \times m]$ we find the numbers $\exists M > 0$ and $\frac{\alpha}{2} \geqslant 0$ such that

$$|f(A)| \leqslant M \cdot e^{\frac{\alpha}{2} \cdot Sp(A)}. \tag{4}$$

According to estimates (3) and (4):

$$\begin{gathered} \left|f(A) \cdot e^{-Sp(ZA)}\right| \leqslant |f(A)| \cdot C \cdot e^{-\left(\sigma - \frac{\alpha}{2}\right) \cdot Sp(A)} \leqslant M \cdot e^{\frac{\alpha}{2} \cdot Sp(A)} \times \\ \times\ C \cdot e^{-\left(\sigma - \frac{\alpha}{2}\right) \cdot Sp(A)} = [M \cdot C = K] = K \cdot e^{-(\sigma - \alpha) \cdot Sp(A)}. \end{gathered} \tag{5}$$

The integral of the majorant function $e^{-(\sigma-\alpha)\cdot Sp(A)}$:

$$\int\limits_{A>0} e^{-(\sigma-\alpha) Sp(A)}\, dA < +\infty$$

is convergent, it follows that the integral (1) is absolutely and uniformly convergent. This allows us to differentiate under the integral sign:

$$\begin{gathered} \frac{\partial}{\partial Z} F(Z) = \frac{\partial}{\partial Z} \left[\int\limits_{A>0} e^{-Sp(ZA)} f(A)\, dA \right] = \int\limits_{A>0} \frac{\partial}{\partial Z} e^{-Sp(ZA)} f(A)\, dA = \\ = \left[\frac{\partial}{\partial Z} Sp(ZA) = A \right] = - \int\limits_{A>0} A \cdot e^{-Sp(ZA)} f(A)\, dA. \end{gathered}$$

Therefore, the matrix image function $F(Z)$ has a derivative in the sense of complex analysis $\forall Z \in \Upsilon_{\Re_{II}} = \{Z \in \Re_{II}(m) :\ \operatorname{Re} Z = X > X_0 > 0\}$, i.e. the function $F(Z)$ is holomorphic in the domain $Z \in \Upsilon_{\Re_{II}}$. The theorem is proved. □

Theorem 2 (The inverse Laplace transform)**.** *If $F(Z)$ is a matrix image function on $\Re_{II}(m)$,*

$$\int\limits_{-\infty}^{+\infty} |F(X+iY)\, dY| < +\infty, \tag{6}$$

and for arbitrary $X > 0$

$$\lim_{X \to 0} \int\limits_{-\infty}^{+\infty} |F(X+iY)\, dY| = 0, \tag{7}$$

and $\mathcal{L}_Z\{f(A)\} = F(Z)$, then at every point where $f(A)$ is differentiable, the following unique inverse Laplace transform relation holds

$$\mathcal{L}^{-1}\{F(Z)\} = f(A) = \frac{2^{\frac{1}{2}m(m-1)}}{(2\pi i)^{\frac{1}{2}m(m+1)}} \int\limits_{\operatorname{Re} Z>0} e^{Sp(ZA)} F(Z)\, dZ\,. \tag{8}$$

Remark 3. In the integral relation (8) we asssume $Z = (\eta_{ij}\, z_{ij})$, $\eta_{ij} = 1$.

Proof. By the condition of the theorem $\mathcal{L}_Z\{f(A)\} = F(Z)$, it follows that the following relation must hold for the inverse Laplace transform relation:

$$\mathcal{L}^{-1}(\mathcal{L}_Z\{f(A)\}) = \mathcal{L}^{-1}(F(Z)) \Rightarrow f(A) = \mathcal{L}^{-1}(F(Z))\,. \tag{9}$$

In view of (9), we obtain

$$f(A) = \mathcal{L}^{-1}\left(\int\limits_{A>0} e^{-Sp(ZA)} f(A)\, dA \right). \tag{10}$$

According to Theorem 1, which was proved above, the matrix image function $F(Z)$ is holomorphic in the matrix right half-plane in the second type classical domain $\Upsilon_{\Re_{II}} = \{Z \in \Re_{II}(m) : \operatorname{Re} Z = X > X_0 > 0\}$, and on the other hand, according to the conditions (6) and (7) of the theorem, it is continuous in the closure of the domain $\overline{\Upsilon}_{\Re_{II}}$. That is $F(Z) \in \mathcal{O}(\Upsilon_{\Re_{II}}) \cap C\left(\overline{\Upsilon}_{\Re_{II}}\right)$. This allows us to apply the *Cauchy inversion formula* for functions with matrix arguments [10, 22, 23]:

$$\frac{1}{(2\pi i)^{\frac{1}{2}m(m+1)}} \int\limits_{\operatorname{Re} Z>0} e^{Sp(ZA)} F(Z)dZ = \begin{cases} f(A), \; if \; A > 0 \\ O, \quad if \; A \leqslant 0. \end{cases} \tag{11}$$

Therefore, from relations (10) and (11) it follows that:

$$f(A) = \frac{1}{(2\pi i)^{\frac{1}{2}m(m+1)}} \int\limits_{\operatorname{Re} Z>0} e^{Sp(ZA)} F(Z)dZ. \tag{12}$$

As $\eta_{ij} = 1$ in formula (2) $Z = (\eta_{ij}\, z_{ij})$, we have a constant coefficient $2^{\frac{m(m-1)}{2}}$ in the integral (12). Therefore, we have the following integral relation:

$$f(A) = \frac{2^{\frac{1}{2}m(m-1)}}{(2\pi i)^{\frac{1}{2}m(m+1)}} \int\limits_{\operatorname{Re} Z>0} e^{Sp(ZA)} F(Z)dZ.$$

The theorem has been proven. □

Remark 4. It should also be noted that the relation $\mathcal{L}^{-1}(F(Z)) = f(A)$ and $\mathcal{L}_Z\{f(A)\} = F(Z)$ are equivalent:

$$\mathcal{L}_Z\{f(A)\} = F(Z) \Leftrightarrow \mathcal{L}^{-1}(F(Z)) = f(A).$$

Lemma 1. *For $F(Z) \in \Re_{II}(m)$ the following linearity relations hold:*

I. if $c = const$ and $\mathcal{L}_Z\{f(A)\} = F(Z)$ then

$$\mathcal{L}_Z\{c\, f(A)\} = c\, F(Z). \tag{13}$$

II. if $\mathcal{L}_Z\{f_1(A)\} = F_1(Z)$ and $\mathcal{L}_Z\{f_2(A)\} = F_2(Z)$ then

$$\mathcal{L}_Z\{f_1(A) + f_2(A)\} = F_1(Z) + F_2(Z). \tag{14}$$

Proof. The proof of the lemma follows directly from the properties of linearity and the exclusion of the constant from the integral sign in integration [14]. □

Theorem 3 (The uniqueness of the original)**.** *If $F(Z) \in \Re_{II}(m)$ is the matrix image function of two matrix original functions $f_1(A)$ and $f_2(A)$, then these originals coincide at all their continuity points. That is, the equality $f_1(A) \equiv f_2(A)$ holds.*

Proof. By the condition of the theorem: $\mathcal{L}_Z\{f_1(A)\} = F(Z)$ and $\mathcal{L}_Z\{f_2(A)\} = F(Z)$. Using the linearity property of the matrix representation function, we obtain the following relation: $\mathcal{L}_Z\{f_1(A) - f_2(A)\} = F(Z) - F(Z) = 0$. It follows that $f_1(A) \equiv f_2(A)$. □

Theorem 4 (Analogs of general properties of the Laplace transform)**.** *If $\mathcal{L}_Z\{f(A)\} = F(Z)$ is the Laplace transform, then the following relations hold between the matrix image and the matrix original:*

1°. If $\forall \beta > 0$, then the following relation holds:

$$\mathcal{L}_Z\{f(\beta A)\} = \frac{1}{\beta} F\left(\frac{1}{\beta} Z\right); \tag{15}$$

2°. *If* $\forall B\,[c] \in S_m \subset \mathbb{R}\,[m \times m]$, $(c = const > 0)$, *then the following relation is satisfied:*

$$\mathcal{L}_Z\left\{f\left(A - B[c]\right)\right\} = e^{-ZB\,[c]}F\left(Z\right); \tag{16}$$

3°. *If* $\forall \Lambda \in S_m^*$, *then the following relation is valid:*

$$\mathcal{L}_Z\left\{e^{\Lambda A}f\left(A\right)\right\} = F\left(Z - \Lambda\right); \tag{17}$$

4°. *If* $f'\left(A\right)$ *exists and forms a matrix original function, then*

$$\mathcal{L}_Z\left\{f'\left(A\right)\right\} = ZF\left(Z\right) - f\left(O\right), \tag{18}$$

here O *is* m*-order zero matrix.*

5°. *If* $Z \in S_m^*$ *for* $\det\left(Z\right) \neq 0$, *then the following relation is valid:*

$$\mathcal{L}_Z\left(\int\limits_{0<B<A} f\left(B\right)dB\right) = Z^{-1}F\left(Z\right). \tag{19}$$

6°. *If* $F'\left(Z\right) \in \Re_{II}\left(m\right)$, *then the following relationship is valid:*

$$\mathcal{L}^{-1}\left(F'\left(Z\right)\right) = Af\left(A\right). \tag{20}$$

7°. *If* $\det\left(A\right) \neq 0$ *and* $A^{-1}f\left(A\right)$ *is a matrix original, then the following relation holds:*

$$\mathcal{L}^{-1}\left(\int\limits_{0<W<Z} F\left(W\right)dW\right) = A^{-1}f\left(A\right). \tag{21}$$

Proof. It is known that the classical domains of second type $\Re_{II}\left(m\right)$ belongs to the class of symmetric matrices. We have proved the above properties for the class of symmetric matrices in our previous article [14]. Therefore, we will not give a proof here. □

3. New properties of the matrix Laplace transform in the classical domains of second type

In this section, we present new properties of the Laplace transform that are important only for matrix-argument functions, which have no analogues in functions with single and multivariable arguments. For this, we introduce the definitions of the necessary concepts that we need.

Let us be given a parameteric matrix $A = (a_{ij}) \in S_m \subset \mathbb{R}\,[m \times m]$.

Definition 8. *The following operators*

$$D_A = (-1)^m \det\left(\frac{\partial}{\partial a_{ij}}\right) \text{ and } \hat{D}_A = (-1)^m \det\left(\eta_{ij}\frac{\partial}{\partial a_{ij}}\right) \tag{22}$$

are called ***matrix differential operators*** *[20, 22, 25].*

The matrix differential operators defined by (22) have the following properties.

Property 1. If $\eta_{ij} = 1$ in the relation (22), then $D_A \equiv \hat{D}_A$.

Property 2. If $A = (a_{ij}) \in S_m \subset \mathbb{R}\,[m \times m]$ and $B = (b_{ij}) \in S_m \subset \mathbb{R}\,[m \times m]$, then the following relation holds:

$$\hat{D}_A e^{-Sp(AB)} = \det\left(B\right)e^{-Sp(AB)}. \tag{23}$$

Property 3. If $A = (a_{ij}) \in S_m \subset \mathbb{R}[m \times m]$ and $Z = (z_{ij}) \in S_m^* \subset \mathbb{C}[m \times m]$, then the following relation holds:

$$D_Z\, e^{-Sp(ZA)} = \det(A)\; e^{-Sp(ZA)}. \tag{24}$$

Property 4. If $A = (a_{ij}) \in S_m \subset \mathbb{R}[m \times m]$ and $Z = (z_{ij}) \in S_m^* \subset \mathbb{C}[m \times m]$, then the following relation holds:

$$\hat{D}_A\, e^{-Sp(ZA)} = \det(Z)\; e^{-Sp(ZA)}. \tag{25}$$

Property 5. If $A = (a_{ij}) \in S_m \subset \mathbb{R}[m \times m]$ and $Z = (z_{ij}) \in S_m^* \subset \mathbb{C}[m \times m]$, then the following relation holds:

$$D_Z^n \left[e^{-Sp(ZA)}\right] = (\det(A))^n\, e^{-Sp(ZA)}. \tag{26}$$

Here, D_Z^n, $(n = 0, 1, 2, \dots)$ is the differential operator applied to itself repeatedly n times.

Proof. All five properties given above follow directly from the definition of the matrix differential operator and the properties of the trace of a symmetric exponential matrix [22]. □

Now we present new properties of the Laplace transform in the classical domains of the second type.

Let us be given a positive definite real symmetric matrix$A$ $(\det|\lambda I - A| = 0 \Rightarrow \forall \lambda_i > 0)$ and its symmetric square root $A^{1/2}$. Let the Laplace transform $\mathcal{L}_Z\{f(A)\} = F(Z)$ relation holds.

Theorem 5. *If* $\forall X = X' > 0,\ Y = Y' > 0$ *we have* $A - X^{-1/2}Y\,X^{-1/2} > 0$*, then*

$$\begin{aligned}\mathcal{L}_Z\left\{f\left(X^{1/2}A\,X^{1/2} - Y\right)\right\} = (\det(X))^{-\frac{m+1}{2}}\mathcal{L}_{X^{-1/2}Z\,X^{-1/2}}(f(A)) \times \\ \times\, e^{-Sp\left(Y^{1/2}X^{-1/2}Z\,X^{-1/2}Y^{1/2}\right)}.\end{aligned} \tag{27}$$

Proof. According to the assumption of the theorem, the inequality $A - X^{-1/2}Y\,X^{-1/2} > 0$ holds. Denote the expression on the left side of this inequality as $T = A - X^{-1/2}Y\,X^{-1/2}$. Differentiating both sides of this equality, we find that $dT = dA$. Also, by definition $A = T + X^{-1/2}Y\,X^{-1/2}$. Now we apply the Laplace transform relation (1) for the matrix original function $f\left(X^{1/2}A\,X^{1/2} - Y\right)$:

$$\mathcal{L}_Z\left\{f\left(X^{1/2}A\,X^{1/2} - Y\right)\right\} = \int\limits_{A - X^{-1/2}Y\,X^{-1/2} > 0} e^{-Sp(ZA)}\, f\left(X^{1/2}A\,X^{1/2} - Y\right) dA =$$

$$= e^{-Sp\left(Z\left(X^{-1/2}Y\,X^{-1/2}\right)\right)} \cdot \int\limits_{T = T' > 0} e^{-Sp(ZT)} f\left(X^{1/2}TX^{1/2}\right) dT. \tag{$*$}$$

In the integral $(*)$, we perform the change of variables: $U = X^{1/2}TX^{1/2} \Rightarrow dU = (\det(X))^{\frac{m+1}{2}} dT$. Then, for the exponential we have: $-Sp(ZT) = -Sp\left(\left(X^{-1/2}ZX^{-1/2}\right)U\right)$. From this equality, it follows that if $Z \in \Re_{II}(m)$, then the transformation $X^{-1/2}ZX^{-1/2}$ preserves the domain, that is $X^{-1/2}ZX^{-1/2} \in \Re_{II}(m)$. This fact is essential, since it guarantees the validity of the Laplace transform in the integral $(*)$:

$$(\det(X))^{-\frac{m+1}{2}} \cdot e^{-Sp\left(Z\left(X^{-1/2}Y\,X^{-1/2}\right)\right)} \int\limits_{U > 0} e^{-Sp\left(\left(X^{-1/2}Z\,X^{-1/2}\right)U\right)} f(U)\, dU =$$

$$= (\det(X))^{-\frac{m+1}{2}} \cdot e^{-Sp\left(Z\left(X^{-1/2}Y\,X^{-1/2}\right)\right)} \mathcal{L}_{X^{-1/2}Z\,X^{-1/2}}(f(A)).$$

Finally, noting that [22, 26, 27]

$$Sp\left(Z\left(X^{-1/2}Y\,X^{-1/2}\right)\right) = Sp\left(Y^{1/2}X^{-1/2}Z\,X^{-1/2}Y^{1/2}\right),$$

we arrive at

$$(\det(X))^{-\frac{m+1}{2}}\mathcal{L}_{X^{-1/2}ZX^{-1/2}}(f(A))\cdot e^{-Sp\left(Y^{1/2}X^{-1/2}ZX^{-1/2}Y^{1/2}\right)}.$$

The theorem is proved. □

Theorem 6. *If $\mathcal{L}_Z\{f(A)\} = F(Z)$ for $F(Z) \in \Re_{II}(m)$, then*

$$\mathcal{L}_Z\{(\det(A))^n f(A)\} = D_Z^n(F(Z)). \tag{28}$$

Here, D_Z is the matrix differential operator (22).

Proof. To prove it, we first take the left-hand side of equality (28):

$$\mathcal{L}_Z\{(\det(A))^n f(A)\} = \int\limits_{A>0} e^{-Sp(ZA)}(\det(A))^n f(A)\,dA. \tag{29}$$

Now we take the right-hand side of equality (28) and use Definition 8 of the matrix differential operator (22):

$$D_Z^n(F(Z)) = D_Z^n\left(\int\limits_{A>0} e^{-Sp(ZA)}f(A)\,dA\right) = \int\limits_{A>0} D_Z^n e^{-Sp(ZA)}f(A)\,dA. \tag{30}$$

In the last equation (30), the integral under $D_Z^n e^{-Sp(ZA)}$ applying property 5 of the matrix differential operator to the expression, we obtain the following relation:

$$\int\limits_{A>0} D_Z^n e^{-Sp(ZA)}f(A)\,dA = \int\limits_{A>0}(\det(A))^n e^{-Sp(ZA)}f(A)\,dA. \tag{31}$$

We can see that equations (29) and (31) are exactly equal to each other. It follows that relation (28) is valid. □

Theorem 7. *If $Z \in \Re_{II}(m)$ and for a positive definite invariant matrix $T \in S_m^*$ such that $Z - T \in \Re_{II}(m)$, the following relation holds:*

$$\int\limits_{Z-T>0}\mathcal{L}_Z\{f(A)\}dZ = \Gamma_m\left(\frac{m+1}{2}\right)\mathcal{L}_T\left\{(\det(A))^{\frac{m+1}{2}}f(A)\right\}. \tag{32}$$

Here $\Gamma_m\left(\dfrac{m+1}{2}\right)$ is the matrix gamma function [22, 24].

Proof. To prove it, we write the left-hand side of equation (32):

$$\int\limits_{Z-T>0}\mathcal{L}_Z\{f(A)\}dZ = \int\limits_{Z-T>0}\left[\int\limits_{A>0} e^{-Sp(ZA)}f(A)\,dA\right]dZ. \tag{33}$$

By introducing the notation $dZ = dW$. In the integral on the right-hand side of the last obtained equation (33) differentiating both sides of the equation, we obtain the following equality: $Z - T =$

W. We take these into the integral (33):

$$\int\limits_{A>0}\left[\int\limits_{W>0} e^{-Sp(WA)}dW\right]e^{-Sp(TA)}f\left(A\right)dA=\begin{bmatrix}WA=B & \Rightarrow & W=A^{-1}B\\ dW=\left(\det\left(A\right)\right)^{-\frac{m+1}{2}}dB & & \\ W>0 & \Rightarrow & B>0\end{bmatrix}=$$

$$=\int\limits_{A>0}\left[\int\limits_{B>0} e^{-Sp(B)}\left(\det\left(A\right)\right)^{-\frac{m+1}{2}}dB\right]e^{-Sp(TA)}f\left(A\right)dA= \tag{34}$$

$$=\int\limits_{A>0}\left[\int\limits_{B>0} e^{-Sp(B)}dB\right]\left(\det\left(A\right)\right)^{-\frac{m+1}{2}}e^{-Sp(TA)}f\left(A\right)dA.$$

The resulting matrix gamma function [22, 27–30] for the inner integral in expression (34) is:

$$\Gamma_m\left(\alpha\right)=\int\limits_{B>0}\left(\det\left(B\right)\right)^{\alpha-\frac{m+1}{2}}e^{-Sp(B)}dB,\quad \mathrm{Re}\left(\alpha\right)>\frac{m-1}{2}.$$

If we assume $\alpha=\dfrac{m+1}{2}$ in this formula, then we get the following relation

$$\Gamma_m\left(\frac{m+1}{2}\right)=\int\limits_{B>0} e^{-Sp(B)}dB.$$

This relation is reduced to expression (34) presented above:

$$\int\limits_{A>0}\Gamma_m\left(\frac{m+1}{2}\right)\left(\det\left(A\right)\right)^{-\frac{m+1}{2}}e^{-Sp(TA)}f\left(A\right)dA=$$

$$=\Gamma_m\left(\frac{m+1}{2}\right)\int\limits_{A>0}\left(\det\left(A\right)\right)^{-\frac{m+1}{2}}e^{-Sp(TA)}f\left(A\right)dA=$$

$$=\Gamma_m\left(\frac{m+1}{2}\right)\mathcal{L}_T\left\{\left(\det\left(A\right)\right)^{\frac{m+1}{2}}f\left(A\right)\right\}.$$

The theorem is proved. □

References

[1] Pierre-Simon Marquis de Laplace, On generating functions, Analytical Probability Theory, 2nd ed., Paris, 1814, chap. I sect. 2–20.

[2] G. Doetsch, Theory and Application of the Laplace Transform, Berlin: Julius Springer, 1937.

[3] L.G. Smyshlyayeva, Laplace transforms of functions of several variables, Leningrad, 1981 (in Russian).

[4] B.Baeumer, On the inversion of the convolution and Laplace transform, *Transactions of the American Mathematical Society*, **355**(2002), no. 3, 1201–1212.
DOI: 10.1090/S0002-9947-02-03174-4

[5] B.F.Marcel, Laplace Transforms: Theory, Problems, and Solutions, Arkansas Tech University, 2013.

[6] Yu.V.Sidorov, M.V.Fedoryuk, M.I.Shabunin, Lectures on the Theory of Functions of a Complex Variable, Moscow, 1982 (in Russian).

[7] P.I. Romanovsky, Fourier Series. Field Theory. Analytic and Special Functions. Laplace Transform, Moscow, 1961 (in Russian).

[8] A. Aghili, H. Zeinali, New trends in Laplace type integral transforms with applications, *Boletim da Sociedade Paranaense de Matematica*, **35**(2017), no. 1, 173–193. DOI: 10.5269/bspm.v35i1.28645

[9] A.K.Gupta, D.K.Nagar, Matrix variate distributions, Monographs and Surveys in Pure and Applied Mathematics, Chapman & Hall/CRC, 2000.

[10] R.M.Joshi, J.M.C.Joshi, Generalized Laplace transform with matrix variables, *Inter. J. Math. Math. Sci.*, **10**(1987), no. 3, 503–512. DOI: 10.1155/S0161171287000590

[11] J.Sastre, E.Defez, L.Jodar, Application of Laguerre matrix polynomials to the numerical inversion of Laplace transforms of matrix functions, *Appl. Math. Lett.*, **24**(2011), 1527–1532. DOI: 10.1016/j.aml.2011.03.039

[12] D.Rani, V.Mishra, C.Cattani, Numerical inversion of Laplace transform based on Bernstein operational matrix, *Math. Methods Appl. Sci.*, (2018), 1–13. DOI: 10.1002/mma.5188

[13] O.E.Yaremko, K.R.Zababurin, Matrix Laplace transform, *Bol. Soc. Mat. Mex.*, **29**(2023), article number 86, 1–21. DOI: 10.1007/s40590-023-00563-7

[14] Sh.Sh.Rajabov, Basic properties of matrix original and matrix image functions, *Scientific Bulletin NamSU.*, (2023) no. 6, 22–29 (in Uzbek).

[15] A.Abdukarimov, U.Rakhmonov, Sh.Rajabov, I.Khaldybaeva, B.Kuralov, Quasi-static problems in the mechanics of hereditarily deformable solids under random loads, ,AIP Conf. Proc., 3244, 060011 (2024). DOI: 10.1063/5.0241526

[16] U.Rakhmonov, A.Abdukarimov, J.Abdullaev, Sh.Rajabov, Siegel domains and Cartan-Siegel homogeneous domains: Siegel disk, AIP Conf. Proc., 3256, 040019 (2025). DOI: 10.1063/5.0267027

[17] U.Rakhmonov, A.Abdukarimov, Sh.Rajabov, Calculation of inherently deformable pipelines lying on a solid viscoelastic base with random characteristics, AIP Conf. Proceedings, 2612, 030016 (2023). DOI: 10.1063/5.0117526

[18] Sh.Sh.Rajabov, The concept of the convolution for functions with symmetric matrix arguments, its main properties, and the analogue of Duhamel's formula, ACTA NUUz no. 2.1.1, 2024. 166–172 (in Uzbek).

[19] E.Cartan, Sur les domaines bornes homogenes de lespace de n variables complexes, *Abh. Math. Sern. Univ. Hamburg*, **11**(1935), 116–162. DOI: 10.1007/BF02940719

[20] Hua Loo-Keng, Harmonic analysis of functions of several complex variables in classical domains, Inostr. Lit., Moscow, 1959 (in Russian).

[21] G.Khudayberganov, A.M Kytmanov, B.A.Shaimkulov, Complex analysis in matrix domains. Monograph, Krasnoyarsk: Siberian Federal University, 2017 (in Russian).

[22] A.M.Mathai, P.Giorgio, Some Properties of Matrix-Variate Laplace Transforms and Matrix-Variate Whittaker Functions *Linear Algebra and its applications. North-Holland*, **2.53**(1997), 209–226.

[23] C.Herz, Bessel Functions of Matrix Argument, *Annals of Mathematics*, 61 (1955), 474–523. DOI: 10.2307/1969810

[24] I.A.Shilin, J Choi, On Some Formulas for Single and Double Integral Transforms Related to the Group SO(2, 2), *Symmetry*, **16**(2024), 1102. DOI: 10.3390/sym16091102

[25] A.I.Nizhnikov, O.E.Yaremko, N.N.Yaremko, Generalized Laplace Transform Based on the Differentiation Operator With Piecewise Constant Coefficients, *Chebyshevskii sbornik*, **22**(2021), no. 5, 174–186. (in Russian). DOI: 10.22405/2226-8383-2021-22-5-172-184

[26] G.Kh.Khudayberganov, J.Sh.Abdullayev, U.S.Rakhmonov, Functional Properties of the Bergman Kernel in the Space $\mathbb{C}^n[m \times m]$, *Lobachevskii Journal of Mathematics*, **46**(2025), 1322-1335. DOI: 10.1134/S1995080225605247

[27] G.Khudayberganov, U.S.Rakhmonov, Z.Q.Matyakubov, Integral formulas for some matrix domains, Contemporary Mathematics, AMS, Volume 662, (2016), 89-95. DOI: 10.1090/conm/662/13318

[28] G.Khudayberganov, U.S.Rakhmonov, The Bergman and Cauchy-Szego kernels for matrix ball of the second type, *Journal of Siberian Federal University. Mathematics and Physics*, **7**(2014), no. 3, 305-310.

[29] U.S.Rakhmonov, J.Sh.Abdullayev, On properties of the second type matrix ball $B^{(2)}_{m,n}$ from space $\mathbb{C}^n[m \times m]$, *Journal of Siberian Federal University. Mathematics and Physics*, **15**(2022), no. 3, 329–342. DOI: 10.17516/1997-1397-2022-15-3-329-342

[30] U.S.Rakhmonov, Z.Q.Matyakubov, Carleman's formula for the matrix domains of Siegel, *Chebyshevskii Sbornik*, **23**(2022), no. 4, 126–135. DOI: 10.22405/2226-8383-2022-23-4-126-135

Новые свойства преобразования Лапласа в классических областях второго типа

Гулмирза Х. Худайберганов
Национальный университет Узбекистана им. М.Улугбека
Ташкент, Узбекистан
Шохрух Ш. Ражабов
Ташкентский Государственный Транспортный Университет
Ташкент, Узбекистан

Аннотация. В данной статье получены матричные аналоги основных свойств преобразования Лапласа для классических областей второго типа. В частности, установлены теорема о голоморфности образа функции, теорема единственности оригинала, а также явная формула обратного преобразования Лапласа в матричной постановке. Кроме того, с использованием прямых матричных дифференциальных операторов доказаны новые свойства, характерные для классических областей второго типа.

Ключевые слова: классические области, Эрмитова матрица, голоморфная функция, преобразования Лапласа, матричная образ- функция, матричные дифференциальные операторы.

EDN: XEGPAK
УДК 510.64, 510.65, 510.66

Bimodal Cluster Temporal Logic: Local Filtration, Stabilization, and Decidability

Kirill A. Petrov*
Vladimir V. Rybakov†
Siberian Federal University
Krasnoyarsk, Russian Federation

Received 13.01.2026, received in revised form 23.02.2026, accepted 17.03.2026

Abstract. We study the satisfiability problem for a bimodal temporal logic interpreted on infinite cluster frames of the form $W = \bigsqcup_{i\in\mathbb{N}} C(i)$. Each cluster $C(i)$ is an arbitrary (possibly infinite) Kripke frame with a reflexive and transitive local relation, while the global relation linearly orders the clusters and represents a discrete macro-time.

We prove decidability of the logic by a two-stage reduction. First, we apply local filtration with respect to the set of subformulas that do not contain global modalities; this compresses the internal structure of each cluster while preserving truth of the local fragment. Second, we establish the existence of a stability index and the correctness of a folding procedure, which replaces an infinite sequence of clusters by a finite lasso-shaped structure without loss of truth for the input formula.

Correctness of folding is proved by induction on the temporal degree of a formula (the nesting depth of the global modality). As a consequence, we obtain a finite-model property with respect to the constructed class of filtered lassos and an effective satisfiability-checking procedure.

Keywords: bimodal logic, local filtration, stability index, temporal degree, model folding, decidability.

Citation: K.A. Petrov, V.V. Rybakov, Bimodal Cluster Temporal Logic: Local Filtration, Stabilization, and Decidability, J. Sib. Fed. Univ. Math. Phys., 2026, 19(3), 417–422. EDN: XEGPAK.

Introduction

Temporal logics are among the standard tools for formalizing and analyzing the behavior of computational systems over time [6, 7]. On the one hand, they serve as specification languages in automatic verification (model checking) [9, 11, 12], where system properties are described by formulas and then checked algorithmically on a model. On the other hand, they form an independent subject of logical research (completeness, expressiveness, decidability, finite-model properties, etc.) [1, 10].

Classical linear temporal logics (e.g. LTL [8]) are usually interpreted over linear sequences of states. However, in many applications each time moment naturally hides an *internal structure*: a set of micro-states, local configurations, or information states with their own reachability, closures, and dependency relations [3, 4]. Typical examples include:

- models with internal transitions inside a single macro-time step (micro-steps, local computations, internal actions);
- abstractions where global time discretizes phases, while many configurations are possible within a phase;

*k1ryushape7rov@yandex.ru
†vrybakov@sfu-kras.ru https://orcid.org/0000-0002-6654-9712

- compositional models where a scenario order is fixed at the top level, and component-local dynamics lives at the lower level.

In this paper we study a formal version of such a two-level picture: the global time axis is a strict linear order of clusters (macro-level), and each cluster is a Kripke frame with a reflexive and transitive relation (micro-level) [2]. The language contains two modalities: a local modality $\Box_l$, quantifying over the local relation inside the current cluster, and a global modality $\Box_g$, quantifying over future clusters.

Our main goal is the *satisfiability problem*: given a formula, does there exist a model of the described kind in which the formula is true at some world? This problem is nontrivial for two reasons [5]. First, the global axis is infinite (a sequence of clusters indexed by $\mathbb{N}$), so filtering the whole model at once does not immediately yield a finite structure. Second, the clusters themselves may be infinite, and local properties (in particular, properties of $R_l^{(i)}$) must be preserved correctly under any compression.

Our approach has two conceptually different steps.

(I) Local filtration. We isolate a sublanguage that does not see the global time: the set Φ_{loc} of local subformulas, i.e. subformulas containing no $\Box_g$. Filtration is performed *separately inside each cluster* by local types, which allows us to replace a (possibly infinite) cluster by a finite structure preserving truth of all formulas from Φ_{loc}.

(II) Global stabilization and folding. After local filtration, the infinite model becomes an infinite sequence of *finite* clusters. We then use finiteness of the set of possible types of filtered clusters (up to isomorphism) and introduce a stability index. This enables construction of a finite lasso model in which truth of *all* formulas of the original language is preserved.

Contribution.

1. We formalize a class of bimodal cluster frames with reflexive-transitive local relations and a global linear order of clusters.
2. We prove a local-filtration lemma preserving truth for all formulas in Φ_{loc}.
3. We introduce a stability index and a folding procedure and prove a truth-preservation theorem for lasso models.
4. We derive a finite reducibility criterion and, consequently, decidability of satisfiability.

Structure of the paper. Section 2 defines syntax and semantics; Section 3 introduces local filtration and proves its correctness; Section 4 describes stabilization and folding; Section 5 derives decidability algorithmically; Section 6 contains concluding remarks.

1. Syntax and semantics

The language $\mathcal{L}$ of our logic contains a countable set of propositional variables Prop, Boolean connectives, and modalities $\Box_l$ (local) and $\Box_g$ (global). We use the standard abbreviations:

$$\Diamond_l\varphi := \neg\Box_l\neg\varphi, \qquad \Diamond_g\varphi := \neg\Box_g\neg\varphi.$$

Intuitively, $\Box_l$ expresses property to be true in all locally reachable states within the current phase (cluster), and $\Box_g$ expresses properties true in all future clusters of the global time.

Definition 1.1 (Bimodal cluster frame)**.** *A frame is a triple $\mathcal{F} = \langle W, \mathcal{R}_l, R_g\rangle$, where:*

1. *$W = \bigsqcup_{i\in\mathbb{N}} C(i)$ is a disjoint union of nonempty clusters.*

2. $\mathcal{R}_l = \bigcup_i R_l^{(i)}$, *where* $R_l^{(i)} \subseteq C(i) \times C(i)$ *is reflexive and transitive.*

3. $R_g \subseteq W \times W$ *is a global strict linear order with total connections:*

$$(w, v) \in R_g \iff \exists i, j \ \big(w \in C(i),\, v \in C(j) \text{ and } i \leqslant j\big).$$

Definition 1.2 (Model)**.** *A model* $\mathfrak{M} = \langle \mathcal{F}, V \rangle$ *consists of a frame and a valuation* $V : \mathrm{Prop} \to 2^W$*. Truth is defined as usual:*

- $\mathfrak{M}, w \models p \iff w \in V(p)$ *for* $p \in \mathrm{Prop}$*;*
- *Truth values for formulas with Boolean connectives have the classical meaning;*
- $\mathfrak{M}, w \models \Box_l \varphi \iff \forall u \in C(i)\big((w, u) \in R_l^{(i)} \Rightarrow \mathfrak{M}, u \models \varphi\big)$*;*
- $\mathfrak{M}, w \models \Box_g \varphi \iff \forall v \in W\big((w, v) \in R_g \Rightarrow \mathfrak{M}, v \models \varphi\big)$.

1.1. How to read formulas

The formula $\Box_g \Diamond_l p$ means: in every future cluster there exists a locally reachable state where p holds. The formula $\Diamond_l \Box_g p$ means: there exists a locally reachable state in the current cluster from which p holds in all future clusters. In general these formulas are not equivalent: the global modality is evaluated from different points (the current world or a locally reachable one).

2. Local filtration

Let φ_0 be some formula. Let Φ_{loc} be the set of all its *local subformulas*, that is the set of all subformulas of φ_0 which do not have occurrences of the logical operation $\Box_g$. Note that Φ_{loc} is finite since the set of subformulas of φ_0 is finite.

2.1. Local types and equivalence

For each world $w \in C(i)$ define its *local type*:

$$\mathrm{tp}_i(w) = \{\psi \in \Phi_{\mathrm{loc}} \mid \mathfrak{M}, w \models \psi\}.$$

Equality of local types yields an equivalence relation on $C(i)$:

$$w \equiv_i w' \iff \mathrm{tp}_i(w) = \mathrm{tp}_i(w').$$

By filtration we mean a procedure starting from exchange each w by $\mathrm{tp}_i(w)$ in each $C(i)$. Filtration identifies all locally indistinguishable worlds within a cluster. The number of distinct local types is at most $2^{|\Phi_{\mathrm{loc}}|}$, hence after filtration each cluster contains at most $2^{|\Phi_{\mathrm{loc}}|}$ states.

After this transformation each modified cluster $C(i)$ looks a

$$\mathcal{C}_i := \{\mathrm{tp}_i(w) \mid w \in C(i)\}.$$

Now we modify our old local accessibility relation on $C(i)$ and define the binary accessibility relation $R_l^{(i),\mathrm{filt}}$ on $\mathcal{C}_i$ by:

$$(\tau, \tau') \in R_l^{(i),\mathrm{filt}} \iff \forall \psi \in \Phi_{\mathrm{loc}} \ \big(\Box_l \psi \in \tau \Rightarrow \psi \in \tau'\big).$$

The reduced (filtered) local model of cluster i is

$$\mathfrak{M}_{\mathrm{filt}}^{(i)} = \langle \mathcal{C}_i, R_l^{(i),\mathrm{filt}}, V_{\mathrm{filt}} \rangle,$$

where the valuation is defined by

$$\tau \in V_{\mathrm{filt}}(p) \iff p \in \tau, \qquad p \in \mathrm{Prop}.$$

Remark 2.1. *We will rely on two standard filtration properties:*

- *(F1)* Preservation of transitions: *if* $(w,u) \in R_l^{(i)}$, *then* $(\mathrm{tp}_i(w), \mathrm{tp}_i(u)) \in R_l^{(i),\mathrm{filt}}$.
- *(F2)* Adequacy for $\Box_l$: *if* $\Box_l\psi \in \mathrm{tp}_i(w)$ *and* $(\mathrm{tp}_i(w), \tau') \in R_l^{(i),\mathrm{filt}}$, *then* $\psi \in \tau'$.

Lemma 2.2 (Local filtration lemma)**.** *For any local formula* $\psi \in \Phi_{\mathrm{loc}}$ *and any world* $w \in C(i)$,

$$\mathfrak{M}, w \models \psi \iff \mathfrak{M}^{(i)}_{\mathrm{filt}}, \mathrm{tp}_i(w) \models \psi.$$

Proof. We proceed by structural induction on ψ.

1. Base case. Let $\psi = p \in \mathrm{Prop}$. By definition of V_{filt},

$$\mathfrak{M}^{(i)}_{\mathrm{filt}}, \mathrm{tp}_i(w) \models p \iff \mathrm{tp}_i(w) \in V_{\mathrm{filt}}(p) \iff p \in \mathrm{tp}_i(w) \iff \mathfrak{M}, w \models p.$$

2. Boolean connectives. For $\neg\alpha$ and $\alpha \wedge \beta$ the claim follows from the induction hypothesis and classical semantics.

3. Modal case $(\Box_l\chi)$**.**

($\Rightarrow$) Assume $\mathfrak{M}, w \models \Box_l\chi$. Then $\Box_l\chi \in \mathrm{tp}_i(w)$. Let $\tau = \mathrm{tp}_i(w)$ and take any τ' with $(\tau, \tau') \in R_l^{(i),\mathrm{filt}}$. By definition of $R_l^{(i),\mathrm{filt}}$, from $\Box_l\chi \in \tau$ we get $\chi \in \tau'$. Since $\tau' \in \mathcal{C}_i$, there exists $u \in C(i)$ such that $\tau' = \mathrm{tp}_i(u)$. Hence $\chi \in \mathrm{tp}_i(u)$, and by the induction hypothesis $\mathfrak{M}^{(i)}_{\mathrm{filt}}, \mathrm{tp}_i(u) \models \chi$. Therefore $\mathfrak{M}^{(i)}_{\mathrm{filt}}, \tau \models \Box_l\chi$.

($\Leftarrow$) Assume $\mathfrak{M}^{(i)}_{\mathrm{filt}}, \mathrm{tp}_i(w) \models \Box_l\chi$. Take any $u \in C(i)$ such that $(w,u) \in R_l^{(i)}$. By (F1), $(\mathrm{tp}_i(w), \mathrm{tp}_i(u)) \in R_l^{(i),\mathrm{filt}}$. Then from truth of $\Box_l\chi$ at $\mathrm{tp}_i(w)$ we get $\mathfrak{M}^{(i)}_{\mathrm{filt}}, \mathrm{tp}_i(u) \models \chi$, and by induction hypothesis $\mathfrak{M}, u \models \chi$. Since u was arbitrary among $R_l^{(i)}$-successors of w, we conclude $\mathfrak{M}, w \models \Box_l\chi$. □

2.2. What local filtration gives

Lemma 2.2 states that within each cluster we can replace the original (possibly infinite) cluster by a finite model whose states correspond to local types, without losing truth of local formulas. Importantly, this step requires no restrictions on the global part of the model: filtration is done independently for each i.

3. Temporal stabilization and folding

After local filtration, the original model can be seen as a sequence of finite cluster models $\Sigma = (\mathcal{K}_0, \mathcal{K}_1, \mathcal{K}_2, \dots)$. Since Φ_{loc} is finite, the set Ω of all possible filtered clusters (up to isomorphism) is also finite. Let $N = |\Omega|$.

3.1. Stability index

Consider the sequence Σ. Because the alphabet Ω is finite and Σ is infinite, repetitions are unavoidable.

Definition 3.1 (Stability index)**.** *A* stability index *is a number* $k \in \mathbb{N}$ *such that after position* k *the sequence* Σ *contains only those cluster types that occur infinitely often in* Σ.

By finiteness of Ω, such a k always exists. Moreover, there exists $m > k$ such that $\mathcal{K}_k \cong \mathcal{K}_m$. We choose m so that the segment $\mathcal{K}_k, \dots, \mathcal{K}_{m-1}$ contains all cluster types occurring in Σ after k.

3.2. Folding construction

We build a finite lasso model $\mathfrak{M}^{fold}$:

1. The handle: $\mathcal{K}_0, \ldots, \mathcal{K}_k$.
2. The loop: $\mathcal{K}_k, \ldots, \mathcal{K}_{m-1}$ with a back edge from $\mathcal{K}_{m-1}$ to $\mathcal{K}_k$.
3. All clusters of the original sequence between k and m are kept, and everything after m is discarded.

Theorem 3.2 (Correctness of folding)**.** *For any formula φ and any $w \in \mathcal{K}_i$ with $i < m$,*

$$\mathfrak{M}, w \models \varphi \iff \mathfrak{M}^{fold}, w \models \varphi.$$

Proof. Define the temporal degree $\deg_\tau(\varphi)$ as the maximal nesting depth of $\Box_g$. We prove the claim by induction on $d = \deg_\tau(\varphi)$.

Base ($d = 0$). Then φ contains no $\Box_g$, so its truth depends only on the local structure of the cluster, which is preserved by Lemma 2.2.

Step. Assume the statement holds for all formulas of degree $< d$. Consider $\Box_g\psi$.

($\Rightarrow$) If $\mathfrak{M}, w \models \Box_g\psi$, then ψ is true at all future worlds in $\mathfrak{M}$. Any future world v' in $\mathfrak{M}^{fold}$ has (by the loop construction) a prototype v in the future of $\mathfrak{M}$ with the same local type. Since $\deg_\tau(\psi) < d$, truth is preserved by the induction hypothesis, hence $\mathfrak{M}^{fold}, v' \models \psi$.

($\Leftarrow$) Assume $\mathfrak{M}^{fold}, w \models \Box_g\psi$ and suppose for contradiction that $\mathfrak{M}, w \not\models \Box_g\psi$. Then there exists v such that wR_gv and $\mathfrak{M}, v \models \neg\psi$. Because the loop in $\mathfrak{M}^{fold}$ contains representatives of all cluster types occurring after the stability index k, there is a world v' in $\mathfrak{M}^{fold}$ matching v by type. By the induction hypothesis (since $\deg_\tau(\neg\psi) < d$), we get $\mathfrak{M}^{fold}, v' \models \neg\psi$, contradicting $\mathfrak{M}^{fold}, w \models \Box_g\psi$. □

4. Decidability

Theorem 4.1 (Decidability)**.** *The satisfiability problem for the bimodal cluster temporal logic is decidable.*

Proof. The folding construction yields a finite lasso model of computable size that preserves truth of the input formula, hence satisfiability reduces to checking a finite class of finite models. Therefore, satisfiability is decidable. □

5. Conclusion

We proved decidability for a temporal logic over infinite cluster structures with a two-level semantics: local dynamics within clusters and a global linear time order of clusters. We showed that a combination of local filtration and folding correctly reduces satisfiability to checking finite models.

This work is supported by the Krasnoyarsk Mathematical Center and financed by the Ministry of Science and Higher Education of the Russian Federation (Agreement no. 075-02-2026-1314).

References

[1] A.Chagrov, M.Zakharyaschev, Modal Logic, Oxford University Press, 1997.

[2] V.V.Rybakov, Admissible Logical Inference Rules, Elsevier, 1997.

[3] R.Fagin, J.Y.Halpern, Y.Moses, M.Y.Vardi, Reasoning About Knowledge, MIT Press, 1995.

[4] S.N.Artemov, The Logic of Justification, *The Review of Symbolic Logic*,**1**(2008), no. 4, 477–513.

[5] V.V.Rybakov, V.R.Kiyatkin, K.V.Grekovich, Problems of decidability and semantic modeling for temporal-modal logics, https://scholar.google.com/citations?view_op=view_citation&hl=ru&user=aXDTJssAAAAJ&sortby=pubdate&citation_for_view=aXDTJssAAAAJ:NJ774b8OgUMC

[6] D.M.Gabbay, I.Hodkinson, M.Reynolds, Temporal Logic: Mathematical Foundations and Computational Aspects, Vol. 1, Oxford University Press, 1994.

[7] R.Goldblatt, Logics of Time and Computation, CSLI Lecture Notes, No. 7, Stanford, 1992.

[8] A.Pnueli, *The Temporal Logic of Programs*. Proceedings of the 18th IEEE Symposium on Foundations of Computer Science (FOCS), 1977, 46–57.

[9] M.Y.Vardi, P.Wolper, An Automata-Theoretic Approach to Automatic Program Verification (Preliminary Report), Proceedings of the 1st Annual IEEE Symposium on Logic in Computer Science (LICS), 1986, 332–344.

[10] P.Blackburn, M. de Rijke, Y.Venema, Modal Logic, Cambridge University Press, 2001.

[11] E.M.Clarke, O.Grumberg, D.A.Peled, Model Checking, MIT Press, 1999.

[12] P.Wolper, Constructing Automata from Temporal Logic Formulas, 2001, https://orbi.uliege.be/bitstream/2268/18956/1/Wol00-fmpa.pdf

Бимодальная кластерная темпоральная логика: локальная фильтрация, стабилизация и разрешимость

Кирилл А. Петров
Владимир В. Рыбаков
Сибирский федеральный университет
Красноярск, Российская Федерация

Аннотация. В работе исследуется задача выполнимости для бимодальной темпоральной логики на бесконечных кластерных шкалах, где глобальное отношение упорядочивает кластеры, а локальное отношение внутри каждого кластера является рефлексивным и транзитивным. Доказывается разрешимость с помощью двухэтапного сведения: локальной фильтрации по подформулам без глобальных модальностей и последующей стабилизации типов кластеров с построением конечной «лассо»-модели. Корректность свёртки доказывается индукцией по темпоральной степени формулы (глубине вложенности глобальной модальности), что даёт конечномодельное свойство и эффективную процедуру проверки выполнимости.

Ключевые слова: бимодальная логика, локальная фильтрация, индекс стабильности, темпоральная степень, свёртка модели, разрешимость.