Generalized Spectral-Analytical Method and Its Applications in Image Analysis and Pattern Recognition Problems

S. A. Makhortykh[a],*, L. I. Kulikova[a], A. N. Pankratov[a], and R. K. Tetuev[a]

[a] *Institute of Mathematical Problems of Biology, Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Pushchino, Moscow oblast, 142290 Russia*

**e-mail: makh@impb.ru*

Abstract—The generalized spectral-analytical method as a new approach to the processing of information arrays is stated. Some theoretical foundations of this method and its applications in different experimental data analysis problems are given. The method is based on the adaptive expansion of initial arrays in the functional bases belonging to the classical algebraic systems of polynomials and functions of continuous and discrete arguments (Jacobi, Chebyshev, Lagrange, Laguerre, Kravchuk, Charlier, and other polynomials). This approach combines analytical and digital data-processing procedures, thus providing a basis for the universal combined technology for the processing of information arrays. An appreciable part of this review is devoted to video data analysis and pattern-recognition problems. In addition, some relevant applications of this method in biomedical and bioinformation data analysis, recognition, classification, and diagnosis problems are described.

Keywords: generalized spectral-analytical method, biomedical data analysis, bioinformatics, genome analysis, search for repeats, magnetic encephalography, recognition and analysis of brain activity pathologies

DOI: 10.1134/S1054661819040102

INTRODUCTION

The trends toward larger volume of information arrays and stricter requirements as to precision and completeness of data processing under time constraints cause the need for information systems with ever-greater computational capacity. The continuous improvement of computer performance speed leads to increased information processing cost and does not always provide the required data-processing conditions. The problem of creating new methods, which would not require an obligatory increase in the capacity of computational complexes, but would still provide the specified precision and speed of processing, remains topical.

The submitted review describes the results of search for possible ways to combine digital calculations and analytical transformations for the purpose of increasing the precision and speed of computations. These results preserve the demonstrativeness of analytical methods and the possibility to have an analytical representation of the sought parameters and estimates at every step of calculations.

The successful implementation of a combined data-processing method directly depends on the form used for the description of the initial digital arrays. Analysis [1–3] has demonstrated that the most completely formulated conditions correspond to the method based on the approximation of data with truncated orthogonal series with the use of classic orthogonal polynomials and functions of continuous and discrete arguments [2]. The application of approximation results in various analytical transforms, making the classical orthogonal bases a promising tool for the operative analytical processing of digital data in on-line and off-line systems (on-line and off-line analytical processing, OLAP) and the creation of data-mining systems.

The theory of classical orthogonal bases is an extension of the theory of Fourier series to algebraic polynomials. Their distinctive feature is that most formulas specifying certain bases have parameters whose change may appreciably vary the properties of the orthogonal polynomials and weight functions composing a certain orthogonal basis. The latter circumstance is especially important in approximation-optimization problems, when the specified precision must be provided by the shortest truncated orthogonal series. The application of special adaptive procedures provides the solution of this problem on a computer in an automatic mode.

Spherical functions are an important class of special functions, which are closely related to the classic orthogonal polynomials. They appear during the solution of a wide variety of problems, e.g., when the Laplace's equation is solved in spherical coordinates. Since the continuous solutions of the Laplace's equation are called harmonic functions, the spherical functions are also called spherical harmonics. The need for

Received January 23, 2019; revised May 13, 2019; accepted July 15, 2019

their use in the problems considered here is associated first of all with the approximation of signals and functions on a sphere. A number of applications leading to this result are considered below.

Hence, the adaptive approximation of data in the bases of classical orthogonal polynomials and other special functions underlies the below-proposed technology for the description, processing, and analysis of information arrays. The set of spectral features adjusted to a class of signals, objects, or systems leads to the efficient procedures of recognition and classification in a number of the below-considered applications. The proposed generalized spectral-analytical method (GSAM) for the processing of information arrays provides the possibility to rationally combine the advantages of both digital calculations and analytical derivations. Its computational procedures can be efficiently implemented on both sequential and parallel computers.

The problem of the diagnosis of biomedical systems usually entails difficulties typical for this field of studies. Among them are the initial parametric complexity of the system, with the resulting incorrect formulation of the inverse problems, and the presence of noise, which decreases the reliability and precision of the parametric identification of the system.

The primary processing of data incorporates the selection of areas with a digital biomagnetic signal characterizing different types of activity. To classify the type of a brain biomagnetic activity signal, the transition to the spectral spherical harmonic representation of the spatial field picture is performed. The use of signal expansions of this kind for practical system diagnosis problems is based on the appreciable sensitivity of the expansion harmonics to the values of the estimated parameters.

An interesting GSAM application field is computer vision, more precisely, contour-analysis-based pattern recognition. The silhouettes of objects are the most informative part of the visible world, and it is believed that our primitive ancestors had no color vision, and shapes were the only source of visual information [4]. Nevertheless, bitmap analyses and similarity tests will never be "yes–no" questions due to the noise effects: objects of the same shape will be similar, but will never be identical in each pixel. For this reason, the similarity of bitmap images is estimated as the amount of pixels in which the images coincide with each other [5]. Many applications are not suitable for brightness-based comparison, when we admit more complicated deformations of contours, such as inflection. Some researchers begin to consider silhouettes as contours, and this means that they are merely two-dimensional closed lines, so we obtain naturally sorted boundary points, which appreciably simplifies the comparison of a pair of rows instead of inner pixels. For example, if there exist square objects to be compared, and their dimensions are 100×100, we must compare 10000 pairs of pixels when following the brightness-based approach, whereas the contour-based method requires only 400 points, or 25 times less data for calculation. We can also accelerate the comparison much more greatly with the use of data decimation [6]: retaining only every tenth boundary point, we obtain merely 40 points and thereby decrease the speed by 10 times. When doing this, we must understand that we do not compare the silhouettes of objects any more, but compare certain polygons with 40 apices and hope that these polygons are rather good for the description of the initial shapes. The resulting boundary curve has been de facto a representation standard for the contour-based approach for decades, but we use here the other boundary data compression method to demonstrate how to estimate the similarity of shapes without the need for reverse data decompression. The feasibility of calculations in the frequency region is one of the most important advantages of GSAM. The authors [7] have demonstrated the technique of calculations in the spectral region, when there is a need for the derivative or integral of spectral functions or algebraic operations over them.

Another GSAM application field is bioinformatics. A variety of algorithms and software for the computer-aided estimation of DNA fragments and derivatives (proteins, RNA) have been developed to date [8–11]. However, the capabilities of computer-aided genetic data analysis have lagged behind the quickly developing experimental facilities in contemporary biology over the last 20 years: most algorithms for the processing of genetic sequences are based on several basic text information-processing principles, such as Hamming or Levenshtein edit distances. The temporal complexity of the algorithms is appreciably non-linear, and the main factor of deceleration in the comparison of similar genetic data are point mutations (such as the replacement, removals, or insertions of letters), whose correction increases the time of analysis. For large compared fragments, it is natural to presume a greater number of mutations and, as a consequence, to observe an abrupt decrease in the efficiency of the algorithms on long sequences (of more than 10000).

The suggestion to use the spectral approach to the problem of searching for repeats in genomes was first made in the works [12, 13], in which the ideas to use the methods of continuous mathematics for the analysis of character sequences were put forward. It was proposed to use the GC-content curves characterizing the force of binding in the double DNA helix as a functional analogue of the genome sequence.

The following stage in the development of this method was the construction of a similarity dot matrix on the basis of the decision-making rule for the recognition of inexact repeats [14]. The estimation of periodicity in a GC-content curve was proposed for searching for extended tandem repeats [15]. As an application of the developed methods, they were used

to reveal some extensive repeats in the genomes [16, 17], thus posing the problem of automating the search for repeated sequences of this type.

In the course of further studies, the method gained a number of improvements [18]: the recognition of repeats was made more stable via the introduction of an additional GA-content curve, making the information description of sequences more complete; an improved decision-making rule invariant to the selection of a scale was derived; the approximation conditions such that the tandem repeats were mapped on the dot matrix in the form of square templates were determined, thus providing the possibility to automate the recognition of repeats over a sample and reveal a new repeat in the genome [19]; and the algorithm of searching for the inverted repeats in the space of expansion coefficients was constructed to present the first results of comparison over the entire genome.

The theoretical foundations and algorithmic implementation of the spectral-analytical method for the recognition of repeats in character sequences were further proposed in this work. The theoretical substantiation is based on the theorem about the equivalent representation of a character sequence with the vector of continuous characteristic functions [20]. The comparison of truncated characteristic functions is performed using the standard metrics in the Euclidian space of coefficients for the Fourier series expansion of orthogonal polynomials. An essential specific feature of this approach is the ability to compare the repeats in different scales. Another important feature is the possibility of efficient data parallelization. When developing the algorithm, we preferred the scheme of calculations with a minimum number of memory references, thus implying repeated calculations and on-demand estimations. In this paradigm, the algorithm for the calculation of orthogonal polynomial expansion coefficients with the use of recurrence equations was proposed. It has been shown that the algorithm for the calculation of orthogonal polynomial expansion coefficients can be efficiently vectorized by means of calculations with a fixed vector length. Parallelization and vectorization were implemented using the standard Open MP extension of the C/C++ language. The developed method can be efficiently scaled depending on the problem parameters and the number of processor cores in shared memory systems. As a result, the SBARS software [21] for searching for inexact repeats of different types (direct, inverted, or tandem) in genomes was developed and compared with other tools [22–25]. Moreover, a unique web service for the global alignment of extensive sequences was also created [26, 27].

Another problem of bioinformatics is the study of structural motifs in protein molecules, being very topical and important for the understanding of the regularities in the packing of a polypeptide chain into spatial structures and the knowledge of all the possible conformations of the studied motifs in the polypeptide chain as a whole. The acquired knowledge is necessary for the establishment of structural organization regularities and may be very useful when solving the problems of the automatic recognition and prediction of different structural motifs in proteins. The comprehensive study of the supersecondary structures of protein molecules is caused by the fact that their structural motifs have unique spatial polypeptide-chain packings, which can play a crucial part in the protein-folding process.

There also exist known structural motifs, which are composed of two and more secondary structure elements and have unique spatial polypeptide chain packings, such as α-α corners, α-α hairpins, L- and V-shaped structures, etc. [28–31]. These motifs are formed by two α helices, which are arranged in the polypeptide chain one after another, linked to each other by means of constrictions, and represent compact spatial structures. It is known that α helices in proteins are closely packed. The most compact packing of two α helices is attained in the case of antiparallel, perpendicular, and so-called oblique slanted orientation between the helices [32–35], and the mentioned supersecondary structures provide some examples of such a packing.

METHODS

Classical Orthogonal Systems

We shall further be interested in the algebraic orthogonal systems (bases) of polynomials of continuous and discrete arguments and some other related functional systems. First of all, the bases of one variable will be considered, but some problems require the use of functional systems depending on a greater number of variables.

The considered functional systems are well studied in the mathematical literature and satisfy a number of important requirements regarding the construction of an analytical description of general form for information arrays. The orthogonal bases of a continuous argument (Table 1) represent three groups of bases related by the formulas and properties generating them. The first group incorporates orthogonal bases, which are specified by the general Jacobi formula and defined on the interval $[-1, 1]$. This formula has the parameters α and β, whose certain values in different combinations lead to both the known orthogonal systems and the other possible systems. Thus, if the parameters $\alpha = \beta = 0$, the spherical orthogonal Legendre polynomials can be derived from the formula specifying the Jacobi basis up to a constant.

At $\alpha = \beta = \pm 0.5$, the general formula gives the expression determining the orthogonal Chebyshev polynomials of the first ($\alpha = \beta = -0.5$) and second ($\alpha = \beta = 0.5$) order, respectively. At ($\alpha = \beta = \sigma - 0.5$), ($\sigma > -0.5$), the ultraspherical

Table 1. Orthogonal classic bases of a continuous argument

Polynomials	General equation	Weight $\rho(x)$	Interval of existence
Jacobi or hypergeometric	$P_n^{\alpha\beta}(x) = \frac{1}{2n}\sum_{k=0}^{n}(x-1)^k(x+1)^{n-k} \times C_n^k \frac{\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}{\Gamma(\alpha+k+1)\Gamma(\beta+n-k+1)}$	$(1-x)^{\alpha}(1+x)^{\beta}$ $\alpha > -1, \beta > -1$	$(-1,1)$
Gegenbauer or ultraspherical	$C_n^{\sigma}(x) = \sum_{k=0}^{n}\frac{(-1)^k\Gamma(\alpha+n-k)(2x)^{n-2k}}{\Gamma(k+1)\Gamma(n-2k+1)}$	$(1-x^2)^{\sigma-0.5}$ $(\alpha=\beta=\sigma-0.5)$	$(-1,1)$
Chebyshev first-kind	$T_n(x) = \frac{2^n n!}{(2n)!}\sqrt{x^2-1}\frac{d^n}{dx^n}[(x^2-1)^{n-2k}]$	$(1-x^2)^{-0.5}$ $\alpha=\beta=-0.5$	$(-1,1)$
Chebyshev second-kind	$U_n(x) = \frac{2^n(n+1)!\frac{d^n}{dx^n}[(x^2-1)^{n+0.5}]}{(2n+1)!\sqrt{x^2-1}}$	$(1-x^2)^{0.5}$ $\alpha=\beta=0.5$	$(-1,1)$
Legendre or spherical	$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}[(x^2-1)^n]$	1 $\alpha=\beta=0$	$(-1,1)$
Sonin-Laguerre	$L_n^{\alpha}(x) = \sum_{k=0}^{n} C_{n+\alpha}^{n-k}\frac{(-x)^k}{k!}$	$x^{\alpha}e^{-x}$	$(0,\infty)$
Laguerre	$L_n(x) = \sum_{k=0}^{n} C_n^{n-k}\frac{(-x)^k}{k!}$	e^{-x}	$(0,\infty)$
Hermite	$H_n(x) = (-1)^n e^{x^2}\frac{d^n}{dx^n}(e^{-x^2})$	e^{-x^2}	$(-\infty,\infty)$

orthogonal polynomials (Gegenbauer polynomials) are obtained. When the values of α and β are randomly selected (their admissible range is $(-1 < \beta < \infty)$, $(-1 < \alpha < \infty)$) in any combinations of α and β, it is possible to obtain a wide variety of orthogonal bases, which belong to the classical ones and have certain common properties, but different formulas. The second group is formed by the orthogonal Sonin–Laguerre bases (generalized Laguerre polynomials). They are defined on the interval $[0,\infty]$ and have the parameter α in their formulas. At $\alpha = 0$, the formula defines the well-known Laguerre polynomials. Specifying α within a range $(-1 < \alpha < \infty)$, it is possible to obtain a set of orthogonal Sonin–Laguerre bases, which differ from each other in the form of the polynomials. The family of orthogonal bases belonging to the Sonin–Laguerre group is frequently modified by introducing the scaling coefficient m. Finally, the last group is formed by the orthogonal Hermite basis specified on the entire number axis $(-\infty < x < \infty)$. This basis is usually applied in statistical studies. The formula specifying this basis does not contain any parameters. For this reason, this group is represented by the only orthogonal Hermite basis.

The common properties of the analytical orthogonal polynomials of a continuous argument from the group of classical ones can be briefly reduced to the following [2].

(1) Any three sequential orthogonal polynomials from the same basis are related to each other by a linear equation [36]; i.e., there exists a recurrence formula, which provides the possibility to use two known polynomials to uniquely determine the third polynomial.

(2) The functions $\varphi_n(x)$ belonging to the classical orthogonal systems satisfy the hypergeometric differential equation like $A(x)y'' + B(x)y' + \lambda_n y = 0$, where $A(x)$ and $B(x)$ are independent of n, and λ_n is independent of x.

(3) The functions $\{\varphi_n'(x)\}$ also form an orthogonal system of polynomials [4].

(4) The generalized Rodrigues formula $\varphi_n(x) = \frac{1}{K_n\rho(x)}\frac{d^n[\rho(x)X_n]}{dx^n}$, where K_n is a constant, and X_n is a polynomial, whose coefficients are independent of n, occurs. The inverse proposition is also true. Any of the last three properties characterizes classical orthogonal polynomials. In other words, any system of orthogonal polynomials which has one of these three properties can be reduced to the family of classical ones.

5) For all orthogonal polynomials from the family of classical ones, the classical weight function $\rho(x)$ is non-negative and integrable on the interval $[a,b]$.

Spherical Functions

Spherical functions are an important class of special functions, which are closely related to the classical orthogonal polynomials. They appear in the solution of a broad spectrum of problems, e.g., when the Laplace's equation is solved in spherical coordinates. Since the continuous solutions of the Laplace's equation are called harmonic functions, the spherical functions are also called spherical harmonics. The process of finding the solution of the Laplace's equation $\Delta u = 0$ in the spherical coordinates r,θ,φ alongside the explicit expression of spherical functions and the properties of the latter are detailed in the works [37–40].

The spherical functions have the general form

$$Y_{lm}(\theta,\varphi) = \frac{1}{\sqrt{2\pi}} e^{im\varphi}\Theta_{lm}(\cos\theta),$$

where the function Θ_{lm} is defined as

$$\Theta_{lm}(x) = \sqrt{\frac{2l+1}{2}\cdot\frac{(l-m)!}{(l+m)!}}p_l^m(x),$$

$$\text{where} \quad p_l^m(x) = (1-x^2)^{m/2}\frac{d^m p_l(x)}{dx^m}.$$

The functions $p_l(x)$ are Legendre functions or polynomials and are also called zonal harmonics. The functions $p_l^m(x)$ are called associated Legendre functions. Since $e^{i\varphi} = \cos\varphi + i\sin\varphi$, the expression for $Y_{lm}(\theta,\varphi)$ will be rewritten as

$$Y_{lm}(\theta,\varphi) = \sqrt{\frac{2l+1}{4\pi}\cdot\frac{(l-m)!}{(l+m)!}}[p_l^m(\cos\theta)\cos m\varphi + ip_l^m(\cos\theta)\sin m\varphi],$$

where the functions $p_n^m(\cos m\varphi, \sin m\varphi)$ are named tesseral harmonics, seemingly, after one of the types of dice games known in Ancient Rome. At m = l, tesseral harmonics are called sectoral harmonics.

Some relations of orthogonality also take place. Each zonal harmonic p_l is orthogonal to the polynomial depending on the argument $x = \cos\theta$ raised to a lower power. The same is also true for the surface harmonics raised to identical powers; any pair of $p_l^m\cos m\varphi$, $p_l^m\sin m\varphi$, $p_l^s\cos s\varphi$, $p_l^s\sin s\varphi$ is orthogonal, and the integral of their product over φ becomes zero except $m = s$, when the integral of a squared harmonic is taken. The linear independence of the standard (normalized) harmonics $Y_{lm}(\theta,\varphi)$ can be easily proven [38].

As for the calculation of the associated polynomials p_n^k, it is possible to use the explicit formula [38–40], but the very high values of n necessitate operations with very large numbers, thus leading to precision loss, arithmetic overflow, and other inconveniences. For this reason, it makes sense to use the recurrence relations for the associated Legendre polynomials [41, 42]

$$(l-m)p_l^m = x(2l-1)p_{l-1}^m - (l+m-1)p_{l-2}^m,$$

$$p_m^m = (-1)^m(2m-1)!!(1-x^2)^{m/2},$$

$$p_{m+1}^m = x(2m+1)p_m^m.$$

The system of spherical harmonics is complete in the space of quadratically integrable functions $f(\theta,\phi)$ (hereinafter, surface harmonics are understood to mean the real and imaginary parts of the function $Y_{lm}(\theta,\varphi)$).

Hence, if a certain function $f(\theta,\varphi)$ is quadratically integrable on a sphere, it can be expanded in a series of spherical harmonics as

$$f(\theta,\varphi) = \sum_{n=0}^{\infty}\sum_{s=0}^{n}(a_{ns}p_n^s\cos s\varphi + b_{ns}p_n^s\sin s\varphi),$$

where a_{ns}, b_{ns} are the expansion coefficients to be determined.

Taking into account the approximate expression, we obtain

$$f(\theta,\phi) \approx \sum_{n=0}^{N}\sum_{k=0}^{n}(a_{nk}p_n^k\cos k\phi + b_{nk}p_n^k\sin k\phi).$$

Taking into consideration the property of orthogonality inherent in the spherical harmonics, the expansion coefficients a_{nk}, b_{nk} can be calculated by the formulas of integrating the product of a function and a spherical harmonic on a sphere with consideration for the orthogonality and normalization conditions.

Generalized Spectral-Analytical Method

The key feature of the proposed method consists in that the process of signal processing represents two interrelated, but largely independent stages. The essence of the first stage is to find the compact analytical description of a studied signal with a required precision in the process of calculations on a computer. At the second stage, the found analytical descriptions are used to derive the analytical formulas necessary to calculate the estimates of the sought parameters and characteristics in the general form. To simplify the process of analytical derivations, it is necessary to strive to a known and constant structure for the analytical descriptions of signals of different nature. The resulting analytical expressions are either input into a computer before the process of signal processing is started in a programmable fashion or are programmed

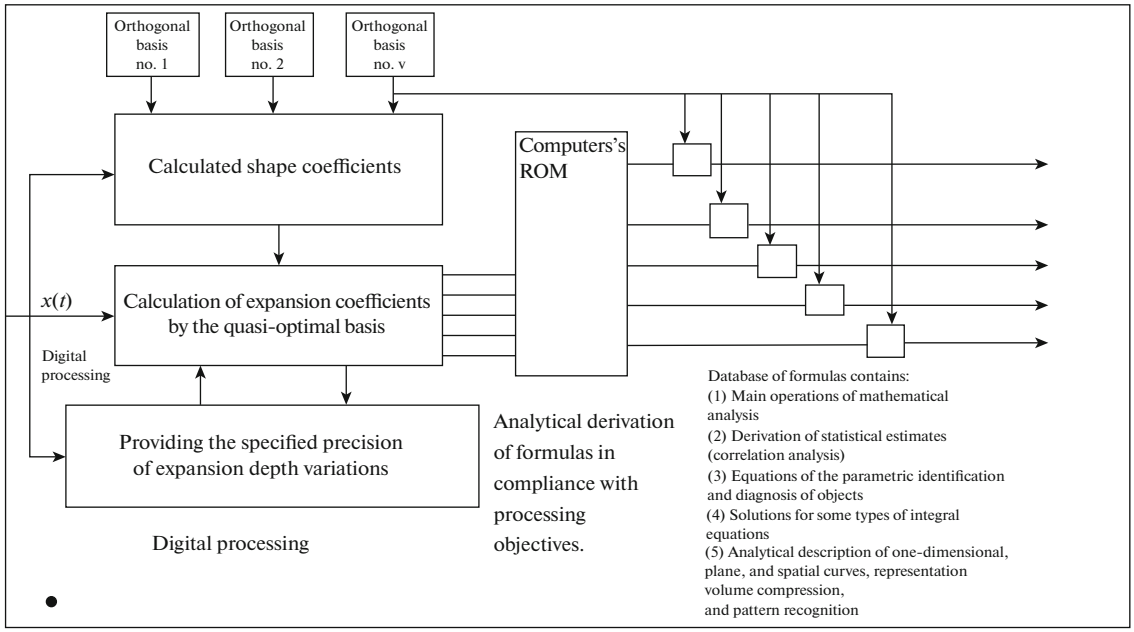


Fig. 1. Functional scheme of the generalized spectral-analytical method.

in the form of a hardware special calculator (the creation of a similar device provides the possibility to appreciably optimize and accelerate such calculations at the signal spectrum calculation stage [43]).

The high efficiency of the described computational scheme will be determined by the fulfillment of the following principal provisions. The process of the automatic description of input signals must be completely automated and have adaptive procedures, which would provide the required precision with the expressions of minimum complexity. The representation of the studied signals in the form of truncated orthogonal series is characterized by the fact that the structure of such descriptions always remains unchanged, and the meaningful information about these signals is contained in the expansion coefficients, which are in turn linearly independent functionals. The mentioned circumstances provide the development of a procedure for the complete processing of signals in the space of expansion coefficients in a "compressed" form, thus promoting an appreciable increase in the precision of the found estimates and a decrease in the volume of computational operations.

The use of expansion coefficients as initial values for the solution of inverse ill-posed problems, in particular parametric identification and diagnosis, increases the precision of estimates for the sought parameters and does not require any additional regularization. Moreover, the stability of the solutions is retained under the conditions of interferences overlapping the useful signal due to the operation of integration in the calculation of the expansion coefficients.

One of the important features of this method is the possibility to choose a data description method via the adequate selection of basis functions from the above-described orthogonal polynomials and functions. This problem is solved by implementing the selection algorithm based on the tool of signal shape coefficients or shape vectors, which are described in [43, 44]. The general GSAM scheme is shown in Fig. 1. The layout of the distributed system for the digital processing of signals and images is illustrated in Fig. 2.

APPLICATIONS

Analysis, Processing, and Classification of Magnetic Encephalography (MEG) Data

The primary processing of data includes the selection of areas with a digital biomagnetic signal characterizing different types of activity. To classify the type of a brain biomagnetic activity signal, the transition to the spectral spherical harmonic representation of the spatial field picture is performed. The use of signal expansions of this kind for practical system diagnosis problems is based on the appreciable sensitivity of expansion harmonics to the values of the estimated parameters.

It is known [45] that the typical case encountered in the MEG records for patients suffering from Parkinson's disease is the spontaneous switches between pathological and normal activity seemingly due to the activation and deactivation of brain excitation areas. Externally, this is exhibited as sporadic attacks of tremor or hallucinations. For this reason, the main problem is to classify the type of activity and select the

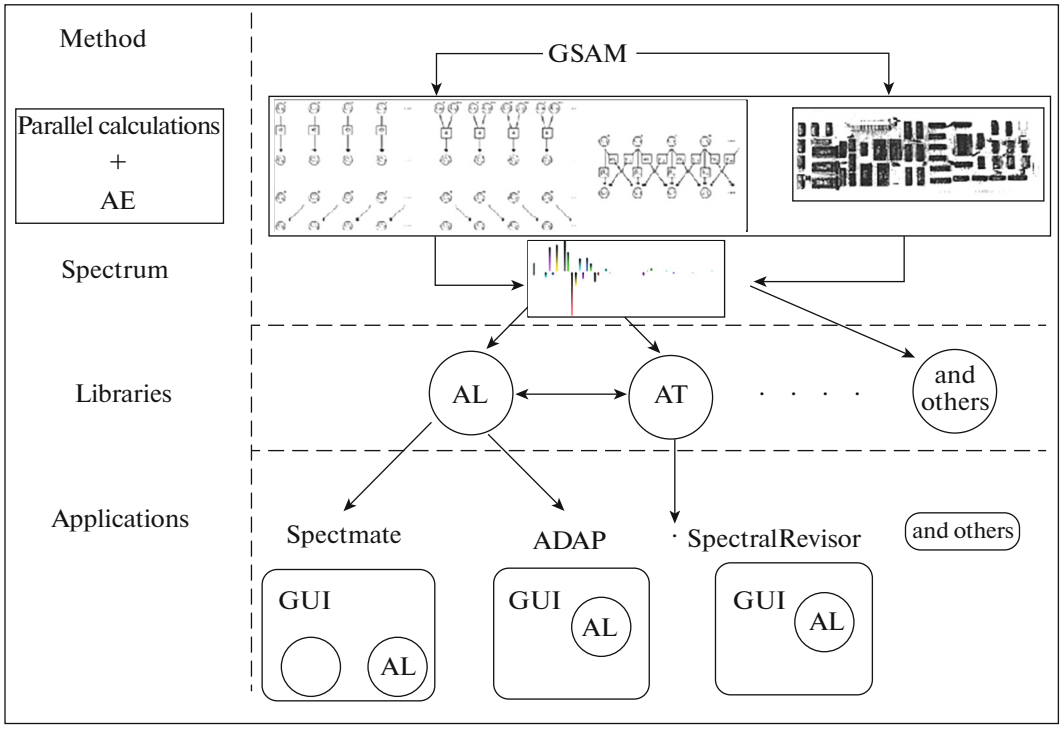


Fig. 2. Layout of the distributed system for the digital processing of signals and images: numerical block + AE (arithmetic expander), which implement the transition to the spectral representation of data. GUI is the graphical user interface, AL is the approximation library, AT is the approximation toolbox. Software implementations ADAP, Spectmate®, and SpectralRevisor®.

time moments to solve the problem of localizing the sources corresponding to the relevant type of activity. Time moments for the solution of the inverse problem are selected with the use of spectral classification results.

Here, the activity type classification approach using the spectral representation of a signal with orthogonal functional expansions [46, 47] is developed. In this case, the attribute space elements are the vectors of Fourier coefficients. The signal classification criterion is the source corresponding to this type of signal. To exclude the effect of source amplitude changes on the picture of the magnetic induction (MI) distribution over the brain case surface, the records were normalized by scaling (all the MI values in the channels of the records were reduced to the average absolute value over the channels):

$$B^j_{norm}(t_i) = 148 \cdot \frac{B^j(t_i)}{\sum\limits_{k=0}^{147} \left| B^k(t_i) \right|},$$

where $B^j(t_i)$ is the initial MI values in the jth channel at the time moment t_i.

The method proposed for the classification of a signal activity type is the following:

(1) The vectorization of MEG data is performed. Their representations are obtained in the spherical coordinate system in the form of a series in orthonormalized spherical functions $Y_l^m(\theta,\phi) = \frac{1}{\sqrt{2\pi}} P_l^{|m|}(\sin\theta) e^{im\phi}$. The existence of a simple analytical relation between the expansion coefficients in the case when the transform function SO(2) is applied to the argument provides the possibility to construct the fast procedure for the enumeration of functions in a given class. Correspondingly, the initial function $f(\theta,\varphi)$ is defined by the formula $f(\theta,\phi) = \sum_{l=0}^{N} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta,\phi)$, where the expansion coefficients are $a_{lm} = \int_0^{2\pi} \int_0^{\pi} f(\theta,\phi) Y(\theta,\phi) \sin\theta d\theta d\phi$.

(2) The selection of the most informative harmonics. The requirement of a maximum value for the ratio of the mathematical expectation to the variance $IN = \max\limits_{i\in[0,n]} \left(\frac{E_i}{D_i} \right)$ is considered as a MEG record-selection criterion.

(3) The removal of noise from the selected harmonics. Discrete wavelet transform is used. The Haar wavelet is used as a parent wavelet. This wavelet forms an orthonormalized basis, and has the property of symmetry.

(4) Cluster analysis is performed by the iterative method of k-means clustering. The found time moments at which the abnormal component exists

Table 2. Classification of MEG signals

Signal name	Signal type	Revealed biomagnetic signal features	Source localization
A	Abnormal	Beginning or termination of increased activity	Cerebellum
B	Abnormal	Intermediate phase, change of the increased activity source	Cerebellum, stem
C	Abnormal	Biomagnetic activity peak, highest signal amplitude	Black substance
D	Normal	No specific features	Generally, cerebrum cortex

become the initial data in the problem of localizing the brain areas associated with the considered pathology.

(5) The localization of the increased biomagnetic activity source.

The classification of a signal type by cluster analysis was performed depending on the spectral characteristic of the signal. In total, four types of signals, such as A, B, C, and D, were revealed (Table 2).

The results of localizing the sources of increased activity in the recorded magnetic encephalogram confirm the existing medical opinion about the relationship of Parkinson's disease to damage to subcortical brain areas. In particular, there exist some data on the relationship of this disease to the death of melanin-containing neurons in one of the subcortical brain ganglia—black matter [48, 49]. The higher level of movement control is localized in the cerebrum cortex, the basal ganglia, and the cerebellum. In this connection, it is interesting that the beginning of a Parkinsonism attack is associated with one of the most important brain areas, i.e., the cerebellum. Although the cerebellum is only 10% of the brain volume, it accommodates more than half of all the central nervous system neurons. It not only provides the continuous control over movement activity, but also participates in the implementation of cognitive encoding and memory [48].

The proposed computational technology is classified among the combined numerical-analytical approaches. It combines computer-aided (digital) data-processing methods with the analytical derivations and transforms in the spectral representation (in the space of Fourier signal coefficients). The formal core of this technology is the generalized spectral-analytical method (GSAM).

The description is based on the expansion of a studied signal in the full system of classic orthogonal polynomials and functions (Jacobi, Chebyshev, Legendre, Laguerre, Gegenbauer, and other polynomials). In this context, the main adaptive procedure is the selection of an optimal basis for the derivation of an optimal spectral representation. In this case, the process of the analytical description of input signals is completely automated. The used adaptive procedures provide the required precision of description with expressions of minimum complexity. As a result, the compact description and constancy of its structure is attained, and this is an essential factor for the further analytical transforms and derivations. Adaptation of the description appreciably facilitates the necessary analytical transforms and derivations in the further processing of signals. The execution of adaptive procedures promotes the efficient compression of the volume of the information array.

The expansion into an orthogonal series is accompanied by the orthogonal projection of unknown signals onto the known functions from the selected basis. The thus-found expansion coefficients characterize the degree of coincidence between the signals and the known orthogonal functions on the considered intervals.

The spectral methods for the analysis of biomagnetic data on the basis of the generalized spectral-analytical approach were implemented. The transition from spatiotemporal MEG records to the spectral representation provides the possibility to appreciably decrease the volume of processed data and increase the precision and stability of calculations. The further improvement of the efficiency of the system-state estimates is performed by taking into account the informativity criteria for the used attributes. As shown by the calculations for clinical Parkinson's disease cases, the precision of the parametric identification of the system with consideration for the preliminary spectral classification substantially increases, providing the source localization precision of 2–5 mm. The results of MEG analysis performed to study the sources of induced and spontaneous activity in cases of normal and pathological biomagnetic activity (Parkinson's disease) are in good agreement with the neurophysiological data on the localization of brain functional and damaged areas in connection with Parkinsonism. In the latter case, the hypothesis about the relationship between the progression of this disease and the degradation of melanin-containing neurons in the *substantia nigra* in the brain stem structures (see Fig. 4) is confirmed.

Contour Recognition of Objects

The solution of the contour-recognition problem on the basis of the Fourier series expansion in harmonic functions is considered.

Let us note that all objects mapped on a plane have a boundary line, which in turn can be represented as a curve S in a certain parametric form with respect to the conditional time t, e.g., as $S(t_i) = \{X(t_i), Y(t_i)\}$, where $X(t)$, $Y(t)$ are the coordinates of the point of "traversal about the boundary" at time t. Such a representation is the absolute path of traversal about an

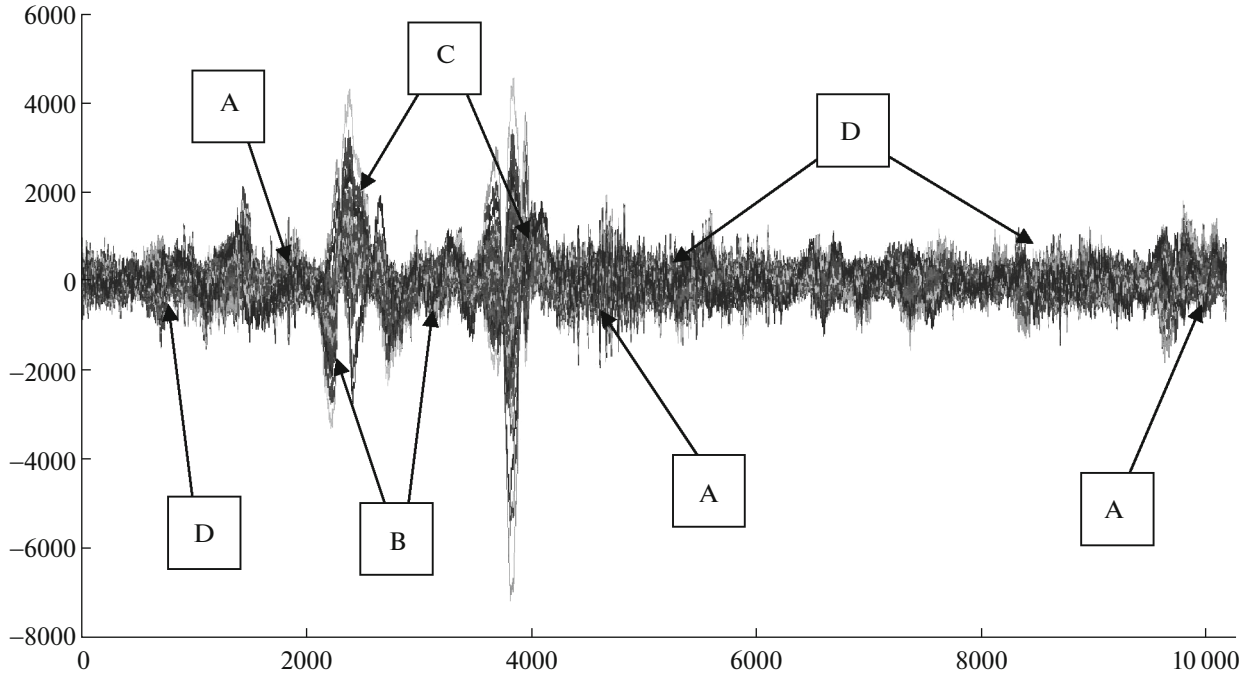


Fig. 3. Classification of different activity types in the case of Parkinsonism.

object, but there exists a good alternative in the form of "relative paths," which are less dependent on the coordinate system and the different transforms of an object on a plane. Thus, if the boundary is described with a pair of other point motion parameters, e.g., the velocity $V(t_i)$ and the path curvature $K(t_i)$, such a record will be independent of the transfer and turn of an object, but these parameters will strictly linearly depend on the dimensions of the object.

Moreover, generally speaking, it does not matter at what velocity a point moves along the boundary of an object in the contour-comparison problem; i.e., the velocity may be taken equal to a certain constant value $V(t_i) = \text{const}$. Indeed, let us presume that we know the onboard readings from the sensors of a Formula-1 race car on its velocity and the position of its steering wheel at time moments t, and it is obvious that these

parameters are sufficient to trace the path of this car with a high precision and thereby to restore the shape of its race track with a quality sufficient for unambiguous recognition. However, if the pilot is required to follow along this track at the same constant velocity, e.g., $V(t) = 60$ km/h, the complete restoration of the track will need only the data on steering wheel positions (i.e., path curvature data), thus reducing the overall volume of compared data by two times.

Hereinafter, we assume that all the object contours considered by us are represented by their curvature functions $K(t)$, which are defined on the interval $[0, T]$ and expressed by truncated Fourier series

$$K(t_i) = A_0 + \sum_{n=1}^{N} A_n \cos\frac{2\pi n t_i}{T} + B_n \sin\frac{2\pi n t_i}{T}.$$

Let us now assume that it is necessary to estimate the degree of similarity between two contours by the available spectral representations, and it is known that the figures have the same size, and it is possible to

Fig. 4. Localization of the MEG signal source upon a Parkinsonian activity burst.

specify the "correct" zero (start) point of traversal about their closed contour. Under these favorable conditions, it remains only to estimate their similarity via the direct comparison of the parametric representations of these curves, e.g., using the mean-square deviation

$$r = \sqrt{\sum_{i=0}^{N} (S(t_i) - s(t_i))^2 \Delta t_i / T},$$

which coincides with the same estimate performed over a pair of their spectral representations

$$r = \sqrt{\sum_{n=0}^{N} ((A_n - a_n)^2 + (B_n - b_n)^2)}$$

according to the Parseval equality.

Since these estimates are equivalent, it is simpler to perform the estimation through the spectral representations, as this appreciably reduces the volume of calculations due to a decrease in the expansion depth. Indeed, as known, the function of N counts requires N spectral coefficients for its precise restoration. However, for many practical problems, it is sufficient to have only an approximate representation of functions. However, it is should be remembered that the scheduled thinning of counts ("decimation") usually gives worse approximation results than the implementation of the simplest filter of low frequencies for the Fourier spectrum with the rejection of the same volume of data alongside the same set of high-frequency noises. When performing the comparison of contours, it is often sufficient to retain only the first 10% of a spectrum and, according to Kotelnikov's theorem, this corresponds the reduction of count points t by ten times as well, but leads to more stable results.

For simplicity, the above-described situation implied that the dimensions of the objects are the same and the selection of a starting point of the contour is not critical. In actual fact, the difference between their dimensions and the incorrect choice of zero points leads to shifts of the compared functions along the ordinate and abscissa axes, thus leading to estimates unsuitable for practical use. Generally, it does not seem difficult to solve the problem of disperancy between the dimensions of objects before the comparison of their shapes; to accomplish this, it is sufficient to reduce all the objects to a certain common (normalized) size. However, the problem of the unambiguous selection of a starting contour point seemed to be rather difficult for a long time. The authors [50] have proposed the alternative amplitude–frequency representation of a harmonic expansion for the solution of this problem. Indeed, if the expansion terms $a_n \cos(nx)$ and $b_n \sin(nx)$ in a spectral series are united into one wave of the same frequency $A_n \cos(nx + W_n)$, it turns out that their overall amplitude A_n is not changed at all when a different start point of traversal about the contour is selected, and only the wave phase W_n is changed. Hence, among all the coefficients of the amplitude–frequency representation of contours ($S = \{A_n, W_n\}$), half are invariant to the selection of a zero point (namely, $\{A_n\}$) and, therefore, the "weak estimation" of similarity between the figures may be performed very quickly, e.g., via the pairwise comparison of the degree of correlation between the first ten coefficients A_n for all of the studied objects. In this case, as shown by practice, good and stable results can be attained even when the similarity between the objects is evaluated with only one such "weak estimation," which can easily be conditioned to the quality of "strong estimation" by involving the phase components W_n into the comparison according to the simple rules described in the same work. Objects of different nature with their own series of the first amplitudes are shown below in Fig. 5 to provide the possibility of estimating the adequacy of the estimates proposed here on well-known contours.

Recognition of Repeats in Genomes

To adapt the spectral-analytical approach to the problems of bioinformatics, it was necessary to generalize the notion of a dot matrix. As a result of generalization, this approach may be applied to the search for repeats in any character or functional sequences. For this reason, the approach is described in the most general form with the retention of certain terminology and examples from the field of bioinformatics.

The method is based on the spectral expansion of the functions which compose the characteristic description of a text sequence, in which the following properties of the functions are important: (1) completeness and (2) smoothness. The completeness of description means that the initial sequence can be restored from the characteristic curves. The property of smoothness for the change of characteristics is necessary to provide their description with truncated orthogonal series.

These conditions are satisfied by the content curves of nucleotide subsets in a window of specified length along the sequence of a macromolecule. This type of curve incorporates the well-known GC content curve, which was studied in bioinformatics. At the same time, the window size, which is a parameter of such a description, actually introduces the notion of scale for a sequence of characters.

In the general case, it is possible to formulate the following theorem.

Theorem about the expansion of a sequence of characters: for a random character sequence in an alphabet M of characters, there exists a sheaf of $Log_2 M$ characteristic functions, from which it is possible to restore the initial sequence, and these functions are discrete and K–digital, where K is the scale parameter.

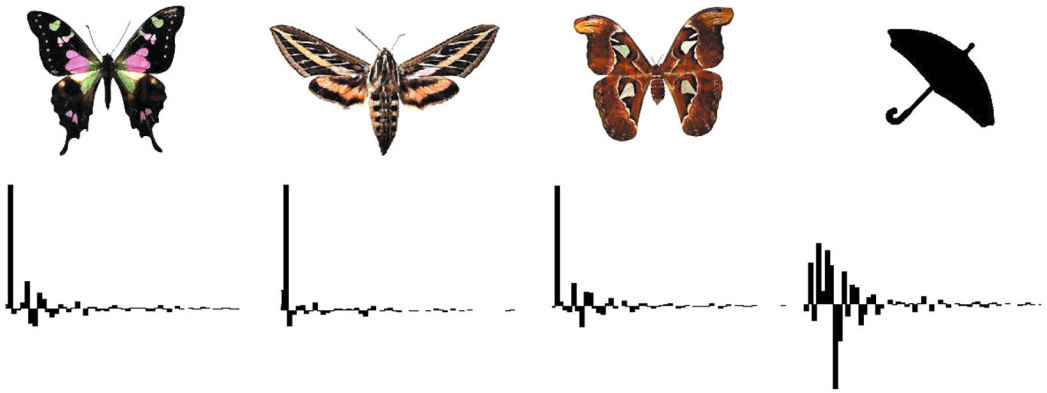


Fig. 5. (Color online) Examples of objects for recognition.

Proof. Let us encode the characters of a sequence with a numerical vector in the binary number system. A vector with a length of no less than $Log_2 M$ will be required. Let us now consider a sliding window with a width K and sum up the number of units with a certain bit for all the sequence characters in this window. Let the thus-determined function depending on the window position be called the characteristic sequence function corresponding to a certain bit in the binary encoding of characters, as every bit of each character in the sequence can be restored from the corresponding characteristic function.

Note. The characteristic functions considered as functions of the beginning coordinate of the summation window are by definition smoothly changing and K-digital functions.

Further, the characteristic functions composing the description of a random sequence in turn is subjected to sliding scanning with a window of width W. In practice, the characteristic functions of a sequence are partitioned into overlapping fragments of length W with a shift d rather than a step of 1. Afterwards, all the fragments f_i, g_i considered as discrete functions with the numbering of samples within the window length are pairwisely compared with each other on the basis of a standard metric in a Euclidian space as

$$\rho(f,g) = (f-g, f-g) = \frac{1}{W}\sum_{i=1}^{W}(f_i - g_i)^2.$$

To reduce the calculations of the distances between the fragments, the approximation of characteristic function fragments with truncated orthogonal series is used. For this reason, the distance is estimated by the formula

$$\rho(f,g) = \sum_{i=1}^{N}(c_i - d_i)^2,$$

where c_i, d_i are the coefficients of the first N Fourier series terms ($N \ll W$). The use of spectral expansion provides the possibility not only to economize on the calculation of distances, but also to perform the transformations for estimating the inverted and complementary sequences in the space of expansion coefficients, thus implying the simultaneous recognition of all the types of repeats without transforming the sequence itself.

The decisive rule for the recognition of repeats is a threshold one: if $\rho < \varepsilon$, where ε is a threshold value, the fragments are considered to be similar, but if $\rho \geq \varepsilon$, the fragments are not similar. If there exist several characteristic curves composing the complete description of an object, they are simultaneously recognized, and this results in the logical multiplication of decision-making rules for each of the characteristic functions. This follows from the fact that the decisive rule is fulfilled in the region of a minimum metric considered as a function of the number of fragments. Hence, the set of minima determines the set of candidates for repeats. In the case of two attributes, e.g., the GC and GA curves, the set of repeats is taken as the overlappings between the sets of repeated candidates found for each of the attributes separately. After these operations, the comparison results are mapped onto the dot matrix, whose one point, however, corresponds to the comparison of two full fragments. The dot matrix is one of the comprehensive standard representations for the results of the comparison of two sequences, thus providing the possibility to map the mutual arrangement of repeats. The generalized dot matrix provides new opportunities for the alignment of inexact repeats. For example, it has been shown that an inexact extensive tandem repeat may be mapped in the form of an ideal square on the dot matrix (Fig. 6). This is attained via the correct selection of the ratio between the window sizes K and W and the shift d. Based on this important result, a completely automatized method for the recognition of tandem repeats was constructed, and repeats not earlier known were found.

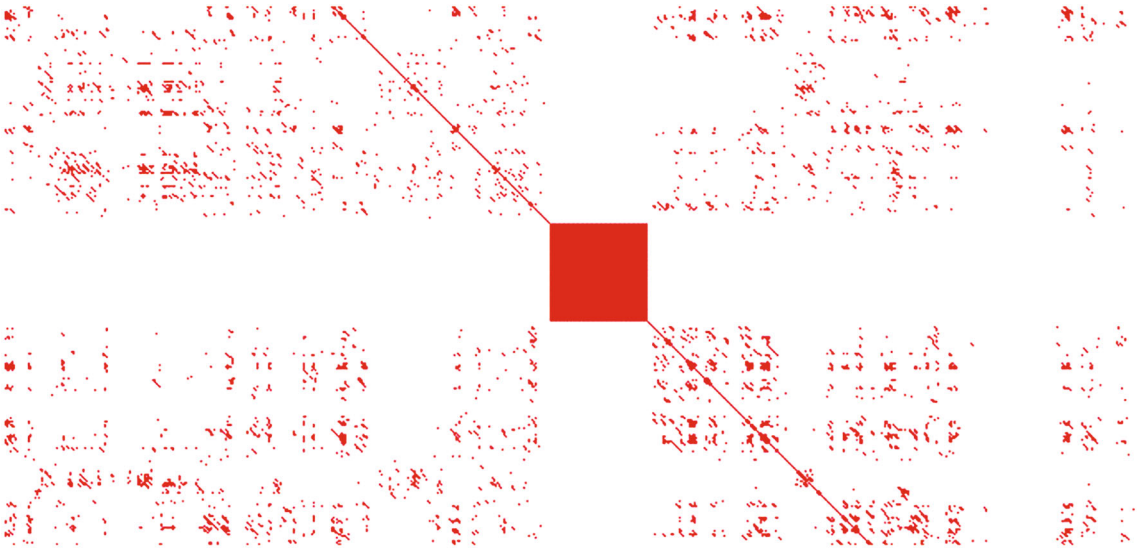


Fig. 6. (Color online) Dot matrix fragment containing a tandem repeat.

The structural scheme composed for this method as a result of the performed studies is the following:

(1) The preliminary processing of a sequence of characters or the formation of an initial alphabet, i.e., the removal of unnecessary characters and the re-encoding of sequence characters;

(2) The conversion of the sequence of characters into the sheaf of characteristic functions on the basis of the proven theorem;

(3) The conversion of the characteristic functions into the spectral representation. In contrast to the previous steps, this stage leads to the irreversible compression of data;

(4) The spectral comparison of sequence fragments;

(5) The mapping and analysis of the dot matrix for the recognition of extensive repeats (direct, tandem, and inverted) and the study of their mutual arrangement; and

(6) The check of repeats by means of alignment with the use of dynamic programming methods.

Recognition and Study of the Properties of Structural Motifs in Protein Molecules

The object of our studies is one of the frequently encountered structural motifs in homologous and nonhomologous proteins, i.e., an $\alpha - \alpha$ corner [38]. This supersecondary structure is formed by two α helices, which are neighboring in a polypeptide chain, linked to each other by means of constrictions, and packed in an orthogonal (crosslike) manner. In proteins, $\alpha - \alpha$ corners are encountered in the form of a left-hand superhelix. Their sequences have a certain location in the chain of hydrophobic, hydrophilic, and glycine residues.

In this work, the problem of the recognition of supersecondary structures in globular proteins with known 3D structures determined by X-ray diffraction and nuclear magnetic resonance is solved on the basis of the generalized spectral-analytical method [3, 11]. Using the analytical description of the coordinates of C^{α} atoms in the main chain of a protein globule, the characteristic profiles of the protein structures represented in the PDB format [51] were obtained, and the spectral algorithm of searching for the specified pattern of a structural motif in the studied proteins was further applied [21–23].

The spatial structure of secondary and supersecondary protein structures is determined by the coordinates of C^{α} atoms in the main chain of the protein globule. Hence, the coordinates of the atoms in the side chains of the protein molecules were not taken into account.

The spatial structure of a protein as a whole and also the spatial structure of secondary and supersecondary structures can be represented via the parametric equation of a curve in a three-dimensional space:

$$\begin{cases} x(t) = \sum\limits_{i=0}^{N} A_i \varphi_i(t) \\ y(t) = \sum\limits_{i=0}^{N} B_i \varphi_i(t) \\ z(t) = \sum\limits_{i=0}^{N} C_i \varphi_i(t), \end{cases}$$

where $\{\varphi_i\}$ is a system of orthogonal polynomials and A_i, B_i, C_i are the coefficients of the expansion of functions in orthogonal bases. The curve was obtained by the methods of the analytical descriptions of the main chain formed by the coordinates of C^α atoms. These methods were implemented with the use of splines and orthogonal polynomials [11].

The most suitable polynomials for the studied functions are the Legendre and Chebyshev polynomials. The expansion coefficients considered as the spectral attributes of a signal are calculated by the general formula

$$Z_i = \int_a^b f(t)\varphi_i(t)\rho(t)dt$$

with a specified weight $\rho(t)$.

The transition from the parametric equation of a curve in a three-dimensional space to the natural equation of a curve in a three-dimensional space gives the parametrized description of the curvature and torsion depending on the natural curve parameter, i.e., the arc length

$$\begin{cases} x = x(t) \\ y = y(t) \\ z = z(t) \end{cases} \Rightarrow \begin{cases} C(s) = \sqrt{\ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2} \\ T(s) = \dfrac{\begin{vmatrix} \dot{x} & \dot{y} & \dot{z} \\ \ddot{x} & \ddot{y} & \ddot{z} \\ \dddot{x} & \dddot{y} & \dddot{z} \end{vmatrix}}{\ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2}, \end{cases}$$

where C is the curvature function, T is the torsion function, and s is the natural curve parameter, i.e., the arc length

$$s(t) = \int_{t_1}^{t_2} \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}$$

Such a description is invariant to the selection of a Cartesian coordinate system. The curvature and torsion functions are the characteristic profiles of the spatial structure of a protein molecule. In this case, the regular fragments, α helices, are represented by areas of constant curvature and torsion values.

Using the specified profiles, it is possible to unambiguously restore the spatial structure of a motif with a precision of up to the selection of a Cartesian coordinate system. The transition from the natural equation of a curve in a three-dimensional space to the parametric equation of a curve in a three-dimensional space is performed by the Frenet formula.

Using the ProteinReviser software analytical complex created by the authors on the basis of the generalized spectral-analytical method for the recognition of supersecondary structures in globular proteins with solved 3D structures, a sample of α – α corners from the PDB database was formed. The sample represents a list of proteins with specified coordinates for the atoms composing these structures. All of the structures found with the software analytical complex were visually revised.

The search for α – α corners in the PDB database was performed by the specified template 1D1L C^α:: 15–37. The band model in the space of a classic α – α corner with a short connection is shown in Fig. 7, and its analytical description is illustrated in Fig. 8. Using the methods of analytical description for the main chain formed by the coordinates of C^α atoms, the parametric equation of a curve in a three-dimensional space was derived. Legendre polynomials were used for the analytical description of this structural motif.

The analytical description of the structure of an α – α corner was used to find the characteristic profiles of the spatial structure of a double helical motif, i.e., the curvature and torsion functions. Using the obtained profiles, it is possible to unambiguously restore the spatial structure of a studied motif with a precision to the selection of a coordinate system.

The curvature and torsion profiles found from the analytical descriptions of the structure of the α – α corner shown in Fig. 7 are illustrated in Figs. 9 and 10.

The developed algorithm of search for α – α corners in protein molecules on the basis of the generalized spectral-analytical method has provided the creation of a sample of structural motifs. Using the mentioned algorithm, 110 α – α corners corresponding to the specified template 1D1L C^α: :15–37 were found in the PDB database.

The hypothesis about the self-sustained stability of α – α corners in an aqueous medium was put forward to provide the basis for further study. In this case, self-sustained stability is understood to mean the stability of the spatial structure of a studied structural motif apart from the protein molecule in which this structure was revealed. Using the method of molecular dynamics, numerical experiment was performed to demonstrate that the α – α corner is a self-sustainably stable structure [52, 53].

It follows from the property of self-sustained stability that all the attributes of α-α corners in the primary structure are localized in the structure of α – α corners themselves. Hence, the analysis of primary structures only for α – α corners apart from the protein molecules in which these structures were revealed will help to reveal the characteristic features of α – α corners. As a result of study, some interesting regularities in the alternation of certain groups of amino acid residues were revealed. In particular, the presence of glycine in the contraction and the alternation of hydrophobic amino acid residues in α helices in certain positions with respect to glycine were statistically confirmed [52].

Hence, the combined approach to the analysis of the spatial structure of proteins on the basis of the analytical description of the main chain of a protein glob-

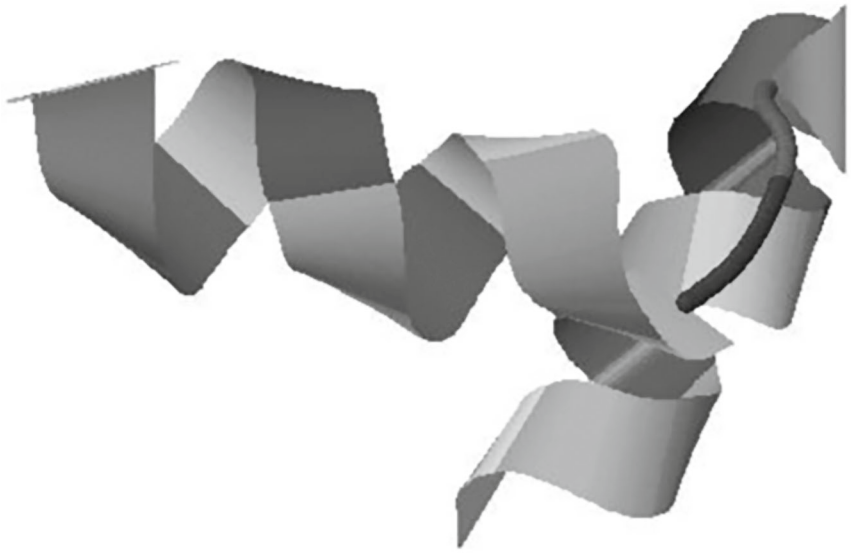


Fig. 7. Band model of an α–α corner with a short constriction.

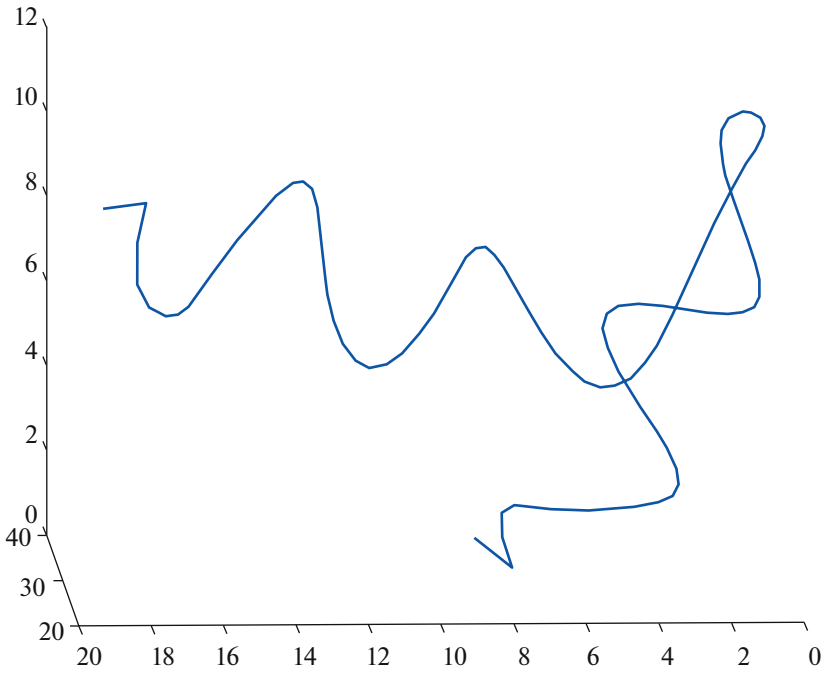


Fig. 8. (Color online) Analytical description of an α–α corner with a short constriction by the Legendre polynomials.

ule and the spectral method of the recognition of repeats was proposed to solve the problem of recognizing the structural motifs of proteins (on the example of $\alpha - \alpha$ corners). In addition, the self-sustained stability of $\alpha - \alpha$ corners was analyzed by the method of molecular dynamics in an aqueous medium. Experiment has demonstrated that the $\alpha - \alpha$ corner is a self-sustainably stable structure, and this property may be considered as an additional verifying attribute in the studies. Some characteristic features of $\alpha - \alpha$ corners in amino acid sequences were revealed, thus providing the recognition of $\alpha - \alpha$ corners in the primary structures of protein molecules also on the basis of these features. Moreover, the geometric characteristics of structural motifs of the $\alpha - \alpha$-corner type, such as the distances and torsion angles between the axes of the helices, the surface areas and perimeters of the helix projection overlapping the polygons, and the dependence of the torsion angles between the axes of helices on their lengths [54–57], were also studied.

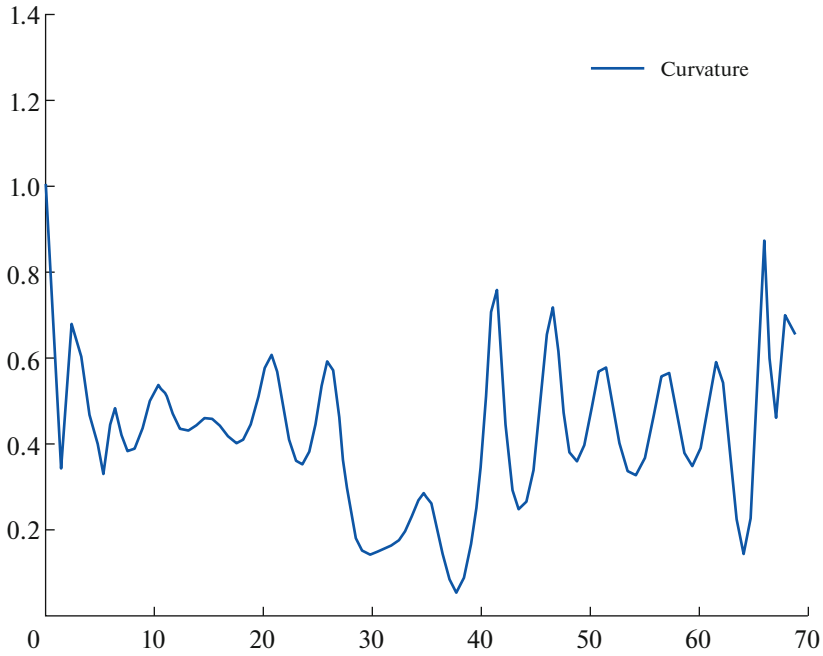


Fig. 9. (Color online) Curvature profile obtained from the analytical description of the α–α corner shown in Fig. 7.

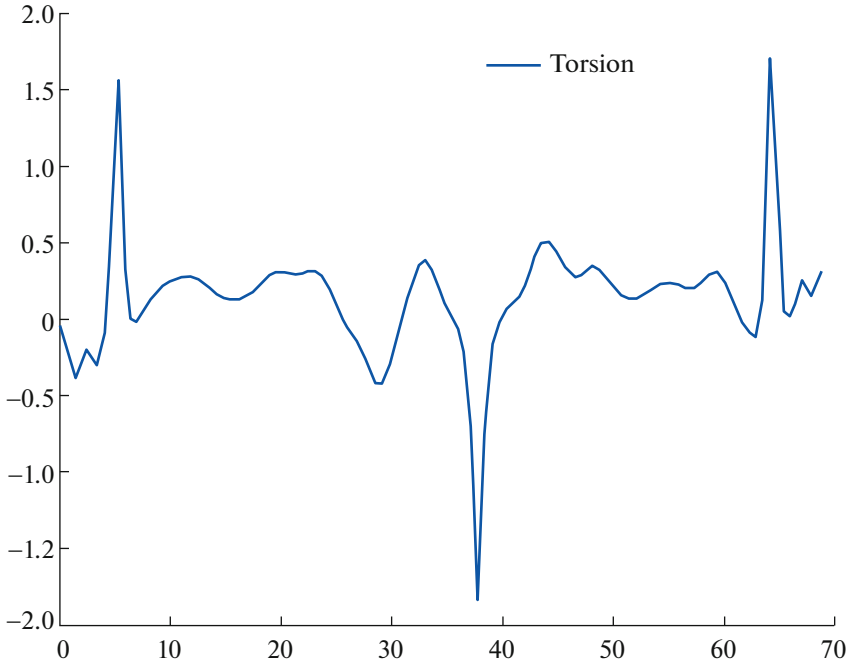


Fig. 10. (Color online) Torsion profile obtained from the analytical description of the α–α corner shown in Fig 7.

CONCLUSIONS

In this work, the generalized spectral-analytical method and its application to the contemporary problems of biomedicine and bioinformatics have been described. Both the general principles of the solution of the recognition problem with the use of GSAM and the specific features of the problem formulations in various fields have been shown. The obtained results argue for the high efficiency and universality of this method. The initial principles underlying the construction of this method lead to the successful and efficient solution of such problems as the classification of biomedical signals, the recognition of repeated structures in bioinformation databases, and the recognition of contour objects in images.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interests.

REFERENCES

1. V. L. Goncharov, *Theory of Interpolation and Approximation of Functions*, 2nd ed. (Gostekhteorizdat, Moscow, 1954) [in Russian].
2. A. F. Nikiforov and V. B. Uvarov, *Special Functions of Mathematical Physics* (Nauka, Moscow, 1978; Birkhäuser, Basel, 1988).
3. F. F. Dedus, A. F. Dedus, and M. N. Ustinin, "A new data processing technology for pattern recognition and image analysis problems," Pattern Recogn. Image Anal. **2** (2), 195–207 (1992).
4. J. K. Bowmaker, "Evolution of color vision in vertebrates," Eye **12** (3), 541–547 (1998).
5. S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Trans. Pattern Anal. Mach. Intell. **24** (4), 509–522 (2002).
6. A. Antoniou, *Digital Signal Processing*: *Signals, Systems, and Filters* (McGraw Hill, New York, 2006).
7. A. N. Pankratov, "On the implementation of algebraic operations on orthogonal function series," Comput. Math. Math. Phys. **44** (12), 2017–2023 (2004).
8. G. Benson, "Tandem repeats finder: a program to analyze DNA sequences," Nucleic Acids Res. **27** (2), 573–580 (1999).
9. R. Kolpakov, G. Bana, and G. Kucherov, "mreps: efficient and flexible detection of tandem repeats in DNA," Nucleic Acid Res. **31** (13), 3672–3678 (2003).
10. A. Y. Ogurtsov, M. A. Roytberg, S. A. Shabalina, and A. S. Kondrashov, "OWEN: aligning long collinear regions of genomes," Bioinformatics **18** (12), 1703–1704 (2002).
11. G. M. Landau, J. P. Schmidt, and D. Sokol, "An algorithm for approximate tandem repeats," J. Comput. Biol. **8** (1), 1–18 (2001).
12. F. F. Dedus, L. I. Kulikova, S. A. Makhortykh, N. N. Nazipova, A. N. Pankratov, and R. K. Tetuev, "Analytical recognition methods for repeated structures in genomes," Dokl. Math. **74** (3), 926–929 (2006).
13. F. F. Dedus, L. I. Kulikova, S. A. Makhortykh, N. N. Nazipova, A. N. Pankratov, and R. K. Tetuev, "Recognition of the structural-functional organization of genetic sequences," Moscow Univ. Comput. Math. Cybern. **31** (2), 49–53 (2007).
14. A. N. Pankratov, M. A. Gorchakov, F. F. Dedus, N. S. Dolotova, L. I. Kulikova, S. A. Makhortykh, N. N. Nazipova, D. A. Novikova, M. M. Olshevets, M. I. Pyatkov, V. R. Rudnev, R. K. Tetuev, and V. V. Filippov, "Spectral analysis for identification and visualization of repeats in genetic sequences," Pattern Recogn. Image Anal. **19** (4), 687–692 (2009).
15. R. K. Tetuev, N. N. Nazipova, A. N. Pankratov, and F. F. Dedus, "Search for megasatellite tandem repeats in eukaryotic genomes by estimation of GC-content curve oscillations," Math. Biolog. Bioinform. **5** (1), 30–42 (2010) [in Russian].
16. R. K. Tetuev and N. N. Nazipova, "Consensus of repeated region of mouse chromosome 6 containing 60 tandem copies of a complex pattern," Repbase Rep. **10** (5), 776 (2010).
17. R. K. Tetuev, F. F. Dedus, and N. N. Nazipova, "Consensus of repeated region of rat chromosome 4 similar to mouse chromosome 6 repeated region, enclosed in the intergenic region between genes Hrh1 and Atg7," Repbase Rep. **10** (8), 1185 (2010).
18. A. N. Pankratov, M. I. Pyatkov, R. K. Tetuev, N. N. Nazipova, and F. F. Dedus, "Search for extended repeats in genomes based on the spectral-analytical method," Math. Biolog. Bioinform. **7** (2), 476–492 (2012) [in Russian].
19. M. I. Pyatkov, V. V. Filippov, and A. N. Pankratov, "Consensus of repeated region of rabbit chromosome 17 containing over 15 huge approximate tandem repeats," Repbase Rep. **12** (3), 256 (2012).
20. A. N. Pankratov, R. K. Tetuev, M. I. Pyatkov, V. P. Toigildin, and N. N. Popova, "Spectral analytical method of recognition of inexact repeats in character sequences," Proc. Inst. Syst. Program. Russ. Acad. Sci., **27** (6), 335–344 (2015) [in Russian].
21. M. I. Pyatkov and A. N. Pankratov, "SBARS: fast creation of dotplots for DNA sequences on different scales using GA-, GC-content," Bioinformatics **30** (12), 1765–1766 (2014).
22. K. Katoh, K. Misawa, K. Kuma, and T. Miyata, "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform," Nucleic Acids Res. **30** (14), 3059–3066 (2002).
23. D. Sharma, B. Issac, G. P. S. Raghava, and R. Ramaswamy, "Spectral Repeat Finder (SRF): Identification of repetitive sequences using Fourier transformation," Bioinformatics **20** (9), 1405–1412 (2004).
24. L. Du, H. Zhou, and H. Yan, "OMWSA: detection of DNA repeats using moving window spectral analysis," Bioinformatics **23** (5), 631–633 (2007).
25. J. Krumsiek, R. Arnold, and T. Rattei, "Gepard: a rapid and sensitive tool for creating dotplots on genome scale," Bioinformatics **23** (8), 1026–1028 (2007).
26. R. K. Tetuev, M. I. Pyatkov, and A. N. Pankratov, "Parallel algorithm for global alignment of long amino-acid and nucleotide sequences," Math. Biolog. Bioinform. **12** (1), 137–150 (2017) [in Russian].
27. A. N. Pankratov, R. K. Tetuev, and M. I. Pyatkov, "LSCGAT: Long sequences customizable global alignment tool," J. Bioinf. Genomics No. 1 (10), 3 pages (2019).
28. E. V. Brazhnikov and A. V. Efimov, "Structure of α-α-hairpins with short connections in globular proteins," Mol. Biol. **35** (1), 89–97 (2001).
29. A. V. Efimov, "A new super-secondary protein structure: the alpha alpha-angle," Mol. Biol. (Mosk.) **18** (6), 1524–1537 (1984) [in Russian].
30. A. V. Efimov, "Standard structures in proteins," Prog. Biophys. Mol. Biol. **60** (3), 201–239 (1993).
31. A. V. Efimov, "L-shaped structure from two alpha-helices with a proline residue between them," Mol. Biol. (Mosk.) **26** (6), 1370–1376 (1992) [in Russian].
32. C. Chothia, M. Levitt, and D. Richardson, "Structure of proteins: Packing of α-helices and pleated sheets,"

Proc. Natl. Acad. Sci. U. S. A. **74** (10), 4130–4134 (1977).

33. C. Chothia, M. Levitt, and D. Richardson, "Helix to helix packing in proteins," J. Mol. Biol. **145** (1), 215–250 (1981).

34. D. Walther, F. Eisenhaber, and P. Argos, "Principles of helix-helix packing in proteins: The helical lattice superposition model," J. Mol. Biol. **255** (3), 536–553 (1996).

35. A. Trovato and F. Seno, "A new perspective on analysis of helix-helix packing preferences in globular proteins," Proteins: Struct., Funct., Bioinf. **55** (4), 1014–102 (2004).

36. H. Bateman and A. Erdélyi, *Higher transcendental functions*, Vol. II (McGraw Hill, New York, 1953; Nauka, Moscow, 1966).

37. A. F. Nikiforov, V. B. Uvarov, and S. K. Suslov, *Classical Orthogonal Polynomials of a Discrete Variable*, in *Springer Series in Computational Physics* (Springer, Berlin, Heidelberg, 1991).

38. H. Jeffreys and B. Swirles, *Methods of Mathematical Physics*, 3rd ed. (Cambridge Univ. Press, Cambridge, 1966; Mir, Moscow, 1970) [Vol. 3 of the Russian translation].

39. E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces* (Princeton Univ. Press, Princeton, NJ, 1971); *Introduction to Harmonic Analysis on Euclidean Spaces* (Mir, Moscow, 1974) [in Russian].

40. N. Ya. Vilenkin, *Special Functions and Theory of Group Representations* (Nauka, Moscow, 1991) [in Russian].

41. W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Chelsea Publ., New York, 1954).

42. M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series, **Vol. 55** (U.S. Government Printing Office, Washington, D.C., 1964) [Chapter 8].

43. F. F. Dedus, S. A. Makhortykh, M. N. Ustinin, and A. F. Dedus, *Generalized Spectral-Analytic Method for Data File Processing. Problems of Image Analysis and Pattern Recognition* (Mashinostroenie, Moscow, 1999) [in Russian].

44. S. A. Makhortykh, "Generalized spectral-analytical method for biomedical data processing," Math. Montisnigri **XXXVI**, 104–113 (2016).

45. M. N. Ustinin, S. A. Makhortykh, A. M. Molchanov, et al., "Problems of the analysis of magnetic encephalography data," in *Computers and Supercomputers in Biology*, Ed. by V. D. Lakhno and M. N. Ustinin (Inst. Komp. Issled., Izhevsk, Moscow, 2002), pp. 327–349 [in Russian].

46. L. I. Kulikova and S. A. Makhortykh, "Mathematical operations on two-dimensional signals in bases of spherical harmonics," Investigated in Russia **9**, 598–608 (2006) [Electronic journal, in Russian].

47. A. V. Derguzov and S. A. Makhortykh, "Spectral analysis and data classification in magnetoencephalography," Pattern Recogn. Image Anal. **16** (3), 497–505 (2006).

48. T. Boraud, E. Bezard, B. Bioulac, and C. E. Gross, "From single extracellular unit recording in experimental and human Parkinsonism to the development of a functional concept of the role played by the basal ganglia in motor control," Prog. Neurobiol. **66** (4), 265–283 (2002).

49. M. B. H. Youdim and P. Riederer, "Understanding Parkinson's disease," Sci. Am. **276** (1), 52–59 (1997).

50. R. K. Tetouev, "Contour recognition based on spectral methods. Solution of the problem of choice of the startpoint," Pattern Recogn. Image Anal. **17** (2), 243–251 (2007).

51. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein Data Bank," Nucleic Acids Res. **28** (1), 235–242 (2000).

52. V. R. Rudnev, A. N. Pankratov, L.I. Kulikova, F. F. Dedus, D. A. Tikhonov, and A. V. Efimov, "Recognition and stability analysis of structural motifs of α-α-corner type in globular proteins," Mat. Biolog. Bioinform. **8** (2), 398–406 (2013) [in Russian].

53. V. R. Rudnev, A. N. Pankratov, L.I. Kulikova, F. F. Dedus, D. A. Tikhonov, and A. V. Efimov, "Conformational analysis of structural motifs of α-α-corner in the computational experiment of molecular dynamics," Mat. Biolog. Bioinform. **9** (2), 575–584 (2014) [in Russian].

54. D. A. Tikhonov, L. I. Kulikova, and A. V. Efimov, "Statistical analysis of internal distances of helical pairs in protein molecules," Mat. Biolog. Bioinform. **11** (2), 170–190 (2016) [in Russian].

55. D. A. Tikhonov, L. I. Kulikova, and A. V. Efimov, "The study of interhelical angles in the structural motifs formed by two helices," Mat. Biolog. Bioinform. **12** (1), 83–101 (2017) [in Russian].

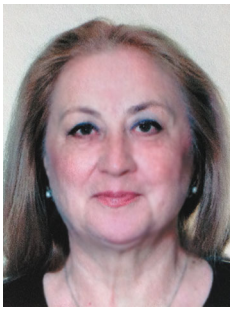
56. D. A. Tikhonov, L. I. Kulikova, and A. V. Efimov, "Analysis of torsion angles between helical axes in pairs of helices in protein molecules," Mat. Biolog. Bioinform. **12** (2), 398–410 (2017) [in Russian].

57. D. A. Tikhonov, L. I. Kulikova, and A. V. Efimov, "Analysis of the areas and perimeters of polygons of the helices projections intersection in helical pairs of protein molecules," Keldysh Institute of Applied Mathematics Preprint No. 59 (2018) [in Russian].

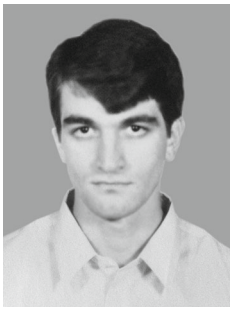
Translated by E. Glushachenkova

Makhortykh Sergei Aleksandrovich. Born in 1963. Graduated from Moscow Physical–Technical Institute (Faculty of Aerophysics and Space Research) in 1986. Academic secretary of the Institute of Mathematical Problems of Biology (Branch of the Keldysh Institute of Applied Mathematics, Russian Academy of Sciences). Candidate in Physics and Mathematics since 1990. Scientific interests: spectral information-processing methods, bioinformatics, mathematical biology, environmental science, pattern recognition. Author of more than 60 reports and papers in peer-reviewed journals.

Pankratov Anton Nikolaevich. Born in 1972. Graduated from Moscow State University (Faculty of Computational Mathematica and Cybernetics) in 1994. Candidate in Physics and Mathematics since 2004. Senior researcher of the Institute of Mathematical Problems of Biology (Branch of the Keldysh Institute of Applied Mathematics, Russian Academy of Sciences). Scientific interests: spectral-analytical methods, algorithms of bioinformatics. Author of more than 20 papers in scientific journals.



Kulikova Ludmila Ivanovna. Born in 1960. Graduated from V.I. Ulyanov Lenin Kazan State University in 1982 (Faculty of Computational Mathematics and Cybernetics). Senior researcher of the Institute of Mathematical Problems of Biology (Branch of the Keldysh Institute of Applied Mathematics, Russian Academy of Sciences). Candidate in Physics and Mathematics, speciality 05.13.17 "Theoretical Foundations of Informatics" since 2007. Scintific interests: spectral-analytical information processing methods, data conversion, bioinformatics, pattern recognition. Author of more than 30 papers in journals.



Tetuev Ruslan Kurmanbievich. Born in 1976. Graduated from the Kh.M. Berbekov Kabardino-Balkarian State University in 1998, speciality Applied Mathematics. Received candidate's degree in Physics and Mathematics at the A.A. Dorodnitsyn Computational Center (Russian Academy of Sciences) by speciality 05.13.17 "Theoretical Foundations of Informatics" in 2007. Scientific interests: algebra of spectral transforms, theoretical informatics. Author of more than 50 papers in Russian and foreign languages in various peer-reviewed journals.