Skolkovo Institute of Science and Technology

as a manuscript

Andrei Chertkov

TENSOR METHODS FOR MULTIDIMENSIONAL DIFFERENTIAL EQUATIONS

Ph.D. Dissertation Summary

for the purpose of obtaining academic degree
Doctor of Philosophy in Computer Science

Moscow — 2023

The Ph.D. dissertation was prepared at Skolkovo Institute of Science and Technology, Center for Artificial Intelligence Technology.

Academic Supervisor: Ivan V. Oseledets,
Doctor of Science, Full Professor, Head of the Center. Skolkovo Institute of Science and Technology, Center for Artificial Intelligence Technology

Introduction

Topic relevance. Tensor train decomposition (hereinafter TT) is a common way for the compact representation of multidimensional arrays (tensors) with the ability to solve the curse of dimensionality in the context of memory consumption and computational complexity. The TT-decomposition makes it possible to obtain a low-parameter representation of a tensor as an ordered set of small three-dimensional tensors, called TT-cores within the framework of computationally efficient procedures. The original tensor, which has N^d elements (d is the number of dimensions, N is the number of elements in each dimension), is transformed with the TT-decomposition into a factorized representation that has only $d \cdot N \cdot R^2$ parameters, where R is the characteristic rank of the TT-decomposition. Many algebraic operations (summation, multiplication, convolution, integration, solving systems of linear equations, etc.) on tensors represented in the TT-format can also be performed with linear complexity in d and N if the TT-rank R is bounded. For vectors and matrices, a similar compact representation can also be obtained by first converting the corresponding one- or two-dimensional tensor to an essentially multidimensional tensor using the quantized tensor train approach (hereinafter QTT). For example, a vector of length $N = 2^d$ can be converted into a d-dimensional tensor with 2 elements in each dimension, and then the TT-decomposition for this tensor has only $2 \cdot log_2 N \cdot R^2$ parameters.

Due to the unique properties noted above, TT-based numerical methods have become extremely popular in a wide range of applications, including computational linear algebra, data analysis, physical process modeling, and machine learning. However, for the successful practical application of these methods, it is necessary to improve the corresponding general algorithms and adapt them for specific subject problems, including in the following important areas, which are considered in this work: the development of new computationally efficient implementations of basic algorithms in the TT-format; the development of new approaches and heuristics for the optimal choice of the initial guess in approximation methods based on the TT-format; the creation of new optimization methods for multidimensional arrays and multivariable functions in the TT-format.

At the same time, an important specific area of application of the TT-decomposition is the solution of differential equations. As shown in this work, it is possible to develop algorithms based on the TT-format for solving partial differential equations, which are efficient both in terms of memory consumption and computational speed. In this case, the differential operator, the coefficients, and the right-hand side of the equation are presented in the compressed form in the TT-format, and the solution, respectively, is also obtained in the compressed form. If by its nature the solution has a low-rank structure, then with the TT-format it becomes possible to solve essentially multidimensional problems or use a fine grid when discretizing the differential operator for one-dimensional and two-dimensional problems within the framework of the QTT-approach.

One-dimensional and two-dimensional partial differential equations considered in this work are used in many problems of physical and chemical modeling, for example, they describe the flow of a liquid in porous media. The corresponding coefficients in such equations often turn out to be multiscale or oscillating, and it becomes necessary to use very fine grids to resolve all scales. Therefore, the application of the approach based on the QTT/TT-decomposition for this class of problems seems to be useful, however, classical discretization schemes, such as finite differences and finite elements, can be unstable on fine computational grids in the TT-format, and the development of alternative specialized discretization schemes most suitable for implementation within the TT-format is required.

As a specific example of a multidimensional differential equation, we consider the Fokker-Planck equation, which plays an important role in the study of the dynamical systems. In recent years this equation has become widespread in the framework of machine learning in the context of problems of estimating the density of data distribution, training generative diffusion models, etc. One of the main difficulties in solving the Fokker-Planck equation is the high dimension of practically significant computational problems. The complexity of using grid representations of the solution increases exponentially with the dimension, so low-parameter representations are required, and the use of the TT-format for this problem is promising.

The goal of this dissertation is to develop new efficient TT-based methods for working with large data arrays and apply them to the solution of differential equations. To achieve the goal, it was necessary to solve the following **tasks**:

1. Develop a new efficient method for approximating multidimensional data arrays and multivariable functions based on the TT-decomposition;
2. Develop a new efficient method for optimizing multidimensional data arrays and multivariable functions based on the TT-decomposition;
3. Develop a new efficient method for solving partial differential equations on fine computational grids based on the QTT/TT-decomposition;
4. Develop a new efficient method for solving the multidimensional Fokker-Planck equation based on the TT-decomposition.

The novelty:

1. A new TT-ANOVA-ALS method based on the TT-decomposition for approximating multidimensional arrays and multivariable functions is proposed;
2. A new Optima-TT method based on the TT-decomposition for optimizing multidimensional arrays and multivariable functions presented in the TT-format is proposed;
3. A new TTOpt method based on the TT-decomposition for optimizing multidimensional arrays and multivariable functions is proposed;
4. A new FS-QTT method based on TT-decomposition for solving one-dimensional and two-dimensional partial differential equations on fine computational grids is proposed;

5. A new FPCross method based on the TT-decomposition for solving the multidimensional Fokker-Planck equation is proposed.

Practical significance of the work lies in the creation of the following open source software products:

1. **teneva**[1] – a framework that implements an extensive set of methods in the TT-format for approximation (including the TT-ANOVA-ALS method), optimization (including the Optima-TT method), analysis and use of multidimensional arrays and multivariable functions;
2. **ttopt**[2] – a library that implements the new TTOpt method for optimizing multidimensional arrays and multivariable functions;
3. **qttpdesolver**[3] – a library that implements the new FS-QTT method for solving partial differential equations;
4. **fpcross**[4] – a library that implements the new FPCross method for solving the multidimensional Fokker-Planck equation.

Note that some of the results on the topic of the dissertation were used with the participation of the author in commercial projects:

1. Brain and information: from natural intelligence to artificial intelligence (2020 - to present, Institute for advanced study of the brain);
2. Acceleration of calculations using tensor methods (2019 – 2022, Gazmpromneft NTC).

Also, some of the results were used in the framework of the author's work on grants from the Ministry of science and higher education of the Russian Federation:

1. Multiscale intelligent neurodynamic systems for multidimensional optimization in machine learning and data processing (2021 – to present, "megagrant");
2. Tensor networks and deep learning for data mining (2016 – 2021, "megagrant");
3. QTT-technology for solving multi-scale problems (2015 – 2016, collaboration with a group of prof. C. Schwab, ETH Zurich).

Main provisions to be defended:

1. TT-ANOVA-ALS method. A first-order ANOVA representation in the TT-format (TT-ANOVA) is constructed and its usage as an initial approximation for the TT-ALS method for approximating multidimensional arrays and multivariable functions in the TT-format is proposed. The performed calculations for a number of model problems, including the problem of approximating a parametric partial differential equation, demonstrate a significant advantage of the proposed method compared to the standard approach based on the TT-ALS method with a random initial approximation;

[1] See the repository `https://github.com/AndreiChertkov/teneva`.
[2] See the repository `https://github.com/AndreiChertkov/ttopt`.
[3] See the repository `https://github.com/AndreiChertkov/qttpdesolver`.
[4] See the repository `https://github.com/AndreiChertkov/fpcross`.

2. Optima-TT method. A new method has been developed that makes it possible to find the minimum and maximum elements of a tensor specified in the TT-format within the framework of successive tensor multiplication of TT-cores with a special selection of potential candidates for the optimum. A probabilistic interpretation of the method has been constructed, theoretical estimates for its complexity and convergence have been formulated, and extensive numerical experiments have been carried out with random tensors and various model functions with input dimensions up to 100;
3. TTOpt method. A new gradient-free optimization method based on a combination of low-rank tensor representation and the principle of maximum volume for matrices has been developed. The applicability of the method for a wide range of model problems is demonstrated and its advantage is shown in comparison with alternative approaches to optimization, including genetic algorithms and evolutionary strategies;
4. FS-QTT method. An efficient solver for 1D and 2D stationary diffusion equations has been developed based on the TT-decomposition and a proposed new robust discretization scheme that allows the use of fine grids with extremely high spatial resolution. Numerical experiments show that this scheme gives high accuracy results and can be used for grids containing up to 2^{60} nodes in the two-dimensional case;
5. FPCross method. A new numerical scheme for solving the multidimensional Fokker-Planck equation based on multidimensional Chebyshev interpolation, spectral differentiation, splitting method and TT-decomposition is proposed. The effectiveness of the proposed approach is demonstrated for a number of applied problems, including the multidimensional Ornstein-Uhlenbeck process, modeling of the molecular structure of liquid polymers, and analysis of transport optimality in diffusion models of machine learning.

Personal contribution. All the main results of the dissertation were obtained personally by the author. However, he acknowledges a valuable contribution of his colleagues: Ph.D. Gleb Ryzhakov (discussion of the ideas behind the TT-ANOVA-ALS and Optima-TT methods), Ph.D. Maxim Rakhuba (discussions on the FS-QTT method's connection with the finite-difference scheme), Ph.D. Valentin Khrulkov (theoretical substantiation of the hypothesis of optimal transport in diffusion models of machine learning), Ph.D. Roman Schutski (discussion of the ways for further development of the TTOpt method), Ph.D. student Konstantin Sozykin (practical implementation of the TTOpt method for reinforcement learning problems) and Ph.D. student Georgii Novikov (discussion of properties of the Optima-TT method).

Publications. The research results are presented in the following works.

First-tier publications.

1. **A. Chertkov**, G. Ryzhakov, I. Oseledets. Black box approximation in the tensor train format initialized by ANOVA decomposition. The work has been accepted for publication in SIAM Journal on Scientific Computing, 2023 (quartile of journal Q1 Scopus; [A1]);

2. V. Khrulkov, G. Ryzhakov, **A. Chertkov**, I. Oseledets. Understanding DDPM latent codes through optimal transport. In Proceedings of the International Conference on Learning Representations, 2023 (conference rating CORE A*; [A2]);
3. K. Sozykin*, **A. Chertkov***, R. Schutski, A. Phan, A. Cichocki, I. Oseledets. TTOpt: a maximum volume quantized tensor train-based optimization and its application to reinforcement learning. In Proceedings of the Advances in Neural Information Processing Systems, 2022 (conference rating CORE A*; the first two authors have an equal contribution to the work; [A3]);
4. **A. Chertkov**, I. Oseledets. Solution of the Fokker–Planck equation by cross approximation method in the tensor train format. Frontiers in Artificial Intelligence, 2021 (conference rating Q2 Scopus; [A4]).

Other publications.

1. I. Oseledets, M. Rakhuba, **A. Chertkov**. Black-box solver for multiscale modelling using the QTT format. In Proceedings of ECCOMAS [A5];
2. **A. Chertkov**, G. Ryzhakov, G. Novikov, I. Oseledets. Optimization of functions given in the tensor train format. arXiv, 2022 [A6];
3. A. Nikitin, **A. Chertkov**, R. Ballester-Ripoll, I. Oseledets, E. Frolov. Are quantum computers practical yet? A case for feature selection in recommender systems using tensor networks. arXiv, 2022 [A7];
4. **A. Chertkov**, I. Oseledets, M. Rakhuba. Robust discretization in quantized tensor train format for elliptic problems in two dimensions. arXiv, 2016 [A8].

Approbation of the work. The results were reported at conferences:

1. Visualization of functioning and stability analysis of artificial neural networks. Lomonosov Readings, Moscow State University. Moscow, 2023;
2. TTOpt: A maximum volume quantized tensor train-based optimization. Fall into ML Conference, HSE. Moscow, 2022;
3. TTOpt: A maximum volume quantized tensor train-based optimization and its application to reinforcement learning. NeurIPS. Online, 2022;
4. Quantized tensor train decomposition for solution of multiscale partial differential equations. Lomonosov Conference, Moscow State University. Moscow, 2017;
5. Tensor methods for multiscale modeling. Gen-Y Conference. Sochi, 2017;
6. Quantized Tensor Train Decomposition for Multiscale Modeling. 1st Annual Workshop of Skoltech/MIT Next Generation Program. Moscow, 2016;
7. Black-box solver for multiscale modelling using the QTT format. ECCOMAS Congress. Crete, 2016;
8. A tensor train approximation in active subspace variables with application to parametric partial differential equations. International Conference on Matrix Methods in Mathematics and Applications. Moscow, 2015.

Volume and structure of the work. The dissertation consists of an introduction, 4 chapters and a conclusion. The total volume of the dissertation is 161 pages, including 27 figures and 13 tables. The list of references contains 162 titles.

Content of the work

Introduction substantiates the relevance of research conducted within the framework of this dissertation work, formulates the goal, sets the tasks of the work, outlines its scientific novelty and practical significance.

The first chapter is an introductory one and is devoted to the study of the TT-decomposition and the main methods of working with multidimensional arrays in the TT-format, which are implemented as part of the teneva framework. In the introductory **Section 1.1**, we show that the TT-decomposition proposed in 2011 in [9] found further applications for approximating multidimensional and parameter-dependent integrals, for calculating multidimensional convolutions, for approximating Green functions of multivariate differential equations, for solving computational problems in the field of financial mathematics, for processing audio and video, for accelerating and compressing artificial neural networks, and even for building new machine learning algorithms directly based on the TT-decomposition. The TT-decomposition has been used by technology companies to speed up artificial neural networks, optimize cell tower placement, speed up field development planning algorithms, and optimize portfolio selection. Note that, unlike many machine learning methods, when using the TT-decomposition, we have mathematically justified algorithms based on proven convergence theorems, and the TT-decomposition has a simple intuitive form, i.e., an ordered set of three-dimensional arrays, each of which represents a factorization for the corresponding dimension, which increases the level of model interpretability.

Section 1.2 introduces the notation used in this work, considers the paradigm of low-rank tensor approximations, provides a brief overview of the relevant methods, and identifies the main advantages of the TT-decomposition: the number of parameters in the TT-decomposition and the complexity of most mathematical operations within the TT-format is linear on the dimension of the tensor and the average size of its modes; there are stable numerical methods for constructing a TT-decomposition for an explicitly given tensor in the full format and for approximate representation of a tensor in the TT-format from a given or dynamically generated training data set; an extensive set of efficient algorithms for working with TT-tensors is available.

Section 1.3 discusses the basic properties of the TT-decomposition and the QTT-decomposition [10], as well as numerical methods based on them, including such operations as get, full, add, sub, mul, const, kron, orthogonalize, truncate, mul_scalar, mul_matvec, mul_matmat, lss, maxvol and maxvol_rect, which will be used in subsequent chapters of the work when describing new developed approaches for approximation, optimization and solution of differential equations. For each operation, an estimate for computational complexity is formulated and implementation features are considered.

Of particular note is the get operation, which returns the value of the TT-tensor for the requested multi-index and represents the actual definition of the TT-decomposition, i.e., a tensor $\mathcal{Y} \in \mathbb{R}^{N_1 \times N_2 \times \ldots \times N_d}$ is in the TT-format if all its

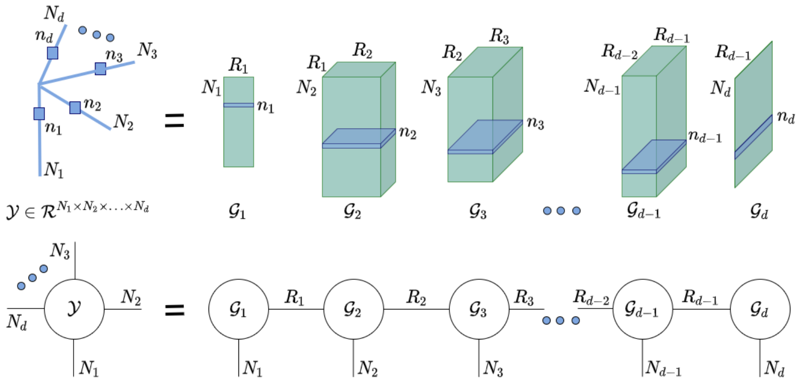

Fig. 1 — Schematic illustration of the low-rank TT-decomposition.

elements ($n_k = 1, 2, \ldots, N_k$; $k = 1, 2, \ldots, d$) are expressed by the following formula (function get):

$$\mathcal{Y}[n_1, n_2, \ldots, n_d] = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \ldots \sum_{r_{d-1}=1}^{R_{d-1}} \mathcal{G}_1[1, n_1, r_1] \mathcal{G}_2[r_1, n_2, r_2] \ldots \mathcal{G}_{d-1}[r_{d-2}, n_{d-1}, r_{d-1}] \mathcal{G}_d[r_{d-1}, n_d, 1], \quad (1)$$

where the three-dimensional tensors $\mathcal{G}_k \in \mathbb{R}^{R_{k-1} \times N_k \times R_k}$ are called TT-cores, and the natural numbers $R_0, R_1, \ldots, R_d$ (with the convention $R_0 = R_d = 1$) are called TT-ranks. In Figure 1 in the upper part we give a graphical illustration of the formula (1), and in the lower part we depict the corresponding tensor diagram. As follows from the above formula (1), storage of TT-cores $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_d$ requires no more than $d \times \max_{1 \le k \le d} \left(N_k R_k^2 \right) \sim d \cdot \overline{N} \cdot \overline{R}^2$ memory cells, where $\overline{N}$ and $\overline{R}$ are the effective ("average") mode size and effective TT-rank, respectively, while storing the full tensor would require $\overline{N}^d$ cells, and as a result, as noted above, the TT-decomposition is free from the curse of dimensionality if the TT-ranks are bounded.

Section 1.4 provides an overview of methods for constructing an approximate TT-decomposition for a given target tensor in some way, including a detailed discussion of the alternating least squares method in the tensor train format (hereinafter TT-ALS) [11] and the cross approximation method in the tensor train format (hereinafter TT-cross) [12], as well as the corresponding practical implementations in the form of functions: als, cross, cross_func and cross_act, which will be actively used further in the work.

Section 1.5 formulates the main conclusions on the first chapter, including a complete list of the considered methods and operations for constructing TT-tensors and working with tensors in the QTT/TT-format.

Chapter Two is devoted to a discussion of the developed new methods for approximating and optimizing multidimensional arrays and multivariable functions in the TT-format. In the introductory **section 2.1**, we show that many physical and engineering models can be represented in the form of a scalar function that depends on a multivariate argument:

$$y = \mathsf{f}(\boldsymbol{x}) \in \mathbb{R}, \quad \boldsymbol{x} = [x_1, x_2, \ldots, x_d]^T \in \Omega \subset \mathbb{R}^d. \tag{2}$$

Such functions often have the form of a black box (hereinafter BB), that is, their internal structure, smoothness properties, etc. turn out to be unknown. The time or amount of resources required to calculate the requested BB values (f function) can be significant, and therefore the approximation of such BB becomes relevant, that is, the construction of simplified (surrogate; low-parametric) models that can be calculated quickly, but with this remain fairly close to the considered BB (2). As argued in this chapter, the TT-decomposition is an efficient way to build surrogate models in the form of low-rank (low-parameter) tensor approximations for a wide class of multidimensional problems, as well as for solving optimization problems for discrete and continuous models of the form (2). The chapter discusses the proposed approximation method (TT-ANOVA-ALS) and two optimization methods (Optima-TT and TTOpt) for multidimensional arrays and multivariable functions given in the BB form. The developed approaches are based on the idea of applying modern algorithms of low-rank tensor approximations and QTT/TT-decomposition to a discretized BB on a multidimensional grid. The conducted numerical experiments demonstrate the significant advantages of these methods in comparison with alternative common approaches both in terms of the accuracy of the result and in terms of computational complexity (i.e., running time). Note that in numerical experiments for all developed methods (TT-ANOVA-ALS, Optima-TT, and TTOpt) we use a single set of model multivariable functions that are often used to test approximation and optimization algorithms, and the section provides a detailed description of them.

Section 2.2 discusses the proposed TT-ANOVA-ALS approach for efficient approximation of multidimensional arrays and multivariable functions in the presence of a training data set of a significantly limited size. In a discrete setting, surrogate modeling is equivalent to restoring a multidimensional array (tensor) from a small part of its elements. The TT-ALS method is currently a widely used approach for effectively solving this problem in the case of training on a given data set. TT-ALS allows to get a low-parameter representation of the tensor in the TT-format, free from the curse of dimensionality, which can then be used to quickly calculate values in arbitrary tensor indices, optimize it, build statistical characteristics, or to effectively implement various algebraic operations with BB (integration, solving systems of equations, convolutions, etc.). However, a good choice of initial approximation is required for high accuracy.

We built an ANOVA representation [13] in the TT-format and proposed its use as an initial approximation for the TT-ALS method. An illustration of the proposed approach is given in Figure 2. The numerical calculations performed for a number of multidimensional model problems, including for a parametric partial differential

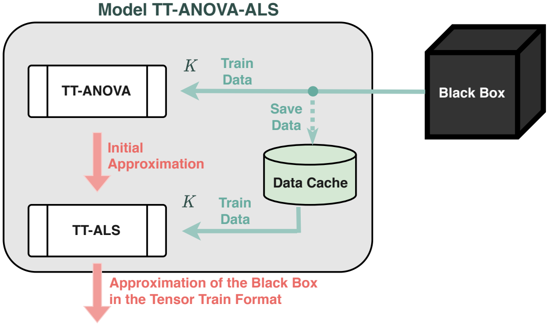

Fig. 2 — TT-ANOVA-ALS method for discrete BB approximation.

equation (hereinafter PDE), demonstrate a significant advantage of TT-ANOVA-ALS in comparison with the commonly used random initial approximation in the basic TT-ALS method. For all considered model problems, an increase in accuracy of at least an order of magnitude relative to the basic method was obtained with the same number of requests to the BB. Note that the proposed TT-ANOVA-ALS approach is very general and can be applied to a wide class of real problems of surrogate modeling and machine learning. The main results are presented in our paper [A1], and the corresponding program code is available in the teneva framework.

Section 2.3 discusses the second very common problem that arises when considering BBs of the form (2), which is the gradient-free optimization of multivariable functions (discrete and continuous). It is assumed that the values of the function (BB) can be calculated at any requested point, or a ready-made training data set is given and it is required to approximately find the minimum or maximum value of the function, spending as few resources as possible, that is, making as few requests to the BB as possible. There is no information about the internal structure of the function within the framework of such a statement, and, for example, the explicit use of classical gradient methods turns out to be difficult. Note that at least two approaches to solving this problem are possible: preliminary approximation of BB with subsequent optimization of the resulting surrogate model, or direct optimization of BB using some intellectual heuristics. The second approach is discussed in the next section, and in this section, the first approach is considered, which assumes the presence of an approximation of BB in the TT-format, that is, the problem of optimizing the TT-tensor itself is posed.

To date, there are practically no methods for finding the minimum and maximum element for a given TT-tensor. We have proposed a new Optima-TT algorithm (see the illustration in Figure 3), which makes it possible to obtain a very accurate approximation for the optimum within the framework of successive tensor multiplication of TT-cores with intelligent selection of potential candidates for the optimum. A probabilistic interpretation of the method was also built, theoretical estimates of its complexity and convergence were made, and extensive numerical

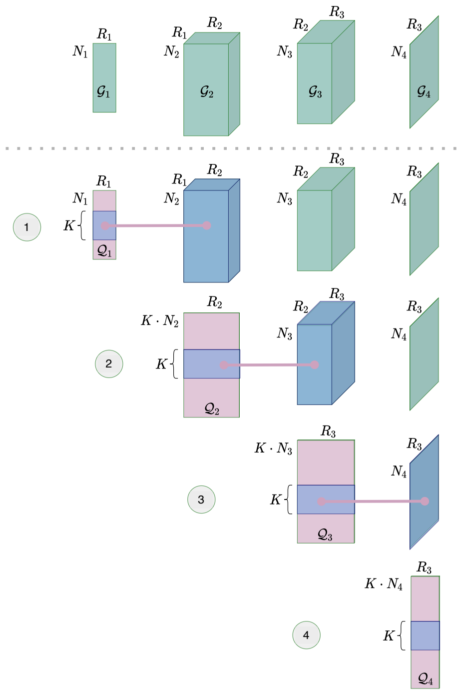

Fig. 3 — Schematic illustration of the proposed Optima-TT method for the case of a 4-dimensional TT-tensor. The rows selected during iterations are depicted as consecutive (in reality, this is not the case).

experiments were carried out with random tensors and various model functions with input dimensions up to 100. As follows from the results of the experiments, our approach leads to a solution close to the exact optimum for all model problems, while the running time of the algorithm is no more than 50 seconds on a regular laptop. The main results are presented in our paper [A6], and the corresponding program code is available in the teneva framework.

In **section 2.4**, we study a more general problem of direct application of low-rank tensor approximation methods to optimization of multidimensional arrays and multivariable functions without preliminary construction of a TT-approximation, which was carried out in the case of the Optima-TT method. The corresponding TTOpt method, developed by us on the basis of the QTT-decomposition and the generalized maximum volume principle for matrices, is discussed in detail in this section (see also the illustration in Figure 4). We demonstrate the efficiency of the new method for a number of multivariable function minimization problems and present the results of its comparison with various popular gradient-free and gradient-based methods. Compared to other approaches, TTOpt turns out to be one of the fastest (even taking into account the use of unoptimized python code in our software implementation) and

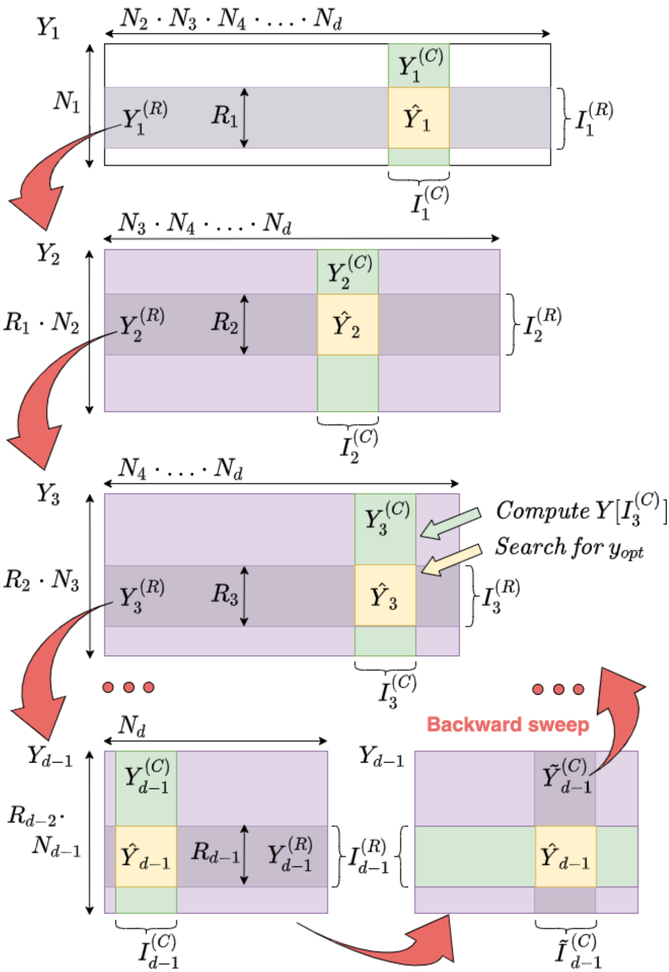

Fig. 4 — Conceptual diagram of the TTOpt algorithm for optimizing multidimensional arrays and multivariable functions, based on the alternating direction method and the maximum volume principle. The algorithm requires only a small part of the tensor elements to be computed (see columns and rows in green). For compactness of the illustration, the columns and rows selected at the iterations of the algorithm are depicted as following directly one after another.

stable (converging for all considered examples). The main results are presented in our work [A3], and the corresponding program code is available in the ttopt software product. Also in the paper [A7] we consider one more specific practical application of the proposed optimization method in the problem of training artificial intelligence recommender systems.

Section 2.5 formulates the main conclusions on the second chapter, and also indicates possible directions for further research, including the development of the TT-ANOVA-ALS method (the construction of a higher-order TT-ANOVA approximation, as well as the use of the TT-ANOVA approximation in other methods, for example, in the TT-cross algorithm), the Optima-TT method (complication of the procedure for bypassing TT-cores, in particular, organizing an iterative bypass of the cores «left to right» and «from center» with the appropriate combination of candidates and refinement of the resulting optimum) and the TTOpt method (combining the results of multiple runs with different initial approximations and additional iterations after combining), as well as the possible joint application of these methods for complex problems of multivariate approximation and optimization.

Chapter Three is devoted to the description of the proposed new FS-QTT method based on the QTT/TT-decomposition for solving D-dimensional ($D = 1, 2$) PDEs with homogeneous Dirichlet boundary conditions in a domain $\Omega = [0, 1]^D$:

$$-\nabla \left(\mathcal{K}(\boldsymbol{x})\nabla \mathsf{u}(\boldsymbol{x})\right) = \mathsf{f}(\boldsymbol{x}), \quad \mathsf{u}(\boldsymbol{x})|_{\partial\Omega} = 0, \tag{3}$$

where f is the given function in Ω, $\mathcal{K}$ is the diagonal diffusion tensor, and u is the desired solution. In the introductory **section 3.1**, we provide a general description of the proposed approach. Within the framework of the FS-QTT scheme, an implicit discretization of the considered PDE (3) with a dense matrix is constructed, which, however, can be compactly represented in the QTT-format. It allows to perform all operations with logarithmic complexity and memory consumption, including solution of corresponding systems of linear algebraic equations in the QTT/TT-format. As a result, we can use very fine grids and get very accurate solutions with extremely high spatial resolution, which is especially relevant for multiscale problems.

Indeed, if, for example, the components of the diffusion tensor $\mathcal{K}$ are rapidly oscillating and/or multiscale, then a dense grid is needed to be able to resolve the smallest scale of coefficient changes. This is a significant problem when solving such an equation explicitly, in particular, using finite element or finite difference grid methods. Today, approaches using methods based on analytical expansions, multiscale finite elements, etc. are widespread, but they usually require knowledge of the analytical properties of the solution, do not have a sufficient degree of generality, and have a high computational complexity.

To solve the indicated problems, low-rank tensor approximations can be used, and, since we usually have a small value of the dimension of the problem ($D = 1, 2, 3$), the quantization of the emerging vectors and matrices is also necessary, that is, the QTT-decomposition should be used. Note that the use of low-rank tensor approximations for solving multidimensional PDEs in the framework of standard discretization schemes is more common. However, the problem of solving PDE using tensor methods for the case of low dimension has also been considered in the literature. In the works [14—17] it was shown that the solution of the equation (3) theoretically allows an efficient representation in the QTT-format. In particular, in the paper [16] it was proved that for the PDE class with piecewise analytic coefficients, the exact solution u can be approximated with accuracy ϵ in the energy norm with $\mathcal{O}\left(\log^{\alpha} \epsilon^{-1}\right)$ degrees of freedom for $\alpha \leq 5$ within the finite element method in the QTT-format. This approach introduces a very fine spatial grid with the number of elements $N = 2^d$ ($d \gg 1$) in each dimension, which is capable of describing the smallest scale of the problem. Then, the corresponding vectors and matrices that appear when discretizing the right-hand side f, the elements of the diffusion tensor $\mathcal{K}$, and the differential operator are transformed into the QTT-format by introducing virtual dimensions. For example, in the one-dimensional case, the vector of discrete values for the right-hand side f can be considered as a d-dimensional tensor with each mode having a size of 2. The formalism described in [16] was successfully applied in [15] for the theoretical analysis of two classes of one-dimensional multiscale problems: two-scale diffusion and the Helmholtz equation with large wave numbers. It was proved that the solutions of these problems can be represented in the QTT-format with a polylogarithmic dependence of the number of parameters on the accuracy. Also in the work [18] a preconditioner in the QTT-format was proposed for these problems.

However, despite the fact that the solution of the equation (3) can be theoretically constructed in the QTT-format with a small TT-rank, real practical implementations, which are based on standard finite element or finite difference methods [14—16] turn out to be numerically unstable on fine grids due to an increase in the condition number and an increase in rounding errors. The FS-QTT discretization scheme proposed by us for one-dimensional and two-dimensional PDEs of the form (3) is free from this problem and preserves the second order of convergence for very fine grids, while the TT-ranks of the resulting vectors and matrices remain limited. The main results for the one-dimensional and two-dimensional problems are presented in our papers [A5] and [A8], respectively, and the program code is available in the qttpdesolver library.

Section 3.2 provides implementation details for the FS-QTT method for PDEs of the form (3) in the one-dimensional case. Using the variational formulation, we obtain an explicit formula for the desired PDE solution, which includes only element-wise operations and matrix multiplications, while all operations can be effectively implemented in the QTT-format. Then we show that in exact arithmetic the solution from the proposed formula is equivalent to the PDE solution obtained in the framework of the standard second-order finite difference scheme on a uniform grid. Next, we describe the practical aspects of implementing the method within the QTT/TT-format and show that the TT-ranks of the solution are limited by the product of the TT-ranks of the right-hand side of the PDE f and the reciprocal of the PDE coefficient $\mathcal{K}$.

Section 3.3 describes the proposed FS-QTT scheme for solving the PDE in the two-dimensional case. As in the previous section, we use the variational formulation of the equation (3) to construct a system of linear equations for the PDE solution, and then show that in exact arithmetic the reduced system corresponds to a second order finite difference scheme. Next, we present algorithms for the practical implementation of the scheme in the QTT/TT-format and formulate estimates for TT-ranks.

Section 3.4 discusses the results of the numerical experiments carried out for 1D and 2D problems, demonstrating the performance, stability, and robustness of the proposed FS-QTT scheme. First, two one-dimensional problems are considered, the first of which corresponds to an artificially generated PDE with a known analytical solution, and the second to a multiscale PDE. The FS-QTT scheme is then applied to solve two examples of 2D PDEs. We compare the results of all numerical calculations performed using the proposed FS-QTT method with the results of two solvers based on the standard finite difference scheme. The first baseline solver is implemented in the framework of the classical sparse format and can only work on moderately fine grids, and the second solver was developed by us using the QTT-decomposition applied directly to the finite difference scheme. The experiments show that the proposed solution method gives accurate results for various PDEs and can be effectively applied with up to 2^{60} grid nodes in the two-dimensional case, while the total computation time is only a few seconds. At the same time, solvers based on a finite difference scheme for sufficiently coarse grids give a result close in accuracy, but they become unstable on fine grids.

In the final **section 3.5**, general conclusions on the third chapter are given and, in particular, it is noted that the discretization scheme used in the FS-QTT solver can naturally be generalized to the three-dimensional case and to some other forms of equations, for example, on a PDE with periodic boundary conditions. Further development of the proposed approach involves a detailed analysis of the forms of the coefficients of the equation and the right-hand sides, which admit an effective low-rank representation. Multiscale problems are promising applications for the FS-QTT solver because they require a very fine grid to resolve all scales.

Chapter Four is devoted to the description of the developed FPCross method for solving the multidimensional Fokker-Planck equation (hereinafter FPE). In the introductory **section 4.1**, we present a general problem statement related to the consideration of a stochastic dynamical system described by a stochastic differential equation (hereinafter SDE) of the form:

$$d\boldsymbol{x} = \mathbf{f}(\boldsymbol{x}, t)\, dt + S(\boldsymbol{x}, t)\, d\boldsymbol{\beta}, \quad d\boldsymbol{\beta}\, d\boldsymbol{\beta}^\top = Q(t)\, dt, \quad \boldsymbol{x} = \boldsymbol{x}(t) \in \mathbb{R}^d, \tag{4}$$

where $d\boldsymbol{\beta}$ is q-dimensional space-time white noise, $\mathbf{f}$ is a known d-dimensional vector function, and $S \in \mathbb{R}^{d \times q}$ and $Q \in \mathbb{R}^{q \times q}$ are known matrices. The evolution of the probability density function (hereinafter PDF) $\rho(\boldsymbol{x}, t)$ for the spatial variable $\boldsymbol{x}$ is described by the corresponding FPE:

$$\frac{\partial \rho(\boldsymbol{x}, t)}{\partial t} = \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial}{\partial \boldsymbol{x}_i} \frac{\partial}{\partial \boldsymbol{x}_j} \left[D_{ij}(\boldsymbol{x}, t) \rho(\boldsymbol{x}, t) \right] - \sum_{i=1}^{d} \frac{\partial}{\partial \boldsymbol{x}_i} \left[\mathbf{f}_i(\boldsymbol{x}, t) \rho(\boldsymbol{x}, t) \right], \tag{5}$$

where $D(\boldsymbol{x}, t) = \frac{1}{2} S(\boldsymbol{x}, t) Q(t) S^\top(\boldsymbol{x}, t)$, is the diffusion tensor. FPE is used to study the properties of various dynamic systems, and in recent years this equation has gained particular popularity in machine learning for recovering distribution densities in the context of neural ordinary differential equations (hereinafter ODE), generative models, diffusion models, etc. The problem of analyzing the propagation of uncertainty in nonlinear dynamic systems subject to stochastic excitation has been actively studied in the literature in recent years. A number of numerical methods have been developed to solve the FPE, such as the method of integrals over directions, finite difference methods, and finite element methods. However, all these methods require the use of spatial grids, and fine grids are needed to obtain a sufficiently accurate solution, which leads to an exponential increase in computational complexity with an increase in the dimension of the problem. Note that alternative approaches based on the Monte Carlo method have a low convergence rate, which also limits the possibility of their use in the multidimensional case.

Thus, to solve the FPE in the multidimensional case, specialized methods are needed that are free from the curse of dimensionality (an exponential increase in computational complexity with an increase in the dimension of the problem d), and it seems promising to use low-rank tensor approximation methods. However, for FPE, this approach is not yet widely adopted. We note the works [19—23], where it is proposed to use the TT-decomposition to solve the multidimensional FPE. In these works, the differential operator from the equation (5) and the right-hand side $\mathbf{f}$ from (4) are represented as a TT-tensor, and in the work [19] joint space-time discretization of the solution in the TT-format is also considered. Accordingly, the equation is discretized on a tensor grid, and the solution $\rho(\boldsymbol{x}, t)$ takes the form of a d-dimensional tensor, which is approximated by a low-rank TT-tensor. Even with this approach, calculations often take too much time, and a restriction is also introduced on the right-hand side $\mathbf{f}$, i.e., it must allow low-rank approximation, which is not always fulfilled in real practical applications, for example, the role of the right-hand side in diffusion probabilistic models are usually played by some artificial neural network.

We propose a new way of using the TT-format to solve multidimensional FPE based on the close connection of this equation with dynamical systems. The difference between our approach and the works mentioned above is its more explicit iterative form for time integration, as well as the absence of the need to represent the right-hand side of the system in a low-rank format, which makes it possible to use this approach in machine learning applications. Let us illustrate the key idea underlying the new method for a special case without a stochastic component ($S \equiv 0$). It can be shown that the evolution of the PDF along the trajectory is given by the equation:

$$\frac{\partial \rho(\boldsymbol{x}, t)}{\partial t} = -\operatorname{Tr}\left(\frac{\partial \mathbf{f}(\boldsymbol{x}, t)}{\partial \boldsymbol{x}}\right) \rho(\boldsymbol{x}, t), \tag{6}$$

where $\operatorname{Tr}(\,\cdot\,)$ denotes the operation of taking the trace of a matrix. To calculate the value of $\rho(\boldsymbol{x}, t)$ at a particular point $\boldsymbol{x} = \widehat{\boldsymbol{x}}$, it suffices to find such a preimage $\widehat{\boldsymbol{x}}_0$ of this point that, when used as an initial condition for (4), the solution at time t will be equal to $\widehat{\boldsymbol{x}}$. To find the pre-image, we need to integrate the equation (4) (without the stochastic part) backward in time, and to find the PDF value, we then jointly integrate the equations (4) and (6). Since we can calculate the value of $\rho(\boldsymbol{x}, t)$ for any $\widehat{\boldsymbol{x}}$, it is possible to use the TT-cross method to recover a low-rank approximation from a small adaptively selected sample. Thus, we only solve the corresponding ODE numerically, without requiring any compact representation for the **f** function. For the case $S \neq 0$, the situation turns out to be more complicated, but the developed FPCross approach based on the splitting scheme, spectral differentiation techniques, multidimensional Chebyshev interpolation in the TT-format, and the TT-cross method allows us to effectively recalculate the density values from a certain moment time t to the next step $t + h$. The combined use of the indicated methods made it possible to create a highly efficient solver that can be used for a wide class of multivariate applications, including the restoration of distribution densities and probabilistic problems in machine learning. The main results are presented in our papers [A4] and [A2], and the corresponding program code is available in the fpcross library.

In **section 4.2** we provide a detailed description of the proposed method for solving the multivariate FPE. We use the standard splitting method for second-order operators, which allows us to independently solve the diffusion and convection parts of the original equation (5). We then discuss a method for interpolating the solution at each time step using Chebyshev grid nodes, Chebyshev polynomials, and the fast Fourier transform. To solve the diffusion equations at each time step on a Chebyshev grid, we discretize the Laplace operator using the differential Chebyshev matrices and compute the corresponding matrix exponential. To solve the convection equations, we use an interpolant for solving the diffusion equation and a special integration scheme along the trajectory of the corresponding ODE.

In **Section 4.3**, we describe the implementation of the proposed approach for solving FPE using the TT-decomposition and the TT-cross method. As it turns out, solving the diffusion equation (which reduces to calculating the matrix exponential) and convection equation (which reduces to iteratively solving the corresponding equation for each grid point), as well as interpolating the solution (which reduces to calculating the fast Fourier transform for each dimension) can be efficiently performed within the framework of the TT-format with linear complexity in terms of the dimension of the problem.

In **section 4.4**, the formulations of specific model problems and the results of numerical experiments with the proposed new FPCross solver in the TT-format for multivariate FPE are presented. First, we consider the solution of the equation with a linear convection part corresponding to the Ornstein-Uhlenbeck process (hereinafter OUP) in one-, three- and five-dimensional cases. Note that the one-dimensional problem is solved without using the TT-format, so the corresponding results are relevant only for checking the general correctness and convergence properties of the scheme, but not its efficiency. In the case of multidimensional problems, we use the proposed tensor-based solver. To check the correctness of the results, we use the well-known analytical stationary solution for OUP, and for the one-dimensional case we also compare it with the constructed analytical solution at an arbitrary time moment. Next, we consider a more complex «dumbbell» [24] problem, which arises, in particular, when considering the molecular structure of liquid polymers and can be represented as a three-dimensional FPE with a nonlinear convection part. For this case, we calculate the Cramer expression and compare our result with the results from [24] and [19]. Then we consider a specific application of the developed FPE solver to investigate the properties of diffusion machine learning models and their relationship with the optimal Monge transport map [A2] using synthetic data sets of dimensions 2, 3 and 7. For all the considered problems, the FPCross method gave a fairly accurate result (the relative error of the solution was about 10^{-3} for most problems).

In the final **section 4.5**, general conclusions are given on the fourth chapter of the work, including the fact that the proposed approach can be used for the numerical analysis of uncertainties in nonlinear dynamic systems subject to stochastic excitations, as well as in a wide range of modern machine learning problems related to with the restoration of probability distributions of dynamically noisy quantities, including neural ordinary differential equations, generative and diffusion models. The following are noted as possible ways for further development of tensor methods for solving multidimensional FPE: the method can be generalized to the case of FPE with an arbitrary diffusion coefficient (in this chapter, only FPEs with a scalar diffusion coefficient were considered); the method can be generalized to fractional FPEs, i.e. to FPEs having a form similar to (5), but with a fractional Laplacian power; additional acceleration and/or accuracy improvement of the proposed method is possible by considering alternative splitting schemes and ODE solution methods, as well as by more accurate selection of heuristics for choosing hyperparameters of the TT-cross method and associated functions in the TT-format.

The **conclusion** to the dissertation work briefly formulates our main results:

1. A new TT-ANOVA-ALS method based on the TT-decomposition for the approximation of multidimensional data arrays and multivariable functions is developed;
2. A new Optima-TT method based on the TT-decomposition is developed for optimizing multidimensional data arrays and multivariable functions presented in the TT-format;
3. A new TTOpt method based on the TT-decomposition is developed for optimizing multidimensional data arrays and multivariable functions;
4. A new FS-QTT method based on the TT-decomposition is developed for solving one- and two-dimensional PDEs on fine computational grids;
5. A new FPCross method based on the TT-decomposition for solving multivariate FPE is developed;
6. Teneva, ttopt, qttpdesolver and fpcross software products have been developed that implement the proposed methods in the python language;
7. Extensive numerical experiments have been carried out for each of the proposed new methods, demonstrating their correctness, efficiency and advantages over alternative approaches.

Author’s publications on the topic of the thesis

A1. *Chertkov*, *A.* Black box approximation in the tensor train format initialized by ANOVA decomposition [Текст] / A. Chertkov, G. Ryzhakov, I. Oseledets // arXiv preprint arXiv:2208.03380 (accepted to SIAM Journal on Scientific Computing). — 2023.

A2. Understanding DDPM latent codes through optimal transport [Текст] / V. Khrulkov [и др.] // 11th International Conference on Learning Representations, ICLR. — 2023.

A3. TTOpt: a maximum volume quantized tensor train-based optimization and its application to reinforcement learning [Текст] / K. Sozykin [и др.] // Advances in Neural Information Processing Systems. — 2022.

A4. *Chertkov*, *A.* Solution of the Fokker–Planck equation by cross approximation method in the tensor train format [Текст] / A. Chertkov, I. Oseledets // Frontiers in Artificial Intelligence. — 2021. — Т. 4.

A5. *Oseledets*, *I.* Black-box solver for multiscale modelling using the QTT format [Текст] / I. Oseledets, M. Rakhuba, A. Chertkov // Proc. ECCOMAS 2016. Crete Island, Greece. — 2016.

A6. Optimization of functions given in the tensor train format [Текст] / A. Chertkov [и др.] // arXiv preprint arXiv:2209.14808. — 2022.

A7. Are quantum computers practical yet? A case for feature selection in recommender systems using tensor networks [Текст] / A. Nikitin [и др.] // arXiv preprint arXiv:2205.04490. — 2022.

A8. *Chertkov*, *A.* Robust discretization in quantized tensor train format for elliptic problems in two dimensions [Текст] / A. Chertkov, I. Oseledets, M. Rakhuba // arXiv preprint arXiv:1612.01166. — 2016.

References

9. *Oseledets*, *I.* Tensor-train decomposition [Текст] / I. Oseledets // SIAM Journal on Scientific Computing. — 2011. — Т. 33, № 5. — С. 2295—2317.

10. *Oseledets*, *I.* Approximation of $2^d \times 2^d$ matrices using tensor decomposition [Текст] / I. Oseledets // SIAM Journal on Matrix Analysis and Applications. — 2010. — Т. 31, № 4. — С. 2130—2145.

11. *Holtz*, *S.* The alternating linear scheme for tensor optimization in the tensor train format [Текст] / S. Holtz, T. Rohwedder, R. Schneider // SIAM Journal on Scientific Computing. — 2012. — Т. 34, № 2. — A683—A713.

12. *Oseledets*, *I.* TT-cross approximation for multidimensional arrays [Текст] / I. Oseledets, E. Tyrtyshnikov // Linear Algebra and its Applications. — 2010. — Т. 432, № 1. — С. 70—88.

13. *Sobol*, *I.* Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates [Текст] / I. Sobol // Mathematics and computers in simulation. — 2001. — Т. 55, № 1—3. — С. 271—280.
14. QTT-FE approximation for multiscale problems [Текст] / V. Kazeev [и др.] // Research Report, Seminar for Applied Mathematics. — 2015.
15. QTT-finite-element approximation for multiscale problems I: model problems in one dimension [Текст] / V. Kazeev [и др.] // Advances in Computational Mathematics. — 2017. — Т. 43, № 2. — С. 411—442.
16. *Kazeev*, *V.* Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions [Текст] / V. Kazeev, C. Schwab // Numerische Mathematik. — 2018. — Т. 138, № 1. — С. 133—190.
17. Quantized tensor FEM for multiscale problems: diffusion problems in two and three dimensions [Текст] / V. Kazeev [и др.] // Multiscale Modeling & Simulation. — 2022. — Т. 20, № 3. — С. 893—935.
18. *Khoromskij*, *B.* A fast iteration method for solving elliptic problems with quasiperiodic coefficients [Текст] / B. Khoromskij, S. Repin // Russian Journal of Numerical Analysis and Mathematical Modelling. — 2015. — Т. 30, № 6. — С. 329—344.
19. *Dolgov*, *S.* Fast solution of parabolic problems in the tensor train/quantized tensor train format with initial application to the Fokker-Planck equation [Текст] / S. Dolgov, B. Khoromskij, I. Oseledets // SIAM Journal on Scientific Computing. — 2012. — Т. 34, № 6. — A3016—A3038.
20. *Sun*, *Y.* Numerical solution of high dimensional stationary Fokker-Planck equations via tensor decomposition and Chebyshev spectral differentiation [Текст] / Y. Sun, M. Kumar // Computers & Mathematics with Applications. — 2014. — Т. 67, № 10. — С. 1960—1977.
21. *Sun*, *Y.* A numerical solver for high dimensional transient Fokker-Planck equation in modeling polymeric fluids [Текст] / Y. Sun, M. Kumar // Journal of Computational Physics. — 2015. — Т. 289. — С. 149—168.
22. *Dolgov*, *S.* A tensor decomposition algorithm for large ODEs with conservation laws [Текст] / S. Dolgov // Computational Methods in Applied Mathematics. — 2019. — Т. 19, № 1. — С. 23—38.
23. Grid methods for Bayes-optimal continuous-discrete filtering and utilizing a functional tensor train representation [Текст] / C. Fox [и др.] // Inverse Problems in Science and Engineering. — 2020. — С. 1—19.
24. *Venkiteswaran*, *G.* A QMC approach for high dimensional Fokker-Planck equations modelling polymeric liquids [Текст] / G. Venkiteswaran, M. Junk // Mathematics and Computers in Simulation. — 2005. — Т. 68, № 1. — С. 43—56.