

На правах рукописи

Оселедец Иван Валерьевич

ВЫЧИСЛИТЕЛЬНЫЕ ТЕНЗОРНЫЕ МЕТОДЫ И ИХ
ПРИМЕНЕНИЯ

01.01.07 — Вычислительная математика

Автореферат
*диссертации на соискание учёной степени
доктора физико-математических наук*

Москва — 2012

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте вычислительной математики РАН.

Научный консультант: чл-корр. РАН, д.ф.-м.н., профессор Тыртышников Евгений Евгеньевич

Официальные оппоненты: Ильин Валерий Павлович, д.ф.-м.н, профессор (Институт вычислительной математики и математической геофизики СО РАН, г.н.с.),

Гутерман Александр Эмилевич д.ф.-м.н., профессор (Механико-математический факультет МГУ, кафедра высшей алгебры),

Хазанов Владимир Борисович, д.ф.-м.н., профессор (Санкт-Петербургский государственный морской технический университет, кафедра прикладной математики и математического моделирования).

Ведущая организация: Вычислительный центр им А.А. Дородницына РАН.

Защита состоится « 16 » марта 20 12 г. в 15 часов на заседании диссертационного совета Д 002.045.01 в Федеральном государственном бюджетном учреждении науки Институте вычислительной математики РАН по адресу: 119333, г. Москва, ул. Губкина, д. 8.

С диссертацией можно ознакомиться в библиотеке Федерального государственного бюджетного учреждения науки Института вычислительной математики РАН.

Автореферат разослан « » февраля 200 12 г.

Учёный секретарь
диссертационного совета Д 002.045.01
д.ф.-м.н.



Г. А. Бочаров

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы

Многомерные задачи возникают во многих приложениях. В вычислительной химии основным уравнением является уравнение Шредингера, в котором число независимых переменных растет линейно по числу электронов. Важными многомерными уравнениями математической физики являются уравнения Фоккера-Планка и уравнение Блэка-Шоулза в финансовой математике. При решении стохастических и многопараметрических задач искомая функция, зависящая от нескольких параметров, должна приближаться с использованием небольшого числа степеней свободы. В задачах сжатия и обработки данных возникает необходимость приближать многомерные массивы данных.

Во всех вышеперечисленных примерах необходимо строить малопараметрические приближения для функций многих переменных и связанных с ними многомерных массивов (тензоров). При этом число координатных осей тензора может быть огромным. Например, в вычислительной химии для молекул с N электронами или атомами функции, подлежащие определению (волновые функции, потенциальные поверхности), описываются $d = 3N$ параметрами. Использование простых сеток сразу приводит к массивам с числом элементов n^d , где n — число точек по одному направлению. Экспоненциальная зависимость числа параметров от числа координатных осей носит название «проклятие размерности». Тем не менее, во многих приложениях удается найти методы представления многомерных массивов. В частности, могут использоваться специальные базисы; например, на основе радиальных базисных функций. Однако такие схемы не являются универсальными, и для каждой конкретной задачи приходится создавать свои методы.

На данный момент не существует единого математического аппарата и подхода к построению надежных и эффективных вычислительных алгоритмов для работы с тензорами (вычислительных тензорных алгоритмов). Из алгебраических подходов можно выделить два. Они основаны на разделении дискретных переменных. Первый подход основан на так называемом каноническом разложении, а второй — на разложении Таккера. Они обладают рядом достоинств и достаточно успешно применялись для различных задач, но при этом обладают и существенными недостатками; разложение Таккера не может быть использовано для большого числа координатных осей, а кано-

ническое разложение нельзя вычислить с помощью надежных алгоритмов.

Алгебраические методы разделения переменных хорошо исследованы в двумерном случае. Это классические результаты матричного анализа, в основе которых лежит *сингулярное разложение матрицы*. Сингулярное разложение является устойчивым и существуют надежные алгоритмы для его вычисления, реализованные в стандартных библиотеках линейной алгебры. Однако, попытки напрямую перенести сингулярное разложение на случай массивов с тремя и более индексами приводят к существенным затруднениям. Поэтому создание новых устойчивых малопараметрических представлений для многомерных массивов и эффективных алгоритмов для работы с такими представлениям, является актуальной проблемой вычислительной математики.

Исследования, отражённые в диссертации, поддерживались грантами РФФИ №02-01-00590-а, 04-07-90336-в, 05-01-00721-а, 06-01-08052-офи, 08-01-00115-а, 09-01-00565-а, 09-01-12058-офи_м, 09-01-91332-ННИО_а, 11-01-00549-а, 11-01-12137-офи-м-2011, Программами ОМН-3 и ОМН-5, грантами МК-127.2009.1 и МК-140.2011.1 президента РФ молодым кандидатам наук, госконтрактами П940, П1178, П1112, 14.740.11.1067, 14.740.11.0345 в рамках ФЦП «Кадры».

Цели работы

Целью работы является создание и обоснование эффективных вычислительных методов для аппроксимации и работы с многомерными массивами (тензорами) и их применение для решения практически важных задач высокой размерности.

Методы исследования

Методами исследования являются классические результаты матричного анализа и вычислительной линейной алгебры (сингулярное разложение, QR-разложение, малоранговые аппроксимации с помощью скелетного разложения матриц), мультилинейная алгебра.

Научная новизна работы

Представленные результаты являются новыми и состоят в следующем.

- Предложено новое представление тензоров — ТТ-формат, которое может быть построено с помощью быстрых устойчивых алгоритмов на основе сингулярного разложения и не содержит экспоненциального по размерности числа параметров.

- Получены базовые алгоритмы для работы с массивами в ТТ-формате: алгоритм перехода от полного массива к ТТ-формату с точностью ε (алгоритм ТТ-SVD), алгоритм округления (приближения) в ТТ-формате. Доказано, что основные операции линейной алгебры (сложение, поэлементное умножение, умножение матрицы на вектор) можно проводить, оставаясь в рамках ТТ-формата; для этих операций получены явные формулы.
- Получено многомерное обобщение скелетного разложения — формула ТТ-интерполяции, которая показывает, что тензор с малыми ТТ-рангами можно восстановить по небольшому количеству его элементов. На основе этой формулы получен многомерный крестовый метод аппроксимации тензоров.
- Введено понятие *тензоризации* или QТТ-формата, когда исходный массив малой геометрической размерности представляется как массив большей размерности за счет введения виртуальных уровней. Доказаны теоремы о QТТ-структуре функций одной переменной, а также о QТТ-структуре характеристической функции многомерного симплекса. Показано, что QТТ-ранги обратной к ленточной теплицевой матрице не зависят от порядка матрицы.
- Предложен и реализован новый метод сжатия данных с потерями на основе интерпретации QТТ-формата как адаптивного вейвлет-преобразования.
- На основе QТТ-формата предложен новый быстрый метод решения многопараметрических задач.
- Получены оценки на QТТ-ранги полиномиальных потенциальных поверхностей квантовой молекулярной динамики, построен линейный по числу степеней свободы алгоритм приближенного нахождения минимального собственного значения на основе метода DMRG (Density Matrix Renormalization Group).

Теоретическая и практическая ценность

Теоретическая значимость работы заключается в создании и обосновании принципиального нового математического аппарата для работы с многомерными массивами. Основой предложенного подхода является Tensor Train -

разложение (ТТ-разложение) и введение дополнительных размерностей (тензоризация, QТТ-формат). Созданы и обоснованы вычислительные алгоритмы для работы с тензорами в ТТ и QТТ форматах.

Практическая значимость состоит в том, что построенный математический аппарат позволил создать новые вычислительные алгоритмы для решения различных практически важных задач (более эффективные, чем используемые ранее), а также позволил решить задачи, для которых применение стандартных методов невозможно в силу экспоненциальной зависимости от числа координатных осей.

В вычислительной химии удалось получить приближенные основные состояния для потенциала Эно-Хайлеса с числом степеней свободы вплоть до 256, а также вычислить основное колебательное состояние для различных молекул более точно, чем с помощью методов, реализованных в современном квантово-химическом пакете GAMESS. При решении многопараметрических и стохастических уравнений методы, основанные на ТТ-разложении, оказались в десятки раз эффективнее, чем стандартные методы (стохастический метод Галеркина и метод стохастической коллокации). Методы аппроксимации многомерных массивов были также применены для сжатия поля температуры, вычисленной с помощью σ -модели ИВМ РАН, удалось получить сжатие в 44 раза при абсолютной погрешности 0.1 градус Цельсия.

Обоснованность и достоверность результатов

Представленные в диссертации результаты имеют строгое математическое обоснование. Вычислительные алгоритмы основаны на применении классических процедур линейной алгебры и матричного анализа. Достоверность результатов основана на известных оценках сложности работы классических алгоритмов линейной алгебры и матричного анализа, а также на сравнении полученных результатов с результатами, полученными с помощью стандартных алгоритмов.

Апробация работы

Основные результаты докладывались на следующих конференциях и семинарах.

- Международные конференции «Матричные методы и операторные уравнения» (ИВМ РАН, 2005 и 2007);

- 3-я международная конференция «Математические идеи П.Л. Чебышёва и их приложение к современным проблемам естествознания» (Обнинск, ИАТЭ, 2006);
- Международные конференции по линейной алгебре ILAS-13 (Амстердам, 2006), ILAS-14 (Шанхай, 2007), ILAS-17 (Пиза, 2010), ILAS-18 (Брауншвейг, 2011);
- Международные конференции SCPDE (Гонконг, 2008, 2010);
- Международная конференция по структурированным матрицам (Кортена, 2008);
- Международная конференция по численному анализу и теории операторов (Хельсинки, 2008);
- Международный симпозиум GAMM (Гданьск, 2009);
- Workshop on Advanced Computer Simulation Methods for Junior Scientists (Санкт-Петербург, 2009);
- Международные семинары GAMM-26, GAMM-27;
- Международная научная конференция «Современные проблемы вычислительной математики и математической физики» памяти А.А. Самарского (Москва, 2009);
- Международная конференция МДОЗМФ (Херсон, 2009);
- Международная конференция NOLTA-2009 (Саппоро, 2009);
- Международная конференция AIM Workshop on computational optimization for tensor decompositions (Пало Альто, 2010);
- Международная конференция «Актуальные проблемы вычислительной математики и математического моделирования» посвященная 85-летию Г.И. Марчука (Новосибирск, 2010);
- Международная конференция Tensor Methods in multi-dimensional boundary-value and eigenvalue problems (Лейпциг, 2010);
- Международная конференция TDA (Монополи, 2010);

- Международная конференция Separation of variables and applications (Ницца, 2010);
- Школа-конференция «Теория и численные методы решения обратных и некорректных задач» (Новосибирск, 2010);
- Международная конференция HDA-2011 (Бонн, 2011);
- Международный симпозиум Householder-18 (Тахо, 2011);
- Всероссийский симпозиум по прикладной и промышленной математике (Кисловодск, 2006, 2008).
- Всероссийские конференции «Научный сервис в сети интернет» (Абрау-Дюрсо, 2006,2007,2008,2009,2011).
- XLIX научная конференция МФТИ (Москва, 2006);
- Ломоносовские чтения (Москва, 2008);
- Тихоновские чтения (Москва, 2006,2007, 2010);
- Семинар «Вычислительные и информационные технологии в математике» (руководители в разные годы Ю.М. Нечепуренко, В.И. Лебедев, Е.Е. Тыртышников, В. И. Агошков, А. Б. Богатырев, Ю.В. Василевский);
- Семинар кафедры теории функций и функционального анализа мехмата МГУ (2009, руководитель В.С. Кашин);
- Семинар кафедры алгебры мехмата МГУ (2009, руководитель В.Н. Латышев);
- Семинар ИПМ РАН им. Бабенко (2009);
- Семинар физфака МГУ и НИВЦ МГУ (2011, руководитель А.Г. Ягола);
- Семинар кафедры химии РХТУ им. Менделеева (2011, руководитель В.Г. Цирельсон);
- Семинар кафедры физической химии химфака МГУ (2011, руководитель А.Ю. Немухин);

Публикации

Основные результаты диссертации отражены в публикациях [1-33], из них 8 работ опубликовано в отечественных рецензируемых изданиях, рекомендованных ВАК, 16 работ опубликовано в международных рецензируемых изданиях из рекомендованного ВАК списка «Web of Science: Citation Index Expanded» (база данных по естественным наукам), 3 работы — в рецензируемых международных изданиях, 1 работа — в сборнике научных трудов. 5 работ опубликовано в других научных изданиях.

Вклад автора в совместные работы заключался: в формировании постановки проблемы [1, 6, 9, 13, 15, 18, 20, 22, 30, 33] предложении идеи решения [2, 6, 9, 10, 11, 12, 13, 16, 17, 18, 19, 20, 22, 24, 25, 26, 27, 31, 30, 33], теоретическом обосновании [6, 11, 12, 16, 17, 18, 27], совместном теоретическом обосновании [1, 2, 9, 10, 13, 15, 20, 24, 25, 26, 30, 33], технической реализации [2, 10, 11, 12, 16, 17, 18, 20, 24, 25, 26, 27], совместной технической реализации [13, 21, 30], постановке численных экспериментов [2, 10, 11, 12, 16, 17, 18, 20, 24, 25, 26, 27].

Благодарности

Автор выражает искреннюю благодарность своему научному консультанту проф., чл.-корр РАН Е.Е. Тыртышникову за постоянную поддержку, академику РАН В.П. Дымникову за поддержку и прекрасную атмосферу, созданную в ИВМ РАН, д.ф.-м.н, проф. В.И. Агошкову, д.ф.-м.н А.Б. Богатыреву, д.ф.-м.н Ю.В. Василевскому, д.ф.-м.н. Ю.М. Нечепуренко за очень полезные и стимулирующие обсуждения и предложения, повлекшие за собой существенную переработку работы, д.ф.-м.н Б.Н. Хоромскому за плодотворное сотрудничество.

Диссертация посвящена памяти моего дедушки, к.ф.-м.н, генерал-майора И.О. Бежаева. (1918-2010).

Структура и объём работы

Работа состоит из введения, четырёх глав и заключения. Объём диссертации — 206 страниц. Библиография включает в себя 158 наименований. Диссертация содержит 20 рисунков и 7 таблиц.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении отображена актуальность темы диссертационной работы, сформулированы основные цели и показана научная новизна и практическая зна-

чимость работы.

В **первой** главе вводится новое тензорное разложение — ТТ-разложение. ТТ-разложение лежит в основе всех разработанных в диссертации вычислительных тензорных методов. В разделе 1.1 формулируются требования к малопараметрическим представлениям тензоров (тензорным форматам). В разделе 1.2 рассматриваются два стандартных разложения тензоров: каноническое разложение и разложение Таккера, приводятся их свойства. Формулируются их недостатки и ставится основной вопрос:

можем ли мы найти такое разложение для d -мерных массивов, которое свободно от экспоненциальной зависимости от d , но при этом операции линейной алгебры и округление можно проводить с помощью устойчивых подходов на основе SVD?

В разделе 1.3 доказываются две основные леммы, которые позволяют построить такое представление.

Лемма 1 *Если тензор $A = [a(i_1, \dots, i_d)]$ имеет каноническое разложение ранга r , то для любой матрицы-развёртки $B = [a(\mathcal{I}_1, \mathcal{I}_2)]$*

$$\text{rank } B \leq r.$$

Лемма 2 *Пусть тензор $A = [A(i_1, \dots, i_d)]$ имеет каноническое представление ранга r и развёртка $B = [A(\mathcal{I}_1, \mathcal{I}_2)]$ определяется «длинными индексами» \mathcal{I}_1 и \mathcal{I}_2 , содержащими соответственно d_1 и d_2 исходных индексов. Рассмотрим произвольное разложение B вида*

$$B = UV^T, \quad U = [U(\mathcal{I}_1, \alpha)], \quad V = [V(\mathcal{I}_2, \alpha)],$$

где U и V имеют $\hat{r} = \text{rank } B$ столбцов. Тогда $U(\mathcal{I}_1, \alpha)$ и $V(\mathcal{I}_2, \alpha)$ могут быть рассмотрены как тензоры порядка $d_1 + 1$ и $d_2 + 1$ соответственно, и для обоих тензоров существуют канонические представления ранга не выше r .

На основании этих двух лемм в разделе 1.4. строится *иерархическое разложение Таккера*. Пример построения приведен на рис. 1.

Предположим, что для 9-мерного массива существует (неизвестное заранее) каноническое разложение ранга r . Тогда на основании леммы 1 его можно заменить на 6-мерный и 5-мерный массивы, которые по лемме 2 имеют каноническое представление ранга не выше r . Продолжая рекурсию, получим вместо исходного 9-мерного массива семь трехмерных массивов. Для числа параметров такого представления верна следующая теорема.

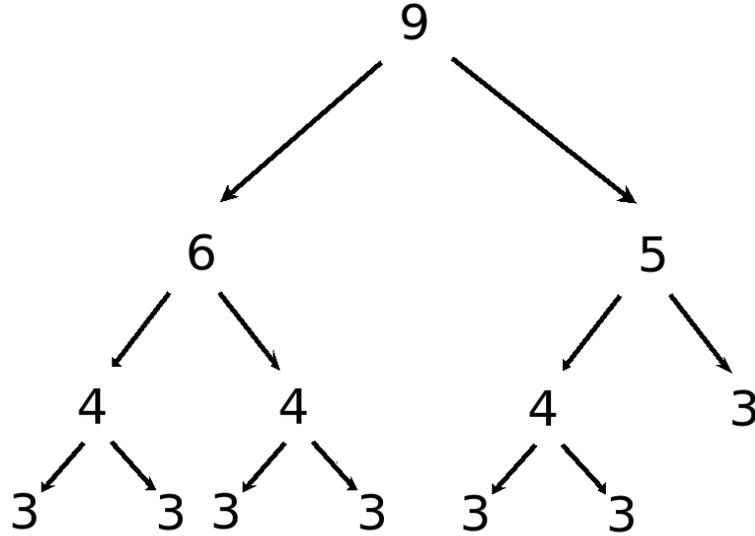


Рис. 1
Расщепление измерений.

Теорема 1 Пусть A — d -мерный тензор с каноническим разложением ранга r . Тогда существует ТТ-разложение с

$$nr d + (d - 2)r^3 \quad (1)$$

параметрами.

В разделе 1.5 все три упомянутых разложения (разложение Таккера, каноническое разложение, иерархическое разложение Таккера) рассматривается с точки зрения формализма тензорных сетей. Как частный случай иерархического разложения Таккера вводится tensor-train разложение (ТТ-разложение). Оно имеет вид

$$A(i_1, \dots, i_d) = \sum_{\alpha_1, \dots, \alpha_{d-1}} G_1(i_1, \alpha_1) G_2(\alpha_1, i_2, \alpha_2) \dots \dots G_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) G_d(\alpha_{d-1}, i_d), \quad (2)$$

Доказывается теорема, показывающая эквивалентность любого иерархического разложения Таккера и ТТ-разложения с точностью до некоторой перестановки индексов.

Теорема 2 Тензорное разложение, заданное произвольным деревом, удовлетворяющим ограничениям 1 и 2, можно представить в виде

$$A(i_1, \dots, i_d) = \sum_{\beta_1, \dots, \beta_{d-1}} G_1(\beta_0, i_{\sigma(1)}, \beta_1) G_2(\beta_1, i_{\sigma(2)}, \beta_2) \dots \dots G_{d-1}(\beta_{d-2}, i_{\sigma(d-1)}, \beta_{d-1}) G_d(\beta_{d-1}, i_{\sigma(d)}, \beta_d). \quad (3)$$

В разделе 1.6 исследуются свойства ТТ-разложения. Вводится компактная матричная форма разложения в виде произведения матриц, зависящих от параметров:

$$A(i_1, \dots, i_d) = G_1(i_1)G_2(i_2) \dots G_d(i_d), \quad (4)$$

где $G_k(i_k)$ — матрица размера $r_{k-1} \times r_k$, $r_0 = r_d = 1$. Доказана теорема, связывающая ТТ-ранги с рангами матриц-разверток тензора A . Матрица A_k называется k -ой разверткой тензора A , если

$$A_k(i_1 \dots i_k; i_{k+1} \dots i_d) = A(i_1, \dots, i_d), \quad (5)$$

т.е. первые k индексов нумеруют строки A_k , а последние $(d - k)$ — столбцы A_k .

Теорема 3 Если для каждой матрицы-развёртки A_k d -мерного тензора A вида (5) выполняется

$$\text{rank } A_k = r_k, \quad (6)$$

то существует разложение (4) с ТТ-рангами не выше r_k .

Доказательство теоремы 3 конструктивно и дает алгоритм построения ТТ-разложения — ТТ-SVD. Разложение является устойчивым, что показывает

Теорема 4 Для заданного тензора $A = [A(i_1, \dots, i_d)]$ существует ТТ-приближение $T = [T(i_1, \dots, i_d)]$ с ТТ-рангами r_k такими, что

$$\|A - T\|_F \leq \sqrt{\sum_{k=1}^{d-1} \varepsilon_k^2}, \quad (7)$$

где ε_k — расстояние (во фробениусовой норме) от A_k до наилучшего приближения ранга r_k :

$$\varepsilon_k = \min_{\text{rank } B \leq r_k} \|A_k - B\|_F.$$

В разделе 1.7 вводится и обосновывается быстрая процедура *тензорного округления*: для заданной точности ε и заданного тензора A в ТТ-формате с рангами $r_k \leq r$ требуется найти тензор B с минимально возможными рангами такими, что $\|A - B\|_F \leq \varepsilon \|B\|_F$. Вычислительная сложность такого алгоритма

— $\mathcal{O}(dnr^3)$ операций. В разделе 1.8 вводится ТТ-формат для матриц и доказывается теорема о связи кронекеровских рангов матрицы с ТТ-рангами. В разделе 1.9 систематически выводятся формулы для основных арифметических операций в ТТ-формате: сложение, умножение на число, поэлементное произведение, скалярное произведение, умножение матрицы на вектор, вычисление многомерной свертки. Например, если заданы два тензора A и B одинаковых размеров в ТТ-формате:

$$A(i_1, \dots, i_d) = A_1(i_1)A_2(i_2) \dots A_d(i_d), \quad B(i_1, \dots, i_d), B_1(i_1)B_2(i_2) \dots B_d(i_d),$$

то их поэлементное произведение $C = A \circ B$ также представимо в ТТ-формате с ядрами

$$C_k(i_k) = A_k(i_k) \otimes B_k(i_k).$$

Базовые процедуры имеют линейную по d сложность и результат получается также в ТТ-формате. В разделе 1.10 вводится понятие *функционального ТТ-разложения*, получены явные формулы для ГТТ-представления нескольких функций многих переменных, которые в дальнейшем будут использованы для получения оценок для числа параметров QТТ-формата. В разделах 1.11-1.12 получена *интерполяционная формула в ТТ-формате*. Доказана следующая теорема

Теорема 5 Пусть A — произвольный d -мерный тензор размера $n_1 \times \dots \times n_d$ с ТТ-рангами

$$r_k = \text{rank } A_k, \quad A_k = A(i_1 i_2 \dots i_k; i_{k+1} \dots i_d).$$

Предположим, что заданы правые индексные множества

$$\mathcal{J}_k = [j_l^{(\beta_l)}], \quad \beta_l = 1, \dots, r_k, \quad l = k, \dots, d-1,$$

и левые индексные множества

$$\mathcal{I}_k = [i_l^{(\alpha_l)}], \quad \alpha_l = 1, \dots, r_k, \quad l = 1, \dots, k-1,$$

так, что левые индексные множества образуют левую вложенную последовательность, и все $r_k \times r_k$ подматрицы на пересечении

$$\widehat{A}_k(\alpha_k, \beta_k) = A(i_1^{(\alpha_k)}, i_2^{(\alpha_k)}, \dots, i_k^{(\alpha_k)}, j_k^{(\beta_k)}, \dots, j_d^{(\beta_k)}), \quad \alpha_k, \beta_k = 1, \dots, r_k$$

невыврождены.

Тогда тензор A может быть восстановлен по трёхмерным тензорам C_k размера $r_{k-1} \times n_k \times r_k$ с элементами

$$C_k(\alpha_k, i_k, \beta_k) = A(i_1^{(\alpha_k)}, i_2^{(\alpha_k)}, \dots, i_{k-1}^{(\alpha_k)}, i_k, j_{k+1}^{(\beta_k)}, \dots, j_d^{(\beta_k)})$$

по интерполяционной формуле «тензорных поездов»

$$A(i_1, i_2, \dots, i_d) = \sum_{\alpha_1, \dots, \alpha_{d-1}} \widehat{C}_1(i_1, \alpha_1) \widehat{C}_2(\alpha_1, i_2, \alpha_2) \dots \widehat{C}_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) \widehat{C}_d(\alpha_{d-1}, i_d),$$

где тензорные вагоны \widehat{C}_k получаются из C_k при помощи умножения тензора на матрицу:

$$\widehat{C}_k = C_k \times_3 \widehat{A}_k^{-1}, \quad k = 2, \dots, d-1,$$

$$\widehat{C}_1 = C_1 \widehat{A}_1^{-1}, \quad \widehat{C}_d = C_d.$$

Смысл этой теоремы состоит в том, что если известно, что тензор имеет ранг r , то его можно точно восстановить по $\mathcal{O}(dnr^2)$ элементам. По аналогии с двумерным случаем, предложен адаптивный многомерный крестовый метод (ТТ-cross) для нахождения интерполяционных индексов, упомянутых в теореме 5. Основным вычислительным инструментом алгоритма ТТ-cross является алгоритм поиска подматрицы максимального объема. В разделе 1.13 приведены краткие выводы.

Глава 2 посвящена *QTT-разложению* (quantized tensor train), которое основано на идее виртуальных уровней или *квантизации*. В разделе 2.1 вводится понятие квантизации и дается определение QTT-формата. При квантизации заданный массив превращается в многомерный. В качестве примера можно рассмотреть «одномерный» случай. Пусть задана некоторая функция f на отрезке $[a, b]$, на котором введена равномерная сетка с 2^d узлами. Тогда вектор значений этой функции представляет собой вектор, т.е. одномерный массив, длины 2^d . Этот массив затем рассматривается как d мерный массив с размерами модовых индексов, равными 2 с помощью стандартного бинарного кодирования одномерного индекса любого элемента. К этому новому массиву применяется ТТ-разложение. Если получившиеся ранги малы, то получаем малопараметрическое представление с количеством параметров $2dr^2$. Число параметров при условии, что r ограничено сверху, растет как логарифм от полного числа элементов, что делает такое представление очень эффективным. В разделе 2.2 доказано, что многие «стандартные» функции имеют

представления с малыми QTT-рангами и получены явные QTT-представления для таких функций. В частности, доказаны соответствующие теоремы для полиномов, тригонометрических функций и функций, для которых функция двух переменных $f(x + y)$ является допускает разделение переменных x и y . Для рациональных функций доказано, что QTT-ранги зависят от шага сетки и требуемой точности логарифмически.

В разделе 2.3 получено явное QTT-представление для характеристической функции K -мерного стандартного симплекса. Такая функция возникает во многих задачах финансовой математики. Для нее показано, что QTT-ранги ограничены $2K - 1$ и не зависят от шага сетки. На рис. 2 приведены QTT-ранги для характеристической функции K -мерного симплекса для $K = 32$ и $n = 2^{10}$ точек по каждому направлению.

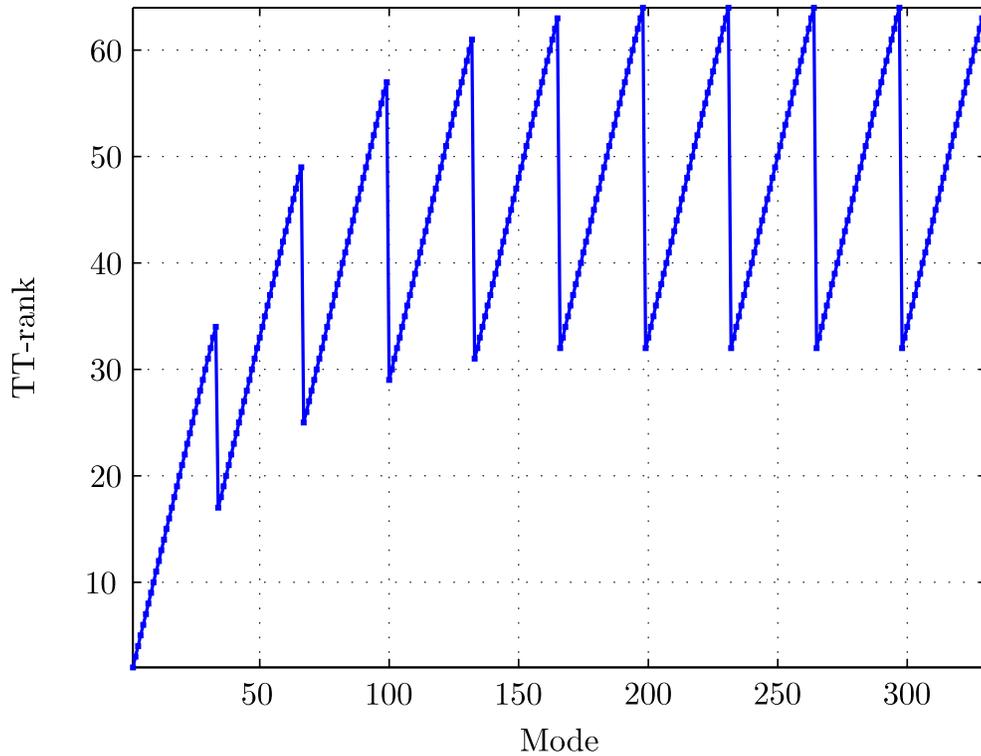


Рис. 2

QTT-ранги характеристической функции 32-мерного симплекса.

В разделе 2.4 приведены QTT-разложения некоторых операторов, в частности, оператора Лапласа, который играет важную роль во многих многомерных задачах (как оператор кинетической энергии). Также в разделе 2.4 доказана следующая теорема о структуре обратной к ленточной теплицевой

матрице.

Теорема 6 Пусть T — ленточная теплицева матрица с верхней шириной ленты s_1 и нижней шириной ленты s_2 размера $N = 2^d$. Тогда, все QTT-ранги ограничены сверху $(s_1 + s_2)^2 + 1$.

В разделе 2.5 выявлена аналогия между QTT-разложением и дискретными вейвлет-преобразованиями. На основе этой связи в разделе 2.6 вводится новая интерпретация QTT как нелинейного адаптивного вейвлет-преобразования с фильтрами, вычисляемыми по самой функции. Построены алгоритмы вычисления фильтров и применения преобразования. В разделе 2.7 показано, как применить WTT для создания новых вейвлет-преобразований.

Глава 3 посвящена применению разработанных в главах 1 и 2 вычислительных тензорных методов для решения практически важных многомерных задач. Раздел 3.1 посвящен перечислению таких задач в следующих областях: вычислительная химия, многопараметрические задачи, сжатие данных. В разделе 3.2 описывается формулировка задачи о решении молекулярного уравнения Шредингера, кратко вводятся основные понятия. В разделе 3.3 формулируется задача на собственные значения многомерного оператора, описывающего колебательные движения молекул. В разделе 3.4 для случая полиномиальной потенциальной поверхности, имеющего важный практический смысл при малых колебаниях, доказаны оценки на TT и QTT ранги вида $r_k = \mathcal{O}(f^{\lceil s/2 \rceil})$, где f — число степеней свободы, а s — порядок полинома. Для молекулы HONO получены QTT-представления высокой точности для потенциальной поверхности с рангами порядка 10 и относительной точностью 10^{-8} .

В разделе 3.5 строится оптимизационный метод нахождения минимального собственного значения матрицы, заданной в QTT формате, основанный на алгоритме Density Matrix Renormalization Group (DMRG), который много лет применялся в физике твердого тела. Оказывается, что этот алгоритм является не чем иным, как методом блочной релаксации в QTT-формате. Получены расчетные формулы для одной итерации алгоритма. В разделе 3.6 приведены численные расчеты для трехмерного потенциала и для потенциала Эно-Хайлеса с числом степеней свободы вплоть до 256. В разделе 3.7 произведено сравнение алгоритма DMRG-QTT на реальных молекулах с использованием данных, полученных с помощью пакета GAMESS. Результаты сравнения по вычислению энергии основного колебательного состояния приведены в таблице 1. Видно, что тензорный метод систематически дает более низкое значение, а так как все рассматриваемые методы являются вариаци-

онными, то и лучше приближает минимальное собственное значение при заданной дискретизации потенциальной поверхности. Время расчета всех трех методов (при вычисленной потенциальной поверхности) составило несколько минут.

Таблица 1

Минимальные собственные значения для различных молекул (в см^{-1}).

Молекула	E(DIAG)	E(VMP2)	E(DMRG-QTT)	Эксперимент
H_2CO	6005.83	5904.54	5808.19	5643.5
CH_3OH	12264.27	12071.90	11187.57	10848.5
C_2H_4	11477.09	11298.93	11263.74	10784.7

В разделе 3.8 рассмотрено вычисление многих собственных значений с помощью решения нестационарной задачи в мнимом времени

$$\frac{d\psi}{dt} = iH\psi, \quad \psi(0) = \psi_0, \quad H = -\frac{1}{2}\Delta + V,$$

и расчет спектра с помощью преобразования Фурье от автокорреляционной функции $\alpha(t) = (\psi(t), \psi(0))$. Используется схема расщепления Марчука-Яненко-Стрэнга и арифметика ТТ-формата с округлением. Результаты вычисления спектра (при разных точностях) приведены на рис. 3 и 4.

В разделе 3.9 описывается применение разработанных вычислительных методов к решению стохастических и многопараметрических задач. Рассматриваются задачи вида

$$\mathcal{L}(a)u = f,$$

где линейный оператор $\mathcal{L}(a)$ зависит от некоторого случайного поля a . Случайное поле предполагается параметризованным с помощью M параметров, после чего задача рассматривается как система линейных уравнений с Np^M неизвестными. Описаны стандартные подходы к решению таких задач, основанные на стохастическом методе Галеркина и методе стохастической коллокации. В диссертации предложено использовать ТТ-разложение для приближенного представления решения. В разделе 3.10 представлен метод решения больших линейных систем, основанный на итерациях Ричардсона с «одноточечным» предобуславливателем. На каждом шаге такого метода требуется решить r систем при некоторых значениях параметров, остальные вычисления производятся с помощью быстрой ТТ-арифметики. В разделе 3.12 приведены численные эксперименты по решению скалярного уравнения диффузии

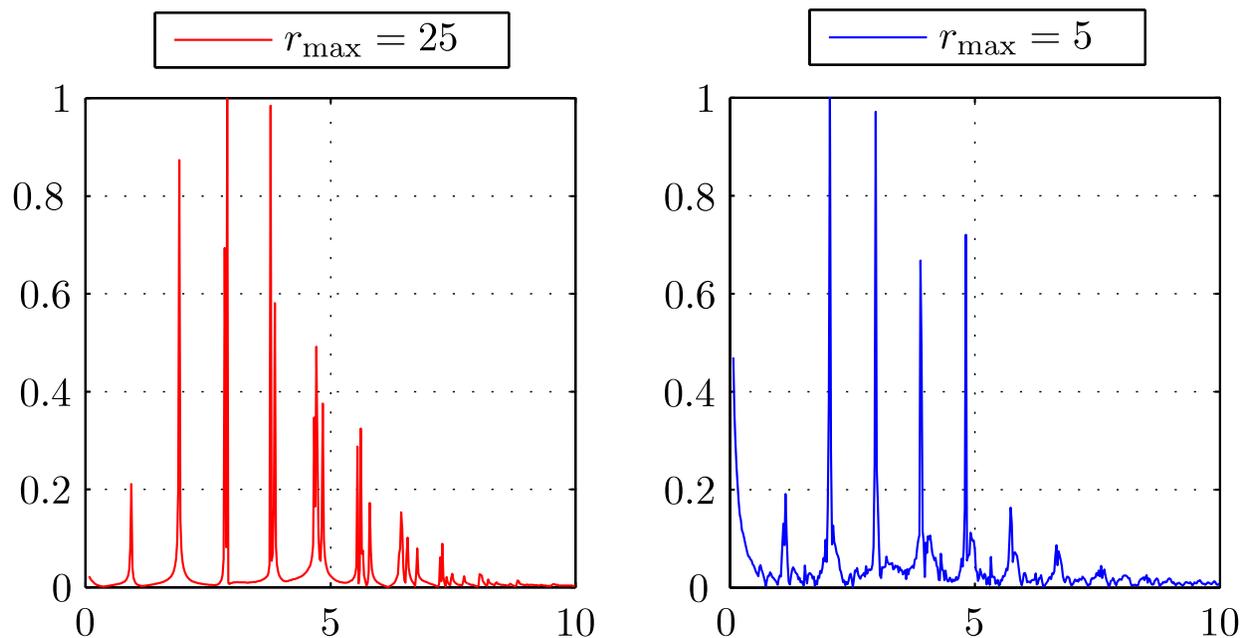


Рис. 3

Спектр для потенциала Эно-Хайлеса ($f = 2$) при использовании различных рангов. По горизонтали — частота, по вертикали — амплитуда

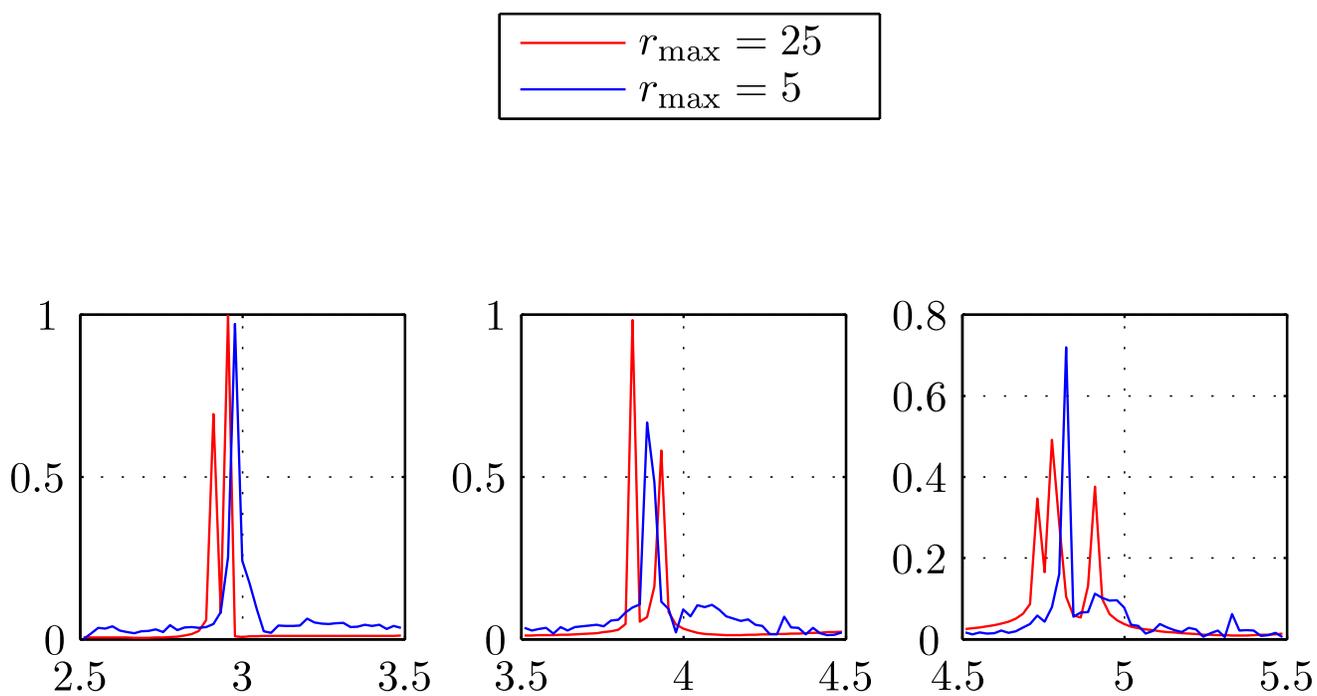


Рис. 4

Некоторые детали спектра

с коэффициентом, зависящим от параметров. Рассмотрены четыре примера. Приведем результаты численных экспериментов для одного из них. Необходимо решить уравнение диффузии с граничными условиями Дирихле на квадрате $[0, 1]^2$. Коэффициент диффузии задан в виде $\alpha(x, y) = 1 + \sum_{i=1}^4 \gamma_i \xi_i(x) y_i$, где $\xi_i(x)$ — характеристические функции четырех кругов (рис.5). Переменные $y_i, i = 1, \dots, 4$ принадлежат интервалу $[-0.99, 0]$. Такая задача может возникнуть, например, при решении обратной задачи о восстановлении коэффициента диффузии в некоторых подобластях: в этом случае решение многопараметрической задачи эквивалентно решению «всех» прямых задач, и эффективно представлению для многомерной параметрической зависимости в QTT-формате. Для дискретизации по физической переменной взята сетка 256×256 . Результаты расчетов приведены на рис. 6.

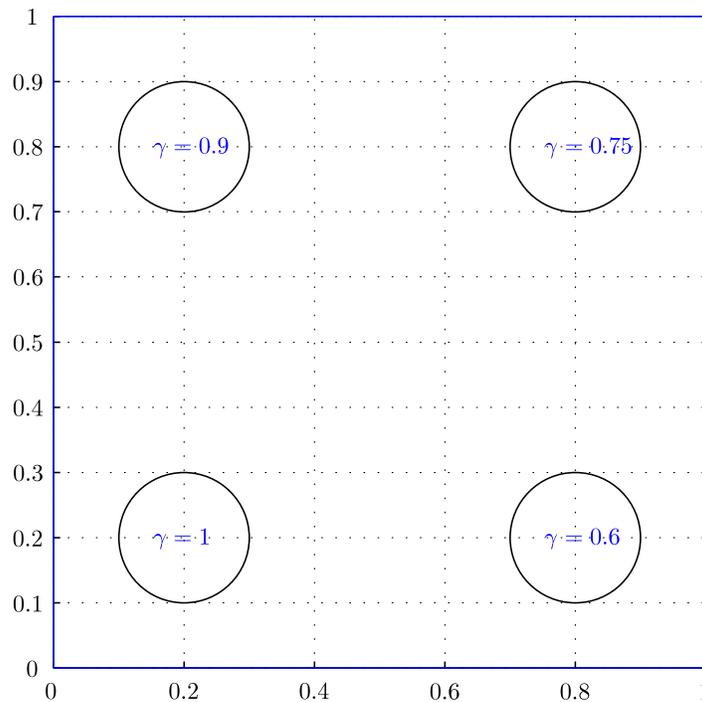


Рис. 5
 Расчетная область: 4 круга

Стандартные методы существенно более трудоемки. Тензорный метод требует всего 100 решений линейных систем для достижения точности 10^{-5} . Метод стохастической коллокации требует решения в 1000 раз больше линейных систем, а стохастический метод Галеркина требует решения системы, в которой примерно 60 миллионов неизвестных.

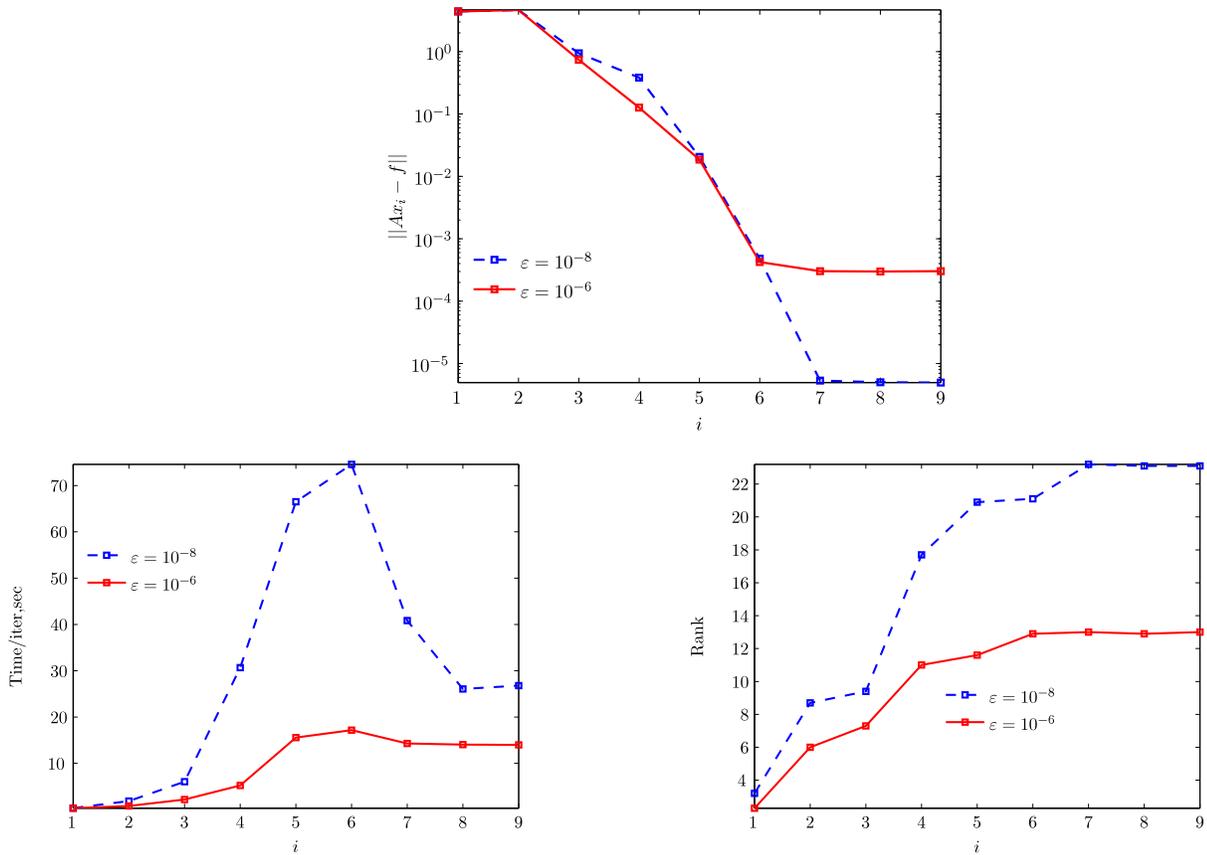


Рис. 6

Задача «4-х кругов». Вверху: невязка в зависимости от итерации. Слева внизу: время на итерацию. Справа внизу: ранги в зависимости от итерации.

В разделе 3.12 рассмотрен пример применения WTT-разложения для сжатия поля температуры, вычисленного с помощью σ -модели общей циркуляции океана, разработанной в ИВМ РАН. Данные представляют собой четырехмерный массив размера $360 \times 337 \times 40 \times 648$ ($N_x \times N_y \times N_z \times N_t$), в котором хранится температура в узлах широтно-долготной сетки, с 40 уровнями по высоте. Интервал по времени — 1 час. Весь массив занимает на диске объём примерно 12 гигабайт, и необходимо приблизить его с потерями, с ошибкой в C-норме не больше 0.1 градуса Цельсия. При таких условиях удалось получить сжатие примерно в 44 раза. Более подробно численные результаты приведены в таблице 2. Отметим, что время сжатия оказалось меньше, чем полное время, требуемое на чтение всего массива в оперативную память.

В заключении сформулированы основные результаты диссертации.

Таблица 2

Сжатие массива температуры с помощью WTT-разложения

Память	Ошибка абсолютная	Ошибка относительная	Время сжатия
497 МВ	0.0392	0.0004	≈ 500 секунд
277 МВ	0.0984	0.0009	≈ 500 секунд

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Основной результат диссертации: *введено и исследовано новое представление тензоров — TT-формат, созданы эффективные вычислительные алгоритмы для работы с тензорами в TT-формате и применены для решения практически важных многомерных задач.* Этот результат состоит в следующем:

- Предложено новое представление тензоров — TT-формат, которое, в отличие от канонического разложения, может быть построено с помощью быстрых устойчивых алгоритмов на основе сингулярного разложения, и, в отличие от разложения Таккера, не содержит экспоненциального по размерности числа параметров
- Получены все базовые алгоритмы для работы с массивами в TT-формате: переход от полного массива к TT-формату с точностью ε (алгоритм TT-SVD), алгоритм округления в TT-формате. Показано, что основные операции линейной алгебры (сложение, поэлементное умножение, умножение матрицы на вектор) можно проводить, оставаясь в рамках TT-формата.
- Получено многомерное обобщение скелетного разложения — формула TT-интерполяции, которая показывает, что тензор с малыми TT-рангами можно восстановить по небольшому количеству его элементов. На основе этой формулы получен многомерный крестовый метод аппроксимации тензоров.
- Введено понятие *тензоризации* или QTТ-формата, когда исходный массив малой геометрической размерности представляется как массив большей размерности за счет введения виртуальных уровней. Доказаны теоремы о QTТ-структуре функций одной переменной, а также о QTТ-структуре характеристической функции многомерного симплекса.

Показано, что QTT-ранги обратной к ленточной теплицевой матрицы не зависят от n .

- Создан новый метод сжатия данных с потерями на основе интерпретации QTT-формата как адаптивного вейвлет-преобразования.
- На основе QTT-формата построен метод решения многопараметрических задач, показана его эффективность на численных экспериментах
- Получены оценки на QTT-ранги полиномиальных потенциальных поверхностей квантовой молекулярной динамики, построен эффективный алгоритм нахождения минимального собственного значения, основанный на методе DMRG, который линеен по числу степеней свободы, вычислены минимальные собственные значения для потенциалов Эно-Хайлеса с числом степеней свободы вплоть до 256.
- На основе решения нестационарной задачи построен метод эффективного нахождения спектра Гамильтонианов квантовой молекулярной механики (колебательных спектров)
- QTT/WTT разложения применены для сжатия поля температуры, вычисленного с помощью σ -модели общей циркуляции океана ИВМ РАН, получено сжатие в 44 раза с точностью 0.1 градус Цельсия.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

- [1] *Оселедец И. В., Савостьянов Д. В.* Методы разложения тензора // Матричные методы и технологии решения больших задач / ред. Е. Е. Тыртышников. — ИВМ РАН, 2005. — С. 51–64.
- [2] *Оселедец И.В., Тыртышников Е.Е.* Приближенное обращение матриц при численном решении гиперсингулярного интегрального уравнения // *Ж. вычисл. матем. и матем. физ.* — 2005. — Т. 45, № 2. — С. 315–326.
- [3] *Оселедец И.В.* Применение разделенных разностей и В-сплайнов для построения быстрых дискретных преобразований вейвлетовского типа на неравномерных сетках // *Матем. заметки.* — 2005. — Т. 77, № 5. — С. 743–752.
- [4] *Оселедец И. В., Савостьянов Д. В.* Минимизационные методы аппроксимации тензоров и их сравнение // *Ж. вычисл. матем. и матем. физ.* — 2006. — Т. 46, № 10. — С. 1752–1734.

- [5] *Оселедец И.В.* Оценки снизу для сепарабельных аппроксимаций ядра Гильберта // *Матем. сб.* — 2007. — Т. 198, № 3. — С. 137–144.
- [6] *Оселедец И. В.* О новом тензорном разложении // *ДАН.* — 2009. — Т. 427, № 2. — С. 168–169.
- [7] *Оселедец И. В.* О приближении матриц логарифмическим числом параметров // *ДАН.* — 2009. — Т. 428, № 1. — С. 23–24.
- [8] *Оселедец И.В., Тыртышников Е.Е.* Рекурсивное приближение многомерных тензоров // *ДАН.* — 2009. — Т. 427, № 1. — С. 14–16.
- [9] *Замарашкин Н.Л., Оселедец И.В., Тыртышников Е.Е.* Тензорная структура обратной к ленточной теплоцевой матрице // *ДАН.* — 2009. — Т. 422, № 2. — С. 168–169.
- [10] *Olshevsky V., Oseledets I. V., Tyrtyshnikov E. E.* Tensor properties of multilevel Toeplitz and related matrices // *Linear Algebra Appl.* — 2006. — Vol. 412, no. 1. — P. 1–21.
- [11] *Oseledets I. V., Tyrtyshnikov E. E.* A unifying approach to the construction of circulant preconditioners // *Linear Algebra Appl.* — 2006. — Vol. 418, no. 2–3. — P. 435–449.
- [12] *Olshevsky V., Oseledets I. V., Tyrtyshnikov E. E.* Superfast inversion of two-level Toeplitz matrices using Newton iteration and tensor-displacement structure // *Operator Theory: Advances and Applications.* — 2008. — Vol. 179. — P. 229–240.
- [13] *Oseledets I. V., Savostianov D. V., Tyrtyshnikov E. E.* Tucker dimensionality reduction of three-dimensional arrays in linear time // *SIAM J. Matrix Anal. Appl.* — 2008. — Vol. 30, no. 3. — P. 939–956.
- [14] *Oseledets I. V.* The integral operator with logarithmic kernel has only one positive eigenvalue // *Linear Algebra Appl.* — 2008. — Vol. 428, no. 7. — P. 1560–1564.
- [15] *Oseledets I. V., Tyrtyshnikov E. E., Zamarashkin N. L.* Matrix inversion cases with size-independent rank estimates // *Linear Algebra Appl.* — 2009. — Vol. 431, no. 5-7. — P. 558–570.

- [16] *Oseledets I. V., Savostyanov D. V., Tyrtyshnikov E. E.* Fast simultaneous orthogonal reduction to triangular matrices // *SIAM J. Matrix Anal. Appl.* — 2009. — Vol. 31, no. 2. — P. 316–330.
- [17] *Oseledets I. V., Savostyanov D. V., Tyrtyshnikov E. E.* Linear algebra for tensor problems // *Computing.* — 2009. — Vol. 85, no. 3. — P. 169–188.
- [18] *Oseledets I. V., Tyrtyshnikov E. E.* Breaking the curse of dimensionality, or how to use SVD in many dimensions // *SIAM J. Sci. Comput.* — 2009. — Vol. 31, no. 5. — P. 3744–3759.
- [19] *Oseledets I. V., Savostyanov D. V., Tyrtyshnikov E. E.* Cross approximation in tensor electron density computations // *Numer. Linear Algebra Appl.* — 2010. — Vol. 17, no. 6. — P. 935–952.
- [20] *Oseledets I. V., Tyrtyshnikov E. E.* TT-cross algorithm for the approximation of multidimensional arrays // *Linear Algebra Appl.* — 2010. — Vol. 432, no. 1. — P. 70–88.
- [21] How to find a good submatrix / S.A. Goreinov, I.V. Oseledets, D.V. Savostyanov et al. // *Matrix Methods: Theory, Algorithms, Applications* / Ed. by V. Olshevsky, E. Tyrtyshnikov. — World Scientific Publishing, 2010. — P. 247–256.
- [22] *Goreinov S. A., Oseledets I. V., Savostyanov D. V.* Wedderburn rank reduction and Krylov subspace method for tensor approximation. Part 1: Tucker case: Preprint 2010-01. — Moscow: INM RAS, 2010. — arXiv:1004.1986. <http://pub.inm.ras.ru>.
- [23] *Oseledets I. V.* Constructive representation of functions in tensor formats: Preprint 2010-04. — Moscow: INM RAS, 2010. <http://pub.inm.ras.ru>.
- [24] *Khoromskij B. N., Oseledets I. V.* QTT-approximation of elliptic solution operators in high dimensions // *Rus. J. Numer. Anal. Math. Model.* — 2011. — Vol. 26, no. 3. — P. 303–322.
- [25] *Khoromskij B. N., Oseledets I. V.* Quantics-TT collocation approximation of parameter-dependent and stochastic elliptic PDEs // *Comput. Meth. Appl. Math.* — 2010. — Vol. 10, no. 4. — P. 376–394.

- [26] *Khoromskij B. N., Oseledets I. V.* DMRG + QTT approach to high-dimensional quantum molecular dynamics: Preprint 68. — Leipzig: MPI MIS, 2010. www.mis.mpg.de/preprints/2010/preprint2010_68.pdf.
- [27] *Oseledets I. V., Tyrtyshnikov E.E.* Algebraic wavelet transform via quantics tensor train decomposition // *SIAM J. Sci. Comput.* — 2011. — Vol. 31, no. 3. — P. 1315–1328.
- [28] *Oseledets I. V.* Approximation of $2^d \times 2^d$ matrices using tensor decomposition // *SIAM J. Matrix Anal. Appl.* — 2010. — Vol. 31, no. 4. — P. 2130–2145.
- [29] *Oseledets I. V.* Tensor-train decomposition // *SIAM J. Sci. Comput.* — Vol. 33, no. 5. — P. 2295–2317.
- [30] *Dolgov S. V., Oseledets I. V.* Solution of linear systems and matrix inversion in the TT-format: Preprint 19 (Submitted to SIAM J. of Sci. Comput.). — Leipzig: MIS MPI, 2011. http://www.mis.mpg.de/preprints/2011/preprint2011_19.pdf.
- [31] Low-rank tensor structure of solutions to elliptic problems with jumping coefficients / Sergey Dolgov, Boris N. Khoromskij, Ivan V. Oseledets, Eugene E. Tyrtyshnikov // *J. Comput. Math.* — 2012. — Vol. 30, no. 1. — P. 14–23.
- [32] *Oseledets I. V.* DMRG approach to fast linear algebra in the TT-format // *Comput. Meth. Appl. Math.* — 2011. — Vol. 10, no. 3. — P. 382–393.
- [33] *Oseledets I. V., Tyrtyshnikov E.E., Zamarashkin N.L.* Tensor-train ranks of matrices and their inverses // *Comput. Meth. Appl. Math.* — 2011. — Vol. 10, no. 3. — P. 394–403.

изд. лиц. №03991 от 12.02.2001. Компьютерный набор. Подписано в печать 05.02.2012. Усл.
печ. л. 1.5. Тираж 100 экз.

Институт вычислительной математики РАН. 119333, Москва, ул. Губкина, 8.