МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Щербакова Елена Михайловна

Матричные и тензорные разложения с условием неотрицательности и их применение

Специальность 1.2.2 – Математическое моделирование, численные методы и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2025

Диссертация подготовлена на кафедре Вычислительных технологий и моделирования факультета Вычислительной математики и кибернетики МГУ имени М.В. Ломоносова.

Научный руководитель:	Тыртышников Евгений Евгеньевич - академик Российской академии наук, доктор физико-математических наук, профессор
Официальные оппоненты:	Крылов Андрей Серджевич – доктор физико-математических наук, профессор, МГУ им. М.В. Ломоносова, профессор кафедры математической физики, заведующий лабораторией математических методов обработки изображений
	Логофет Дмитрий Олегович – доктор физико-математических наук, профессор, Федеральное государственное бюджетное учреждение науки Институт физики атмосферы им. А.М. Обухова Российской академии наук, главный научный сотрудник лаборатории математической экологии
	Рахуба Максим Владимирович – кандидат физико-математических наук, Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Национальный исследовательский университет «Высшая школа экономики», доцент департамента больших данных и информационного поиска факультета компьютерных наук

Защита диссертации состоится «31» марта 2025 г. в 15 часов 30 минут на заседании диссертационного совета МГУ.012.1 Московского государственного университета имени М.В. Ломоносова по адресу: 119991, Москва, Ленинские горы, д. 1, стр. 52, факультет ВМК.

E-mail (диссертационного совета): ds@cs.msu.ru

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский просп., д. 27) и на портале: `https://dissovet.msu.ru/dissertation/3351`

Автореферат разослан «_____» _______________ 2025 г.

Ученый секретарь
диссертационного совета МГУ.012.1,
чл.-корр. Российской академии наук,
доктор физико-математических наук — А.В. Ильин

Актуальность работы.

Данная работа посвящена разработке и программной реализации эффективных алгоритмов для неотрицательной факторизации матриц и тензоров. Задача неотрицательной матричной факторизации (Nonnegative Matrix Factorization – NMF) формулируется следующим образом: имея $m \times n$ матрицу V с $V_{ij} \geq 0$, $i = \overline{1,m}$, $j = \overline{1,n}$ (в дальнейшем неотрицательность элементов матрицы будет обозначаться как $V \geq 0$) и натуральное число $r < min(m,n)$, требуется найти такие матрицы $W \in \mathbb{R}^{m \times r}$ и $H \in \mathbb{R}^{r \times n}$ с неотрицательными элементами, что $V \approx WH$. Минимальное r такое, что $V = WH$, называется неотрицательным рангом V.

Задача построения разложения матрицы является хорошо изученной, однако добавление требования неотрицательности кардинально меняет ситуацию. В общем случае неотрицательная матричная факторизация считается NP-трудной проблемой даже при условии, что неотрицательный ранг матрицы известен. То же самое можно сказать о сложности поиска неотрицательного ранга матрицы. Если W, H - решение задачи, то пара $W_1 = WS, H_1 = S^{-1}H$ тоже является решением при условии, что матрица перехода S невырожденная и сохраняет неотрицательность компонент разложения.

Обычно для поиска решения проблему NMF формулируют как задачу оптимизации:

$$(W^*, H^*) = \underset{W \geq 0, H \geq 0}{\arg\min} D(V||WH),$$

где $D(P||Q) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(p_{ij}, q_{ij}): \ d(p,q) \geq 0, d(p,q) = 0 \Leftrightarrow p = q$. В качестве функции цены D чаще всего выбирают норму Фробениуса или обобщенную дивергенцию Кульбака-Лейблера. Сформулированная задача оптимизации не является выпуклой и может иметь несколько локальных минимумов.

Матрицы – один из простых способов представления информации, однако во многих прикладных задачах чаще приходится работать с многомерными массивами, или тензорами. И здесь мы сталкиваемся с другой проблемой: совершать численные операции с ними при большом числе измерений представляется невозможным из-за "проклятия размерности". Имеется в виду, что память, требуемая для хранения элементов, и число действий, необходимых для выполнения базовых манипуляций с тензорами, растет экспоненциально с увеличением ранга d. Таким образом, возникает потребность в неких малопараметрических представлениях, одним из которых стал тензорный поезд - TT-разложение.

Однако часто существует ограничение на неотрицательность элементов тензоров, и это условие должно выполняться и для их приближений. Данное требование возникает в задачах машинного обучения и обработки изображений или при проведении вычислений для физических моделей и позволяет дать интерпретацию полученным результатам. Отсюда появляется необходимость в развитии такой технологии, как неотрицательная тензорная факторизация - NTF.

Ранее в литературе уже были рассмотрены алгоритмы NTF для канонического разложения и разложения Таккера. Но к моменту начала работы автора над диссертацией почти не было методов для построения неотрицательной аппроксимации исходного тензора тензорным поездом, когда вагоны заполнены только неотрицательными числами.

Основной целью данной работы является построение эффективных методов для неотрицательной факторизации матриц и тензоров на основе малоранговых аппроксимаций и быстрых алгоритмов линейной алгебры. Также целью является теоретическое исследование предложенных методов, их реализация в виде программ и рассмотрение приложений. В качестве приложений были рассмотрены гиперспектральные изображения, тензор ERP (event related potential – электрофизиолигический отклик на стимул), ядра уравнения Смолуховского и ранжирование вершин многомерного графа.

Научная новизна. В работе предложены методы для неотрицательной факторизации матриц и тензоров, указаны оценки алгоритмической сложности, приведены теоретические результаты, доказывающие эффективность алгоритмов для ряда задач. Разработанные методы реализованы программно и протестированы на ряде примеров в сравнении с другими алгоритмами.

Теоретическая ценность.

В данной работе предлагается использовать для неотрицательной факторизации матриц и тензоров двухэтапный подход, где на первом шаге строится малоранговая аппроксимация, далее алгоритм неотрицательного разложения работает с вычисленной аппроксимацией вместо исходного объекта, что при работе с большими данными позволяет получить ускорение в тысячи раз.

Для определенных классов матриц, а именно неотрицательных сепарабельных матриц и неотрицательных матриц ранга 2, предлагаются методы решения задачи неотрицательной факторизации за время, линейно зависящее от размеров матриц. Исследуется случай возмущенных сепарабельных матриц и доказывается теорема о том, как зависит точность аппроксимации от величины возмущений.

Предлагаются алгоритмы для построения неотрицательного тензорного поезда, канонического полиадического разложения и разложения Таккера на основе двухэтапного подхода, позволяющие снизить сложность итерации с $O(n^d)$ до линейной зависимости от n. Дополнительно рассматривается возможность комбинации алгоритма факторизации в неотрицательный тензорный поезд с методом DMRG, чтобы обеспечить возможность выбора рангов динамически.

Предлагается подход точечной коррекции элементов разложения неотрицательного тензора в формате тензорного поезда с целью получения неотрицательной факторизации.

Рассмотрено применение модели неотрицательного тензорного поезда при вычислении собственных векторов Перрона, которые используются, в частности, для моделирования степени важности узлов многомерной сети.

Практическая ценность заключается в программной реализации предложенных алгоритмов, позволяющих ускорить расчеты за счет использования малоранговых приближений и быстрых методов линейной алгебры, что подтверждается результатами численных экспериментов для ряда задач.

Полученные теоретические результаты и разработанный комплекс программ позволяют качественно расширить круг задач, доступных для детального изучения методами математического моделирования.

На данный момент алгоритмы, полученные в диссертации, уже успешно применяются для вычисления неотрицательного решения уравнения Смолуховского, сжатия видео и изображений с запуском на нескольких процессорах.

На защиту выносятся следующие результаты и положения. Основные результаты:

- Обоснован эффективный редуцированный двухэтапный алгоритм неотрицательной факторизации сепарабельных матриц и матриц ранга 2, исследована применимость метода к возмущенным сепарабельным матрицам, выведена оценка зависимости точности приближения от величины возмущения;

- Разработаны методы построения неотрицательного тензорного поезда для аппроксимации исходного тензора с неотрицательными элементами с использованием двухэтапного подхода. Рассмотрена комбинация одного из алгоритмов с техникой DMRG, что позволяет подбирать также ранги аппроксимации, которые в большинстве алгоритмов считаются входными данными;

- Предложен подход к построению неотрицательного TT-приближения (без требования неотрицательности элементов ядер), основанный на точечной коррекции элементов, данный метод успешно применяется для решения мультикомпонентного уравнения Смолуховского (математической модели процессов коагуляции при неупругих соударениях огромного числа частиц) с сохранением неотрицательности;

- Разработанные методы реализованы в виде комплекса программ, проведен ряд численных экспериментов, иллюстрирующих эффективность и точность предложенных методов, описано применение модели неотрицательного TT при ранжировании узлов многомерного графа.

Достоверность полученных результатов обеспечивается большим количеством дополняющих друг друга теоретических результатов и численных экспериментов.

Апробация работы. Основные результаты работы докладывались автором и обсуждались на следующих конференциях:

- “The Sixth China-Russia Conference on Numerical Algebra with Applications” (Москва, 2017)
- “Ломоносов” (МГУ имени М. В. Ломоносова, Москва, 2018)
- “Тихоновские чтения” (МГУ имени М. В. Ломоносова, Москва, 2018)
- “Large-Scale Scientific Computations” (Созополь, Болгария, 2019)
- “SIAM Conference on Applied Algebraic Geometry” (Берн, Швейцария, 2019)
- “Методы вычислений и математическая физика” (Сочи, НТУ Сириус, 2020)
- “The 6th international conference on matrix methods and machine learning in mathematics and applications”' (Москва, научный парк МГУ Lomonosov Hall и ИВМ РАН, 2023)
- “Матричные методы и интегральные уравнения” (НТУ Сириус, 2023)

Личный вклад автора.

Работа [1] полностью выполнена автором. В статье [2] идея редуцированного алгоритма для сепарабельных матриц была предложена автором, автор участвовала в доказательстве лемм и теорем, программная реализация и проведение численных экспериментов были выполнены автором полностью самостоятельно. В работе [3] автор предложила алгоритм, дала оценку точности аппроксимации, выполнила программную реализацию и провела численные эксперименты. В статье [4] автор реализовала и развила идею точечной корректировки элементов в аппроксимации тензорным поездом с целью получения неотрицательного приближения, предложила двухэтапный подход к построению неотрицательного канонического разложения и разложения Таккера, снизив сложность одной итерации с $O(n^d)$ до

$O(dnr^3)$, реализовала все методы программно и провела численные эксперименты. В работе [5] автор провела численное исследование двух подходов малоранговой неотрицательной тензорной факторизации, которое на практике демонстрирует превосходство предложенных ей ранее алгоритмов.

Диссертационное исследование является законченным и самостоятельным трудом автора.

Публикации. Основные результаты по теме диссертации изложены в 5 печатных изданиях, изданных в журналах Scopus, WoS, RSCI, а также в изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности.

Структура работы.

Диссертационная работа состоит из введения, четырёх глав, заключения, списка литературы и списка публикаций автора. Общий объем диссертации 117 страниц, включая 13 рисунков, 17 таблиц и список литературы из 73 наименований.

Благодарности. Автор выражает благодарность академику РАН Тыртышникову Евгению Евгеньевичу за научное руководство и постоянную поддержку в исследованиях.

Содержание работы.

Во введении обоснована актуальность темы диссертационной работы, дается краткая характеристика диссертации: изложены цель работы, её научная новизна и практическая ценность, сформулированы положения, выносимые на защиту, отмечен личный вклад автора. Кроме того, представлен отчёт об апробации диссертационной работы и публикациях, содержащих основные результаты.

В первой главе исследуется возможность построения неотрицательной факторизации матрицы размера $m \times n$, зная ее неотрицательный ранг, используя лишь несколько ее строк и столбцов. Предлагаются методы решения этой задачи для определенных классов матриц: неотрицательных сепарабельных матриц – тех, для которых существует конус, натянутый на несколько столбцов исходной матрицы и содержащий все ее столбцы; неотрицательных сепарабельных матриц с возмущениями; неотрицательных матриц ранга 2. На практике предложенные алгоритмы используют число операций и объем памяти, линейно зависящие от $m + n$.

На первый взгляд условие сепарабельности может показаться несколько искусственным, однако существуют приложения, где оно является естественным и разумным. Например, данное условие широко распространено при работе с гиперспектральными изображениями. В данном случае каждый столбец M интерпретируется как "спектральная подпись" смеси материалов для одного элемента изображения, и условие сепарабельности означает, что для каждого материала существует элемент изображения, содержащий исключительно его. При работе с текстами столбцы M могут описывать слова. Тогда условие сепарабельности будет означать, что для каждой темы существует слово, относящееся только к ней.

Теорема 1. *Пусть матрица M является неотрицательной сепарабельной матрицей ранга r. Если известно некоторое разложение $M = UV$, $U \in R^{m\times r}$, $V \in R^{n\times r}$, то число действий для построения неотрицательной матричной факторизации будет линейно зависеть от размеров матрицы.*

Исследован случай возмущенных сепарабельных матриц. В частности, выведена теоретическая оценка точности работы алгоритма в зависимости от возмущения.

Теорема 2. *Пусть $A \in \mathbb{R}^{m\times n}$ – нормированная сепарабельная матрица ранга r и рассматривается возмущенная матрица $\widetilde{A} = A + E$, где $||E||_F \le \varepsilon$. Тогда существуют такие r столбцов C и r строк R матрицы $\widetilde{A}$, дающие в пересечении $r \times r$-матрицу G, что применение редуцированного алгоритма к факторизации $CG^{-1}R$ находит столбцы $\widetilde{a}_{i_1}, \ldots, \widetilde{a}_{i_r}$, которые при некоторой нумерации определяющих столбцов матрицы A при всех достаточно малых ε удовлетворяют неравенствам*

$$||\widetilde{a}_{i_k} - a_{j_k}|| \le O\left(r\frac{\mu^2}{\sigma_r^2}\varepsilon\right), \quad k = 1, \ldots, r,$$

где μ – наибольшая длина столбцов матрицы A, σ_r – минимальное сингулярное число $m \times r$-матрицы, составленной из столбцов определяющей системы матрицы A.

Во второй главе вводится модель неотрицательного тензорного поезда. Элементы неотрицательного тензорного поезда записываются как:

$$a(i_1, \ldots, i_d) = \sum_{\alpha_1, \ldots, \alpha_{d-1}} G_1(i_1, \alpha_1) G_2(\alpha_1, i_2, \alpha_2) \ldots$$

$$\ldots G_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) G_d(\alpha_{d-1}, i_d),$$

с вагонами $G_1, ..., G_d$, среди которых любые два соседних имеют общий индекс суммирования. Индексы суммирования α_k принимают значения от 1 до r_k и называются вспомогательными индексами. Величины r_k называются рангами тензорного поезда (ТТ ранги). Легче предполагать, что G_1, G_d не двумерные, а трехмерные с дополнительными вспомогательными индексами $\alpha_0 = \alpha_d = 1$ и ТТ рангами $r_0 = r_d = 1$. Это предположение помогает упростить некоторые алгоритмы. Тензорные вагоны $G_1, \ldots, G_d \in \mathbb{R}_+^{r_{k-1} \times n_k \times r_k}$.

Предлагаются методы построения неотрицательного тензорного поезда для аппроксимации исходного тензора с неотрицательными элементами с использованием двухэтапного подхода.

Первый алгоритм – NTTF (Nonnegative Tensor Train Factorization) основан на последовательном применении двухэтапной неотрицательной факторизации матриц к матрицам-разверткам. Исследован вопрос точности разложения (ε – параметр, подаваемый на вход метода):

Теорема 3. *Алгоритм NTTF строит тензорный поезд B так, что*

$$\|A - B\|_F \leq \varepsilon.$$

Второй алгоритм – NTT-MU (Nonnegative Tensor Train - Multiplicative Updates). На первом шаге исходный тензор факторизуется в тензорный поезд, все последующие операции проводятся с этим разложением, за счет чего снижается общая сложность алгоритма. Далее формируется начальное неотрицательное приближение в формате тензорного поезда, затем на каждой итерации все вагоны, кроме одного, фиксируются, значения оставшегося ядра заменяются по правилу мультипликативных обновлений, чтобы улучшить аппроксимацию. Сложность обновления вагона $O(nr^4)$, то есть зависимость от n линейная. Рассмотрена комбинация NTT-MU с техникой DMRG, что позволяет подбирать также ранги аппроксимации, которые в большинстве алгоритмов считаются входными данными.

Предлагается использование двухэтапного подхода для построения других разложений, в частности неотрицательного канонического разложения и разложения Таккера.

Каноническое полиадическое разложение для тензора $Y \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ состоит из матриц-факторов $U^{(i)} = [u_1^{(i)} \dots u_R^{(i)}] \in \mathbb{R}^{n_i \times R}$ и определяется как

$$y(i_1, \dots, i_d) = \sum_{r=1}^{R} u_r^1(i_1) u_r^2(i_2) \dots u_r^d(i_d),$$

где R - ранг разложения.

Каноническое полиадическое разложение (CPD) очень популярно и было использовано для многих приложений, например обработка сигналов и задачи машинного обучения, такие как восстановление изображений, извлечение признаков и т.д. Данная модель и ее свойства, включая единственность, хорошо изучены. Однако, несмотря на все преимущества, у нее есть существенный недостаток: вычислительные затраты большинства существующих алгоритмов для CPD растут экспоненциально с порядком тензора. Но существует способ существенно снизить сложность итерации с $O(dRN^d)$ до $O(dNR^3)$, когда $n_1 = \dots = n_d = N$. Для этого предлагается сначала построить разложение в формате тензорного поезда и далее работать с данной аппроксимацией вместо исходного тензора при построении неотрицательного канонического полиадического разложения.

Разложение Таккера для тензоров высокого порядка представляется как декомпозиция исходного тензора $Y \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ в тензор-ядро $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \cdots \times R_d}$, умноженное на набор d матриц-факторов, $A^{(i)} = \left[a_1^{(i)}, \dots, a_{R_i}^{(i)} \right] \in \mathbb{R}^{n_i \times R_i} (i = 1, 2, \dots, d)$, в результате:

$$y(i_1, \dots, i_d) = \sum_{j_1, j_2, \dots, j_d} g(j_1, j_2, \dots, j_d) a_{j_1}^1(i_1) a_{j_2}^2(i_2) \dots a_{j_d}^d(i_d).$$

Классическое неотрицательное разложение Таккера (NTD) должно иметь $\mathcal{G} \in \mathbb{R}_{+}^{R_1 \times R_2 \cdots \times R_d}$ и $A^{(i)} = \left[a_1^{(i)}, a_2^{(i)}, \ldots, a_{R_i}^{(i)}\right] \in \mathbb{R}_{+}^{n_i \times R_i} (i = 1, 2, \ldots, d)$.

Применение двухэтапного подхода позволяет снизить сложность итерации в существующих алгоритмах с $O(dRN^d)$ до $O(dNR^3)$.

Также в главе предлагается новый подход к неотрицательной тензорной факторизации. Когда строится тензорное разложение с заданной точностью, ожидается, что в итоге все элементы аппроксимации будут близки к исходным, и если исходные данные были неотрицательными, то естественно предположить, что большая часть элементов аппроксимации таковыми и останутся. Тогда, если число отрицательных элементов мало, то можно найти их и точечно поправить. Данный метод, в частности, успешно применяется для решения мультикомпонентного уравнения Смолуховского с сохранением неотрицательности. В случае этой задачи неотрицательные сдвиги применяются к численной аппроксимации решения в формате тензорного поезда. Уже известно, что предложенная методика может быть полезна как в режиме регулярных поправок при интегрировании по времени, так и при постобработке стационарных решений. Предлагаемые методы позволяют решать задачи с решением, содержащим 400^4 элементов в случае баллистического ядра коагуляции и 3200^4 в случае постоянного ядра коагуляции.

Одной из главных задач специалистов по обработке данных является выявление релевантных компонентов в сети или графе. Сети часто представляют в виде матриц; следовательно, инструменты матричного анализа оказываются полезными при решении различных проблем, таких как выявление сообществ и антисообществ, разделение сети на кластеры или идентификация центральных узлов, ребер. За последние годы несколько авторов, сосредоточив внимание на этой последней проблеме и, в частности, на проблеме определения наиболее центральных узлов, работали над определением различных показателей центральности. Мера центральности - это вещественнозначная функция набора узлов, которая является инвариантной при перемаркировке узлов и, таким образом, может быть использована для их ранжирования в соответствии с их важностью. Меры, основанные на собственных векторах или сингулярных векторах подходящих матриц, называемые (линейными) центральностями собственных векторов, являются одними из наиболее популярных мер центральности. Соответствующие примеры включают, например, индекс Боначича и рейтинг PageRank; последний широко известен во всем мире благодаря использованию поисковой системой Google.

В литературе также предлагалось использовать собственный вектор Перрона простого мультиоднородного отображения для ранжирования узлов важности многоуровневой сети, в частности мультиплекса - набора графов. Все слои мультиплекса содержат один и тот же набор узлов, при этом нет никаких связей между узлами, которые принадлежат двум разным слоям. Мультиплекс можно описать как набор графов:

$$\mathcal{G} = \left\{ G^{(\ell)} = \left(V_n, E^{(\ell)} \right) \right\}_{\ell \in V_L}$$

где $V_n = \{1, 2, \ldots, n\}$ - множество узлов, $V_L = \{1, \ldots, L\}$ - множество слоев, $E^{(\ell)} \subset V_n \times V_n$ - множество ребер на слое ℓ. Для каждого $\ell \in V_L$ граф $G^{(\ell)}$ описывается неотрицательной матрицей смежности $A^{(\ell)} = \left(A_{ij}^{(\ell)} \right) \in \mathbb{R}^{n \times n}$. Поэтому мультиплекс можно описать тензором третьего порядка $\mathcal{A} = (\mathcal{A}_{ij\ell})$, называемым тензором смежности:

$$\mathcal{A}_{ij\ell} = A_{ij}^{(\ell)} = \begin{cases} w_\ell(i,j) & \text{если } (i,j) \in E^{(\ell)} \\ 0 & \text{иначе} \end{cases}$$

где $w_\ell(i,j)$ - положительное число, отражающее силу связи между узлами i и j в слое $\ell \in V_L$. Мультиплексы являются частным случаем многомерных графов, на практике мы сталкиваемся с более сложными объектами.

Будем использовать V, V_+ и V_{++} для обозначения пространства $V = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_d}$, конуса $V_+ = \mathbb{R}_+^{n_1} \times \cdots \times \mathbb{R}_+^{n_d}$ и $V_{++} = \mathbb{R}_{++}^{n_1} \times \cdots \times \mathbb{R}_{++}^{n_d}$, где $\mathbb{R}_+$ – все неотрицательные действительные числа, а $\mathbb{R}_{++}$ – все положительные действительные числа.

Рассмотрим функцию $\mathcal{F} = (F_1, \ldots, F_d) : V \longrightarrow V$ со следующими свойствами:

1. $\mathcal{F}$ непрерывна

2. Существует $d \times d$ матрица $A = (a_{ij}) \geq 0$ такая, что

$$F_i(\boldsymbol{x}_1, \dots, \lambda \boldsymbol{x}_j, \dots \boldsymbol{x}_d) = \lambda^{a_{ij}} F_i(\boldsymbol{x}_1, \dots, \boldsymbol{x}_j, \dots \boldsymbol{x}_d), \qquad \forall i, j = 1, \dots, d$$

3. $\mathcal{F}(\underline{\boldsymbol{x}}) \in V_+$ для любого вектора $\underline{\boldsymbol{x}} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_d) \in V_+$

4. $\mathcal{F}(\underline{\boldsymbol{x}}) \in V_{++}$ для любого вектора $\underline{\boldsymbol{x}} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_d) \in V_{++}$

5. Если $\boldsymbol{x}_i \geq \boldsymbol{y}_i$ (поэлементно) для любого $i = 1, \dots, d$, то $F_i(\underline{\boldsymbol{x}}) \geq F_i(\underline{\boldsymbol{y}})$ для любого $i = 1, \dots, d$, где $\underline{\boldsymbol{x}} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_d)$ и $\underline{\boldsymbol{y}} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_d)$.

Скажем, что $\mathcal{F} \in \mathsf{MH}_d$, если она удовлетворяет всем вышеперечисленным свойствам. Каждой $\mathcal{F} \in \mathsf{MH}_d$ соответствует матрица A, определенная как в пункте 2, которую назовем матрицей однородности $\mathcal{F}$. Отметим, что свойство пункта 2 можно сократить:

$$\mathcal{F}(\boldsymbol{\lambda} \odot \underline{\boldsymbol{x}}) = \boldsymbol{\lambda}^A \odot \mathcal{F}(\underline{\boldsymbol{x}}),$$

где $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$, $\underline{\boldsymbol{x}} \in V$, $\boldsymbol{\lambda} \odot \underline{\boldsymbol{x}} = (\lambda_1 \boldsymbol{x}_1, \dots, \lambda_d \boldsymbol{x}_d)$, а "матричное возведение в степень" $\boldsymbol{\lambda}^A$ — это вектор $\mathbb{R}^d$, определяемый следующим образом

$$\boldsymbol{\lambda}^A = (\lambda_1^{a_{11}} \cdots \lambda_d^{a_{1d}}, \cdots, \lambda_1^{a_{d1}} \cdots \lambda_d^{a_{dd}}) = (\textstyle\prod_{j=1}^d \lambda_j^{a_{1j}}, \cdots, \prod_{j=1}^d \lambda_j^{a_{dj}}).$$

Пусть $\mathbf{T} = (t_{i_1,\dots,i_d}) \geq 0$ — тензор с неотрицательными элементами. Для $\underline{\boldsymbol{x}} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_d) \in V$ и $j = 1, \dots, d$ определяется $T_j : V \to \mathbb{R}^{n_j}$ как

$$T_j(\underline{\boldsymbol{x}})_{i_j} = \sum_{i_1,\dots,i_{j-1},i_{j+1},\dots,i_d} t_{i_1,\dots,i_d} (\boldsymbol{x}_1)_{i_1} \cdots (\boldsymbol{x}_{j-1})_{i_{j-1}} (\boldsymbol{x}_{j+1})_{i_{j+1}} \cdots (\boldsymbol{x}_d)_{i_d} \tag{1}$$

для $i_j = 1, \dots, n_j$. Конечно, приведенная выше сумма справедлива для $i_1 = 1, \dots, n_1$, $i_2 = 1, \dots, n_2$ и так далее.

Теперь, учитывая $\alpha_1, \dots, \alpha_d > 0$, определяется $\mathcal{F} : V \to V$ как

$$\mathcal{F}(\underline{\boldsymbol{x}}) = \left(T_1(\underline{\boldsymbol{x}})^{1/\alpha_1}, \cdots, T_d(\underline{\boldsymbol{x}})^{1/\alpha_d} \right) \tag{2}$$

где степень берется поэлементно, т.е. $\underline{\boldsymbol{x}}^\beta = (x_1^\beta, x_2^\beta, \dots)$. Тогда $\mathcal{F} \in \mathsf{MH}_d$ со следующей матрицей однородности

$$A = \begin{pmatrix} 0 & 1/\alpha_1 & 1/\alpha_1 & 1/\alpha_1 & 1/\alpha_1 & \cdots \\ 1/\alpha_2 & 0 & 1/\alpha_2 & 1/\alpha_2 & 1/\alpha_2 & \cdots \\ 1/\alpha_3 & 1/\alpha_3 & 0 & 1/\alpha_3 & 1/\alpha_3 & \cdots \\ 1/\alpha_4 & 1/\alpha_4 & 1/\alpha_4 & 0 & 1/\alpha_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}. \tag{3}$$

Предположим, что параметры α_i выбраны так, что $\rho(A) < 1$. Тогда для любого $\underline{\boldsymbol{x}}^{(0)} \in V_{++}$ можно построить сходящуюся последовательность

$$\left(\frac{T_1(\underline{\boldsymbol{x}}^{(k)})^{1/\alpha_1}}{\|\cdots\|_{p_1}}, \cdots, \frac{T_d(\underline{\boldsymbol{x}}^{(k)})^{1/\alpha_d}}{\|\cdots\|_{p_d}}\right) \xrightarrow{k\to\infty} \underline{\boldsymbol{u}} \in V_{++} \tag{4}$$

где $\boldsymbol{u}$ — единственный положительный вектор такой, что $\mathcal{F}(\underline{\boldsymbol{u}}) = \lambda \odot \underline{\boldsymbol{u}}$.

Когда $t_{i_1,\dots,i_d}$ описывает свойства сети (например, это может быть тензор смежности), тогда вектор $\boldsymbol{u}$ можно использовать для присвоения определенной важности узлам и слоям (и временным интервалам) в многоуровневой сети. Это приложение для сетей особенно интересно для нас, поскольку в этом случае нам не нужна высокая точность. В том смысле, что нам необязательно знать значение всех элементов $\underline{\boldsymbol{u}} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_d)$ с высокой точностью. Что важно, так это относительный порядок самых больших элементов каждого $\boldsymbol{u}_i$, которые обычно являются элементами, которые сходятся быстрее в последовательности (4).

Поясним это на примере. Предположим, у нас есть последовательность $\mathcal{S}_t$ многослойных графов с $t = 1, \dots, \bar{t}$, где для каждого t $\mathcal{S}_t$ - это множество графов $\mathcal{S}_t = (S_1, \dots, S_{\bar{\ell}})$ на одном и том же наборе узлов $\{1, \dots, n\}$. Параметр t является меткой времени, т.е. предполагается, что многоуровневый граф/сеть меняется во времени. Весь этот объект может быть однозначно описан следующим тензором смежности

$$\mathbf{A} = (a_{ij\ell t}) = \begin{cases} \omega > 0 & \text{, если существует ребро от узла } i \text{ до узла } j \text{ в слое } \ell \\ & \text{ в момент времени } t \\ 0 & \text{, иначе} \end{cases}$$

Определим разные виды векторов “важности” (влиятельности) для последовательности $\mathcal{S}_1, \ldots, \mathcal{S}_{\bar{t}}$: $\boldsymbol{x} = (x_1, \ldots, x_n)$ — вектор влиятельности узлов, которые можно назвать *трансляторами*, $\boldsymbol{y} = (y_1, \ldots, y_n)$ — вектор влиятельности узлов известных как *получатели*, $\boldsymbol{w} = (w_1, \ldots, w_{\bar{\ell}})$ — вектор влиятельности слоев, а $\boldsymbol{z} = (z_1, \ldots, z_{\bar{t}})$ — вектор влиятельности меток времени. Тогда возможная модель влиятельности в точности определяется выражением $\mathcal{F}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{z}) = (\lambda, \mu, \nu, \eta) \odot (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{z})$, где $\mathcal{F}$ определяется как в (2), с $\mathbf{A}$ вместо $\mathbf{T}$.

Нам интересно узнать, какие k (скажем, 10) наиболее важных узлов, слоев и меток времени. Итак, фиксируются $\alpha_1, \ldots, \alpha_d$ так, чтобы $\rho(A) < 1$; вычисляется собственный вектор Перрона $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{z})$ $\mathcal{F}$, а затем отмечаются 10 крупнейших элементов (и их порядок) $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{w}$ и $\boldsymbol{z}$. Однако на практике мы хотим выбрать $\rho(A) \approx 1$, и это может замедлить сходимость. Более того, если у нас много слоев и большое временное окно, умножение тензора смежности, как в (1), является трудоемкой операцией, так как алгоритм вычисления векторов включает операцию умножения исходного тензора, описывающего многоуровневую сеть, на вектор. В главе предлагается использовать неотрицательную аппроксимацию в формате тензорного поезда с целью ускорения вычислений.

Предположим, что нам даны два неотрицательных тензора T и $\widetilde{T}$ такие, что $(1-\rho)T_{ijk} \leq \widetilde{T}_{ijk} \leq (1+\rho)T_{ijk}$. Тогда можно определить гильбертово расстояние между T и $\widetilde{T}$ как

$$\mu(T, \widetilde{T}) = \log \left(\max_{\substack{i,j,k \\ a,b,c}} \frac{T_{ijk}\widetilde{T}_{abc}}{\widetilde{T}_{ijk}T_{abc}} \right)$$

Если $(1-\rho)T_{ijk} \le \widetilde{T}_{ijk} \le (1+\rho)T_{ijk}$ для ненулевых элементов T и $\widetilde{T}$, то

$$\mu(T,\widetilde{T}) \le \log\left(\frac{1+\rho}{1-\rho}\right).$$

Теорема 4. *Пусть A — общая матрица однородности $\mathcal{F}_T$ и $\mathcal{F}_{\widetilde{T}}$. Если $\rho(A) < 1$, тогда существуют единственные $\underline{\boldsymbol{u}} = (\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) > 0$ и $\widetilde{\underline{\boldsymbol{u}}} = (\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}},\widetilde{\boldsymbol{z}}) > 0$ такие, что*

$$\mathcal{F}_T(\underline{\boldsymbol{u}}) = \boldsymbol{\lambda} \odot \underline{\boldsymbol{u}} \qquad \text{и} \qquad \mathcal{F}_{\widetilde{T}}(\widetilde{\underline{\boldsymbol{u}}}) = \widetilde{\boldsymbol{\lambda}} \odot \widetilde{\underline{\boldsymbol{u}}},$$

а также имеет место следующая оценка:

$$\|\underline{\boldsymbol{u}} - \widetilde{\underline{\boldsymbol{u}}}\|_{\boldsymbol{b},\infty} \le \left(\frac{1}{d-1}\right)\left(\frac{\rho(A)}{1-\rho(A)}\right)\mu(T,\widetilde{T}),$$

$\|\underline{x}\|_{b,\infty} = \sum_{i=1}^{d} b_i \|x_i\|_\infty$.

Четвёртая глава посвящена описанию численных экспериментов и их результатов, полученных с помощью разработанного программного комплекса. На практике подтверждаются теоретические оценки сложности и точности предложенных алгоритмов, проводится сравнение с существующими методами, демонстрирующее ускорение вычислений при сравнимой точности решения. Объекты исследования - синтетические и реальные данные, такие как изображения, данные отклика мозга в ответ на сенсорное, когнитивное либо двигательное явление, ядра уравнения Смолуховского (уравнения, описывающего математическую модель процессов коагуляции при неупругих соударениях огромного числа частиц), транспортная модель. Отдельное внимание уделяется сравнению сравнению двухэтапного подхода с подходом чередующихся проекций, общая идея которого заключается в построении итеративной последовательности матриц или тензоров, сходящихся к некоторой точке от пересечения двух множеств: множества неотрицательных матриц (тензоров) $M_{\ge 0}$ и множества матриц / тензоров низкого ранга $M_{\le \mathbf{r}}$. Во всех экспериментах двухшаговая факторизация позволяет значительно ускорить вычисления, являясь наилучшим выбором при работе с большими данными.

В заключении формулируются основные результаты работы.

Публикации автора по теме диссертации

[1] Shcherbakova E. Nonnegative Tensor Train Factorization with DMRG Technique // Lobachevskii Journal of Mathematics. – 2019. – V. 40, № 11. – P. 1863-1872. – (Web of Science, Scopus: SJR 2023 – 0.453) [0.7143/0.7143].

Работа полностью выполнена автором.

[2] Тыртышников Е.Е., Щербакова Е.М. Методы неотрицательной матричной факторизации на основе крестовых малоранговых приближений // Журнал вычислительной математики и математической физики. – 2019. – Т. 59, № 8. – С. 1314-1330. - (RSCI, двухлетний импакт-фактор РИНЦ – 1.115) [1.2/1.1] . Перевод:

Tyrtyshnikov E.E., Shcherbakova E.M. Methods for Nonnegative Matrix Factorization Based on Low-Rank Cross Approximations // Computational Mathematics and Mathematical Physics. – 2019. – V. 59. – P. 1251–1266. – (Web of Science, Scopus: SJR 2023 – 0.429) [1.2/1.1].

В статье идея редуцированного алгоритма для сепарабельных матриц была предложена автором, автор участвовала в доказательстве лемм и теорем, программная реализация и проведение численных экспериментов были выполнены автором полностью самостоятельно.

[3] Shcherbakova E., Tyrtyshnikov E. Nonnegative Tensor Train Factorizations and Some Applications // Lecture Notes in Computer Science. – 2020. – V. 11958 – P. 156-164. – (Scopus: SJR 2023 – 0.606) [0.643/0.637].

В работе автор предложила алгоритм, дала оценку точности аппроксимации, выполнила программную реализацию и провела численные эксперименты.

[4] Shcherbakova E., Tyrtyshnikov E. Fast Nonnegative Tensor Factorizations with Tensor Train Model // Lobachevskii Journal of Mathematics. – 2022.

– V. 43, № 4. – P. 882–894. – (Web of Science, Scopus: SJR 2023 – 0.453) [0.929/0.92].

В статье автор реализовала и развила идею точечной корректировки элементов в аппроксимации тензорным поездом с целью получения неотрицательного приближения, предложила двухэтапный подход к построению неотрицательного канонического разложения и разложения Таккера, снизив сложность одной итерации с $O(n^d)$ до $O(dnr^3)$, реализовала все методы программно и провела численные эксперименты.

[5] Shcherbakova Elena M., Matveev Sergey A., Smirnov Alexander P., Tyrtyshnikov Eugene E. Study of performance of low-rank nonnegative tensor factorization methods // Russian Journal of Numerical Analysis and Mathematical Modelling. – 2023. – V. 38, № 4. – P. 231-239. – (Web of Science, Scopus: SJR 2023 – 0.244) [0.643/0.637].

В работе автор провела численное исследование двух подходов малоранговой неотрицательной тензорной факторизации, которое на практике демонстрирует превосходство предложенных ей ранее алгоритмов.