

АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ
ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«СКОЛКОВСКИЙ ИНСТИТУТ НАУКИ И ТЕХНОЛОГИЙ»

УДК 004.896

УТВЕРЖДАЮ

Старший вице-президент
по связям с промышленностью

_____ А. Б. Начальников

«___» _____ 2022 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ
СПАСЕНИЕ МИРА С ИСПОЛЬЗОВАНИЕМ ТЕНЗОРНЫХ ПОЕЗДОВ
(промежуточный, этап 1)

Руководитель работы,
д. ф.-м. н., проф.

_____ А. Б. Тензорев

Москва 2022

СПИСОК ИСПОЛНИТЕЛЕЙ

д-р физ.-мат. наук проф.	<hr/> (подпись, дата)	А. Б. Тензорев (введение, заключение)
канд. физ.-мат. наук ст. преп.	<hr/> (подпись, дата)	В. Г. Матричный (введение, разделы 4, 5)
стажер-иссл.	<hr/> (подпись, дата)	Д. Е. Векторович (разделы 1, 2)
инж.-иссл.	<hr/> (подпись, дата)	Б. Ъ. Человеков (разделы 2, 3)
инж.	<hr/> (подпись, дата)	Б. Ъ. Человеков-Два (раздел 3)
магистр	<hr/> (подпись, дата)	Б. Ъ. Человеков-Три (раздел 1)
Нормоконтроль	<hr/> (подпись, дата)	Н. О. рмоконтролер

РЕФЕРАТ

Отчет 15 с., 1 кн., 4 рис., 1 табл., 5 источн., 0 прил.

МАЛОРАНГОВАЯ АППРОКСИМАЦИЯ, ТЕНЗОРНОЕ РАЗЛОЖЕНИЕ, ТЕНЗОР-
НЫЙ ПОЕЗД, ТТ-РАЗЛОЖЕНИЕ.

Здесь стоит написать 3 абзаца текста, содержащих: 1) общее содержание проводимых работ, их актуальность и значимость; 2) описание проведенных на данном этапе работ; 3) описание основных полученных результатов по проведенным работам. Во многих случаях удобнее писать реферат на основе ранее завершенных Введения и Заключения к работе. Отметим, что в текущей версии шаблона приложения не поддерживаются, соответственно счетчик приложений выше всегда будет указывать нуль.

/ Данный шаблон отчета (репозиторий; архив) может быть загружен и собран в системе «overleaf», либо локально на компьютере при наличии стандартной сборки «tex/latex». Вы можете свободно использовать данный шаблон для любых своих отчетов при указании в основном LaTeX файле проекта и в README файле проекта (при наличии) ссылки на репозиторий шаблона <https://github.com/AndreiChertkov/temscirep>. На основе предлагаемого шаблона было успешно подготовлено множество различных отчетов по коммерческим контрактам и государственным грантам, однако текущая версия шаблона в репозитории является черновой и будет в дальнейшем уточняться (пожалуйста, проверяйте наличие обновлений; можно даже поставить звездочку репозиторию, для того чтобы не потерять его в дальнейшем, а также морально поддержать авторов). Использование данного шаблона не гарантирует отсутствие замечаний заказчика по оформлению, но вероятность их отсутствия довольно высока. */*

СОДЕРЖАНИЕ

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ	5
ВВЕДЕНИЕ	6
1 Рекомендации по оформлению	7
1.1 Описание редактируемых файлов	7
1.2 Примеры оформления различных элементов	8
1.2.1 Оформление формул	8
1.2.2 Оформление рисунков	10
1.2.3 Оформление таблиц	10
1.2.4 Вспомогательные команды	11
1.2.5 Разное	12
1.3 Общие комментарии	12
1.4 Выводы по разделу	13
ЗАКЛЮЧЕНИЕ	14
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	15

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящем отчете о научно-исследовательской работе применяются следующие обозначения и сокращения.

API – программный интерфейс приложения (Application Programming Interface).

GUI – графический интерфейс пользователя (Graphical User Interface).

ML – машинное обучение (Machine Learning).

SVD – сингулярное разложение (Singular Value Decomposition).

ТТ-разложение – разложение тензорного произведения (Tensor Train decomposition).

ТТ-тензор – тензор, представленный в форме разложения тензорного произведения.

Вектор – одномерный упорядоченный массив вещественных или комплексных чисел (является также одномерным тензором).

ИИ – Искусственный Интеллект.

ИНС – Искусственная Нейронная Сеть.

Матрица – двумерный упорядоченный массив вещественных или комплексных чисел (является также двумерным тензором).

Отчет – данный отчет о НИОКР.

ПО – Программное Обеспечение.

ТЗ – Техническое Задание.

Тензор – одномерный или многомерный упорядоченный массив вещественных или комплексных чисел.

ЭО – Экспериментальный Образец.

$I_J \in \mathbb{R}^{J \times J}$ – диагональная единичная матрица размера $J \times J$.

$\mathcal{U}[:, i, :]$ – матрица, полученная как срез трёхмерного тензора \mathcal{U} при фиксированной второй размерности.

$\mathcal{A} \circ \mathcal{B}$ – Адамарово (поэлементное) произведение тензоров \mathcal{A} и \mathcal{B} .

$A \otimes B$ – Кронекерово произведение матриц A и B .

A^T – матрица, транспонированная к A .

ВВЕДЕНИЕ

Приводим абзац текста, с описанием текущего состояния дел по нашему направлению исследований, и формулируем целевую масштабную нерешенную научно-техническую проблему.

Соответственно основными целями проведения данной научно-исследовательской работы являются (обычно можно скопировать цель или цели из ТЗ):

- а) Первая цель работы;
- б) Вторая цель работы;
- в) Третья цель работы.

Основанием для проведения работ является ТЗ на научно-исследовательскую работу по теме «Спасение мира с использованием тензорных поездов», оформленное как приложение к Договору от 01.01.22 № 424242, заключенному между Министерством Добрых Дел и АНОО ВО «Сколковский институт науки и технологий». При этом, для успешного достижения поставленных целей на данном, первом, этапе работ по Проекту, было запланировано решение следующих задач (обычно можно скопировать задачи из ТЗ):

- а) Первая задача работы на этапе;
- б) Вторая задача работы на этапе;
- в) Третья задача работы на этапе;
- г) Четвертая задача работы на этапе;
- д) Пятая задача работы на этапе.

Все поставленные задачи были выполнены нами в полном объеме, и далее в отчете мы описываем соответствующие основные полученные результаты. В разделе 1 мы приводим разработанный В разделе Наконец, в Заключение формулируются основные выводы по проведенной на данном этапе работе и обсуждаются возможные пути ее дальнейшего развития.

1 Рекомендации по оформлению

В данном разделе мы приводим основные рекомендации по оформлению отчета (в будущем они будут уточнены). Отметим, что некоторые из рекомендаций могут оказаться нелогичными и / или неверными (если вы заметили подобную рекомендацию, то, пожалуйста, сообщите об этом разработчикам шаблона).

Если раздел содержит подразделы, то в разделе должно также обязательно присутствовать заключение (соответствующий подраздел может называться «Выводы» или «Выводы по разделу»), а также должно быть введение (вместо создания специального подраздела «Введение», мы приводим для компактности вводный текст непосредственно в начале раздела до первого подраздела – это не является грубым нарушением ГОСТа, однако при желании можно формально добавлять подразделы «Введение»). См. в качестве примера раздел 1 и раздел 1.4.

В качестве названия разделов во многих случаях стоит указывать в точности названия пунктов плана работ из Соглашения по гранту или контракту (часто заказчики это требуют явно), чтобы проверяющий мог осуществить формальную проверку соответствия отчета пунктам официального ТЗ. Во введении к разделу стоит указать общие мысли по теме, также можно описать, что будет содержаться в последующих подразделах (см. пример далее).

В разделе 1.1 мы детально опишем структуру данного проекта, приведем перечень редактируемых файлов в репозитории и их назначение. Далее в разделе 1.2 мы рассмотрим примеры оформления различных элементов отчета, таких как таблицы и графические изображения. Затем в разделе 1.3 мы сформулируем общие рекомендации по оформлению отчета о научно-исследовательской работе.

1.1 Описание редактируемых файлов

В процессе работы над отчетом могут (и должны) редактироваться следующие файлы:

- а) «main.tex» – основной файл проекта. При добавлении нового раздела отчета в папку «content», ссылка на раздел должна быть добавлена в данный файл;
- б) «content/» – папка для хранения основных разделов отчета. Ссылки на разделы должны быть указаны в файле «main.tex». В репозитории данная папка содержит раздел с рекомендациями по оформлению («content/99_demo.tex» и файлы с подразделами);
- в) «images/» – папка для хранения графических изображений. Отметим, что в latex-команде «includegraphics» достаточно указывать только имя файла (без указания папки). В репозитории данная папка содержит несколько примеров изображений;
- г) «biblio.bib» – файл с библиографией в bibtex-формате. Поля библиографической записи (англоязычные) автоматически приводятся к нижнему регистру. При необходимости

сохранения заглавных букв, их нужно заключать в фигурные скобки;

д) «commands.tex» – файл содержит различные полезные команды (выделение векторов, матриц, тензоров и т.п.). В конце данного файла могут быть заданы собственные команды и сокращения;

е) «special/title.tex» – файл содержит титульный лист отчета. См. также титульный лист для мегагранта в файле «special/title_mega.tex» (переключение на мегагрант осуществляется посредством задания команды «mega» в файле «main.tex»);

ж) «special/performers.tex» – файл содержит список исполнителей отчета;

з) «special/abstract.tex» – файл содержит реферат отчета;

и) «special/definition.tex» – файл содержит перечень обозначений и сокращений;

к) «special/intro.tex» – файл содержит введение к отчету;

л) «special/conclusion.tex» – файл содержит выводы по отчету.

Также проект содержит следующие файлы, редактирование которых не предполагается:

а) «README.md» – файл с общим описанием репозитория в markdown формате;

б) «gost.bst» – стандартный стилевой файл для оформления библиографии по ГОСТу;

в) «style.sty» – наш набор стилей для оформления отчета в соответствии с ГОСТ.

1.2 Примеры оформления различных элементов

В данном разделе мы приводим рекомендации и примеры оформления различных типовых элементов отчета: формул (см. раздел 1.2.1), рисунков (см. раздел 1.2.2), таблиц (см. раздел 1.2.3), различных вспомогательных команд (см. раздел 1.2.4) и иных элементов (см. раздел 1.2.5).

1.2.1 Оформление формул

Формулы оформляются стандартным образом. Отметим, что все формулы (даже те, на которые нет ссылок в тексте) стоит нумеровать. Перед формулой ставится двоеточие, а после формулы необходимо вставить знак препинания (запятая или точка).

Для произвольного элемента (i_1, \dots, i_d) тензора $\mathcal{A} \in \mathbb{R}^{N_1 \times N_2 \times \dots \times N_d}$, разложение тензорного произведения (ТТ-разложение) может быть записано как:

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



Рисунок 1 – Схематическая форма представления разложения тензорного поезда

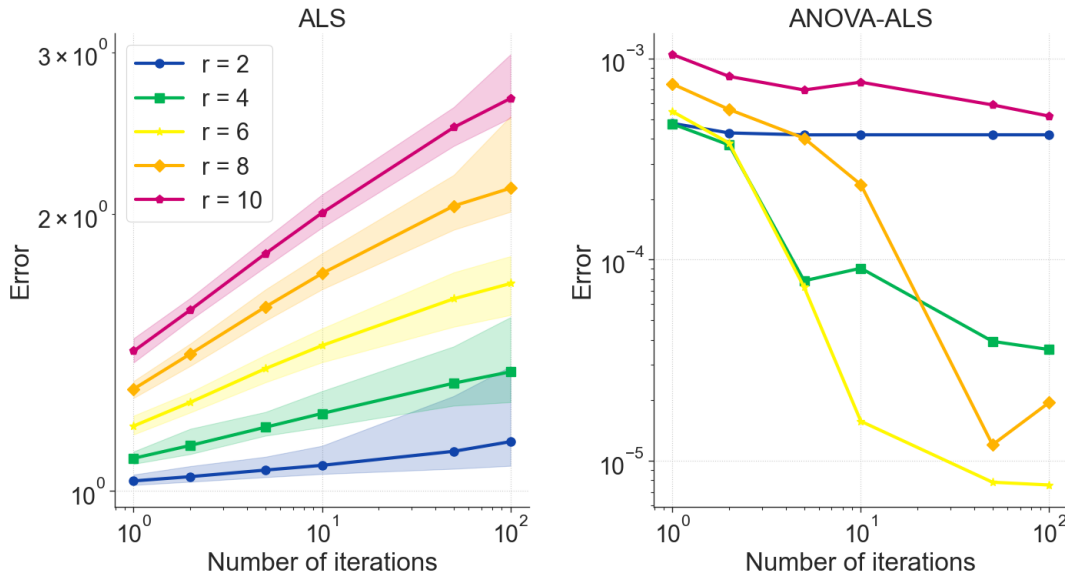


Рисунок 2 – Ошибка приближения целевой величины стандартным методом ТТ-ALS (на графике слева) и предлагаемым методом ТТ-ANOVA-ALS (на графике справа)

где числа (R_1, R_2, \dots, R_d) — это ТТ-ранги (обычно для упрощения записи формально вводятся также ранги $R_0 = R_d = 1$), а трехмерные тензоры $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_d$ — ядра ТТ-разложения (\mathcal{G}_1 и \mathcal{G}_d являются двухмерными тензорами, но формально могут рассматриваться как трехмерные тензоры с единичным размером первой и последней моды соответственно). ТТ-разложение (1) также может быть записано в более компактной матричной форме:

$$\mathcal{A}[i_1, i_2, \dots, i_d] = G_1(i_1)G_2(i_2) \dots G_d(i_d). \quad (2)$$

В последней формуле $G_k(i_k)$ — это ядра ТТ-разложения, записанные в виде матриц размера $R_{k-1} \times R_k$.

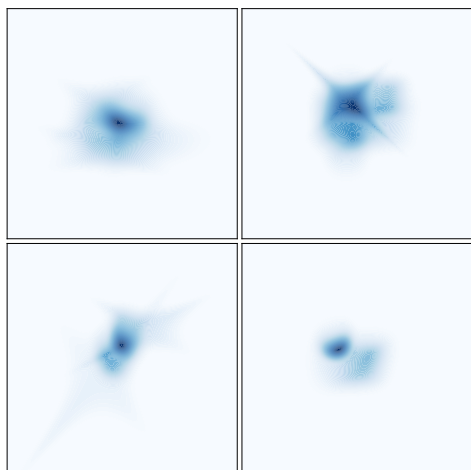


Рисунок 3 – Пример двумерного распределения, соответствующего какой-то ерунде

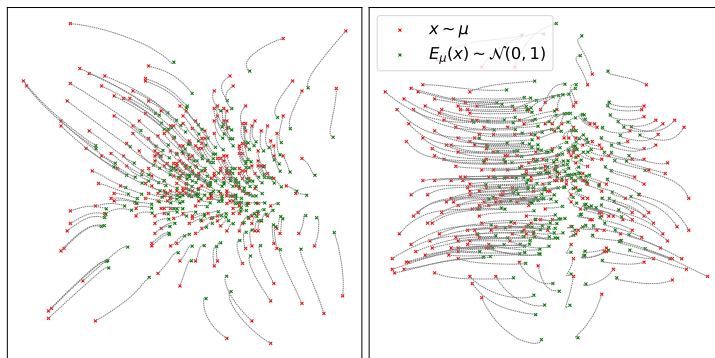


Рисунок 4 – Пример траекторий каких-то там потоков, текущих куда-то

Таблица 1 – Модельные функции из работы [2] для анализа методов аппроксимации

Функция	Ссылка	Область	Аналитическая формула
Rosenbrock	[4]	$[-2.048, 2.048]$	$f(\mathbf{x}) = \sum_{i=1}^{d-1} (100 \cdot (x_{i+1} - x_i^2)^2 + (1 - x_i)^2)$
Schaffer	[4]	$[-100, 100]$	$f(\mathbf{x}) = \sum_{i=1}^{d-1} (0.5 + \frac{\sin^2(\sqrt{x_i^2 + x_{i+1}^2}) - 0.5}{(1 + 0.001(x_i^2 + x_{i+1}^2))^2})$
Schwefel	[3]	$[-500, 500]$	$f(\mathbf{x}) = 418.9829 \cdot d - \sum_{i=1}^d x_i \cdot \sin(\sqrt{ x_i })$

1.2.2 Оформление рисунков

На рисунке 1 в схематической форме приводится общий вид ТТ-разложения. Можно делать подпись сразу к двум картинкам как показано на рисунке 2. Если «очень приспичит», то можно и вот так оформлять рисунки: на рисунке 3 изображена одна фигня, а на рисунке 4 – другая фигня.

Файлы с графическими изображениями должны храниться в папке «images». Рисунки удобно размещать в верхней части листов, «разбавляя» пространство между последовательными рисунками текстом. Отметим, что в конце подписи к рисунку точка не ставится.

1.2.3 Оформление таблиц

Таблицы оформляются стандартным образом. В качестве примера см. таблицу 1. Отметим, что в конце подписи к таблице точка не ставится.

1.2.4 Вспомогательные команды

В файле «commands.tex» приводятся различные удобные вспомогательные команды для отображения различных математических объектов и операций, в том числе:

- а) Жирный шрифт в тексте или в формуле $x^2 = \mathbf{y}^4$;
- б) Множество $\in \mathbb{N}$ или $\in \mathbb{R}$;
- в) Вектор $\mathbf{x} \in \mathbb{R}^{N_1}$;
- г) Матрица $X \in \mathbb{R}^{N_1 \times N_2}$;
- д) Тензор $\mathfrak{X} \in \mathbb{R}^{N_1 \times N_2 \times \dots \times N_d}$;
- е) Функция $\mathbf{f}(\mathbf{x})$;
- ж) Вектор-функция $\mathbf{f}(\mathbf{x})$;
- з) Функционал $\widehat{f}(\cdot)$;
- и) Оператор \widehat{W} ;
- к) След матрицы $\text{Tr}(A)$;
- л) Порядок $\mathcal{O}(n^2)$;
- м) Дивергенция $\text{div}[\mathbf{x}]$;
- н) Производная $\frac{df}{dx}$;
- о) Производная по времени $\frac{df}{dt}$;
- п) Производная по x $\frac{df}{dx}$;
- р) Частная производная $\frac{\partial f}{\partial z}$;
- с) Частная производная по времени $\frac{\partial f}{\partial t}$;
- т) Частная производная по x $\frac{\partial f}{\partial x}$;
- у) Частная производная по компоненте координаты $\frac{\partial f}{\partial x_3}$;
- ф) Векторизация $\mathbf{vec}(\mathfrak{X})$;
- х) Построение вектора $[1, 2, \dots, d]^\top$;
- ц) Построение индекса из мульти-индекса $i = \text{ind}(i_1, i_2, \dots, i_d)$;
- ч) Построение мульти-индекса из индекса $\mathbf{i} = \mathbf{mind}(i)$;
- ш) Точка $\widehat{f}(\cdot)$;
- щ) Заметка /* Надо будет все уточнить */ .

1.2.5 Разное

Можно и нужно использовать сноски для вспомогательных комментариев, например, для пояснения, что такое тензор⁽¹⁾ и т.п. Ссылки на статьи оформляются стандартным образом, см., например, работу [5] или работы [3, 4, 1]. /* На данный момент русскоязычные записи в библиографии часто приводят к ошибкам, будет исправлено. */

Отметим, что малоранговая аппроксимация на основе ТТ-разложения [5] представляет один из распространенных подходов для компактного приближения функций многих переменных, включая приложения в области анализа данных и машинного обучения. В работе [5] доказывается теорема существования точного ТТ-представления с рангами R_k не выше, чем $\text{rank } A^{(k)}$, где $A^{(k)}$ — это k -ая матричная развертка тензора.

1.3 Общие комментарии

- а) Списки стоит употреблять только типа «enumerate», начинать пункты с заглавной буквы, в конце всех пунктов (кроме последнего) ставить точку с запятой;
- б) Кавычки стоит делать вот так «привет»;
- в) Абзацы текста должны содержать не менее трех предложений;
- г) Числа в тексте стоит оформлять как инлайновые формулы, например 42;
- д) Ссылки стоит соединять с текстом тильдами для неразрывности, то есть не «рисунок 1», а «рисунок 1» или «см. раздел 1.2»;
- е) Стандартные сокращения стоит использовать с тильдами: «и т.д.», «и т.п.»;
- ж) Для именования разделов в папке «content» удобно использовать систему следующего типа: «99_3 demo_comments», то есть последовательно указывается номер раздела, нижнее подчеркивание, номер подраздела, пробел, имя раздела, нижнее подчеркивание, имя подраздела. В данном случае в качестве метки на подраздел можно указать «sec:demo_comments» (см. раздел 1.3). Подобное единообразие позволит быстро указывать верные метки в тексте, а также визуально различать разделы и подразделы в файловом менеджере. Отметим, что мы используем номер «99» для данного раздела, чтобы его всегда можно было включить в черновую версию отчета в качестве последнего раздела (для справки по оформлению);
- з) Предложения в latex файле удобно оформлять с новой строки — это в ряде случаев позволяет быстрее локализовать ошибку, а также упрощает процесс комментирования ненужных предложений;

⁽¹⁾Здесь и далее (в соответствии с нормами современной вычислительной математики и машинного обучения) термин *тензор* является фактически синонимом для термина *многомерный массив*.

- и) Стоит не забывать использовать « \rightarrow » вместо «-», когда это требуется;
- к) Стоит аккуратно оформлять библиографические записи. В качестве тегов стоит использовать (как и в системе Google Scholar) «фамилияГОДслово» (где «слово» – это первое слово названия статьи; если в слове имеется дефис, то его следует убрать), например, «oseledets2011tensortrain» (см. [5]). Стоит использовать единый способ представления ФИО (например, «Фамилия, И.»). Иногда удобно располагать библиографические записи в файле «biblio.bib» в алфавитном порядке;
- л) Для «строковых» символов стоит использовать нижнее троеточие: « a, b, \dots, z », а для «центрированных» символов можно вот так: « $a \times b \times \dots \times z$ »;
- м) Стоит использовать единую систему абзацных отступов в latex коде внутри блоков (графические изображения, таблицы и т.п.) для единообразия;
- н) Перед формулами ставим двоеточие, а после формулы точку или запятую;
- о) Подписи к рисункам и таблицам начинаем с большой буквы, в конце точку не ставим, в тексте ссылаемся без сокращений: «см. рисунок 1» или «приводится в таблице 1»;
- п) Хорошо если на последней странице раздела остается в конце мало пустого места;
- р) Букву «ё» употреблять в тексте не стоит.

1.4 Выводы по разделу

Как было показано в данном разделе, оформлять отчет не так уж и сложно. Нами были описаны основные файлы и папки репозитория с данным шаблоном отчета, а также были сформулированы общие рекомендации по оформлению текста отчета. Если это не последний раздел, то здесь можно кратко описать, что будет рассмотрено далее, например, «в следующем разделе мы перейдем к рассмотрению более конкретных ...».

ЗАКЛЮЧЕНИЕ

Можно привести абзац общего текста, кратко повторяющего (другими словами) описание актуальности и новизны работы из Введения.

На данном, первом, этапе работ по проекту нами были получены следующие основные результаты (обычно можно скопировать ожидаемые результаты из ТЗ, либо использовать перечень задач из Введения к работе с очевидным переводом «задача → результат»):

- а) Первый результат работы на этапе;
- б) Второй результат работы на этапе;
- в) Третий результат работы на этапе;
- г) Четвертый результат работы на этапе;
- д) Пятый результат работы на этапе.

Таким образом, все поставленные в соответствии с ТЗ задачи были выполнены нами в полном объеме. В силу универсальности и уникальности предложенных нами подходов, разработанные алгоритмы являются расширяемыми и масштабируемыми, а научно-технический задел, полученный на данном этапе работ позволит успешно реализовать задачи следующих этапов работ по проекту, а также может быть использован и для ряда иных практически значимых задач в области ...

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] An abstract domain for certifying neural networks / G. Singh, T. Gehr, M. Püschel, M. Vechev // *Proceedings of the ACM on Programming Languages*. — 2019. — Vol. 3, no. POPL. — Pp. 1–30.
- [2] *Chertkov A., Ryzhakov G., Oseledets I.* Black box approximation in the tensor train format initialized by ANOVA decomposition // *arXiv preprint arXiv:2208.03380*. — 2022.
- [3] *Dieterich J., Hartke B.* Empirical review of standard benchmark functions using evolutionary global optimization // *Applied Mathematics*. — 2012. — Vol. 3, no. 10. — Pp. 1552–1564.
- [4] *Jamil M., Yang X.* A literature survey of benchmark functions for global optimization problems // *Journal of Mathematical Modelling and Numerical Optimisation*. — 2013. — Vol. 4, no. 2. — Pp. 150–194.
- [5] *Oseledets I.* Tensor-train decomposition // *SIAM Journal on Scientific Computing*. — 2011. — Vol. 33, no. 5. — Pp. 2295–2317.