

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
НАУКИ
ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ
ИМЕНИ Г.И. МАРЧУКА
РОССИЙСКОЙ АКАДЕМИИ НАУК

На правах рукописи

Петров Сергей Владимирович

**ЭФФЕКТИВНЫЕ МЕТОДЫ ПРИБЛИЖЕНИЯ
МАТРИЦ И ТЕНЗОРОВ В УСЛОВИЯХ НЕПОЛНЫХ И
ЗАШУМЛЕННЫХ ДАННЫХ**

Специальность 1.1.6 —
«Вычислительная математика»

Диссертация на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
Академик РАН, профессор
Тыртышников Евгений Евгеньевич

Москва — 2023

Оглавление

	Стр.
Введение	5
Глава 1. Восполнение матриц малого ранга	13
1.1 Метод проекции градиента для общей задачи минимизации при условии ограниченной изометрии	14
1.1.1 Случай точного проектирования	14
1.1.2 Случай приближенного проектирования	16
1.2 Частный случай оператора восполнения матриц	23
1.3 Практические модификации метода проекции градиента	26
1.3.1 Последовательный набор ранга приближения	26
1.3.2 Выбор параметра шага	28
1.4 Методы приближенного вычисления проекции на множество матриц малого ранга	30
1.4.1 Проектирование на случайные пространства	30
1.4.2 Псевдоскелетный метод на матрицах большого проективного объема	32
1.5 Численные результаты	35
Глава 2. Приближение матриц в виде суммы матрицы малого ранга и разреженной	44
2.1 Теория возмущений сингулярных подпространств	45
2.1.1 Углы между подпространствами	45
2.1.2 Теорема о возмущениях	46
2.2 Анализ сходимости восполнения матриц с неточными входными данными на основе теории возмущений	48
2.2.1 Оценки возникающего возмущения	48
2.2.2 Более строгое условие приближенного проектирования	50
2.2.3 Общий вид теорем о сходимости методов восполнения с неточными данными	54
2.3 Частный случай разреженных ошибок	57

2.3.1	Пересечение шаблонов разреженности	59
2.3.2	Нормы разреженных ошибок	61
2.4	Метод парного восполнения	64
2.5	Эксперименты на искусственных данных	66
Глава 3. Тензорные приближения малого ранга в условиях зашумленных данных		
3.1	Обозначения тензорных операций	74
3.2	Аппарат гауссовой ширины	76
3.2.1	Общие рассуждения	76
3.2.2	Аналитический вид введенной нормы для множества матриц ограниченного ранга	79
3.2.3	Связь введенной величины с задачей приближения тензора шума тензором малого ранга	80
3.3	Гауссова ширина множества тензоров ранга 1 ограниченной нормы	81
3.4	Обобщение на случай тензоров большего ранга и различных тензорных форматов	87
3.4.1	Канонический формат	87
3.4.2	Формат Таккера	90
3.4.3	Формат тензорного поезда	92
Глава 4. Восполнение тензоров в формате Таккера		
4.1	Обобщение метода проекции градиента на восполнение тензоров в формате Таккера	94
4.1.1	Сведение к матрицам развертки	94
4.1.2	Снижение сложности метода проекции градиента на развертках	96
4.2	Замена оператора восполнения	98
Глава 5. Практические приложения построенной теории и алгоритмов		
5.1	Автомобильные радары	102
5.1.1	Структура возникающих массивов данных	102
5.1.2	Квантизация	105
5.1.3	Комбинирование моделей	106

5.2	Распространение сигналов по каналам беспроводной связи MIMO	107
5.3	Численные методы для интегральных и дифференциальных уравнений	110
5.3.1	Нелинейное уравнение магнитостатики	110
5.3.2	Уравнения рассеяния на метаповерхностях	116
	Заключение	119
	Список сокращений и условных обозначений	120
	Список литературы	121
	Публикации автора по теме диссертации	126
	Список рисунков	127
	Список таблиц	129

Введение

Актуальность темы исследования.

Для большинства современных математических алгоритмов, в частности, численных методов линейной алгебры, их сложность имеет первоочередное значение для практических приложений. Она определяет либо применимость алгоритма в режиме реального времени, либо допустимые размеры задачи, при которых алгоритм сможет завершить работу на современном вычислительном устройстве за разумное время, используя при этом ограниченные объемы памяти. В связи с этим, современная вычислительная математика часто опирается на различные малопараметрические представления данных: вводится предположение, что фигурирующие в алгоритмах входные данные и/или неизвестные с контролируемой погрешностью подчиняются некоторой наперед заданной модели, характеризующейся сравнительно небольшим числом параметров. На основе этого предположения строится упрощенный алгоритм, вычислительная сложность и необходимая величина объема памяти которого либо, в идеальном случае, зависит только от числа параметров модели, либо просто снижена. При успешном построении такого алгоритма с сохранением его практической эффективности говорят о редукции размерности модели.

В случае данных, организованных в виде матрицы, наиболее типичной и известной малопараметрической моделью является модель матрицы малого ранга. Эта модель хорошо изучена: в частности, известно, что оптимальное приближение любой матрицы во фробениусовой, спектральной и любой другой унитарно инвариантной норме [1] можно получить с помощью сингулярного разложения матрицы за (примерно) кубическое относительно линейного размера матрицы число операций. Недавние результаты [2] объясняют, почему многие матрицы, встречающиеся в приложениях, хорошо приближаются матрицами малого ранга: в частности, показано, что *любые* матрицы большого размера обладают приближениями малого ранга в чебышевской норме, качество которых зависит от ранга как обратная величина квадратного корня.

Несмотря на существование широкой теоретической базы для модели матриц малого ранга, исследования практически значимых задач, связанных с этой моделью, продолжают.

Распространенной темой исследований последних десятилетий являются быстрые алгоритмы построения приближений малого ранга, близких к оптимальным. Погрешность такого приближения не должна достигать точного минимума по всевозможным матрицам заданного ранга, но должна быть близка к этому минимуму, а вычислительная сложность соответствующего алгоритма должна быть ниже сложности построения сингулярного разложения. Для такой задачи в литературе исследованы, в частности, алгоритмы на основе крестовой аппроксимации [3–6], и алгоритмы, основанные на проектировании на случайные подпространства [7; 8].

Другим современным направлением исследований, связанным с моделью матриц малого ранга, является задача восполнения матриц. Задачей матричного восполнения называют задачу поиска приближений малого ранга для матриц в случае неполных данных, т.е. в случае, когда известна только небольшая часть элементов матрицы. Приложения задачи матричного восполнения включают в себя рекомендательные системы [9], обработка коррелирующих сигналов [10–12], машинное обучение [13; 14] и другие области.

Еще одной развивающейся темой исследований, связанной с моделью матриц малого ранга, является задача приближения матриц в формате суммы матрицы малого ранга и разреженной. Известными приложениями данной модели являются, например, задача усвоения данных измерений, или задача выделения движущихся объектов на видеозаписях [15; 16]; также в рамках данной работы будет рассмотрено применение этой модели к решению линейных систем, связанных с интегральными уравнениями. Задача представления матрицы в виде суммы матрицы малого ранга и разреженной матрицы некорректна без дополнительных предположений; в частности, без условия 'неразреженности' малорангового слагаемого. В работе [17] предложены формализация такого условия и алгоритм, обладающий гарантиями сходимости; однако, этот алгоритм обладает существенной вычислительной сложностью, так как используют выпуклую оптимизацию с m^2 неизвестными, где m - линейный размер матрицы.

В современных приложениях также возникают структуры малого ранга большей размерности, или тензоры малого ранга. Модель тензоров малого ранга имеет большую актуальность, в частности, в задачах, связанных с обработкой сигналов беспроводных сетей или обнаружением объектов на автомобильном радаре. Физическая модель сигнала в этих приложениях описывается много-

мерным комплекснозначным тензором, размерности которого соответствуют частоте, времени, индексу принимающей антенны и индексам двумерной решетки отправляющих антенн базовой станции. Согласно распространенной „многолучевой“ модели такой тензор допускает каноническое разложение с небольшим рангом, соответствующим числу различных путей от источника сигнала к приемнику, которые появляются за счет переотражений сигнала от объектов среды, например, зданий [18; 19]. В приложениях, связанных с беспроводными сетями, актуален широкий спектр математических задач, связанных с тензором канала [20]. Так, построение приближения малого ранга к тензору канала соответствует сжатию и очистке от шума канала беспроводной связи. Если при этом в качестве входных данных является не весь тензор канала, а лишь небольшой набор его измерений, то говорят о задаче оценки канала. На основе приближений тензора канала в реальном времени проводится построение матриц прекодирования, непосредственно описывающих настройку передаваемых с антенн базовой станции сигналов. Точные теоретические оценки погрешности приближений зашумленного канала при этом могут улучшить прекодирование и результирующую пропускную способность сети [0]. Исследование устойчивости тензорной структуры малого ранга к шуму актуально и в других приложениях: извлечение малорангового тензора из шума большой нормы соответствует задаче обнаружения объектов на автомобильных радарх [19]. При таком исследовании отдельную сложность представляет свойство незамкнутости множества тензоров ограниченного ранга. Так, *оптимального* приближения наперед заданного тензора с малым каноническим рангом может не существовать, а даже если оно существует, его нахождение затруднительно на практике. Практически используемые алгоритмы поиска разложений тензора в формате Таккера [21], формате тензорного произведения [22] и каноническом формате [23] оптимальное разложение на выходе не дают, а в случае канонического формата - и вовсе не имеют теоретических гарантий сходимости. Поэтому, построение теории устойчивости тензорных приближений не должно опираться на *оптимальность* таких приближений.

Наконец, методы редукции размерности, основанные на структурах малого ранга, применимы и к нелинейным задачам, в частности, к численным схемам, соответствующим нелинейным дифференциальным уравнениям с параметрами. Матрицы малого ранга можно использовать для ограничения про-

странства, в котором проводится поиск решений, а также для ускорения вычислений, связанных с нелинейной функцией, путем интерполяции [24; 25]. Такой подход применим для различных дифференциальных уравнений с параметрами, в частности, уравнений магнитостатики, связанных с моделированием электродвигателей [26—28].

Целью данной работы является разработка алгоритмов малой вычислительной сложности, связанных с поиском малоранговых приближений матриц и тензоров в условиях неполных данных и/или высокого уровня шума, а также построения теории их сходимости.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Разработать и аналитически исследовать возможность понижения сложности известных алгоритмов [29] восполнения матриц, основанных на методах проекции градиента, с помощью приближенных методов вычисления частичного сингулярного разложения.
2. Исследовать возможность адаптации алгоритмов восполнения к случаю разреженных ошибок во входных данных. Связать эту задачу с аппроксимацией матрицы в виде суммы матрицы малого ранга и разреженной.
3. Теоретически исследовать близость малорангового приближения наперед заданного тензора, подверженного возмущению, к исходному тензору. Проверить гипотезу об улучшении свойств такой близости у тензорных приближений по сравнению с матричными при том же ранге и числе элементов.
4. Применить разработанные алгоритмы и теоретические результаты к практическим задачам обработки сигналов из областей беспроводной связи и автомобильных радаров.

Научная новизна:

1. Впервые предложен и теоретически обоснован метод проекции градиента для задачи восполнения матриц для случая приближенного проектирования, вычисляемого с помощью проектирования на случайные подпространства или крестовой аппроксимации.
2. Впервые предложен метод аппроксимации матриц в виде суммы матрицы малого ранга и разреженной матрицы, основанный на матричном

- восполнении, имеющий меньшую сложность по сравнению с известными в литературе и сходящийся геометрически на искусственных данных. Доказан ряд утверждений для теоретического обоснования алгоритма.
3. Впервые предложена теория точности тензорных приближений малого ранга в условиях зашумленных данных. Получены асимптотические оценки близости малорангового приближения наперед заданного тензора, подверженного возмущению большой нормы, к исходному тензору.
 4. Впервые предложен метод приближения тензоров в формате Таккера с малыми рангами, основанный на методах восполнения и быстром преобразовании Джонсона-Линденштраусса.

Теоретическая и практическая значимость. На основе проведенных исследований разработаны методы обработки сигналов в приложениях, связанных с автомобильными радарными и станциями беспроводной связи, основанные на тензорных разложениях малого ранга.

Методология и методы исследования. В работе используются математические методы линейной алгебры, методов оптимизации и теории вероятностей.

Основные положения, выносимые на защиту:

1. Для задачи восполнения матриц предложен и теоретически обоснован метод приближенной проекции градиента.
2. Метод аппроксимации матриц в виде суммы матрицы малого ранга и разреженной матрицы, основанный на матричном восполнении, имеющий меньшую сложность по сравнению с известными в литературе и сходящийся геометрически на искусственных данных.
3. Теорема о точности тензорных приближений в условиях зашумленных данных.
4. Метод приближения тензоров в формате Таккера с малыми рангами малой сложности.

Степень достоверности и апробация работы. Достоверность полученных результатов подкреплена согласованностью выводов аналитического исследования и численного моделирования. Основные результаты работы докладывались на следующих научно-исследовательских школах, конференциях и семинарах.

1. Конференция SIAM Conference on Applied Linear Algebra (LA21), 17-21 мая 2021, онлайн-формат,
2. Конференция Matrix Equations and Tensor Techniques IX (METTIX), Perugia, Италия, 9-10 сентября 2021 (гибридный формат),
3. Конференция The 5th International Conference on Matrix Methods in Mathematics and applications (МММА-2019, Москва),
4. Конференция Large-Scale Scientific Computations, 10-14 июня 2019 (LSSC'19, Созополь, Болгария),
5. Конференция Random Matrix Theory and Beyond Workshop, Сочи, 8-9 августа 2022,
6. Конференция Electromagnetic Simulation Workshop 2021, Москва, 17 декабря 2022,
7. Конференция Huawei Russian Wireless Workshop 2020, Москва, 21-23 октября 2020,
8. Летняя школа «Римско-Московская школа по матричным методам и прикладной линейной алгебре» (Москва, Россия, 20 августа - 3 сентября 2016, Рим, Италия, 4-18 сентября 2016).

Личный вклад. Все результаты работы получены автором лично под научным руководством академика РАН, профессора Тыртышников Е.Е, при консультациях к.ф.м.н. Замарашкина Н.Л.. В работах, написанных в соавторстве, вклад автора диссертации состоит в следующем: предложены методы построения приближения матриц в формате суммы матрицы малого ранга и разреженной матрицы, основанные на методах восполнения с одной или двумя масками; доказаны теоретические утверждения о сходимости обоих алгоритмов, а также проведены численные эксперименты; доказана теорема о сходимости приближенного метода проекции градиента общего вида и подробно рассмотрен частный случай построения приближенных проекций с помощью случайных подпространств.

Публикации. Основные результаты по теме диссертации изложены в 3 печатных изданиях, 3 из которых изданы в журналах Scopus, WoS, RSCI, а также в изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности.

Объем и структура работы. Диссертация состоит из введения, пяти глав, и заключения. Полный объём диссертации составляет 129 страниц с 25 рисунками и 4 таблицами. Список литературы содержит 56 наименований.

В **первой главе** исследуется модель матриц малого ранга в специальных условиях: рассматривается задача построения приближения малого ранга в условиях неполных данных, известная также как задача восполнения матрицы. В частности, рассматривается возможность понижения вычислительной сложности существующего глобально сходящегося метода восполнения матриц с помощью использования приближенного проектирования на множество матриц малого ранга. В качестве такого приближения можно использовать методы, основанные, например, на проектировании на случайные подпространства. Предлагается теорема о сходимости метода восполнения с таким приближенным проектированием.

Вторая глава посвящена возможности решения задачи восполнения матриц в случае входных данных с разреженной ошибкой; такая задача также может быть рассмотрена как задача приближения матрицы в формате суммы матрицы малого ранга и разреженной матрицы. Предлагается новый алгоритм, основанный на методах восполнения, и позволяющий эффективно решать задачу приближения матрицы в виде суммы матрицы малого ранга и разреженной матрицы, не обладая информацией о позициях элементов разреженной части.

В **третьей главе** рассматриваются многомерные модели тензоров малого ранга и вопросы устойчивости модельных приближений к большому белому шуму, который по норме может в несколько раз превышать норму самого приближения. В главе вводится набор предположений о понятии „приближения“, позволяющих получить теоремы об асимптотических характеристиках устойчивости тензорных приближений к шуму, подтверждаемых численными экспериментами.

В **четвертой** главе рассмотрено обобщение алгоритмов восполнения на случай тензоров в формате Таккера, а также рассмотрена возможность построения алгоритмов аппроксимации по модели Таккера низкой сложности на основе алгоритмов восполнения общего вида.

В **пятой** главе рассмотрены практические приложения построенной в работе теории и алгоритмов. Рассмотрены приложения теории устойчивости тензорных приближений из области автомобильных радаров и обработки сигналов

беспроводной связи, а также приложения построенного в работе алгоритма, связанного с моделью матриц малого ранга, к численному методу решения интегрального уравнения. Также отдельно рассмотрена возможность приложения модели матриц малого ранга к численному решению нелинейного дифференциального уравнения, используемого при моделировании электродвигателей.

Глава 1. Восполнение матриц малого ранга

Данную диссертационную работу начнем с рассмотрения классической малопараметрической модели матрицы малого ранга. Структура оптимальных малоранговых матричных приближений широко известна в случае, если качество приближения измеряется в любой унитарно инвариантной норме, и полностью описывается с помощью частичного сингулярного разложения, поэтому в настоящее время задача построения матричных приближений малого ранга представляет интерес только в присутствии дополнительных ограничений или условий.

Один из возможных вариантов такого дополнительного условия будет рассмотрен в этой главе: будет дополнительно предполагаться, что входные данные являются неполными. Иными словами, пусть известна лишь часть элементов матрицы, удовлетворяющей модели матрицы малого ранга, а требуется найти ее целиком. Такая задача будет называться задачей восполнения матрицы [29; 30].

Данная глава будет посвящена вычислительному аспекту решения такой задачи. Алгоритмы для решения задачи восполнения матриц уже были предложены ранее в литературе. Часто [29; 31; 32] такие алгоритмы предполагают итерационную схему с использованием частичного сингулярного разложения на каждой итерации, вычисление которого является операцией наиболее высокой сложности, на практике являющейся кубической относительно размеров матрицы, и может быть неприемлема для приложений. С учетом того, что объемы входных и выходных данных задачи восполнения близки к линейным, естественно исследовать возможность снижения вычислительной сложности алгоритмов восполнения. В этой главе такое снижение будет проведено с помощью замены точного вычисления частичного сингулярного разложения приближенным. В качестве алгоритма построения такого приближенного разложения будут рассмотрены методы

- метод проектирования на случайные пространства [7; 8; 33];
- метод псевдо-крестовой аппроксимации, основанной на принципе большого проективного объема [34–36].

При этом не предполагается высокой точности получаемых приближений. Напротив, достаточно такой точности, при которой алгоритм восполнения сохра-

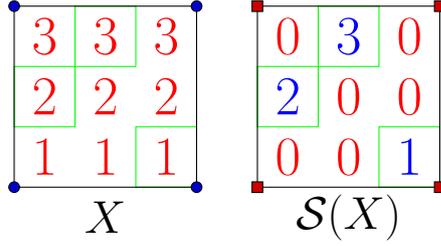


Рисунок 1.1 — Иллюстрация входных данных задачи восполнения матрицы

имеет свойство глобальной сходимости к искомой матрице X , а возможное увеличение общего числа итераций незначительно и компенсируется скоростью построения приближениями малого ранга на каждой итерации.

1.1 Метод проекции градиента для общей задачи минимизации при условии ограниченной изометрии

1.1.1 Случай точного проектирования

В рамках данной работы сконцентрируемся на методе „Singular Value Projection“ [29; 31], который по своей структуре относится к классу методов проекции градиента. Следуя [29], рассмотрим теорию такого метода в более общей постановке задачи восполнения. А именно, пусть на множестве $m_1 \times m_2$ матриц задан аффинный оператор $\mathcal{A} : \mathbb{R}^{m_1 \times m_2} \rightarrow \mathbb{R}^{m_1 \times m_2}$, где образ оператора \mathcal{A} имеет размерность M , строго меньшую, чем $m_1 m_2$. Тогда для произвольной матрицы $B \in \mathbb{R}^{m_1 \times m_2}$ обобщенную задачу малорангового восполнения будем рассматривать как задачу оптимизации

$$\psi_{\mathcal{A}}(X) = \frac{1}{2} \|\mathcal{A}(X) - B\|_F^2 \rightarrow \inf, \quad \text{rank}(X) \leq r, \quad (1.1)$$

с квадратичным функционалом $\psi_{\mathcal{A}}(X)$. Заметим, что градиент $\nabla\psi_{\mathcal{A}}(Y)$ функционала (1.1) в произвольной точке $Y \in \mathbb{R}^{m_1 \times m_2}$ может быть записан в виде

$$\nabla\psi_{\mathcal{A}}(Y) = \mathcal{A}^*(\mathcal{A}(Y) - B).$$

где оператор $\mathcal{A}^* : \mathbb{R}^{m_1 \times m_2} \rightarrow \mathbb{R}^{m_1 \times m_2}$ является сопряженным к \mathcal{A} (предполагается, что в пространстве $\mathbb{R}^{m_1 \times m_2}$ задано стандартное скалярное произведение, порожденное фробениусовой нормой).

Итерация метода проекции градиента в общем виде состоит из двух частей: шага градиентного спуска и проекции нового приближения на многообразие матриц ранга не выше r . Другими словами, если X_k – приближение, полученное на шаге k алгоритма, то следующее приближение X_{k+1} получается в виде

$$\begin{aligned} X_{k+1} &= \mathcal{P}_r(X_k - \tau \nabla\psi_{\mathcal{A}}(X_k)) \\ &= \mathcal{P}_r(X_k - \tau \mathcal{A}^*(\mathcal{A}(X_k) - B)) \end{aligned} \quad (1.2)$$

где $\tau \in \mathbb{R}$ – некоторый шаг градиентного спуска, который мы определим позже, а \mathcal{P}_r – наилучший во фробениусовой норме проектор на множество матриц ранга не выше r . Для сходимости метода проекции градиента в [29] доказана следующая теорема.

Теорема 1.1.1. Пусть X_* – решение (1.1), для которого $\psi_{\mathcal{A}}(X_*) = 0$, а оператор \mathcal{A} удовлетворяет условию ограниченной изометрии вида

$$(1 - \sigma)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \sigma)\|X\|_F^2, \quad (1.3)$$

с параметром $0 < \sigma < 1$ для всех матриц X ранга не выше $2r$. Если дополнительно

$$\frac{2\sigma}{1 - \sigma} < 1,$$

то метод проекции градиента с постоянным шагом $\tau = 1/(1 + \sigma)$ сходится к решению X_* , а скорость сходимости определяется соотношением

$$\psi_{\mathcal{A}}(X_{k+1}) \leq \frac{2\sigma}{(1 - \sigma)} \psi_{\mathcal{A}}(X_k).$$

Из (1.2) следует, что алгоритмическая сложность метода проекции градиента определяется сложностью вычисления наилучшей проекции \mathcal{P}_r на многообразии матриц ранга не выше r . Если $m_1 = m_2 = m$, то практическая сложность такого вычисления, основанного на сингулярном разложении, имеет вид $\mathcal{O}(m^3)$. Цель настоящей работы состоит в исследовании возможности ускорения метода путем замены оптимального проектора \mathcal{P}_r на быстро вычисляемый приближенный $\hat{\mathcal{P}}_r$.

1.1.2 Случай приближенного проектирования

Формализуем понятие приближенного проектирования на многообразии матриц ранга не выше r . Пусть оператор $\hat{\mathcal{P}}_k : \mathbb{R}^{m_1 \times m_2} \rightarrow \mathbb{R}^{m_1 \times m_2}$, заданный произвольным образом, всегда возвращает матрицу ранга не выше r . Будем называть $\hat{\mathcal{P}}_r$ *оператором приближенного проектирования*, если для любой матрицы Y выполнено

$$\|\hat{\mathcal{P}}_r(Y) - Y\|_F^2 \leq (1 + \varepsilon) \|\mathcal{P}_r(Y) - Y\|_F^2, \quad (1.4)$$

с некоторой константой $\varepsilon > 0$. При этом подразумевается, что вычисление $\hat{\mathcal{P}}_r(Y)$ занимает меньше арифметических операций, чем вычисление $\mathcal{P}_r(Y)$. По аналогии с алгоритмом точной проекции градиента определим итерацию метода приближенной проекции градиента равенством

$$X_{k+1} = \hat{\mathcal{P}}_r(X_k - \tau \mathcal{A}^*(\mathcal{A}(X_k) - B)).$$

Покажем, что для определенного таким соотношением алгоритма ASVP справедлива теорема, аналогичная теореме 1.1.1.

Теорема 1.1.2. *Пусть на матрице X_* ранга не выше r достигается минимум функционала*

$$\psi_{\mathcal{A}}(X) = \frac{1}{2} \|\mathcal{A}(X) - B\|_F^2 \rightarrow \inf, \quad \text{rank}(X) \leq r, \quad (1.5)$$

с оператором \mathcal{A} , для которого условие ограниченной изометрии

$$(1 - \sigma)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \sigma)\|X\|_F^2, \quad (1.6)$$

выполняется на всех матрицах ранга не выше $2r$.

В этом случае для метода приближенной проекции градиента с постоянным шагом $1/(1 + \sigma)$ и оператором $\hat{\mathcal{P}}_r$ приближенного проектирования, удовлетворяющим условию (1.4) с константой ε , справедливо неравенство

$$\begin{aligned} \psi_{\mathcal{A}}(X_{k+1}) &\leq (1 + \varepsilon)\psi_{\mathcal{A}}(X_*) - \varepsilon\psi_{\mathcal{A}}(X_k) \\ &\quad + (1 + \varepsilon)\frac{\sigma}{1 - \sigma}\|\mathcal{A}(X_* - X_k)\|_F^2 \\ &\quad + \frac{\varepsilon\|\mathcal{A}^T\|^2}{2(1 + \sigma)}\|\mathcal{A}(X_t) - B\|_F^2 \end{aligned}$$

Доказательство. Используя арифметическое тождество

$$(b - a)^2 - (c - a)^2 = (b - c)^2 + 2(c - a)(b - c),$$

и символьную подстановку

$$a \leftrightarrow B, \quad b \leftrightarrow \mathcal{A}(X), \quad c \leftrightarrow \mathcal{A}(X_k),$$

для произвольной матрицы X можно записать

$$\begin{aligned} \psi_{\mathcal{A}}(X) - \psi_{\mathcal{A}}(X_k) &= \frac{1}{2}\|\mathcal{A}(X) - B\|_F^2 - \frac{1}{2}\|\mathcal{A}(X_k) - B\|_F^2 \\ &= (\mathcal{A}^*(\mathcal{A}(X_k) - B), X - X_k)_F + \frac{1}{2}\|\mathcal{A}(X - X_k)\|_F^2, \quad (1.7) \end{aligned}$$

где скалярное произведение $(U, V)_F$ для матриц $U, V \in \mathbb{R}^{m \times n}$ определяется соотношением $(U, V)_F = \text{tr}(V^T U)$. Символ F подчеркивает, что такое скалярное произведение порождается нормой Фробениуса, так как выполнено $(U, U)_F = \text{tr}(U^T U) = \|U\|_F^2$. Для произвольной матрицы X ранга не выше r воспользуемся во втором слагаемом (1.7) свойством ограниченной изометрии оператора \mathcal{A}

и установим верхнюю оценку для разности:

$$\begin{aligned}\psi_{\mathcal{A}}(X) - \psi_{\mathcal{A}}(X_k) &= (\mathcal{A}^*(\mathcal{A}(X_k) - B), X - X_k)_F + \frac{1}{2}\|\mathcal{A}(X - X_k)\|_F^2, \\ &\leq (\mathcal{A}^*(\mathcal{A}(X_k) - B), X - X_k)_F + \frac{(1 + \sigma)}{2}\|X - X_k\|_F^2.\end{aligned}\quad (1.8)$$

Правую часть (1.8) обозначим через $f_k(X)$. Таким образом,

$$f_k(X) = (\mathcal{A}^*(\mathcal{A}(X_k) - B), X - X_k)_F + \frac{(1 + \sigma)}{2}\|X - X_k\|_F^2, \quad (1.9)$$

и

$$\psi_{\mathcal{A}}(X) - \psi_{\mathcal{A}}(X_k) \leq f_k(X), \quad (1.10)$$

для всех матриц X , ранг которых не превосходит r .

Значения $f_k(X)$ дают оценку на падение функционала невязки для на всем множестве матриц ранга не выше r , а вид $f_k(X)$ позволяет найти в явном виде матрицу ранга не выше r , на которой падение невязки является значительным.

Действительно, используем арифметическое тождество

$$\frac{1 + \sigma}{2}(2pq + p^2) = \frac{1 + \sigma}{2}((p + q)^2 - q^2),$$

вместе с соответствием

$$p \leftrightarrow X - X_k, \quad q \leftrightarrow \frac{1}{1 + \sigma}\mathcal{A}^*(\mathcal{A}(X_k) - B), \quad f_k(X) \leftrightarrow \frac{1 + \sigma}{2}(2pq + p^2).$$

После соответствующих замен $f_k(X)$ принимает вид

$$\begin{aligned}f_k(X) &= \frac{(1 + \sigma)}{2}\|X - \left(X_k - \frac{1}{1 + \sigma}\mathcal{A}^*(\mathcal{A}(X_k) - B)\right)\|_F^2 \\ &\quad - \frac{1}{2(1 + \sigma)}\|\mathcal{A}^T(\mathcal{A}(X_k) - B)\|_F^2.\end{aligned}$$

Определяя Y_{k+1} равенством,

$$Y_{k+1} = X_k - \frac{1}{1 + \sigma}\mathcal{A}^*(\mathcal{A}(X_k) - B),$$

представим $f_k(X)$ в виде суммы

$$f_k(X) = \frac{(1 + \sigma)}{2} \|X - Y_{k+1}\|_F^2 - \frac{1}{2(1 + \sigma)} \|\mathcal{A}^*(\mathcal{A}(X_k) - B)\|_F^2. \quad (1.11)$$

Так как второе слагаемое в (1.11) не зависит от X , то минимум $f_k(X)$ на множестве всех матриц ранга не выше r достигается на матрице $Z_{k+1} = \mathcal{P}_r(Y_{k+1})$. При этом сама матрица Y_{k+1} формально совпадает с матрицей, полученной градиентным спуском из точки X_k с величиной шага $\tau = 1/(1 + \sigma)$, не зависящей от его номера k .

Как несложно видеть, Z_{k+1} совпадает с приближением на шаге $k+1$ метода точной проекции градиента. Так как Z_{k+1} дает минимум $f_k(X)$, то значение $f_k(Z_{k+1})$ можно оценить сверху значением $f_k(X_*)$. В таком случае

$$\begin{aligned} f_k(Z_{k+1}) &\leq f_k(X_*) \\ &= (\mathcal{A}^*(\mathcal{A}(X_k) - B), X_* - X_k)_F + \frac{(1 + \sigma)}{2} \|X_* - X_k\|_F^2. \end{aligned}$$

Используя свойство ограниченной изометрии оператора \mathcal{A} на матрице $(X_* - X_k)$, ранг которой не превосходит $2r$, преобразуем последнее выражение к виду

$$f_k(Z_{k+1}) \leq (\mathcal{A}^*(\mathcal{A}(X_k) - B), X_* - X_k)_F + \frac{1 + \sigma}{2(1 - \sigma)} \|\mathcal{A}(X_* - X_k)\|_F^2.$$

Теперь воспользуемся формулой (1.7) для первого слагаемого, чтобы получить

$$f_k(Z_{k+1}) \leq \psi(X_*) - \psi(X_k) + \frac{\sigma}{(1 - \sigma)} \|\mathcal{A}(X_* - X_k)\|_F^2. \quad (1.12)$$

Положим $X_{k+1} = \hat{\mathcal{P}}_r(Y_{k+1})$. В силу определения $\hat{\mathcal{P}}_r$ выполняется неравенство

$$\|X_{k+1} - Y_{k+1}\|_F^2 \leq (1 + \varepsilon) \|Z_{k+1} - Y_{k+1}\|_F^2 \quad (1.13)$$

Рассмотрим разность $f_k(X_{k+1}) - f_k(Z_{k+1})$. Подставляя для $f_k(X_{k+1})$ и $f_k(Z_{k+1})$ их представление из (1.11), сокращая постоянное слагаемое и учитывая (1.13),

получим

$$\begin{aligned} f_k(X_{k+1}) - f_k(Z_{k+1}) &= \frac{(1+\sigma)}{2} (\|X_{k+1} - Y_{k+1}\|_F^2 - \|Z_{k+1} - Y_{k+1}\|_F^2) \\ &\leq \varepsilon \frac{(1+\sigma)}{2} \|Z_{k+1} - Y_{k+1}\|^2. \end{aligned} \quad (1.14)$$

Еще раз применим (1.11) теперь к правой части (1.14). Тогда

$$\begin{aligned} f_k(X_{k+1}) - f_k(Z_{k+1}) &\leq \varepsilon f_k(Z_{k+1}) + \varepsilon \frac{1}{2(1+\sigma)} \|\mathcal{A}^*(\mathcal{A}(X_k) - B)\|_F^2 \\ &\leq \varepsilon f_k(Z_{k+1}) + \varepsilon \frac{\|\mathcal{A}^*\|^2}{2(1+\sigma)} \|\mathcal{A}(X_k) - B\|_F^2, \end{aligned}$$

где $\|\mathcal{A}^*\|$ – операторная норма \mathcal{A}^* , подчиненная фробениусовой норме в пространстве $\mathbb{R}^{m_1 \times m_2}$. Теперь мы готовы оценить разность $\psi(X_{k+1}) - \psi(X_k)$. Действительно, как следует из (1.10),

$$\begin{aligned} \psi_{\mathcal{A}}(X_{k+1}) - \psi_{\mathcal{A}}(X_k) &\leq f_k(X_{k+1}) \\ &= f_k(X_{k+1}) - (f_k(Z_{k+1}) - f_k(Z_{k+1})) \\ &= (f_k(X_{k+1}) - f_k(Z_{k+1})) + f_k(Z_{k+1}) \\ &\leq (1+\varepsilon)f_k(Z_{k+1}) + \varepsilon \frac{\|\mathcal{A}^*\|^2}{2(1+\sigma)} \|\mathcal{A}(X_k) - B\|_F^2. \end{aligned}$$

Заменяя $f_k(Z_{k+1})$ выражением (1.12), преобразуем последнее соотношение к окончательному виду

$$\begin{aligned} \psi_{\mathcal{A}}(X_{k+1}) - \psi_{\mathcal{A}}(X_k) &\leq (1+\varepsilon)(\psi_{\mathcal{A}}(X_*) - \psi_{\mathcal{A}}(X_k)) \\ &\quad + (1+\varepsilon) \frac{\sigma}{1-\sigma} \|\mathcal{A}(X_* - X_k)\|_F^2 \\ &\quad + \frac{\varepsilon \|\mathcal{A}^*\|^2}{2(1+\sigma)} \|\mathcal{A}(X_k) - B\|_F^2 \end{aligned}$$

Откуда после сокращения слагаемого $\psi(X_k)$ в левой и правой частях неравенства заканчиваем доказательство. \square

Следствие 1.1.3. Пусть существует матрица X_* ранга не выше k , для которой $\psi_{\mathcal{A}}(X_*) = 0$ и выполняется условие ограниченной изометрии (1.3). Метод приближенной проекции градиента с постоянным шагом $\tau = \frac{1}{1+\sigma}$ сходится к

решению X_* , а скорость сходимости определяется соотношением

$$\psi_{\mathcal{A}}(X_{k+1}) \leq \psi_{\mathcal{A}}(X_k) \left(\frac{2\sigma}{1-\sigma} + \varepsilon \frac{\|\mathcal{A}^*\|^2}{(1+\sigma)} \right). \quad (1.15)$$

Доказательство. Из предыдущей теоремы и в силу соотношений

$$\begin{aligned} \psi_{\mathcal{A}}(X_*) &= 0, \\ \mathcal{A}(X_*) &= B, \\ \psi_{\mathcal{A}}(X_k) &= \frac{1}{2} \|\mathcal{A}(X_k - X_*)\|_F^2 = \frac{1}{2} \|\mathcal{A}(X_k) - B\|_F^2, \end{aligned}$$

имеем

$$\begin{aligned} \psi_{\mathcal{A}}(X_{k+1}) &\leq \left(\frac{2\sigma(1+\varepsilon)}{1-\sigma} + \varepsilon \frac{\|\mathcal{A}\|_F^2}{(1+\sigma)} - \varepsilon \right) \psi_{\mathcal{A}}(X_k) = \\ &= \left[\frac{2\sigma}{1-\sigma} + \varepsilon \left(\frac{2\sigma}{1-\sigma} + \frac{\|\mathcal{A}^*\|^2}{2(1+\sigma)} - 1 \right) \right] \psi_{\mathcal{A}}(X_k). \quad (1.16) \end{aligned}$$

В предположении $\sigma < 1/3$, аналогичном тому, что используется для доказательства сходимости метода проекции градиента с точным проектированием в [29], видим, что ε можно выбрать достаточно малой константой, чтобы из неравенства (1.16) следовала геометрическая сходимость приближенного варианта метода проекции градиента. \square

Как следует из формулы (1.15), использование приближенного проектора $\hat{\mathcal{P}}_r$ в методе приближенной проекции градиента ухудшает коэффициент линейной сходимости, что, конечно, не является неожиданным. Однако, как будет показано далее, выигрыш, получаемый за счет упрощения \mathcal{P}_r , превосходит проигрыш, связанный с увеличением числа итераций в приближенном варианте метода проекции градиента.

В следствии 1.1.3 предполагается существование решения X_* , для которого $\psi_{\mathcal{A}}(X_*) = 0$. Однако на практике для заданной правой части $B \in \mathbb{R}^{m_1 \times m_2}$ минимум функционала может быть не равен нулю, например, в случае $B = \mathcal{A}(X) + \Delta$, где X - матрица малого ранга, а Δ имеет малую норму. Оценки для такого случая представлены в следствии 1.1.4.

Следствие 1.1.4. Пусть на матрице X_* достигается минимум функционала $\psi_{\mathcal{A}}(X)$ среди всех матриц ранга не выше k , а $\Delta = B - \mathcal{A}(X_*)$ имеет норму $\delta = \|\Delta\|_F$.

Если для приближения X_k выполняется неравенство $\psi_{\mathcal{A}}(X_t) \geq C^2 \frac{\delta^2}{2}$ с константой $C > 0$, то для следующего приближения, получаемого в методе приближенной проекции градиента справедливо соотношение

$$\psi_{\mathcal{A}}(X_{k+1}) < \kappa \psi_{\mathcal{A}}(X_k),$$

с константой κ , удовлетворяющей неравенству,

$$\kappa \leq \theta_0 + \varepsilon (\theta_0 + \theta_1), \quad (1.17)$$

где

$$\begin{aligned} \theta_0 &= \frac{1}{C^2} + \frac{2\sigma}{(1-\sigma)} \left(1 + \frac{1}{C}\right)^2, \\ \theta_1 &= \frac{\|\mathcal{A}^*\|^2}{(1+\sigma)} - 1. \end{aligned}$$

Доказательство. Из теоремы 1.1.2,

$$\begin{aligned} \psi_{\mathcal{A}}(X_{k+1}) &\leq (1+\varepsilon) \frac{\delta^2}{2} \\ &+ (1+\varepsilon) \frac{\sigma}{1-\sigma} \|B - \mathcal{A}(X_k) - \Delta\|_F^2 \\ &+ \varepsilon \left(\frac{\|\mathcal{A}^*\|^2}{(1+\sigma)} - 1 \right) \psi_{\mathcal{A}}(X_k) \end{aligned}$$

Учитывая, что $\delta^2 \leq \frac{2}{C^2}\psi(X_k)$ и раскрывая второе слагаемое, получим

$$\begin{aligned}
\psi_{\mathcal{A}}(X_{k+1}) &\leq \frac{(1+\varepsilon)}{C^2}\psi_{\mathcal{A}}(X_k) \\
&+ (1+\varepsilon)\frac{2\sigma}{1-\sigma}\left(\psi_{\mathcal{A}}(X_k) + \frac{2}{C}\psi_{\mathcal{A}}(X_k) + \frac{1}{C^2}\psi_{\mathcal{A}}(X_k)\right) \\
&+ \varepsilon\left(\frac{\|\mathcal{A}\|^2}{(1+\sigma)} - 1\right)\psi_{\mathcal{A}}(X_k) \\
&\leq \left(\frac{1}{C^2} + \frac{2\sigma}{(1-\sigma)}\left(1 + \frac{1}{C}\right)^2\right)\psi_{\mathcal{A}}(X_k) \\
&+ \varepsilon\left(\frac{1}{C^2} + \frac{2\sigma}{(1-\sigma)}\left(1 + \frac{1}{C}\right)^2 + \frac{\|\mathcal{A}^*\|^2}{(1+\sigma)} - 1\right)\psi_{\mathcal{A}}(X_k)
\end{aligned}$$

Полагая $\theta_0 = \frac{1}{C^2} + \frac{2\sigma}{(1-\sigma)}\left(1 + \frac{1}{C}\right)^2$ и $\theta_1 = \frac{\|\mathcal{A}^*\|^2}{(1+\sigma)} - 1$, окончательно получаем

$$\psi_{\mathcal{A}}(X_{k+1}) \leq (\theta_0 + \varepsilon(\theta_0 + \theta_1))\psi_{\mathcal{A}}(X_k). \quad (1.18)$$

□

1.2 Частный случай оператора восполнения матриц

Вернемся к задаче восполнения матрицы малого ранга по значениям ее элементов на случайном разреженном шаблоне. Сведем задачу к рассмотренной выше обобщенной постановке путем специального способа выбора аффинного преобразования \mathcal{A} .

Пусть $X_* \in \mathbb{R}^{m_1 \times m_2}$ – неизвестная матрица ранга r , причем $r \ll \min(m_1, m_2)$, а Ω – случайный набор пар индексов (j_1, j_2) , где $i_1 \in 1, \dots, m_1$ и $i_2 \in 1, \dots, m_2$, причем любая пара (j_1, j_2) может быть выбрана равновероятно среди всех $m_1 m_2$ пар с вероятностью q , которую мы определим позже. Будем также для простоты считать, что выполнено в точности $|\Omega| = q m_1 m_2$, а q будем также называть плотностью известных элементов матрицы.

Определим линейный оператор $\mathcal{A}_\Omega : \mathbb{R}^{m_1 \times m_2} \rightarrow \mathbb{R}^{m_1 \times m_2}$ по правилу

$$(\mathcal{A}_\Omega(X))_{j_1 j_2} = \begin{cases} \frac{1}{\sqrt{q}} X_{j_1 j_2}, & (j_1, j_2) \in \Omega, \\ 0, & (j_1, j_2) \notin \Omega. \end{cases} \quad (1.19)$$

Определим B формулой $B = \mathcal{A}_\Omega(X_*) \in \mathbb{R}^{m_1 \times m_2}$ и зададим функционал $\psi_{\mathcal{A}_\Omega}(X) = \frac{1}{2} \|\mathcal{A}_\Omega(X) - B\|_F^2$. Задача восполнения матрицы X_* по части ее элементов записывается в виде

$$\psi_{\mathcal{A}_\Omega}(X) = \frac{1}{2} \|\mathcal{A}_\Omega(X) - B\|_2^2 \rightarrow \inf, \quad \text{rank}(X) \leq r. \quad (1.20)$$

Для упрощения записи в дальнейшем индекс \mathcal{A}_Ω в обозначении функционала ψ опускается. Нетрудно видеть, что в частном случае оператора восполнения $\mathcal{A}_\Omega^* = \mathcal{A}_\Omega$.

В качестве оператора $\hat{\mathcal{P}}_r$ при этом можно использовать произвольный алгоритм построения малоранговой аппроксимации $\tilde{Y} = \hat{\mathcal{P}}_r Y$, достаточно близкой к проекции с использованием сингулярного разложения

$$\|Y - \tilde{Y}\|_F \leq (1 + \varepsilon) \|Y - Y_r\|_F.$$

Нижний индекс r у матрицы здесь и далее обозначает наилучшее приближение ранга r по норме Фробениуса, которое можно получить с помощью сингулярного разложения.

При $\sigma \approx 0$ из следствия 1.1.3 подтсановкой $\mathcal{A} = \mathcal{A}_\Omega$, используя тривиальную оценку $\|\mathcal{A}_\Omega\|^2 \leq \frac{1}{q}$, мы получаем условие

$$\begin{aligned} \varepsilon \|\mathcal{A}_\Omega^T\|^2 &< 1, \\ \frac{\varepsilon}{q} &< 1, \end{aligned}$$

а потому для сходимости метода восполнения матриц достаточно $\varepsilon = O(q)$.

Заметим, что для задачи восполнения матрицы малого ранга по значениям ее элементов на разреженном шаблоне Ω напрямую воспользоваться результатами разделов 1.1 и 1.1.2 нельзя.

Так например в случае произвольного шаблона Ω оператор $\mathcal{A}_\Omega(X)$, определенный ранее, не удовлетворяет условию ограниченной изометрии, которое

является ключевым при доказательстве сходимости метода. Тем не менее, как показано в [29], если Ω имеет случайный характер, то условие ограниченной изометрии для оператора \mathcal{A}_Ω выполняется с вероятностью, неотличимой от 1, на почти всем множестве матриц ранга не превосходящего $2r$. Чтобы сформулировать соответствующее утверждение нам потребуется понятие μ -некогерентных матриц.

Определение 1.2.1. Пусть для матрицы $X \in \mathbb{R}^{m_1 \times m_2}$ задано ее сингулярное разложение $X = U\Sigma V^T$. Матрица X называется μ -некогерентной, если для матриц сингулярных векторов U и V справедливы неравенства

$$\max_{i,j} |U_{ij}| \leq \frac{\sqrt{\mu}}{\sqrt{m_1}}, \max_{i,j} |V_{ij}| \leq \frac{\sqrt{\mu}}{\sqrt{m_2}}$$

Определение некогерентности предполагает неразреженность сингулярных факторов матрицы; можно показать, что из этого определения следует и неразреженность самой матрицы. Так, выполнена

Лемма 1.2.1. Если $X \in \mathbb{R}^{m \times m}$ - матрица ранга r , удовлетворяющая свойству некогерентности (1.2.1) с константой μ , то выполнено

$$\|X\|_C \leq \frac{\mu\sqrt{r}\|X\|_F}{m}.$$

Доказательство. Доказательство приведено в [29], Лемме 4.7. □

Используя определение 1.2.1, сформулируем следующий результат о выполнении свойства ограниченной изометрии для оператора \mathcal{S} .

Теорема 1.2.2 (Теорема 4.2, [29]). Существует константа $C \geq 0$ такая, что для любого $0 < \sigma < 1$, любых $\mu \geq 1$, и $m_2 \geq m_1 \geq 3$, и для любого шаблона разреженности Ω , пары индексов которого выбираются случайно в соответствии с моделью Бернулли с параметром

$$q \geq C \frac{\mu^2 r^2 \log(m_2)}{\sigma^2 m_1},$$

определяющим вероятность для пары индексов (j_1, j_2) принадлежать маске Ω , оператор \mathcal{A}_Ω удовлетворяет на множестве всех μ -некогерентных матриц

свойству ограниченной изометрии вида

$$(1 - \sigma)\|X\|_F^2 \leq \|\mathcal{A}_\Omega(X)\|_F^2 \leq (1 + \sigma)\|X\|_F^2,$$

с вероятностью не меньше $1 - \exp(-n \log(n))$.

Из теорем 1.1.2 и 1.2.2 следует сходимость метода проекции градиента с точным и приближенным проектированием в случае, если на всех итерациях (то есть для всех k) матрицы $X_{k+1} - X_k$ и $X_k - X_*$ обладают ограниченной константой некогерентности μ . Отметим, что для разности $X_{k+1} - X_k$ формальная проверка μ -некогерентности является возможной и вычислительно недорогой процедурой. В тоже время для разности $X_k - X_*$ на практике оценить константу некогерентности невозможно.

Из сказанного выше следует, что для задачи восполнения матриц малого ранга, заданных на случайном шаблоне, обоснование метода проекции градиента не является безусловным. Формальные реализации этих методов могут и проявляют нерегулярное поведение, то есть расходятся при выборе теоретически обоснованного шага градиентного спуска τ (см., например, [12]) или в том случае, когда сингулярные числа решения X_* быстро убывают.

На практике получение (быстро) сходящихся реализаций алгоритмов возможно только с использованием дополнительных приемов, среди которых мы выделяем два: (а) последовательный набор ранга приближения; (б) выбор шага τ градиентного спуска.

1.3 Практические модификации метода проекции градиента

1.3.1 Последовательный набор ранга приближения

Чтобы избежать нерегулярного поведения метода проекции градиента на практике, предлагается использовать последовательный набор ранга приближения. Другими словами, вместо того, чтобы решать задачу восполнения сразу для предписанного ранга r , задачу восполнения предлагается последовательно

решать для рангов $1, 2, \dots, r$, и для каждого последующего ранга решение, соответствующее предыдущему рангу, использовать в качестве начального приближения.

Наиболее примитивный подход к набору ранга решения состоит в том, чтобы заранее фиксировать время работы алгоритма и отвести константные отрезки времени на работу алгоритма с каждым рангом приближения r . При этом можно, например, отвести первую половину всего времени работы алгоритма на стадию набора ранга, а вторую - на итерации с необходимым конечным рангом приближения.

Другой, более обоснованный способ набора ранга может быть основан на проверке выполнения условия μ -некогерентности для разности $X_k - X_{k-1}$. Анализ численных экспериментов показал, что при нерегулярном поведении точного и приближенного вариантов метода проекции градиента для восполнения матриц параметр μ -некогерентности последовательности матриц $X_{k+1} - X_k$ увеличивается от итерации к итерации и стремится к максимально возможному значению $\max(\sqrt{m_1}, \sqrt{m_2})$. Таким образом, на каждой итерации можно вычислять текущее значение некогерентности $X_{k+1} - X_k$, и проводить увеличение ранга приближения на единицу только в случае, если на протяжении нескольких итераций эта величина оказывается достаточно малой; для определения допустимой величины некогерентности полезно иметь априорную оценку на эту величину для искомой матрицы.

Сложность вычисления μ для матрицы $X_k - X_{k-1}$ определяется сложностью вычисления сингулярного разложения объединенных левых и объединенных правых факторов матриц X_k, X_{k-1} , что в случае $m_1 = m_2 = m$ требует $O(mr^2) \ll m^3$ операций.

При этом точно указать необходимое число последовательных шагов алгоритма, при которых величина некогерентности $X_t - X_{k-1}$ остается ниже допустимой границы - затруднительно; на основе проведенных экспериментов установлено, что в случае матриц порядка $m = 1000$ и случайных скелетных факторов достаточно около пяти шагов. Это число последовательных шагов можно оценивать по числу предшествующих шагов, при которых величины некогерентности были наоборот, большими, например, используя следующую процедуру:

Такой способ набора ранга оказался особенно эффективен в случае использования приближенных проекций, основанных на псевдоскелетном методе

Algorithm 1: Последовательный набор ранга приближения

Data: Граничное значение μ_{crit} , по которому определяется, что некогерентность на текущей итерации достаточно мала. Константа минимального числа последовательных итераций с низкой некогерентностью s_{crit} . Вводятся переменные, сохраняющие число предшествующих итераций метода точной (приближенной) проекции градиента с большими и малыми величинами μ : $s_{high} = 0$, $s_{low} = 0$.

Result: Ранг k увеличивается, если на протяжении нескольких итераций $\mu < \mu_{crit}$.

```

1 if  $\mu(X_k - X_{k-1}) > \mu_{crit}$  then
2   |  $s_{high} := s_{high} + 1$  ;
3   |  $s_{crit} := \max(s_{high}, s_{crit})$  ;
4 else
5   |  $s_{low} := s_{low} + 1$  ;
6 if  $s_{low} > s_{crit}$  then
7   |  $r := r + 1$  ;
8   |  $s_{low} := 0$  ;
9   |  $s_{high} := 0$  ;

```

с использованием подматриц большого объема, о котором рассказано в последующем разделе 1.4.2.

1.3.2 Выбор параметра шага

Численные эксперименты показывают, что нерегулярного поведения метода проекции градиента можно избежать путем понижения параметра шага градиентного метода τ . Так, теорема 1.1.2 гарантирует, что в условиях теоремы метод проекции градиента сходится геометрически при $\tau = \frac{1}{1+\sigma} \approx \frac{3}{4}$; с учетом того, что, согласно определению оператора задачи восполнения матриц \mathcal{A}_Ω , все элементы на маске Ω скалируются на константу $\frac{1}{\sqrt{q}}$, использование такого шага интуитивно неустойчиво при критически малых значениях плотности известных элементов q . На практике оказывается, что при таком шаге в случае малых q и быстром падении сингулярных чисел искомой матрицы алгоритмы точной и приближенной проекции градиента могут расходиться, но при тех же

условиях и малом шаге порядка q расходимости не наблюдается. Экспериментально установлено, что при шаге τ в пределах от $[q, 2q]$ при достаточно малых q сходимость точного и приближенного вариантов метода проекции градиента наблюдается всегда.

Случаи расходимости метода проекции градиента с большим шагом были замечены уже в работе [12]. Для повышения устойчивости алгоритма ее авторы также предложили выбирать τ из интервала $[q, 2q]$. Не проводя формального доказательства сходимости метода с шагом $\tau = q$, приведем некоторые соображения в пользу его устойчивости.

Действительно, если X_k и X_{k+1} приближения с номерами k и $k + 1$, и $B = \mathcal{A}_\Omega(X_*)$, то шаг градиентного метода $\tau = q$ соответствует присваиванию каждому элементу текущего приближения, лежащему на маске известных элементов, соответствующего известного элемента. Это значит, что

$$\|X_k - X_*\|_{F,\Omega} = \|X_k - Y_{k+1}\|_F \geq \|Y_{k+1} - X_{k+1}\|_F \geq \|X_{k+1} - X_*\|_{F,\Omega}, \quad (1.21)$$

где $\|X\|_{F,\Omega} = \|\mathcal{A}_\Omega(X)\|_F$. Неравенство (1.21) можно записать в эквивалентной форме

$$\psi(X_{k+1}) \leq \psi(X_k).$$

откуда имеем, что невязка на итерациях алгоритма с шагом $\tau = q$ монотонно не возрастает.

Тем не менее, эксперименты показывают, что в случаях, когда при использовании шага $\tau \approx \frac{3}{4}$ на практике не нарушаются гипотезы о некогерентности, сходимость метода проекции градиента с шагом $\tau \approx \frac{3}{4}$ значительно быстрее, чем с шагом $\tau \approx q$, то есть на практике использование большого шага является предпочтительным. Чтобы на практике использовать преимущество большого шага в скорости сходимости, но не допускать нерегулярного поведения алгоритмов, предлагается использовать адаптивные процедуры изменения шага τ по ходу алгоритма. Для этого предлагается рассматривать изменение невязки $\|\mathcal{A}_\Omega(X_k - X_*)\|_F$ между итерациями. Если отношение невязок на итерациях $k + 1$ и k

$$\alpha_{k+1,k} := \frac{\|\mathcal{A}_\Omega(X_{k+1} - X_*)\|_F}{\|\mathcal{A}_\Omega(X_k - X_*)\|_F} > \alpha_{crit} > 1$$

то делается вывод о том, что алгоритм начал вести себя нерегулярно, в связи с чем шаг градиентного метода понижается, а последняя итерация пересчиты-

вается с уменьшенным шагом. Формально, после выполнения итерации $k + 1$ метода проекции градиента выполняется процедура

Algorithm 2: Адаптивный выбор шага

Data: Пограничная величина $\alpha_{crit} > 1$; стартовое значение шага $\tau = \frac{3}{4}$; константы увеличения и уменьшения шага $\beta_{inc} < 1, \beta_{dec} < 1$.

Result: Шаг τ увеличивается или уменьшается в зависимости от роста или падения погрешности.

```

1 if  $\alpha_{k+1,k} > \alpha_{crit}$  then
2   |  $\tau := \tau + \beta_{dec}(q - \tau)$  ; Итерация  $k + 1$  проводится заново с
   |   уменьшенным шагом ;
3 else
4   |  $\tau := \tau + \beta_{inc}(\frac{3}{4} - \tau)$  ;

```

Этот адаптивный алгоритм выбора шага оказался эффективен на практике для метода приближенной проекции градиента с обоими вариантами метода приближенного проектирования, которые будут рассмотрены далее. Однако, вычисление невязки может быть асимптотически сложной процедурой, так как требует не менее $O(m_1 m_2 q r)$ операций, поэтому возможно использовать эту процедуру обновления величины шага не после каждой итерации, а, например, после каждой десятой итерации.

1.4 Методы приближенного вычисления проекции на множество матриц малого ранга

1.4.1 Проектирование на случайные пространства

В этом разделе рассмотрим один из возможных методов приближенного проектирования, который может лежать в основе метода приближенной проекции градиента. Будем использовать способ приближенного вычисления частичного сингулярного разложения на основе проектирования на случайные подпространства. Такой способ введен и исследован в [7]. Будем использовать

вспомогательную матрицу

$$J \in \mathbb{R}^{m_2 \times l}, \quad l = r + p, \quad p > 0$$

все элементы которой выбираются независимо и случайно по стандартному нормальному распределению. Тогда в качестве проектора $\hat{\mathcal{P}}$ предлагается использовать

$$\hat{\mathcal{P}}_r(Y) = Q (Q^T Y)_r, \quad Y \in \mathbb{R}^{m_1 \times m_2}$$

Где матрица Q определяется как ортогональный базис столбцов YJ : $YJ = QR$, а $(Q^T Y)_k$ - наилучшее приближение матрицы $(Q^T Y)$ матрицей ранга r , вычисляемое с помощью сингулярного разложения вытянутой матрицы. Из леммы 6.1 в [8], которая обобщает результаты [7], следует, что для любой матрицы $Y \in \mathbb{R}^{m_1 \times m_2}$

Лемма 1.4.1 ([8]).

$$\mathbb{E}_J \|Y - Q (Q^T Y)_r\|_F^2 \leq (1 + \frac{k}{p-1}) \|Y - Y_r\|_F^2 \quad (1.22)$$

В работах [7], [8] приведены оценки и на отклонение от матожидания, которые опустим в силу их громоздкости. Таким образом, в [7] предложен следующий алгоритм вычисления приближенного частичного сингулярного разложения матрицы Y :

1. Создается случайная матрица $J \in \mathbb{R}^{n \times l}$.
2. Вычисляется произведение YJ .
3. Для произведения YJ вычисляется QR - разложение: $YJ = QR$.
4. Вычисляется произведение $Q^T Y$.
5. Вычисляется сокращенное сингулярное разложение для вытянутой матрицы:

$$(Q^T Y)_r = U \Sigma V^T, \quad \text{где } U \in \mathbb{R}^{l \times r}, \quad V^T \in \mathbb{R}^{r \times m_2}.$$

6. Так как

$$(QU) \Sigma V^T = Q (Q^T Y)_r,$$

то матрицы QU и V^T являются искомым приближением сингулярных векторов матрицы Y .

Согласно оценке (1.22), введенный таким образом оператор $\hat{\mathcal{P}}$ удовлетворяет определению оператора приближенного проектирования (1.4) с точностью $\varepsilon = \frac{r}{p-1}$ в среднем. Оценим сложность одной итерации полученного приближенного алгоритма:

- Умножение на случайную матрицу: т.к. приближенное сингулярное разложение вычисляется для суммы матрицы малого ранга (матрица предыдущей итерации) и разреженной матрицы (градиент), имеем $O((m_1 + m_2)rl + m_1m_2lq)$ операций.
- QR - разложение YJ : $O(m_1l^2)$ операций.
- Умножение на матрицу Q^* слева: аналогично первому пункту, $O((m_1 + m_2)rl + m_1m_2lq)$ операций.
- Сингулярное разложение вытянутой матрицы: $O(m_2l^2)$ операций.
- Восстановление левого сингулярного базиса: $O(m_1l^2)$ операций.

Таким образом, имеем суммарную сложность одной итерации приближенного алгоритма $O((m_1 + m_2)l^2 + m_1m_2lq)$. С учетом оценки $\varepsilon = O(\frac{k}{p})$ можно считать что при малых ε выполнено $p \gg r, l \approx p$, а вычислительная сложность одной итерации соответственно оценивается величиной

$$O((m_1 + m_2)p^2 + m_1m_2pq)$$

1.4.2 Псевдоскелетный метод на матрицах большого проективного объема

Здесь мы опишем метод приближенной проекции градиента с использованием в качестве проектора $\hat{\mathcal{P}}_r$ аппроксимации на основе крестового (псевдоскелетного) CGR приближения, в котором используются строки $R \in \mathbb{R}^{p \times m_2}$ и столбцы $C \in \mathbb{R}^{m_1 \times p}$ приближаемой матрицы

$$Y = X_k - \tau \mathcal{S}^T(\mathcal{S}(X_k) - B).$$

Один из самых быстрых способов построения аппроксимации – построение Fast CGR из [3]. Построение аппроксимации состоит из следующих этапов:

1. **Использование алгоритма `maxvol` [6] для поиска подматрицы $\hat{Y} \in \mathbb{R}^{2r \times 2r}$.**

Увеличенный в 2 раза ранг позволяет позже использовать усеченное сингулярное разложение для увеличения точности аппроксимации (по сравнению с прямым поиском аппроксимации ранга r), а также позволяет алгоритму не „зацикливаться“ на одной и той же подматрице.

Эксперименты в [3] показывают, что в случае прямого построения аппроксимации $\tilde{Y} = C\hat{Y}^{-1}R$, коэффициент ε растет линейно с ростом r , поэтому нельзя гарантировать произвольную точность приближения ε , ограничиваясь лишь алгоритмом `maxvol`.

При фиксированном числе шагов алгоритм требует $O((m_1 + m_2)r^2)$ операций и знания $O(r)$ строк и столбцов матрицы.

2. **Увеличение числа строк и столбцов до p .**

Новые строки и столбцы выбираются с помощью ускоренного варианта алгоритма `maxvol2`. Расширение требует лишь знания уже найденных с помощью `maxvol` строк и столбцов и занимает $O((m_1 + m_2)rp)$ операций.

Согласно гипотезе из [3] для аппроксимации ранга $2r$ с p столбцами справедлива оценка коэффициента

$$1 + \varepsilon \leq 1 + \frac{r}{p - r + 1}, \quad (1.23)$$

а потому достаточно набрать $p = O(r/\varepsilon)$ строк и столбцов. При этом в качестве ядра CGR аппроксимации выбирается матрица $G = \hat{Y}_{2r}^+$, для чего потребуется сингулярное разложение матрицы \hat{A} , занимающее $O(p^3)$ операций.

3. **Сингулярное разложение аппроксимации**

Найденная на предыдущем шаге CGR аппроксимация ранга $2r$ подвергается сингулярному разложению с сохранением лишь r максимальных сингулярных чисел.

Сингулярное разложение произведения $C\hat{Y}_{2r}^+R$ требует $O((m_1 + m_2)rp)$ операций и выполняется в следующем порядке:

- (a) Факторы $\hat{U} \in \mathbb{R}^{p \times 2r}$ и $\hat{V}^T \in \mathbb{R}^{2r \times p}$ сингулярного разложения матрицы \hat{Y}_{2r}^+ умножаются на C и R соответственно.

- (b) Для матрицы $C\hat{U}$ вычисляется представление в виде произведения матрицы с ортогональными столбцами и верхней треугольной матрицы; для матрицы $\hat{V}^T R$ вычисляется представление в виде произведения нижней треугольной матрицы и матрицы с ортогональными строками.
- (c) Выполняется сингулярное разложение произведения двух полученных треугольных матриц, с помощью которого находится оптимальное приближение этого произведения матрицей ранга r .

Полученная таким образом аппроксимация ранга r найденной на предыдущем шаге матрицы CGR считается искомым приближением для Y_r .

Таким образом, полная сложность алгоритма аппроксимации составляет $O((m_1 + m_2)rp + p^3) = O((m_1 + m_2)p^2q + p^3)$. Данная аппроксимация может быть использована в качестве оператора приближенного проектирования $\hat{\mathcal{P}}_r$ для алгоритма приближенной проекции градиента.

Полученное CGR разложение записывается в виде $CGR = X_{k+1} = U_{k+1}V_{k+1}^T$ с $U_{k+1} \in \mathbb{R}^{m \times k}$, $V_{k+1}^T \in \mathbb{R}^{r \times n}$. Напомним, что приближенное проектирование (аппроксимация) производится для матрицы вида

$$Y = X_k - \tau \mathcal{S}^T(\mathcal{S}(X_k) - B).$$

Всего нам требуется $O(p)$ ее строк и столбцов, которые легко найти на основе матрицы $X_k = U_k V_k^T$ за $O((m_1 + m_2)rp)$ операций, что не увеличивает асимптотическую сложность алгоритма.

Заметим, что на практике число шагов алгоритма `maxvol` можно считать константой, так как по ходу итераций алгоритма приближение каждой итерации меняется медленно, откуда имеем, что и объем подматриц приближения меняется слабо. Алгоритм `maxvol` основан на максимизации объема, а потому будет слабо реагировать на небольшие изменения входных данных. Более того, на основе сингулярных чисел, которые отбрасываются на стадии сингулярного разложения аппроксимации CGR , можно судить о скорости сходимости алгоритма и иногда вообще не изменять выбранные ранее строки и столбцы.

В численных экспериментах использовалось

$$p = 2r + \left\lceil 0.7 \frac{r}{q} \right\rceil$$

строк и столбцов. Это число намеренно ниже оценки (1.23): численные эксперименты из [3] показали, что реальные значения погрешности гораздо ниже верхних оценок, хотя те и предсказывают верную асимптотическую зависимость.

1.5 Численные результаты

В этом разделе рассмотрим соотношения сложностей метода проекции градиента для задачи восполнения матриц в случаях использования точной и приближенной проекций. Для анализа отметим следующие особенности алгоритмов:

1. Доказательство сходимости методов проекции градиента базируется на использовании свойства ограниченной изометрии (1.3) рассматриваемого оператора A .
2. В случае, если $A = A_\Omega$ - оператор задачи матричного восполнения, существует асимптотическая оценка на минимальное количество известных элементов матрицы, которого достаточно для полного восполнения. Эта оценка определяется размерами неизвестной матрицы, рангом неизвестной матрицы, и „некогерентностью“ неизвестной матрицы, численной величиной, характеризующей разреженность сингулярных факторов матрицы.
3. Не обязательно использовать именно такой порядок числа известных элементов матрицы, допустимо использовать и большие порядки.
4. В случае метода приближенной проекции градиента, гарантируется геометрическая сходимость алгоритма с порядком не менее $O(\frac{\epsilon}{q})$, где q - плотность известных элементов матрицы, а ϵ - относительная ошибка приближенного проектирования. Таким образом, при уменьшении числа известных элементов матрицы требуется большая точность приближенного проектирования.

5. Как в случае использования проектирования на случайные подпространства, так и в случае использования псевдоскелетного метода для построения приближенных проекций вводится дополнительный параметр p , определяющий некую вспомогательную размерность метода, причем с увеличением p увеличивается и точность, и вычислительная сложность приближенного проектирования, и наоборот.
6. Таким образом, с уменьшением плотности известных элементов матрицы требуется большая точность приближения, что приводит к увеличению вспомогательных размерностей рассматриваемых методов случайного проектирования, что может привести к увеличению сложности алгоритма.

Проведем формальное сравнение сложностей; их удобно выразить через параметры m_1, m_2, r, q, p . Кроме того, $q \geq q_{\min}(m_1, m_2, r, \mu)$; также, и для псевдоскелетного метода, и для случайного проектирования выполнено $\epsilon = O(\frac{r}{p})$; считая множитель геометрической сходимости константой, чтобы обеспечить константный порядок числа итераций алгоритма, имеем $\epsilon = O(q)$, откуда $p = O\left(\frac{r}{q}\right)$.

Метод проекции градиента состоит из двух основных операций, оценим их сложность:

1. Вычисление невязки на маске известных элементов, покомпонентно: $O(r)$ операций на каждый элемент, $O(m_1 m_2 r q)$ операций всего.
2. Вычисление сингулярного разложения для получения приближения следующей итерации: $O(m_1 m_2 \min(m_1, m_2))$.

Так как $q < 1, r < \min(m_1, m_2)$, имеем сложность одной итерации алгоритма $O(m m_2 \min(m_1, m_2))$, кубическую в случае квадратной матрицы.

Рассмотрим полученные в разделах выше оценки на сложность итерации метода приближенной проекции градиента:

- $O((m_1 + m_2)p^2) + O(m_1 m_2 p q)$ для случайного проектирования,
- $O((m_1 + m_2)rp) + O(p^3)$ для псевдоскелетного метода.

С учетом $q = O(\epsilon) = O\left(\frac{r}{p}\right)$ для метода случайного проектирования, отсюда имеем, что асимптотика сложности одной итерации псевдоскелетного метода в $O(\min(m_1, m_2)/p)$ раз меньше сложности одной итерации метода случайного проектирования.

Теперь рассмотрим различные значения плотности известных элементов q . Пусть $m_2 \geq m_1$. В случае минимально возможного $q = q_{min} = O\left(\frac{r^2 \mu^2 \log n}{m_1}\right)$ имеем сложность порядка

- $O\left(\frac{(m_1+m_2)m_1^2}{r^2 \mu^4 \log^2 m_2}\right) + O(m_1 m_2 r)$ для случайного проектирования;
- $O\left(\frac{m_1^3}{\mu^6 r^3 \log^3 m_2}\right) + O\left(\frac{(m_1+m_2)m_1}{\mu^2 \log m_2}\right)$ для псевдоскелетного метода.

Покажем, что при больших порядках q можно достичь и более низкой сложности метода приближенной проекции градиента. Пусть $q = m_1^{-1/2} \gg \frac{r^2 \mu^2 \log n}{m}$ (при малых порядках μ , близких к константе, или к логарифму от размеров матрицы); тогда подстановкой имеем сложность

- $O((m_1 + m_2)m_1 r^2)$ для случайного проектирования;
- $O((m_1 + m_2)m_1^{\frac{1}{2}} r^2) + O(m_1^{\frac{3}{2}} r^3)$ для псевдоскелетного метода.

Проведем экспериментальное сравнение метода проекции градиента с точным проектированием (в подписях для краткости - 'SVP') и двумя рассмотренными вариантами методов приближенного проектирования (в подписях - 'ASVP').

- Рассматриваются квадратные матрицы размера $m_1 = m_2 = 1000$. Рассматриваются ранги матриц r от 5 до 25 и плотности распределения известных элементов q от 0.1 до 0.4.
- Матрицы формируются случайно следующим образом:
 - Формируются случайные матрицы $\hat{U}_r \in \mathbb{R}^{m_1 \times r}$ и $\hat{V}_r \in \mathbb{R}^{m_2 \times r}$, каждый элемент которых распределен случайно по стандартному нормальному распределению.
 - Вычисляются ортогональные базисы U_r и V_r в пространствах столбцов \hat{U}_r и \hat{V}_r , соответственно, например, с помощью QR -разложения. Экспериментально установлено, что некогерентность факторов U_r, V_r , выбранных таким образом, растет логарифмически от m_1, m_2 .
 - Искомая матрица X_* задается в виде $X_* = U_r \Sigma_r V_r^T$, с матрицей Σ_r сингулярных значений $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots$. Экспериментально установлено, что сходимость зависит от характера убывания сингулярных чисел σ_j ; в экспериментах рассмотрены следующие законы σ_j :
 - * $\sigma_j = 1$;
 - * $\sigma_j = 1/j$;

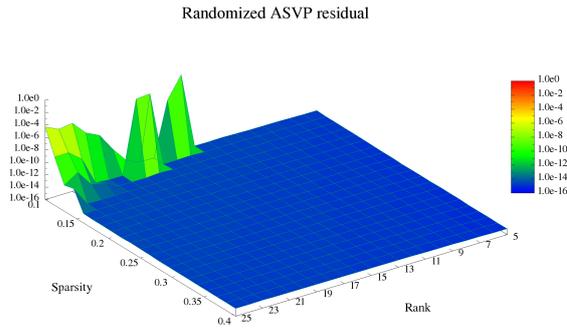
$$* \sigma_j = 1/j^2;$$

$$* \sigma_j = 1/2^j.$$

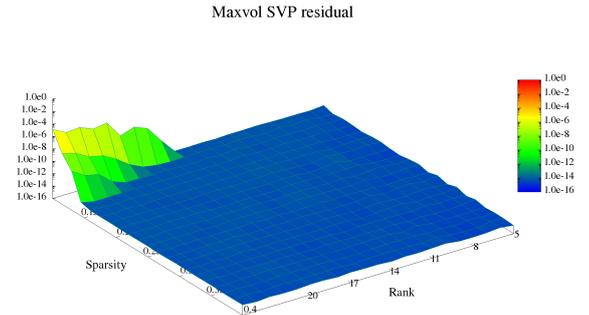
- Сравнение проводится с ограничением по времени практической работы алгоритмов. Ограничение задается константой, которая определяется временем выполнения десяти или пятидесяти итераций метода проекции градиента с полным SVD; предполагается, что за это время приближенный вариант алгоритма выполнит большее число итераций. При этом для сокращения записи полный вариант метода проекции градиента помечен как „SVP“ а метод приближенной проекции градиента помечен как „ASVP“ („Approximate Singular Value Projection“).

Ниже приведены графики зависимостей невязки от q и r при фиксированном размере матрицы, при различных законах падения сингулярных чисел неизвестной матрицы. Значения полных относительных ошибок по всей матрице на практике во всех экспериментах отличаются от относительных невязок на маске известных элементов не более, чем в два раза.

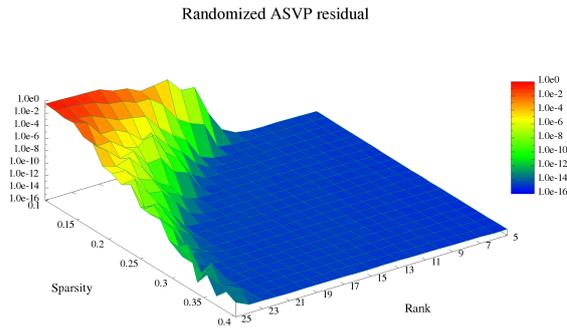
Графики сходимости в случае использования точной проекции приведены в двух вариантах: с шагом q и с шагом $\frac{3}{4}$. Так как рассматривалось ограничение в 50 итераций алгоритма с точной проекцией, вариант с шагом $\frac{3}{4}$ на большей части графиков не достиг сходимости, так как такого числа итераций оказалось недостаточно для завершения последовательного набора ранга.



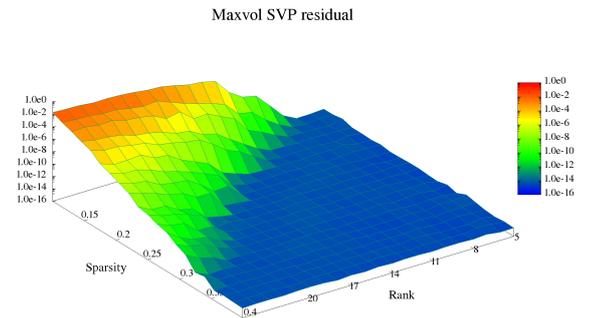
а) ASVP со случайным проектированием, ограничение по времени в 50 итераций SVP.



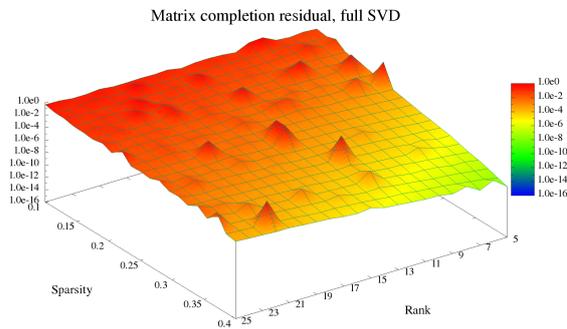
б) Псевдоскелетный ASVP, ограничение по времени в 50 итераций SVP.



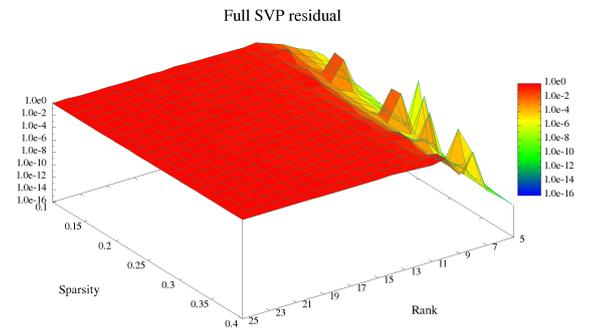
в) ASVP со случайным проектированием, ограничение по времени в 10 итераций SVP.



г) Псевдоскелетный ASVP, ограничение по времени в 10 итераций SVP.

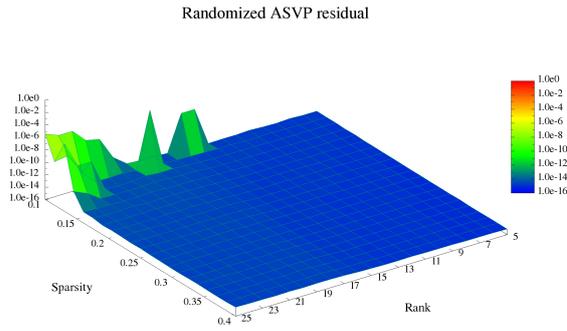


д) Полный SVP, 50 итераций, параметр шага $\tau = q$.

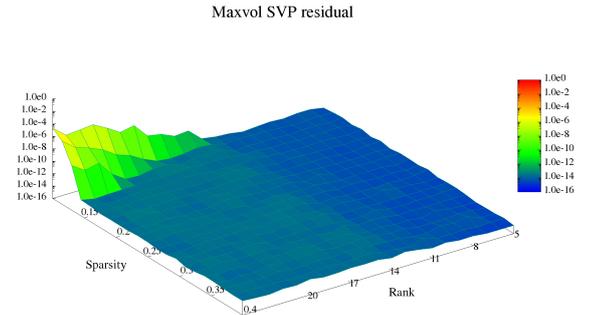


е) Полный SVP, 50 итераций, параметр шага $\tau = \frac{3}{4}$.

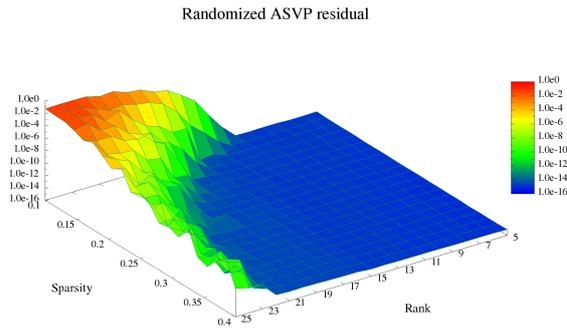
Рисунок 1.2 — Графики итоговой относительной невязки $\|\mathcal{A}_\Omega(X_t - X_*)\|_F / \|\mathcal{A}_\Omega(X_*)\|_F$ восполнения матриц с сингулярными числами вида $\sigma_i = 1$.



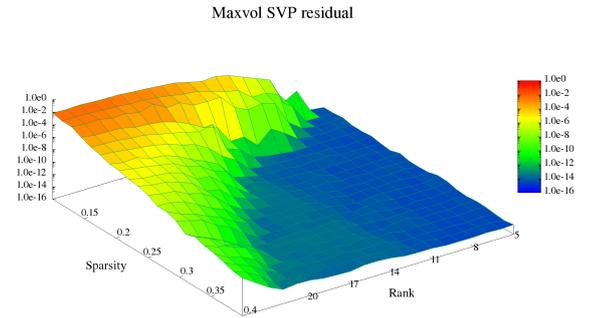
а) ASVP со случайным проектированием, ограничение по времени в 50 итераций SVP.



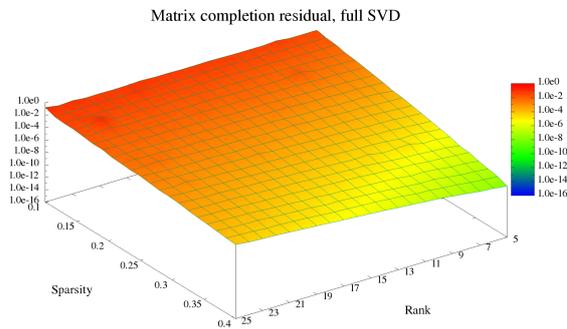
б) Псевдоскелетный ASVP, ограничение по времени в 50 итераций SVP.



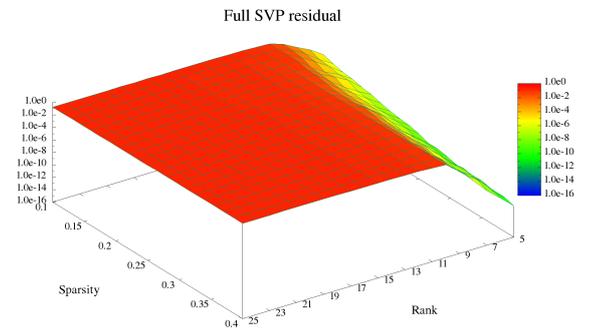
в) ASVP со случайным проектированием, ограничение по времени в 10 итераций SVP.



г) Псевдоскелетный ASVP, ограничение по времени в 10 итераций SVP.

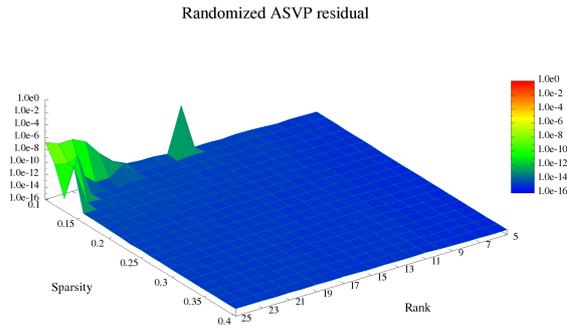


д) Полный SVP, 50 итераций, параметр шага $\tau = q$.

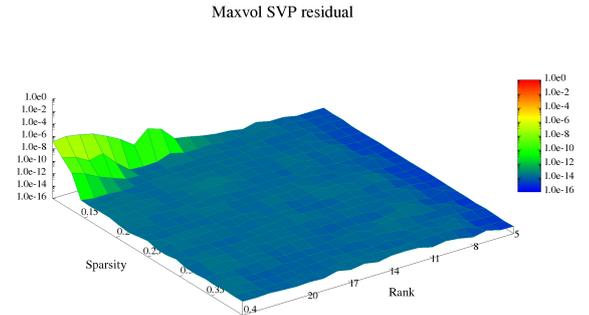


е) Полный SVP, 50 итераций, параметр шага $\tau = \frac{3}{4}$.

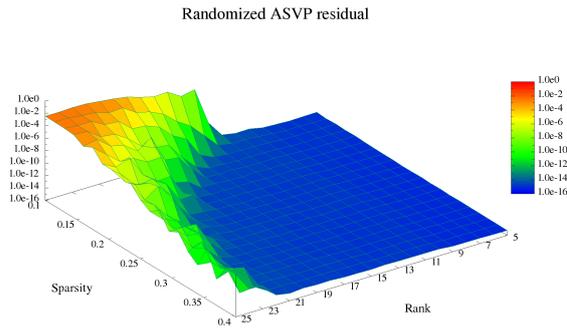
Рисунок 1.3 — Графики итоговой относительной невязки $\|\mathcal{A}_\Omega(X_k - X_*)\|_F / \|\mathcal{A}_\Omega(X_*)\|_F$ восполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{i}$.



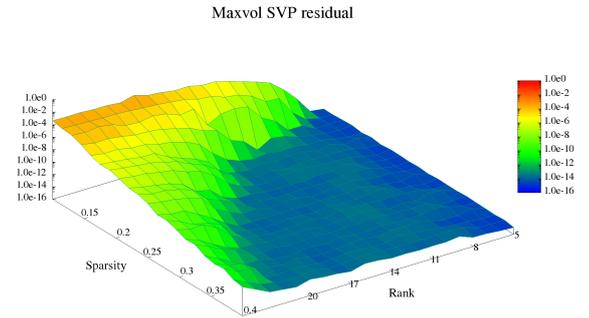
а) ASVP со случайным проектированием, ограничение по времени в 50 итераций SVP.



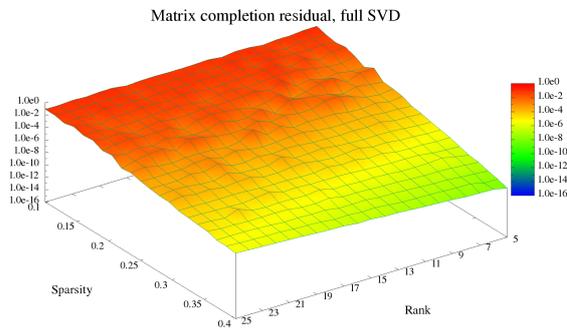
б) Псевдоскелетный ASVP, ограничение по времени в 50 итераций SVP.



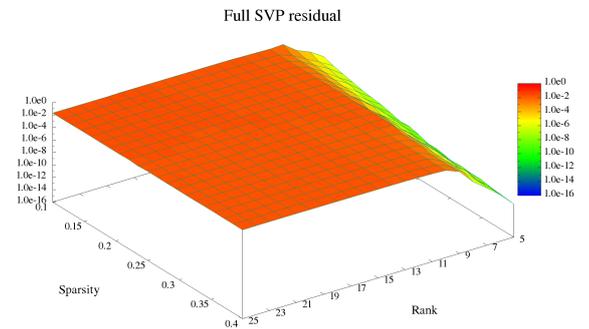
в) ASVP со случайным проектированием, ограничение по времени в 10 итераций SVP.



г) Псевдоскелетный ASVP, ограничение по времени в 10 итераций SVP.

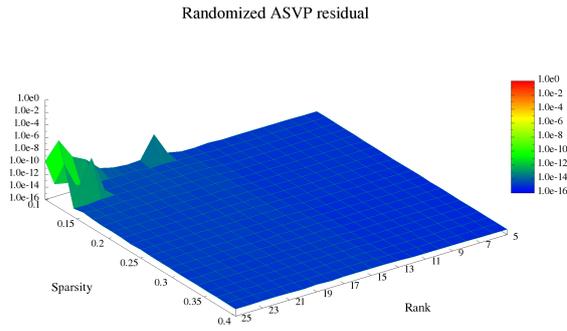


д) Полный SVP, 50 итераций, параметр шага $\tau = q$.

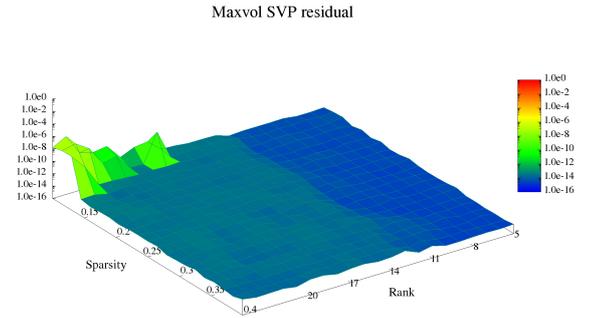


е) Полный SVP, 50 итераций, параметр шага $\tau = \frac{3}{4}$.

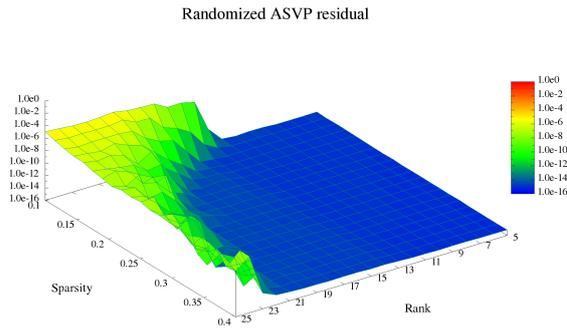
Рисунок 1.4 — Графики итоговой относительной невязки $\|\mathcal{A}_\Omega(X_k - X_*)\|_F / \|\mathcal{A}_\Omega(X_*)\|_F$ восполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{i^2}$.



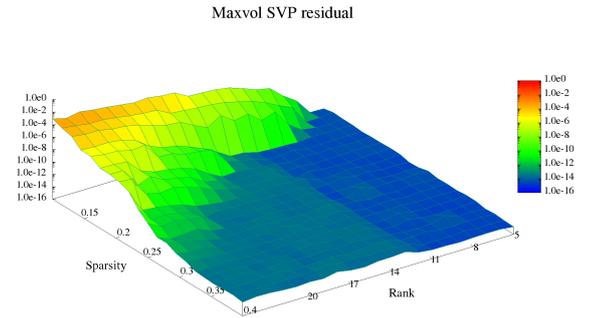
а) ASVP со случайным проектированием, ограничение по времени в 50 итераций SVP.



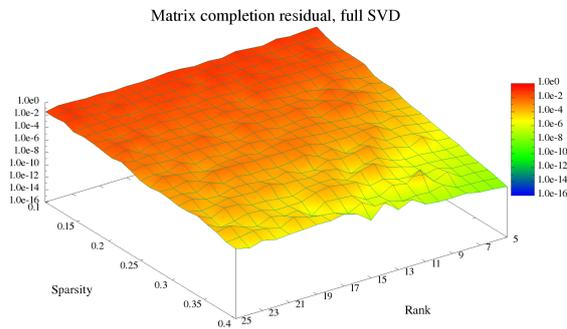
б) Псевдоскелетный ASVP, ограничение по времени в 50 итераций SVP.



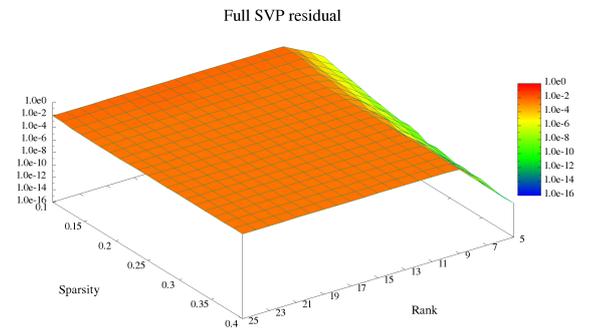
в) ASVP со случайным проектирование, ограничение по времени в 10 итераций SVP.



г) Псевдоскелетный ASVP, ограничение по времени в 10 итераций SVP.



д) Полный SVP, 50 итераций, параметр шага $\tau = q$.



е) Полный SVP, 50 итераций, параметр шага $\tau = \frac{3}{4}$.

Рисунок 1.5 — Графики итоговой относительной невязки $\|\mathcal{A}_\Omega(X_k - X_*)\|_F / \|\mathcal{A}_\Omega(X_*)\|_F$ восполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{2^i}$.

Подведем итоги этой главы. В главе рассмотрено применение методов приближенного вычисления частичного сингулярного разложения к существующему итерационному алгоритму восполнения матриц малого ранга, сходящемуся геометрически, основной операцией которого является сингулярное разложение матрицы с последующим отсечением всех неглавных компонент (SVD-проекция) на каждой итерации. Для этого алгоритма получен теоретический

результат, при определенных предположениях гарантирующий сохранение геометрической сходимости даже в случае замены точного вычисления SVD-проекции на приближенное. Сформулировано условие приближения, выполнения которого достаточно для геометрической сходимости при произвольном методе получения самой приближенной проекции. Рассмотрено два возможных варианта таких методов, удовлетворяющих этому условию в среднем, или с высокой вероятностью. Для обоих этих методов проведены численные эксперименты и получены похожие результаты, показывающие большую вычислительную эффективность использования приближенных SVD-проекций в методе проекции градиента для задачи восполнения матриц малого ранга.

Глава 2. Приближение матриц в виде суммы матрицы малого ранга и разреженной

В этой главе рассмотрим другую малопараметрическую модель матриц, связанную с матрицами малого ранга: модель, представляющую собой сумму матрицы малого ранга и разреженной матрицы. Число элементов разреженной матрицы будем считать линейным относительно числа строк/столбцов матриц; таким образом общее число параметров, описывающих малоранговую часть и общее число параметров, описывающих разреженную часть будут соразмерны. Позиции ненулевых элементов разреженной части будем считать неизвестными.

В этой главе предлагается применить к задаче поиска представления наперед заданной матрицы в виде суммы матрицы малого ранга и разреженной матрицы алгоритмы восполнения. Пусть входная матрица удовлетворяет модели в точности. В случае выбора искусственным образом случайной маски элементов такой матрицы пересечение шаблона разреженности разреженной части входной матрицы и маски будет с большой вероятностью иметь ненулевое, но достаточно малое число индексов. Таким образом, предлагается свести задачу приближения матрицы в формате суммы матриц малого ранга и разреженной к задаче восполнения с малым числом ошибочных данных. Для этого потребуются дополнительные модификации алгоритма восполнения; более того, в конечном итоге получено, что более эффективно рассматривать не одну, а две случайные маски одновременно. Для удобства в этой главе будем считать неизвестные матрицы квадратными.

Для построения теоретического анализа предложенного подхода структура рассмотренного ранее доказательства сходимости приближенного SVP оказывается неудобной. В этой главе будет предложен альтернативный подход к теоретическому анализу методов восполнения на основе теории возмущений сингулярных подпространств матриц малого ранга. Такой анализ будет требовать более жестких условий на приближенную операцию проектирования на множество матриц малого ранга по сравнению с анализом предыдущей главы, однако будет более общим и окажется удобным для рассмотрения случая разреженных ошибок во входных данных.

2.1 Теория возмущений сингулярных подпространств

Рассмотрим в этом разделе известные в литературе утверждения об устойчивости сингулярных подпространств к почти ортогональным возмущениям. Эта теория использует инструмент канонических углов между подпространствами в качестве численной характеристики таких возмущений.

2.1.1 Углы между подпространствами

Пусть даны два линейных подпространства $\mathcal{L} \in \mathbb{R}^m$ и $\mathcal{M} \in \mathbb{R}^m$ одинаковой размерности $\dim(\mathcal{L}) = \dim(\mathcal{M}) = r$. Пусть $P_{\mathcal{L}}, P_{\mathcal{M}}$ обозначают матрицы соответствующих ортогональных проекторов. Тогда [37; 38],

- Сингулярные числа σ_k произведения $P_{\mathcal{L}}P_{\mathcal{M}}$ лежат в интервале $[0,1]$. Отсортированные наборы левых и правых сингулярных векторов матрицы $P_{\mathcal{L}}P_{\mathcal{M}}$ называются „каноническими векторами“ между подпространствами \mathcal{L}, \mathcal{M} . Отсортированный набор значений $\phi_k \in [0, \frac{\pi}{2}]$, такой что $\cos \phi_k = \sigma_k(P_{\mathcal{L}}P_{\mathcal{M}})$, известен как набор „канонических углов“ между подпространствами \mathcal{L}, \mathcal{M} .
- Первая пара канонических векторов между подпространствами имеет смысл наиболее взаимно сонаправленной пары векторов из этих подпространств:

$$\{\vec{x}_1, \vec{y}_1\} = \arg \max_{\vec{x} \in \mathcal{L}, \vec{y} \in \mathcal{M}} (\vec{x}, \vec{y})_2.$$

Все остальные пары канонических векторов могут быть определены рекурсивно:

$$\{\vec{x}_k, \vec{y}_k\} = \arg \max_{\substack{\vec{x} \in \mathcal{L}, x \perp x_1 \dots x_{k-1}; \\ \vec{y} \in \mathcal{M}, y \perp y_1 \dots y_{k-1}}} (\vec{x}, \vec{y})_2, k \leq r.$$

- Собственные значения самосопряженной матрицы $P_{\mathcal{L}} - P_{\mathcal{M}}$ равны $\pm \sin \phi_k$ (и не более $2r$ из них не равны нулю), где величины ϕ_k обозначают канонические углы между $P_{\mathcal{L}}, P_{\mathcal{M}}$.

2.1.2 Теорема о возмущениях

Теперь приведем оценку возмущений сингулярных подпространств, доказанную в [39], схожую с известной оценкой Дэвиса-Кахана. Лемма гарантирует, что в случае если возмущение почти ортогонально сингулярным базисам наперед заданной матрицы малого ранга, то добавление такого возмущения слабо изменяет сингулярные подпространства этой матрицы. Здесь и далее будем считать что $P_U := UU^*$ обозначает матрицу ортогонального проектирования на подпространство столбцов базиса U , а $\mathcal{P}_r(X) \in \mathbb{R}^{m \times m}$, как и ранее, обозначает точную проекцию матрицы $X \in \mathbb{R}^{m \times m}$ на множество матриц ранга не выше r , задающуюся с помощью сингулярного разложения.

Лемма 2.1.1 (Wedin, [39; 40]). Пусть $X \in \mathbb{R}^{m \times m}$ - матрица ранга r , а $X = U\Sigma_r V^*$ - ее сокращенное сингулярное разложение с базисами $U, V \in \mathbb{R}^{m \times r}$. Пусть $N \in \mathbb{R}^{m \times m}$ - аддитивное возмущение, удовлетворяющее $\max\{\|P_U N\|_F, \|N P_V\|_F\} \leq \eta$. Пусть $\gamma \in \mathbb{R}$ - положительная константа. При условии $\gamma \leq \sigma_r(X) - \|N\|_2$ и обозначении $\mathcal{P}_r(X + N) = \hat{U}_r \hat{\Sigma}_r \hat{V}_r^*$ выполнено

$$\max\{\|P_U - P_{\hat{U}_r}\|_F, \|P_V - P_{\hat{V}_r}\|_F\} \leq 2\frac{\eta}{\gamma}.$$

Доказательство. Утверждение доказано в статье [39] и расширено в работе [40]. Приведенный вариант утверждения получается путем подстановки в Теорему 4 из [40] обозначений $\Sigma_r \leftrightarrow \tilde{\Sigma}_1$, $U_r^* N \leftrightarrow S$, $N V_r \leftrightarrow R$. \square

Приведенная теорема характеризует сингулярные подпространства и, в частности, синусы углов между ними. Для дальнейшего использования докажем лемму, в которой на основе рассмотренных выше оценок возмущений сингулярных подпространств построена оценка на близость оптимального малорангового приближения возмущенной матрицы к исходной матрице.

Лемма 2.1.2. Пусть $X \in \mathbb{R}^{m \times m}$ - матрица ранга r , и пусть $U\Sigma V^*$ - ее сокращенное сингулярное разложение, где $U, V \in \mathbb{R}^{m \times r}$. Пусть $N \in \mathbb{R}^{m \times m}$ - аддитивное возмущение, такое что выполнено $\max\{\|P_U N\|_F, \|N P_V\|_F\} \leq \eta$, $\|N\|_2 \leq \frac{\sigma_r(X)}{2}$.

Пусть $\kappa := \frac{\sigma_1(X)}{\sigma_r(X)}$ - характеристика обусловленности малоранговой матрицы X . Тогда,

$$\|\mathcal{P}_r(X + N) - X\|_F \leq \eta(8\kappa + 3).$$

Доказательство. Обозначая $\mathcal{P}_r(X + N) := \tilde{U}\tilde{\Sigma}\tilde{V}^*$, $\Delta P_U := P_{\tilde{U}} - P_U$, $\Delta P_V := P_{\tilde{V}} - P_V$, где $\tilde{U}, \tilde{V} \in \mathbb{R}^{n \times r}$, и используя предыдущее утверждение, раскроем скобки:

$$\begin{aligned} \|\mathcal{P}_r(X + N) - X\|_F &= \|P_{\tilde{U}}(X + N)P_{\tilde{V}} - P_U X P_V\|_F \\ &= \|(P_U + \Delta P_U)(X + N)(P_V + \Delta P_V) - P_U X P_V\|_F \\ &\leq \|(P_U + \Delta P_U)X(P_V + \Delta P_V) - P_U X P_V\|_F \end{aligned} \quad (2.1)$$

$$+ \|(P_U + \Delta P_U)N(P_V + \Delta P_V)\|_F. \quad (2.2)$$

Обозначим $\gamma := \sigma_r(X) - \|N\|_2$; так как $\|N\|_2 < \frac{\sigma_r(X)}{2}$, $\gamma > \frac{\sigma_r(X)}{2}$. Тогда, Лемма 2.1.1 позволяет оценить $\|\Delta P_U\|_F \leq \frac{2\eta}{\gamma}$, $\|\Delta P_V\|_F \leq \frac{2\eta}{\gamma}$, и для двух слагаемых (2.1), (2.2) выполнены оценки

$$\begin{aligned} \|(P_U + \Delta P_U)(X + N)(P_V + \Delta P_V) - P_U X P_V\|_F &\leq \\ &\leq \|P_U X \Delta P_V\|_F + \|\Delta P_U X P_{\tilde{V}}\|_F \\ &\leq (\|\Delta P_V\|_F + \|\Delta P_U\|_F) \|X\|_2 \\ &\leq \frac{4\eta}{\gamma} \|X\|_2 \leq 8\eta \frac{\sigma_1(X)}{\sigma_r(X)} = 8\eta\kappa \end{aligned}$$

и

$$\begin{aligned} \|(P_U + \Delta P_U)N(P_V + \Delta P_V)\|_F &\leq \|P_U E P_{\tilde{V}}\|_F + \|\Delta P_U N P_{\tilde{V}}\|_F \\ &\leq \|P_U N\|_F + \|\Delta P_U\|_F \|N\|_2 \\ &\leq \eta + \frac{2\eta}{\gamma} \|N\|_2 \leq 3\eta, \end{aligned}$$

где несколько раз использовано $\|AB\|_F \leq \|A\|_F \|B\|_2$, а последнее неравенство следует из $\gamma > \frac{\sigma_r(X)}{2} \geq \|N\|_2$. Суммируя два слагаемых, имеем требуемое. \square

2.2 Анализ сходимости выполнения матриц с неточными входными данными на основе теории возмущений

2.2.1 Оценки возникающего возмущения

Вспомним рассмотренный в предыдущей главе метод выполнения матриц „Singular Value Projection“ [29], представляющий собой метод проекции градиента для функционала

$$\begin{aligned} \Psi(Y) = \|\mathcal{A}(Y) - \mathcal{A}(X)\|_F^2 \rightarrow \min, \\ \text{rank}(Y) \leq r, \end{aligned} \quad (2.3)$$

где \mathcal{A} - линейный оператор, удовлетворяющий свойству ограниченной изометрии. Одна итерация метода проекции градиента тогда может быть записана в виде уравнений

$$\begin{aligned} W_k &= X - X_k, \\ X_{k+1} &= \mathcal{P}_r(X_k - \tau \nabla \Psi(X_k)) \\ &= \mathcal{P}_r(X_k + \tau \mathcal{A}^* \mathcal{A}(W_k)) \\ &= \mathcal{P}_r(X - W_k + \tau \mathcal{A}^* \mathcal{A}(W_k)) \\ &= \mathcal{P}_r(X + (\tau \mathcal{A}^* \mathcal{A} - \mathcal{I})(W_k)); \\ X_{k+1} &= \mathcal{P}_r(X + N_k), \\ N_k &:= (\tau \mathcal{A}^* \mathcal{A} - \mathcal{I})(W_k). \end{aligned} \quad (2.4)$$

где k обозначает номер итерации, X_k обозначает текущее приближение к искомой матрице X , W_k обозначает текущую ошибку, $\tau > 0$ - параметр шага, а оператор $\mathcal{I} \in \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$ - единичный в линейном подпространстве квадратных матриц.

В [29] доказано что если оператор задачи $\mathcal{A} : \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$ удовлетворяет условию ограниченной изометрии (1.3) с константой $0 < \delta_{2r} < 1$ для всех матриц ранга не выше $2r$, тогда метод SVP с константой шага $\tau = \frac{1}{1+\delta_{2r}} \approx 1$ сходится геометрически с точки зрения функционала (2.3), вне зависимости от

начального приближения X_0 . Также, в главе 1 эти результаты были обобщены на случай приближенного проектирования. В этой же главе, рассмотрим случай константного параметра шага $\tau = 1$ и воспользуемся теорией возмущений сингулярных подпространств. Сперва покажем, что ограниченная изометрия, определяемая через норму (1.3), влечет за собой и определенные свойства скалярного произведения:

Лемма 2.2.1. *Если линейный оператор $\mathcal{A} \in \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$ удовлетворяет условию ограниченной изометрии (1.3) для всех матриц $Z : \text{rank}(Z) \leq 2r$, то*

$$(X, Y)_F - \delta_{2r} \|X\|_F \|Y\|_F \leq (\mathcal{A}(X), \mathcal{A}(Y))_F \leq (X, Y)_F + \delta_{2r} \|X\|_F \|Y\|_F, \\ X, Y \in \mathbb{R}^{n \times n} : \text{rank}(X) \leq r, \text{rank}(Y) \leq r.$$

Доказательство. Используя равенство параллелограмма для нормы Фробениуса, имеем

$$(X, Y)_F = \frac{1}{4} (\|X + Y\|_F^2 - \|X - Y\|_F^2) \\ (\mathcal{A}(X), \mathcal{A}(Y))_F = \frac{1}{4} (\|\mathcal{A}(X + Y)\|_F^2 - \|\mathcal{A}(X - Y)\|_F^2)$$

Теперь, если $\|X\|_F = \|Y\|_F = 1$, имеем

$$|(\mathcal{A}(X), \mathcal{A}(Y))_F - (X, Y)_F| \leq \frac{1}{4} | \|\mathcal{A}(X + Y)\|_F^2 - \|X + Y\|_F^2 | + \\ + \frac{1}{4} | \|\mathcal{A}(X - Y)\|_F^2 - \|X - Y\|_F^2 | \\ \leq \frac{\delta_{2r}}{4} \|X + Y\|_F^2 + \frac{\delta_{2r}}{4} \|X - Y\|_F^2 \\ \leq \delta_{2r} \frac{4(X, Y)}{4} \leq \delta_{2r}. \quad (2.5)$$

В общем случае произвольных норм $\|X\|_F, \|Y\|_F$, утверждение следует из применения рассуждений выше к векторам единичной длины $\frac{X}{\|X\|_F}, \frac{Y}{\|Y\|_F}$, и последующего скалирования результата (2.5) на $\|X\|_F \|Y\|_F$. \square

Теперь оценим свойства возмущения N_k (2.4), возникающие в методе проекции градиента:

Лемма 2.2.2 (Оценка проекции невязки $\mathcal{A}^* \mathcal{A} - \mathcal{I}$). *Пусть $U, V \in \mathbb{R}^{n \times r}$ - произвольные матрицы с ортогональными столбцами, и пусть $Y \in \mathbb{R}^{n \times n}$,*

$\text{rank}(Y) \leq r$. Пусть \mathcal{A} - оператор, удовлетворяющий условию ограниченной изометрии с константой δ_{2r} на всех матрицах ранга не выше $2r$. Тогда

$$\max\{\|P_U((\mathcal{A}^*\mathcal{A} - \mathcal{I})Y)\|_F, \|((\mathcal{A}^*\mathcal{A} - \mathcal{I})Y)P_V\|_F\} \leq \delta_{2r}\|Y\|_F,$$

где P_U, P_V обозначают матрицы ортогональных проекторов UU^*, VV^* соответственно.

Доказательство. Используя $\text{rank}(P_U(\mathcal{A}^*\mathcal{A} - \mathcal{I})Y) \leq r$, и раскрывая норму Фробениуса как максимум скалярного произведения, имеем

$$\begin{aligned} \|P_U(\mathcal{A}^*\mathcal{A} - \mathcal{I})Y\|_F &= \max_{L,R \in \mathbb{R}^{m \times r}, \|LR^*\|_F \leq 1} (P_U(\mathcal{A}^*\mathcal{A} - \mathcal{I})Y, LR^*)_F \\ &= \max_{\|LR^*\|_F \leq 1} ((\mathcal{A}(Y), \mathcal{A}(P_U LR^*))_F - (Y, P_U LR^*)_F) \\ &\leq \delta_{2r}\|Y\|_F \max_{\|LR^*\|_F \leq 1} \|P_U LR^*\|_F \\ &= \delta_{2r}\|Y\|_F. \end{aligned}$$

где неравенство следует из Леммы 2.2.1 о сохранении скалярных произведений. \square

Лемма 2.2.2 имеет следующее значение: хотя $N_k = (\mathcal{A}^*\mathcal{A} - \mathcal{I})W_k$ вообще говоря имеет „большую“ норму по сравнению с ошибкой W_k текущей итерации, она почти ортогональна подпространствам искомой X . Тогда, по лемме 2.1.1, возмущение N_k слабо изменяет сингулярные базисы X , и выполнено $\|\mathcal{P}_r(X + N_k) - X\|_F \ll \|W_k\|_F$.

2.2.2 Более строгое условие приближенного проектирования

Возвращаясь к рассуждениям предыдущей главы, вспомним что использование точного проектирования \mathcal{P}_r с помощью сингулярного разложения обладает высокой вычислительной сложностью и желательно заменить его приближенным. Теория, которая будет построена в этой главе, будет учитывать эту возможность, однако будет использовать более строгое условие приближе-

ния. Для построения теории Главы 1 достаточно было использовать условие 1.4, которое для удобства продублируем:

Определение 2.2.1 (Приближенное проектирование, [41], [42]). *Оператор $\hat{\mathcal{P}}_r$ будем называть ϵ -приближенным оператором проектирования, если*

$$\|\hat{\mathcal{P}}_r(Y) - Y\|_F^2 \leq (1 + \epsilon)\|Y - \mathcal{P}_r(Y)\|_F^2, \quad Y \in \mathbb{R}^{m \times m}. \quad (2.6)$$

В этой главе будем использовать другое условие, которое, однако, позволит впоследствии свести все варианты восполнения в единую теорему:

Определение 2.2.2 (Приближенное проектирование, более строгое). *Оператор $\hat{\mathcal{P}}_r$ будем называть сильно ϵ -приближенным оператором проектирования, если*

$$\|\hat{\mathcal{P}}_r(Y) - \mathcal{P}_r(Y)\|_F \leq \epsilon\|Y - \mathcal{P}_r(Y)\|_F, \quad Y \in \mathbb{R}^{m \times m}. \quad (2.7)$$

Определения 2.2.1, 2.2.2 неэквивалентны: если выполнено (2.7), то

$$\begin{aligned} \|\hat{\mathcal{P}}_r(Y) - Y\|_F &\leq \|\hat{\mathcal{P}}_r(Y) - \mathcal{P}_r(Y)\|_F + \|\mathcal{P}_r(Y) - Y\|_F \\ &\leq (1 + \epsilon)\|\mathcal{P}_r(Y) - Y\|_F, \end{aligned}$$

и (2.6) выполнено с константой $1 + \tilde{\epsilon} = \sqrt{1 + \epsilon}$. Если выполнено (2.6), определение 2.2.2 вообще говоря не выполняется, что можно увидеть на следующем примере. Пусть

$$Y = \begin{bmatrix} 1.001 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathcal{P}_r(Y) = \begin{bmatrix} 1.001 & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{\mathcal{P}}_r(Y) := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix};$$

Тогда, полагая $\epsilon = 0.001$ имеем

$$\|Y - \hat{\mathcal{P}}_r(Y)\|_F = (1 + \epsilon)\|Y - \mathcal{P}_r(Y)\|_F, \quad \|\mathcal{P}_r(Y) - \hat{\mathcal{P}}_r(Y)\|_F = \|Y\|_F.$$

В контексте этой работы, однако, подоператорное выражение Y обладает определенными частными свойствами. Лемма 2.1.2, в частности, требует наличия зазора сингулярных чисел матрицы $Y = X + E$ в форме $\sigma_r(Y) - \sigma_{r+1}(Y) > \text{const}$. Покажем, что при определенных свойствах сингулярных чисел матрицы Y определения 2.2.1, 2.2.2 близки к эквивалентным.

Сначала отметим, что если матрица ранга r имеет ортогональный базис столбцов вида $\hat{U} \in \mathbb{R}^{m \times r}$, $\hat{\mathcal{P}}_r(Y) = \hat{U}Z$, тогда $\|\hat{\mathcal{P}}_r(Y) - Y\|_F \geq \|P_{\hat{U}}Y - Y\|_F$, где $P_{\hat{U}} = \hat{U}\hat{U}^*$ - ортогональный проектор. Так как вычисление \hat{U}^*Y имеет низкую сложность по сравнению с вычислением сингулярного разложения, естественно считать что приближенный проектор $\hat{\mathcal{P}}_r(Y)$ всегда имеет вид $P_{\hat{U}(Y)}Y$, где $\hat{U}(Y)$ - некое приближение базиса из r сингулярных столбцов Y .

Лемма 2.2.3 (Эквивалентность условий приближения). Пусть задана матрица $Y \in \mathbb{R}^{m \times n}$, и пусть $\mathcal{P}_r(Y) = P_UY = U\Sigma V^*$, $U, V \in \mathbb{R}^{m \times r}$ - ее оптимальная проекция на множество матриц ранга не выше r . Пусть $P_{\hat{U}}Y$ - приближенная проекция, удовлетворяющая

$$\|Y - P_{\hat{U}}Y\|_F^2 \leq (1 + \epsilon)\|Y - P_UY\|_F^2. \quad (2.8)$$

Пусть сингулярные числа Y удовлетворяют условиям

$$\sum_{i=1}^r \sigma_i^2(Y) \leq 2\epsilon \sum_{i=r+1}^n \sigma_i^2(Y), \quad (2.9)$$

$$\sum_{i=r+1}^{2r} \sigma_i^2(Y) \leq \epsilon \sum_{i=r+1}^n \sigma_i^2(Y). \quad (2.10)$$

$$(2.11)$$

Тогда,

$$\|P_{\hat{U}}Y - P_UY\|_F \leq (\sqrt{2} + \sqrt{6})\sqrt{\epsilon}\|Y - P_UY\|_F.$$

Доказательство. По свойству аппроксимации 2.8, можно видеть, что

$$(1 + \epsilon) \sum_{i=r+1}^n \sigma_i^2(Y) \geq \|(I - P_{\hat{U}})Y\|_F^2 = \|P_U(I - P_{\hat{U}})Y\|_F^2 + \|(I - P_U)(I - P_{\hat{U}})Y\|_F^2,$$

$$\|(I - P_U)(I - P_{\hat{U}})Y\|_F^2 = \|(I - P_U)Y - (I - P_U)P_{\hat{U}}Y\|_F^2 \geq \sum_{i=2r+1}^n \sigma_i^2(Y),$$

где последнее неравенство следует из оценки на оптимальную фробениусову невязку между матрицей $(I - P_U)Y$ и другой матрицей ранга r . Вычитая, имеем

$$\begin{aligned} \|P_U(I - P_{\hat{U}})Y\|_F^2 &\leq \sum_{i=r+1}^{2r} \sigma_i^2(Y) + \epsilon \sum_{i=r+1}^n \sigma_i^2(Y) \\ &= \beta^2(Y) \|P_U Y\|_F^2, \\ \beta^2(Y) &:= \frac{\sum_{i=r+1}^{2r} \sigma_i^2(Y) + \epsilon \sum_{i=r+1}^n \sigma_i^2(Y)}{\sum_{i=1}^r \sigma_i^2(Y)}. \end{aligned}$$

С другой стороны, по свойствам ортогонального проектирования, и используя оптимальность сингулярного пространства U , имеем

$$\begin{aligned} \|(I - P_{\hat{U}})Y\|_F^2 &= \|Y\|_F^2 - \|P_{\hat{U}}Y\|_F^2 = \|Y\|_F^2 - \|(I - P_U)P_{\hat{U}}Y\|_F^2 - \|P_U P_{\hat{U}}Y\|_F^2, \\ \|(I - P_{\hat{U}})Y\|_F^2 &\geq \|(I - P_U)Y\|_F^2. \end{aligned}$$

Переставляя слагаемые, имеем

$$\|(I - P_U)P_{\hat{U}}Y\|_F^2 \leq \|Y\|_F^2 - \|(I - P_U)Y\|_F^2 - \|P_U P_{\hat{U}}Y\|_F^2 = \|P_U Y\|_F^2 - \|P_U P_{\hat{U}}Y\|_F^2.$$

Теперь оценим величину $\|P_U Y\|_F^2 - \|P_U P_{\hat{U}}Y\|_F^2$ с помощью полученного ранее $\|P_U Y - P_U P_{\hat{U}}Y\|_F^2 \leq \beta^2(Y) \|P_U Y\|_F^2$. Оценка была бы тривиальной в отсутствие квадратов; чтобы обработать квадраты, введем арифметическую подстановку $a \leftrightarrow \|P_U P_{\hat{U}}Y\|_F, b \leftrightarrow \|P_U Y\|_F, b \geq a \geq 0, (b - a)^2 \leq \beta^2 b^2$. Тогда,

$$\begin{aligned} b^2 - a^2 &= b^2 - 2ab + a^2 + 2ab - 2a^2 = (b - a)^2 + 2a(b - a) \leq (\beta^2 + 2\beta)b^2. \\ \|(I - P_U)P_{\hat{U}}Y\|_F^2 &\leq (\beta^2 + 2\beta) \|P_U Y\|_F^2. \end{aligned}$$

Продолжим оценку

$$\begin{aligned} \|P_{\hat{U}}Y - P_U Y\|_F &\leq \|P_U Y - P_U P_{\hat{U}}Y\|_F + \|P_{\hat{U}}Y - P_U P_{\hat{U}}Y\|_F \\ &= \|P_U(I - P_{\hat{U}})Y\|_F + \|P_{\hat{U}}(I - P_U)Y\|_F \\ &\leq (\beta + \sqrt{\beta^2 + 2\beta}) \|P_U Y\|_F. \end{aligned}$$

Для завершения доказательства, используем условия (2.9), (2.10) чтобы оценить $\beta(Y)$. Неравенство (2.10) дает

$$\sum_{i=r+1}^{2r} \sigma_i^2(Y) + \epsilon \sum_{i=r+1}^n \sigma_i^2(Y) \leq 2\epsilon \sum_{i=r+1}^n \sigma_i^2(Y),$$

откуда по (2.9) $\beta(Y) > 1$. Следовательно,

$$\begin{aligned} \|P_{\hat{U}}Y - P_UY\|_F &\leq (1 + \sqrt{3})\beta\|P_UY\|_F = (1 + \sqrt{3})\sqrt{\sum_{i=r+1}^{2r} \sigma_i^2(Y) + \epsilon \sum_{i=r+1}^n \sigma_i^2(Y)} \\ &\leq (1 + \sqrt{3})\sqrt{2\epsilon \sum_{i=r+1}^n \sigma_i^2(Y)} = (\sqrt{2} + \sqrt{6})\sqrt{\epsilon}\|Y - P_UY\|_F. \end{aligned}$$

□

2.2.3 Общий вид теорем о сходимости методов восполнения с неточными данными

Теперь докажем общую теорему о сходимости метода восполнения матриц; эта теорема будет использовать Определение 2.2.2 и будет требовать выполнение условия ограниченной изометрии (1.3) для всех матриц ранга не выше $4r$. Эти условия являются более строгими по сравнению с условиями, использованными в [29], [41] и Главе 1; однако, в этих условиях удастся значительно сократить доказательство, используя теорию возмущений.

Теорема 2.2.4 (Сходимость метода проекции градиента). *Пусть X - матрица ранга r и пусть $\kappa = \frac{\sigma_1(X)}{\sigma_r(X)}$. Пусть \mathcal{A} - линейный оператор, удовлетворяющий условию ограниченной изометрии с константой δ_{4r} на всех матрицах ранга не выше $4r$. Тогда метод проекции градиента (2.4) обладает локальной геометрической сходимостью с константой $\delta_{4r}(8\kappa + 3)$.*

Пусть $\|\mathcal{A}^\|_F = \max_{X \neq 0} \frac{\|\mathcal{A}^*(X)\|_F}{\|X\|_F}$. Дополнительно, пусть $\hat{\mathcal{P}}_r$ - ϵ -приближенный оператор проектирования в смысле Определения 2.2.2, и пусть $\epsilon\|\mathcal{A}^*\|_F(1 + \delta_{4r}) < 1 - \delta_{4r}(8\kappa + 3)$. Тогда, метод проекции градиента*

с заменой точной проекции \mathcal{P}_r на приближенную $\hat{\mathcal{P}}_r$ обладает локальной геометрической сходимостью.

Доказательство. В случае точного вычисления проекции, принимая во внимание $\text{rank}(W_k) \leq 2r$, и предполагая $\|N_k\|_2 \leq \frac{\sigma_r(X)}{2}$, Леммы 2.1.2, 2.2.2 дают оценки

$$\begin{aligned} \|P_U N_k\|_F &\leq \delta_{4r} \|W_k\|_F, \quad \|N_k P_V\|_F \leq \delta_{4r} \|W_k\|_F \\ \|W_{k+1}\|_F &= \|\mathcal{P}_r(X + N_k) - X\|_F \leq \delta_{4r}(8\kappa + 3) \|W_k\|_F. \end{aligned}$$

Предположение $\|N_k\|_2 \leq \frac{\sigma_r(X)}{2}$ выполнено локально при достаточно малых значениях ошибки текущей итерации: так как

$$\begin{aligned} \|N_k\|_2 &= \|(\mathcal{I} - \mathcal{A}^* \mathcal{A})(W_k)\|_2 \leq \|\mathcal{I}\|_F \|W_k\|_F + \|\mathcal{A}^*\|_F \|\mathcal{A}(W_k)\|_F \\ &\leq (1 + (1 + \delta_{4r}) \|\mathcal{A}^*\|_F) \|W_k\|_F, \end{aligned}$$

достаточно предположить

$$\|W_k\|_2 \leq \frac{1}{2\|\mathcal{A}^*\|_F} \sigma_r(X).$$

В случае замены точной операции проектирования \mathcal{P}_r приближенной операцией $\hat{\mathcal{P}}_r$, будем использовать

$$\begin{aligned} W_{k+1} &= X - \hat{\mathcal{P}}_r(X + N_k) \\ &= X - \mathcal{P}_r(X + N_k) + \mathcal{P}_r(X + N_k) - \hat{\mathcal{P}}_r(X + N_k), \end{aligned}$$

и достаточно оценить величину $\|\mathcal{P}_r(X + N_k) - \hat{\mathcal{P}}_r(X + N_k)\|_F$ с помощью

$$\begin{aligned} \|\mathcal{P}_r(X + N_k) - \hat{\mathcal{P}}_r(X + N_k)\|_F &\leq \epsilon \|\mathcal{P}_r(X + N_k) - (X + N_k)\|_F \\ &= \epsilon \|\mathcal{P}_r(X_k + \mathcal{A}^* \mathcal{A}(W_k)) - (X_k + \mathcal{A}^* \mathcal{A}(W_k))\|_F \\ &\leq \epsilon \|X_k - (X_k + \mathcal{A}^* \mathcal{A}(W_k))\|_F \\ &= \epsilon \|\mathcal{A}^* \mathcal{A}(W_k)\|_F \leq (1 + \delta_{4r}) \epsilon \|\mathcal{A}^*\|_F \|W_k\|_F. \end{aligned}$$

где была произведена замена оптимального приближения на неоптимальное X_k ранга r с целью получения верхней оценки. \square

По сравнению с результатами [29] и Теоремы 1.1.2, Теорема 4.2 оценивает сходимость ошибки, а не сходимость функционала (невязки), и предполагает явную зависимость сходимости от числа обусловленности задачи κ , наблюдаемую в численных экспериментах.

Проведем обобщение этого теоретического результата на случай аддитивной ошибки градиента, актуальный для построения алгоритма представления матрицы в виде суммы матрицы малого ранга и разреженной матрицы. Результат представлен в следующей Лемме.

Лемма 2.2.5 (Восполнение с аддитивной ошибкой). *Пусть выполнены условия предыдущей теоремы, но на каждой итерации добавляется слагаемое ошибки градиента \hat{S}_k :*

$$\begin{aligned} X_{k+1} &:= \mathcal{P}_r(X_k + \mathcal{A}^* \mathcal{A}(W_k + \hat{S}_k)) \\ &= \mathcal{P}_r(X + N_k + S_k), \\ S_k &:= \mathcal{A}^* \mathcal{A}(\hat{S}_k). \end{aligned}$$

Пусть ошибочное слагаемое удовлетворяет оценкам:

$$\|S_k\|_F \leq \|\mathcal{A}^*\|_F \|W_k\|_F, \forall k \quad (2.12)$$

$$\|P_U S_k\|_F \leq \delta_{4r} \|W_k\|_F,$$

$$\|S_k P_V\|_F \leq \delta_{4r} \|W_k\|_F, \forall k. \quad (2.13)$$

Тогда,

- Если выполнено $2\delta_{4r}(8\kappa + 3) < 1$, метод проекции градиента с точной проекцией и ошибочным слагаемым в градиенте обладает локальной геометрической сходимостью с константой $2\delta_{4r}(8\kappa + 3)$.
- Если выполнено $\epsilon \|\mathcal{A}^*\|_F (2 + \delta_{4r}) < 1 - 2\delta_{4r}(8\kappa + 3)$, метод проекции градиента с ϵ -приближенной проекцией и ошибочным слагаемым в градиенте обладает локальной геометрической сходимостью.

Доказательство. Снова, как и в случае точного градиента, Лемма 2.1.2 дает оценку

$$\|W_{k+1}\|_F = \|\mathcal{P}_r(X + N_k + S_k) - X\|_F \leq \max\{\|P_U(N_k + S_k)\|_F, \|(N_k + S_k)P_V\|_F\}(8\kappa + 3).$$

По Лемме 2.2.2 и ограничениям на ошибочное слагаемое градиента имеем

$$\begin{aligned}\|P_U(N_k + S_k)\|_F &\leq \|P_U N_k\|_F + \|P_U S_k\|_F \leq 2\delta_{4r}\|W_k\|_F, \\ \|(N_k + S_k)P_V\|_F &\leq \|N_k P_V\|_F + \|S_k P_V\|_F \leq 2\delta_{4r}\|W_k\|_F.\end{aligned}$$

Эти оценки верны при условии локальности $\|N_k + S_k\|_2 \leq \sigma_r(X)$, для которого, принимая во внимание условие (2.12), достаточно предположить, что ошибка текущей итерации оценивается величиной

$$\|W_k\|_2 \leq \frac{1}{4\|\mathcal{A}^*\|_F} \sigma_r(X).$$

В случае использования ϵ -приближенного проектирования, имеем

$$\begin{aligned}W_{k+1} &= X - \hat{\mathcal{P}}_r(X + N_k + S_k) \\ &= X - \mathcal{P}_r(X + N_k + S_k) + \mathcal{P}_r(X + N_k + S_k) - \hat{\mathcal{P}}_r(X + N_k + S_k),\end{aligned}$$

и достаточно оценить $\|\mathcal{P}_r(X + N_k + S_k) - \hat{\mathcal{P}}_r(X + N_k + S_k)\|_F$ используя определения ϵ -приближенной проекции (2.7)

$$\begin{aligned}\|\mathcal{P}_r(X + N_k + S_k) - \hat{\mathcal{P}}_r(X + N_k + S_k)\|_F &\leq \\ &\leq \epsilon \|\mathcal{P}_r(X + N_k + S_k) - (X + N_k + S_k)\|_F \\ &= \epsilon \|\mathcal{P}_r(X_k + \mathcal{A}^* \mathcal{A}(W_k) + S_k) - (X_k + \mathcal{A}^* \mathcal{A}(W_k) + S_k)\|_F \\ &\leq \epsilon \|X_k - (X_k + \mathcal{A}^* \mathcal{A}(W_k) + S_k)\|_F = \epsilon (\|\mathcal{A}^* \mathcal{A}(W_k)\|_F + \|S_k\|_F) \\ &\leq \epsilon (2 + \delta_{4r}) \|\mathcal{A}^*\|_F \|W_k\|_F.\end{aligned}$$

□

2.3 Частный случай разреженных ошибок

Теперь вернемся к исходному плану: применить методы восполнения к задаче разложения матрицы в виде суммы матрицы малого ранга и разреженной. Будем предполагать, что матрица $Y \in \mathbb{R}^{m \times m}$ может быть представлена в виде

СУММЫ

$$Y = X + \hat{S}, \quad (2.14)$$

где первое слагаемое имеет наперед известный ранг, $\text{rank}(X) = r \ll n$, а второе слагаемое \hat{S} разрежено и имеет $O(m)$ ненулевых элементов. Рассмотрим задачу одновременного поиска отдельных компонентов X и \hat{S} при известном Y , но неизвестном шаблоне разреженности \hat{S} . Для построения теории будем предполагать, что позиции ненулевых элементов \hat{S} случайно и равномерно распределены по всевозможному набору индексов элементов матрицы. Будем предполагать, что вероятность присутствия каждого элемента в шаблоне разреженности константна и имеет вид $\frac{\beta}{m}$, $\beta = \text{const}$, что соответствует матожиданию в βm ненулевых элементов во всем \hat{S} .

Возвращаясь к анализу сходимости методов восполнения 1.2.2, будем предполагать что X является μ -некогерентной. Алгоритмы поиска точного представления матрицы в виде суммы матрицы малого ранга и разреженной с теорией сходимости в литературе существуют [17], однако основаны на выпуклой релаксации и обрабатывают неизвестную матрицу X как набор из $O(m^2)$ неизвестных, которые в конечном итоге сходятся к матрице малого ранга. Такой подход обладает высокой сложностью итерации. В этой работе предложим итерационный алгоритм, поддерживающий малопараметрическую структуру приближений к обоим неизвестным X и \hat{S} . Для этого, будем использовать следующую постановку задачи:

$$\|(\mathcal{I} - \mathcal{P}_\Lambda)(Y - X)\|_F \rightarrow_{Y, \Lambda} \inf, Y \in \mathbb{R}^{m \times m}, |\Lambda| \leq C\beta n, C > 1. \quad (2.15)$$

Здесь, $\Lambda \subset \{1 \dots m\} \times \{1 \dots m\}$, и $\mathcal{P}_\Lambda \in \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$ обозначает оператор в пространстве матриц, зануляющий все элементы входной матрицы, индексы которых не входят во множество Λ . Функционал (2.15) предполагает, что строится разложение, удовлетворяющее модели $Y = X + \hat{S}$, но не являющееся оптимальным: разреженная часть приближения может содержать число элементов, в константное число раз большее минимально возможного.

Рассмотрим итерационный алгоритм, инициализируемый посредством $k = 0, X_0 = 0, W_0 = X$, пустого множества Λ и случайной маски Ω_0 , выбран-

ной случайно с разреженностью $\rho \geq C_{RIP} \frac{\mu^2 r^2 \log(m)}{\delta_{4r}^2 m}$, по теореме 1.2.2. Итерации предлагаемого алгоритма выглядят следующим образом:

Algorithm 3: Одномасочный алгоритм восполнения по входным данным с разреженной ошибкой

```

1 while  $\|(\mathcal{I} - \mathcal{P}_\Lambda)(Y - X)\|_F \geq \varepsilon$  do
2   Увеличивается номер итерации  $k := k + 1$ .;
3   Находятся  $C\beta n, C > 1$  наибольших элементов текущей невязки
    $W_{k-1} + \hat{S} = X + \hat{S} - X_{k-1}$  по модулю. Множество соответствующих
   индексов обозначается как  $\Upsilon_k$ .;
4    $\Lambda := \Lambda \cup \Upsilon_k, \Omega_k := \Omega_{k-1} / \Upsilon_k$ .;
5    $X_k = \mathcal{P}_r(X_{k-1} + \mathcal{A}_{\Omega_k}^* \mathcal{A}_{\Omega_k}(W_{k-1} + \hat{S}))$ .;

```

Значение ε соответствует наперед заданной желаемой величине невязки в смысле корня из суммы квадратов элементов, вычисленной только по элементам матрицы, соответствующим по позиции нулевым элементам разреженной части приближения. Логика алгоритма в том, что по построению слагаемое ошибки $S_k := \mathcal{A}_{\Omega_k}^* \mathcal{A}_{\Omega_k}(\hat{S})$, возникающее на итерациях, будет попадать по условию утверждения о сходимости восполнения с ошибкой градиента 2.2.5. Для обоснования этого докажем несколько лемм о структуре S_k .

2.3.1 Пересечение шаблонов разреженности

Во-первых, множество ненулевых элементов S_k лежит в пересечении двух независимо выбранных случайных разреженных шаблонов \hat{S} и Ω_k ; это означает, что с высокой вероятностью S_k является сильно разреженной:

Лемма 2.3.1 (Пересечение шаблонов разреженности). Пусть $C \in [e^2, \frac{m}{\beta})$ - константа, и каждая позиция элемента матрицы (j_1, j_2) имеет одинаковую вероятность $\frac{\beta}{m}$ принадлежать множеству $\text{supp}(\hat{S})$ ненулевых элементов \hat{S} . Тогда:

- С вероятностью не меньше $1 - e^{-(C+1)\beta m}$, $|\text{supp}(\hat{S})| \leq C\beta m$.
- С вероятностью не меньше $1 - e^{-(C+1)\beta \rho m}$, $|\text{supp}(S_k)| \leq C\rho\beta m, \rho \in (0,1)$.

Доказательство. Отметим что $|supp(S_k)| \leq |supp(S_0)|$, и обе величины $|supp(\hat{S})|$ и $|supp(S_0)|$ могут быть рассмотрены как биномиально-распределенные случайные величины с параметром количества экспериментов n^2 и параметром „успеха“ одного эксперимента $\frac{\beta}{m}$ и $\frac{\beta\rho}{m}$, соответственно. Лемма тогда соответствует оценке на биномиальное распределение вида

$$\begin{aligned}\mathbb{P}(|supp(\hat{S})| > C\beta m) &= F(m^2 - C\beta m; m^2, 1 - \frac{\beta}{m}), \\ \mathbb{P}(|supp(S_0)| > C\beta\rho m) &= F(m^2 - C\beta\rho m; m^2, 1 - \frac{\beta\rho}{m}),\end{aligned}$$

где $F(s; l, p) = \mathbb{P}(X_{l,p} \leq s)$ - функция распределения биномиальной случайной величины $X_{l,p}$ с параметрами l и p . Оценку на „хвосты“ такого распределения возьмем из, например, [43], где показано

$$\begin{aligned}F(s; l, p) &\leq e^{-lD(\frac{s}{l}||p)}; \\ D(q||p) &:= q \log \frac{q}{p} + (1 - q) \log \frac{1 - q}{1 - p}; q, p \in (0, 1),\end{aligned}$$

где $D(q||p)$ обозначает так называемое расстояние Кульбака-Лейблера, имеющее явный аналитический вид для конкретного случая двух броска монетки с вероятностями q, p „орла“ и „решки“. Тогда, требуемая оценка следует из

$$\begin{aligned}F(m^2 - C\beta m; m^2, 1 - \frac{\beta}{m}) &\leq e^{-m^2(\frac{m^2 - C\beta m}{m^2} \log \frac{\frac{m^2 - C\beta m}{m^2}}{\frac{m^2 - \beta m}{m^2}} + \frac{C\beta m}{m^2} \log \frac{\frac{C\beta m}{m^2}}{\frac{\beta m}{m^2}})} \\ &= e^{-(m^2 - C\beta m) \log \frac{m^2 - C\beta m}{m^2 - \beta m}} e^{-C \log C\beta m} \\ &= e^{(m^2 - C\beta m) \log \frac{m^2 - \beta m}{m^2 - C\beta m}} e^{-C \log C\beta m} \\ &= e^{(m^2 - C\beta m) \log(1 + \frac{(C-1)\beta m}{m^2 - C\beta m})} e^{-C \log C\beta m} \\ &\leq e^{(C-1)\beta m} e^{-C \log C\beta m} \leq e^{-(C+1)\beta m}.\end{aligned}$$

где мы использовали $\log(1 + x) \leq x$ и $\log C \geq 2$. Вторая оценка следует из замены $\beta \leftrightarrow \beta\rho$ в приведенных рассуждениях. Условие $\log C \geq 2$ является избыточным и используется для упрощения вида итоговых оценок. \square

2.3.2 Нормы разреженных ошибок

Теперь, построим оценки на нормы $\|S_k\|_F, \|P_U S_k\|_F, \|S_k P_V\|_F$, где $U, V \in \mathbb{R}^{m \times r}$ - сингулярные базисы μ -некогерентной малоранговой части решения X . Оценка основана на следующей идее: по построению множество Υ_k исключает все „выраженно большие“ элементы разреженной ошибки и оставляет лишь те, что по модулю соразмерны норме ошибки текущего итерационного приближения. Для этого будем использовать следующую Лемму:

Лемма 2.3.2 (Исключение разреженных ошибок). *Пусть $C \in [e^2, \frac{m}{\beta})$. Предположим, что W_{k-1} является μ -некогерентной. Тогда, с вероятностью не ниже $1 - e^{-C\beta\rho m}$,*

$$\|S_k\|_F \leq 2\mu \sqrt{\frac{C\beta r}{\rho m}} \|W_{k-1}\|_F,$$

$$\max\{\|P_U S_k\|_F, \|S_k P_V\|_F\} \leq 2C\beta\mu r \sqrt{\frac{\mu}{m}} \|W_{k-1}\|_F.$$

Доказательство. Будем исходить из того, что $|supp(\hat{S})| \leq C\beta m$ и $|supp(S_k)| \leq C\beta\rho m$; по предыдущей Лемме, вероятность этого не ниже $1 - e^{-C\beta\rho m}$. Тогда, принимая во внимание $S_k = \mathcal{A}_{\Omega_k}^* \mathcal{A}_{\Omega_k}(\hat{S})$, имеем

$$\begin{aligned} \|S_k\|_C &= \|\mathcal{A}_{\Omega_k}^* \mathcal{A}_{\Omega_k}(\hat{S})\|_C = \|\mathcal{A}_{\Omega_k}^* \mathcal{A}_{\Omega_k}(W_{k-1} + \hat{S}) - \mathcal{A}_{\Omega_k}^* \mathcal{A}_{\Omega_k}(W_{k-1})\|_C \leq \\ &\leq \|\mathcal{A}_{\Omega_k}^* \mathcal{A}_{\Omega_k}(W_{k-1} + \hat{S})\|_C + \|\mathcal{A}_{\Omega_k}^* \mathcal{A}_{\Omega_k}(W_{k-1})\|_C \\ &= \max_{(i,j) \in \Omega_k} \frac{|W_{k-1} + \hat{S}|_{i,j}}{\rho} + \max_{(i,j) \in \Omega_k} \frac{|W_{k-1}|_{i,j}}{\rho} \\ &\leq \min_{(i,j) \in \Upsilon_k} \frac{|W_{k-1} + \hat{S}|_{i,j}}{\rho} + \max_{(i,j) \in \Omega_{k-1}} \frac{|W_{k-1}|_{i,j}}{\rho} \\ &\leq 2 \max_{(i,j) \in \Omega_{k-1}} \frac{|W_{k-1}|_{i,j}}{\rho}, \end{aligned}$$

где последнее неравенство следует из того, что в множестве Υ_k больше элементов, чем в шаблоне разреженности \hat{S} , по этому хотя бы для одной позиции $(i,j) \in \Upsilon_k$ выполнено $(W_{k-1} + \hat{S})_{i,j} = (W_{k-1})_{i,j}$. Используя свойство некогерент-

ности и Лемму 1.2.1, продолжим оценку:

$$\|S_k\|_C \leq 2 \frac{\|W_{k-1}\|_C}{\rho} \leq \frac{2\mu\sqrt{r}\|W_{k-1}\|_F}{\rho m}.$$

Теперь, совместим оценку $|supp(S_k)| \leq C\beta\rho m$ и поэлементную оценку на S_k для получения оценок на нормы Фробениуса для $S_k, P_U S_k, S_k P_V$:

$$\begin{aligned} \|S_k\|_F &\leq \sqrt{C\beta\rho m} \|S_k\|_C \leq 2\mu \sqrt{\frac{C\beta r}{\rho m}} \|W_{k-1}\|_F. \\ \|P_U S_k\|_F^2 &= \|U^* S_k\|_F^2 = \left\| \sum_{(i,j) \in supp(S_k)} s_{i,j} U^* \vec{e}_i \vec{e}_j^T \right\|_F^2 \\ &\leq |supp(S_k)| \sum_{(i,j) \in supp(S_k)} \|s_{i,j} U^* \vec{e}_i \vec{e}_j^T\|_F^2 \\ &\leq |supp(S_k)|^2 \|S_k\|_C^2 \max_i \|U^* \vec{e}_i\|_2^2 \leq \\ &\leq |supp(S_k)|^2 \|S_k\|_C^2 \frac{\mu r}{m} \\ &\leq \frac{(C^2 \beta^2 \rho^2 m^2) (4r\mu^2) (\mu r)}{(\rho^2 m^2) (m)} \|W_{k-1}\|_F^2 \\ &= \frac{4C^2 \beta^2 \mu^3 r^2}{m} \|W_{k-1}\|_F^2; \\ \|P_U S_k\|_F &\leq 2C\beta\mu r \sqrt{\frac{\mu}{m}} \|W_{k-1}\|_F. \end{aligned}$$

Вывод оценки на норму $\|S_k P_V\|_F$ опущен и полностью повторяет вывод для $\|P_U S_k\|_F$. \square

Лемма 2.3.2 в сочетании с доказательством Леммы 2.2.5 дает следующий вывод: предложенный алгоритм 3 сходится геометрически по величине $\|(\mathcal{I} - \mathcal{P}_\Lambda)(Y - X)\|_F$ если на каждой итерации k выполнены условия:

- И текущее приближение X_k , и ошибка текущего приближения W_k являются μ -некогерентными.
- Оператор \mathcal{A}_{Ω_k} удовлетворяет условию ограниченной изометрии (1.3) на всех μ -некогерентных матрицах ограниченного ранга с большой вероятностью.

Доказательство первого условия затруднительно: в то время как матрица X_k гарантированно неразрезана в случае $\|W_k\|_2 \ll \sigma_r(X)$, известно, что

ошибка W_k может терять свойства некогерентности даже в классическом методе проекции градиента без ошибочных слагаемых (для борьбы с чем в предыдущей главе были рассмотрены эвристики, такие как последовательный набор ранга приближения). Второе же утверждение может быть доказано на основе следующего рассуждения: несмотря на то что множество Ω_k теряет случайность и равномерность (так как исключенные элементы множеств Υ_j не случайны и зависят от X, \hat{S}), Ω_k получается из случайной маски Ω_0 удалением относительно небольшого количества элементов, а оператор \mathcal{A}_{Ω_0} условию ограниченной изометрии удовлетворяет по Теореме 1.2.2. Поэтому, завершим теоретический анализ Алгоритма 3 следующей Леммой.

Лемма 2.3.3 (Сохранение ограниченной изометрии). *Пусть исходное множество Ω_0 Алгоритма 3 выбрано случайно по равномерному распределению индексов*

$$\mathbb{P}\{(i, j) \in \Omega_0\} = \rho, \forall i, j,$$

и пусть $\rho \geq C_{RIP} \frac{\mu^2 r^2 \log(m)}{\delta_r^2 m}$. Тогда, с вероятностью не меньше $1 - \exp(-m \log(m))$, оператор \mathcal{A}_{Ω_k} (с тем же скалированием ρ) возникающий на итерации k Алгоритма 3 - удовлетворяет условию ограниченной изометрии с константой $\delta_{r,k}$ на всех μ -некогерентных матрицах ранга не выше r , и

$$\delta_{r,k} = \delta_r + \frac{C\beta k}{C_{RIP} r \log m}.$$

Доказательство. По Теореме 1.2.2, оператор \mathcal{A}_{Ω_0} с вероятностью не ниже $1 - \exp(-m \log(m))$ удовлетворяет условию ограниченной изометрии на всех μ -некогерентных матрицах ранга не выше k . Теперь рассмотрим произвольную μ -некогерентную $Y \in \mathbb{R}^{m \times m}$, $\text{rank}(Y) \leq r$. Отметим что $\|\mathcal{A}_{\Omega_0}(Y)\|_F^2 =$

$\|\mathcal{A}_{\Omega_k}(Y)\|_F^2 + \frac{\|P_{\Omega_0/\Omega_k}(Y)\|_F^2}{\rho}$, и $|\Omega_0/\Omega_k| \leq Ck\beta m$. Тогда,

$$\begin{aligned} \|\mathcal{A}_{\Omega_k}(Y)\|_F^2 &\leq \|\mathcal{A}_{\Omega_0}(Y)\|_F^2 \leq (1 + \delta_r)\|Y\|_F^2 \leq (1 + \delta_{r,k})\|Y\|_F^2; \\ \|\mathcal{A}_{\Omega_k}(Y)\|_F^2 &= \|\mathcal{A}_{\Omega_0}(Y)\|_F^2 - \frac{\|P_{\Omega_0/\Omega_k}(Y)\|_F^2}{\rho} \\ &\geq (1 - \delta_r)\|Y\|_F^2 - \frac{C\beta km\|Y\|_C^2}{\rho} \\ &\geq (1 - \delta_r)\|Y\|_F^2 - \frac{C\beta k\mu^2 r}{\rho m}\|Y\|_F^2 \\ &\geq (1 - \delta_r)\|Y\|_F^2 - \frac{C\beta k}{C_{RIPr} \log m}\|Y\|_F^2 \\ &= (1 - \delta_{r,k})\|Y\|_F^2. \end{aligned}$$

□

Предполагая геометрическую сходимость, $k = O(\log \varepsilon)$, где ε - желаемый параметр сходимости относительной ошибки, k невелико. При желании константу δ_k можно улучшить повышая плотность ρ выше теоретического минимума, как видно из доказательства.

2.4 Метод парного выполнения

Предложенный Алгоритм 3 имеет следующие заметные недостатки:

1. Число элементов разреженной части вывода Алгоритма 3 может быть в константу раз выше настоящего числа ненулевых элементов искомой разреженной матрицы, где константа соответствует числу итераций. Хотя и анализ сходимости алгоритма (Лемма 2.2.5), и численные результаты (см. следующий параграф) показывают геометрическую сходимость алгоритма, скорость сходимости алгоритма зависит от числа обусловленности задачи $\kappa := \frac{\sigma_1(X)}{\sigma_r(X)}$, где X - малоранговая часть неизвестного. Таким образом, плохая обусловленность задачи может привести к существенному росту числа итераций, и, соответственно, числу элементов разреженной части вывода алгоритма.

2. На практике, если $\kappa \gg 1$, на ранних (в частности, первой) итерациях k Алгоритма 3 возможно, что $\|S_k\|_2$ больше $\sigma_1(X)$ но меньше $\sigma_r(X)$ (что противоречит условию локальности Леммы 2.2.5 о сходимости). В таком случае, матрица $X + N_k + S_k$ может иметь набор сингулярных векторов, схожих с сингулярными векторами искомой X , не все из которых соответствуют первым r сингулярным числам. Матрица $W_{k+1} = \mathcal{P}_r(X + N_k + S_k) - X$ тогда часто теряет свойства некогерентности, потому что один из сингулярных векторов $\mathcal{P}_r(X + N_k + S_k)$ существенно определяется слагаемым S_k , и Алгоритм 3 может расходиться. Та же практическая проблема наблюдается для плохо обусловленных X даже в случае классического метода проекции градиента в отсутствие разреженных ошибок ([41]).

В качестве средства борьбы с описанными проблемами, предлагается идея „Парного восполнения“. Подход основан на следующих идеях:

- Если случайная разреженная маска Ω применяется к уже разреженной матрице \hat{S} , результат $\mathcal{A}_\Omega(\hat{S})$ может считаться матрицей малого ранга, потому что содержит очень малое число элементов.
- Если две скалированные проекции на две случайные разреженные маски $\mathcal{A}_{\Omega_1}, \mathcal{A}_{\Omega_2}$ применяются к уже разреженной матрице, результаты $\mathcal{A}_{\Omega_1}(\hat{S}), \mathcal{A}_{\Omega_2}(\hat{S})$ будут сильно разреженными матрицами малого ранга, и с большой вероятностью будут иметь взаимно ортогональные пространства строк и столбцов.
- Линейные подпространства, имеющие базис из разреженных векторов, почти ортогональны линейным подпространствам, имеющим базис из некогерентных векторов. Поэтому, сумма некогерентной матрицы и почти разреженной матрицы, скорее всего, имеет набор некогерентных сингулярных векторов и набор разреженных сингулярных векторов.
- Если параллельно выполнены два шага метода проекции градиента, соответствующие двум различным разреженным маскам, можно ожидать, что два результата малого ранга будут в своих подпространствах строк и столбцов иметь пары взаимно близких подпространств, схожих с базами неизвестного X , и пары взаимно ортогональных подпространств.

На основании этих рассуждений предлагается обобщить Алгоритм 3 таким образом, чтобы он использовал две маски. Пусть на каждой итерации одновременно

проводится два шага градиента вдоль двух различных разреженных масок, и пусть затем два результата проектируются на множество матриц ранга не выше $r+p$, где p - параметр, соответствующий оценке на число элементов пересечения разреженной матрицы и разреженной маски.

Алгоритм инициализируется посредством $k = 0, X_0 = 0, W_0 = X$, пустого множества Λ и двух масок $\Omega_{a,0}, \Omega_{b,0}$, выбранных случайно и независимо с одинаковой плотностью, равной $\rho \geq C_{RIP} \frac{\mu^2 k^2 \log(m)}{\delta_r^2 m}$, в соответствии с теоремой 1.2.2. Итерации алгоритма описаны в Алгоритме 4.

Предложенный Алгоритм 4 более не опирается напрямую на конструкцию $\mathcal{P}_r(X + N_k + S_k)$, включающую в себя информацию только о первых r сингулярных векторах, поэтому является более устойчивым в случае плохо обусловленных решений. Более того, на практике алгоритм теперь позволяет ослабить требования на норму $\|S_k\|_C$, которая контролируется шириной множества исключения Υ_k . Иными словами, исключения можно проводить реже, чем на каждой итерации (или исключать меньше элементов на каждой итерации), что приводит к меньшему размеру множества Λ по окончании работы алгоритма, что соответствует меньшему числу элементов в разреженной части итоговой аппроксимации. Теоретический анализ сходимости Алгоритма 4 при условии дополнительных предположений проведен и опубликован в [44] и будет опущен в этой работе в силу громоздкости; эффективность Алгоритма 4 будет продемонстрирована численными экспериментами.

2.5 Эксперименты на искусственных данных

Чтобы численно проверить построенную теорию сходимости и сравнить результаты работы предложенных одномасочного и двухмасочного Алгоритмов 3, 4, оба алгоритма были протестированы на матрицах искусственных данных, которые в точности удовлетворяли модели суммы матриц малого ранга и разреженной. Этим матрицы были построены следующим образом:

- Два базиса ортогональных столбцов $U, V \in \mathbb{R}^{m \times r}$ были выбраны как набор из r сингулярных векторов случайной матрицы, с независимым выбором каждого элемента по стандартному нормальному распределе-

Algorithm 4: Парное восполнение, двухмасочный алгоритм восполнения по входным данным с разреженной ошибкой

Data: Инициализируются $k = 0, X_0 = 0, W_0 = X$; случайно выбираются маски $\Omega_{a,0}, \Omega_{b,0}$.

- 1 **while** $\|(\mathcal{I} - \mathcal{P}_\Lambda)(Y - X)\|_F \geq \varepsilon$ **do**
- 2 Увеличивается номер итерации $k := k + 1$;
- 3 Находятся $C\beta m, C > 1$ наибольших по модулю элементов текущей невязки $W_{k-1} + \hat{S} = X + \hat{S} - X_{k-1}$. Множество соответствующих индексов обозначается Υ_k ;
- 4 $\Lambda := \Lambda \cup \Upsilon_k, \Omega_{a,k} := \Omega_{a,k-1}/\Upsilon_k, \Omega_{b,k} := \Omega_{b,k-1}/\Upsilon_k$;
- 5 Выполняется два шага метода проекции градиента вдоль двух выбранных масок:
- 6 $W_{k-1} := X - X_{k-1}$;
- 7 $N_{a,k} := (\mathcal{A}_{\Omega_{a,k}}^* \mathcal{A}_{\Omega_{a,k}} - \mathcal{I})(W_{k-1})$,
- 8 $N_{b,k} := (\mathcal{A}_{\Omega_{b,k}}^* \mathcal{A}_{\Omega_{b,k}} - \mathcal{I})(W_{k-1})$;
- 9 $S_{a,k} := \mathcal{A}_{\Omega_{a,k}}^* \mathcal{A}_{\Omega_{a,k}} \hat{S}$;
- 10 $S_{b,k} := \mathcal{A}_{\Omega_{b,k}}^* \mathcal{A}_{\Omega_{b,k}} \hat{S}$;
- 11 $X_{a,k} = \mathcal{P}_p(X + N_{a,k} + S_{a,k})$
- 12 $= U_{a,k} \Sigma_{a,k} V_{a,k}^*, U_{a,k}, V_{a,k} \in \mathbb{R}^{m \times p}$,
- 13 $X_{b,k} = \mathcal{P}_p(X + N_{b,k} + S_{b,k})$
- 14 $= U_{b,k} \Sigma_{b,k} V_{b,k}^*, U_{b,k}, V_{b,k} \in \mathbb{R}^{m \times p}$;
- 15 Вычисляются r первых канонических углов и векторов между парами подпространств, задаваемыми базисами $\{U_{a,k}, U_{b,k}\}$ и $\{V_{a,k}, V_{b,k}\}$:
- 16 $\mathcal{P}_r(U_{b,k}^* U_{a,k}) = U_{col} \Sigma_{col} V_{col}^*; U_{col}, V_{col} \in \mathbb{R}^{p \times r}$;
- 17 $\mathcal{P}_r(V_{b,k}^* V_{a,k}) = U_{row} \Sigma_{row} V_{row}^*; U_{row}, V_{row} \in \mathbb{R}^{p \times r}$;
- 18 $\bar{U}_{a,k} := U_{a,k} V_{col}, \bar{V}_{a,k} = V_{a,k} V_{row}$;
- 19 $\bar{U}_{b,k} := U_{b,k} U_{col}, \bar{V}_{b,k} = V_{a,k} U_{row}$;
- 20 Результаты двух шагов метода проекции градиента далее проектируются на подпространства, соответствующие первым r канонических векторов. Оптимальное приближение ранга r для полусуммы результатов далее выбирается в качестве X_k .
- 21 $X_k = \mathcal{P}_r \left(\frac{P_{\bar{U}_{a,k}} X_{a,k} P_{\bar{V}_{a,k}} + P_{\bar{U}_{b,k}} X_{b,k} P_{\bar{V}_{b,k}}}{2} \right)$;

нию. Эмпирически такие базисы имеют хорошую характеристику некогерентности, логарифмически зависящую от размеров матриц, поэтому хорошо подходят для экспериментов с восполнением.

– Малоранговые части решений X формировались как произведения $U\Sigma V^*$, где Σ - диагональная вещественная матрица, определяющая по-

ведение сингулярных чисел решений X по одному из следующих законов:

- $\sigma_j = \frac{1}{j}$, что моделирует хорошо обусловленную задачу;
 - $\sigma_j = \frac{1}{3^j}$, что моделирует плохо обусловленную задачу.
- Разреженная ошибка \hat{S} также выбиралась случайно, таким образом что каждая позиция матрицы имела равную вероятность $\frac{\beta}{n}$ содержать ненулевой элемент; соответствующие ненулевые элементы также выбирались случайно по нормальному распределению со стандартным отклонением выбранным таким образом, что $\frac{\|\hat{S}\|_F}{\|X\|_F} \approx 0.3$.

Для дополнительной устойчивости, эвристические техники, описанные в Главе 1 и [41], такие как постепенный набор ранга и контроль шага градиента применялись как в одномасочном, так и в двухмасочном алгоритме. Это значит, что приближение строилось начиная с однорангового X_0 , и ранг приближения X_k постепенно поднимается вплоть до r с номером итерации k . Кроме того, использовалось и отслеживание падения невязки: например, если относительное падение невязки $\frac{\|W_k + S_k\|_F}{\|W_{k-1} + S_{k-1}\|_F} \leq \lambda$, принимается решение о либо увеличении ранга, либо расширении множества Λ посредством непустого Υ_k . Двухмасочный алгоритм позволяет удобный механизм принятия такого решения: если первый „хвостовой“ канонический угол ϕ_{r+1} меньше наперед заданно константы (в экспериментах использовалось $\frac{\pi}{4}$), тогда повышается ранг; иначе классическим образом множество наибольших элементов невязки Υ_k добавляется во множество разреженного исключения Λ .

С использованием таких эвристических процедур рассмотренные алгоритмы являются более гибкими с точки зрения выбора параметра C , определяющего размер множества Υ_k индексов исключенных элементов, добавляемого в Λ одновременно. В то время как теоретический анализ предполагает $C > 1$, на практике можно использовать меньшие значения вида 0.2 – 0.5 без каких-либо потерь (однако, по определению C , множество Λ должно быть расширено хотя бы на $O(\frac{1}{C})$ итерациях, поэтому слишком малые значения C недопустимы).

Так как одномасочный алгоритм не может опираться на информацию о канонических углах, доступную двухмасочному алгоритму, некогерентность текущего приближения X_k использовалась как критерий принятия решений о повышении ранга или расширении Λ в случае использования одной маски.

Ниже детально описаны параметры и результаты разнообразных экспериментов.

– *Обычный сценарий.*

$$m = 1024, r = 10, \sigma_k = \frac{1}{1+k}, p = 3, \rho = 0.25, \beta = 10.0, C = 1.$$

Малоранговая часть имеет ранг 10, разреженная часть имеет 10 элементов на каждой строке, и коэффициент расширения шаблона исключенных элементов C равен единице, что означает что число элементов разреженного неизвестного \hat{S} один к одному соответствует размеру множества Υ_k , добавляемого к множеству исключенных элементов Λ одновременно. Результаты приведены на рис. 2.1. На графике приведены невязка и синусы возникающих канонических углов: видно, что канонические углы с номерами от 1 до r сходятся с той же скоростью, с которой и невязка, а последующий угол ϕ_{r+1} остается большим на протяжении всех итераций, как и предполагает идея алгоритма. Острые пики в начале графиков величин $\sin(\phi_r(U_{a,k}, U_{b,k}))$ обусловлены эвристическими процедурами увеличения ранга (значение ранга r инициализируется единицей и постепенно повышается до правильной величины).

Свойства сходимости одномасочного и двухмасочного Алгоритмов 3, 4 в этом сценарии неразличимы. В этом и всех последующих экспериментах канонические углы между парами базисов столбцов и парами базисов строк ведут себя одинаково.

– *Сглаженный сценарий.*

$$m = 1024, r = 10, \sigma_k = \frac{1}{1+k}, p = 3, \rho = 0.25, \beta = 10.0, C = 0.01.$$

График невязки предыдущего сценария является кусочно-гладким, а интервалы гладкости соответствуют наборам итераций, на протяжении которых множество исключения Λ константно - при каждом проведении исключения, невязка резко падает небольшое число итераций и затем фиксируется на новом меньшем значении.

Такое поведение можно контролировать с помощью параметра C , определяющим число позиций, добавляемых в множество Λ одновременно. На рис. 2.2, результаты приведены для $C = 0.01$: график невязки является гораздо более гладким, но общая скорость сходимости при этом снижается.

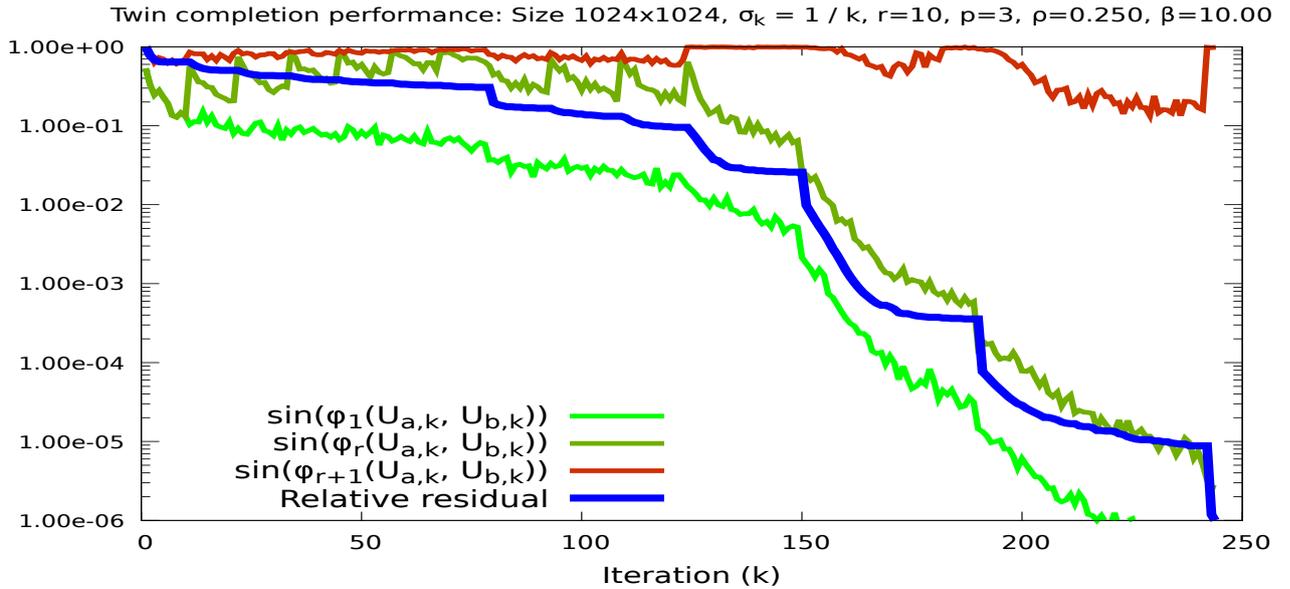


Рисунок 2.1 — Поведение методов проекции градиента для восполнения матрицы с разреженной ошибкой данных в обычном сценарии.

– *Сценарий плохой обусловленности.*

$$m = 1024, r = 10, \sigma_k = \frac{1}{3^k}, p = 3, \rho = 0.25, \beta = 10.0, C = 1.$$

В этом сценарии, неизвестная матрица построены таким образом, что ее сингулярные числа быстро убывают, что вызывает расхождения в поведении одномасочного и двухмасочного алгоритмов. В этом случае, факторы строк и столбцов итерационных приближений одномасочного Алгоритма 3 как правило теряют свойства некогерентности, что приводит к потере сходимости: на каждой итерации проводится пополнение множества исключения Λ пока количество оставшихся элементов не станет слишком малым для возможности восполнения. Соответствующие графики представлены на рис. 2.3, где некогерентность факторов измеряется в процентах по отношению к максимально допустимому определением значению. График „Matrix locked“ показывает размер Λ по отношению к числу m^2 всех элементов матрицы. Двухмасочный алгоритм 4 при этом поддерживает неразреженность итерационных приближений и сохраняет сходимость. Соответствующие графики приведены на рис. 2.4.

– *Сценарий большого ранга.*

$$m = 1024, r = 30, \sigma_k = \frac{1}{1+k}, p = 7, \rho = 0.25, \beta = 10.0, C = 1.$$

В случае хорошо обусловленной матрицы большого ранга сходимость

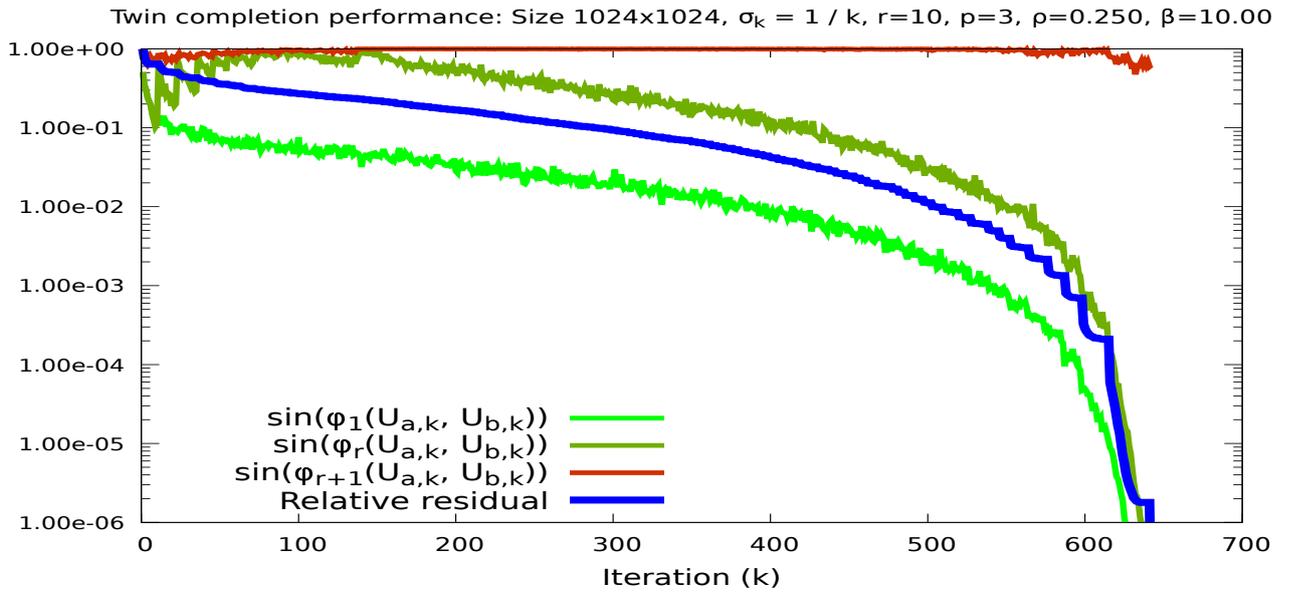


Рисунок 2.2 — Поведение методов проекции градиента для восполнения матрицы с разреженной ошибкой данных в сглаженном сценарии.

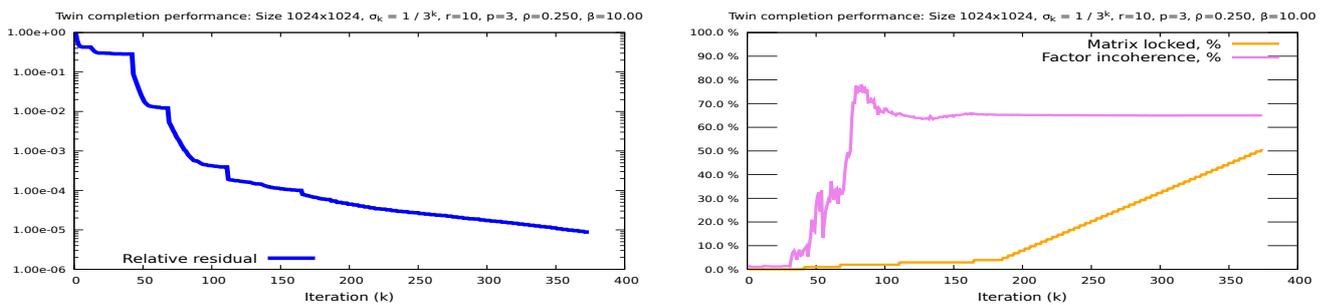


Рисунок 2.3 — Поведение одномасочного варианта метода проекции градиента для восполнения матрицы с разреженной ошибкой данных в сценарии плохой обусловленности.

и одномасочного, и двухмасочного вариантов алгоритма стабильна, но процедуры поэтапного ранга могут занимать существенную часть итераций. Соответствующий график представлен на рис. 2.5.

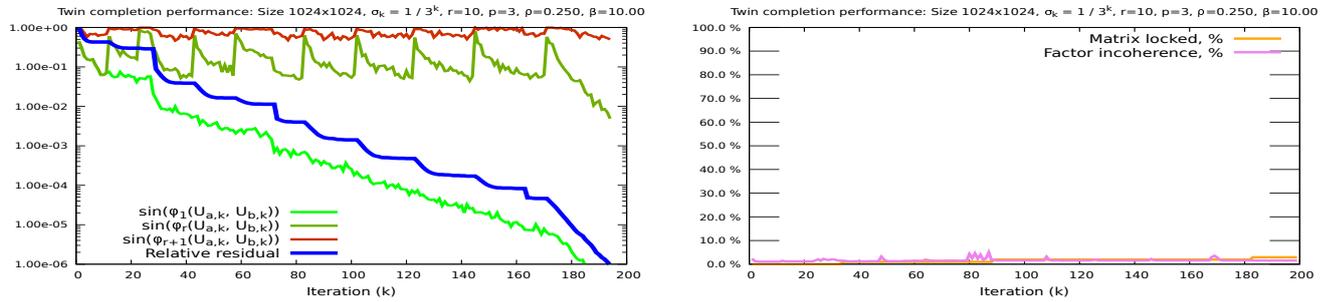


Рисунок 2.4 — Поведение двухмасочного варианта метода проекции градиента для восполнения матрицы с разреженной ошибкой данных в сценарии плохой обусловленности.

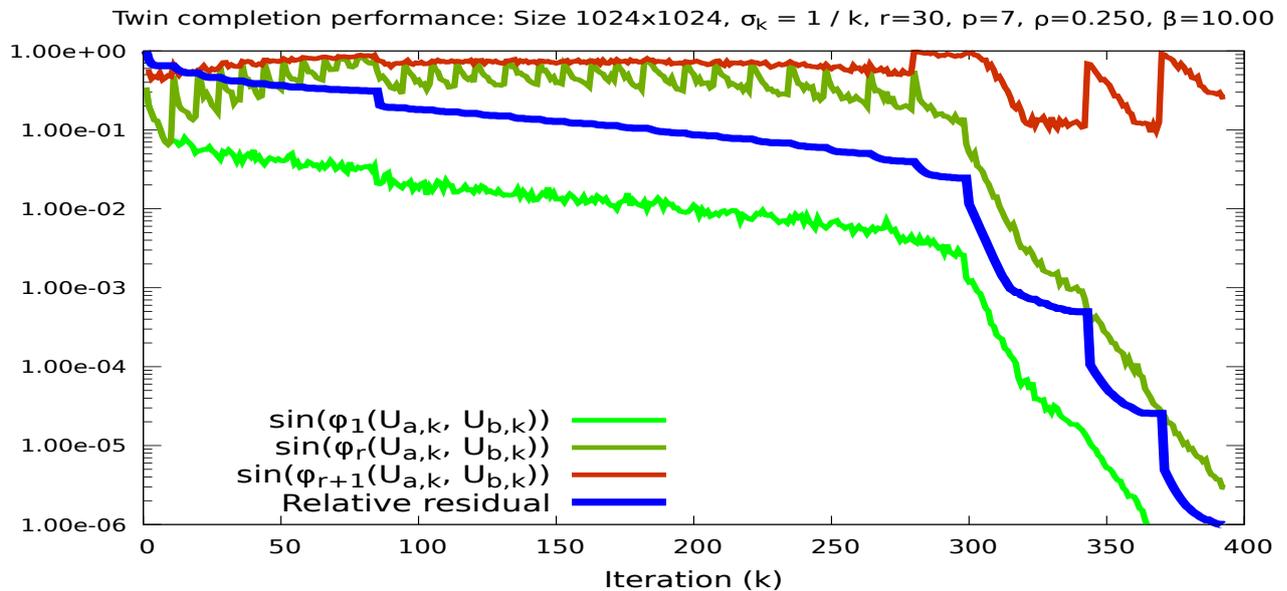


Рисунок 2.5 — Поведение методов проекции градиента для восполнения матрицы с разреженной ошибкой данных в сценарии большого ранга.

Глава 3. Тензорные приближения малого ранга в условиях зашумленных данных

В этом разделе будет рассмотрена проблема устойчивости тензорных разложений разных видов к случайному 'белому' (поэлементно-гауссовому) шуму. Общая цель исследования этой задачи - численно и теоретически проверить/подтвердить гипотезу следующего вида: тензорные приближения малого ранга гораздо более устойчивы к возмущениям вида белого шума, чем матричные, при одинаковом числе элементов у матрицы и тензора. При этом под „устойчивостью“ в этом контексте будем понимать близость малорангового приближения наперед заданного тензора, подверженного возмущению, к исходному тензору. Близость при этом будем характеризовать корнем из суммы квадратов модулей поэлементных отклонений; другие метрики в этой главе не рассматриваются.

Понятие „тензорное приближение малого ранга“ допускает несколько трактовок: в этой работе будут рассмотрены канонический формат разложения тензора, формат Таккера и формат тензорного поезда ('Tensor-Train', ТТ).

В рамках исследования необходимо принять во внимание, что тензорные разложения малого ранга обладают качественными отличиями от матричных: для наперед заданного тензора может не существовать 'наилучшего' приближения тензором наперед заданного ранга. В частности, если речь идет о каноническом формате тензора, то множество тензоров канонического ранга, не превосходящего наперед заданной константы r , может быть незамкнуто в случае $r > 1$. Даже если *оптимальное* приближение тензором малого ранга и определенного формата существует, его поиск часто нереализуем на практике в силу высокой вычислительной сложности. Более того, даже алгоритмы поиска неоптимальных приближений имеют ограниченную теорию сходимости - так, наиболее распространенный алгоритм построения тензорных приближений в каноническом формате ALS ('Alternating Least Squares', [23]) может сходиться к локальным минимумам, далеким от оптимальных по значению невязки. Теоретические оценки ошибок, полученных по алгоритмам HOSVD [21] для построения приближений в формате Таккера и TT-SVD [22] для построения приближений в формате тензорного поезда существуют, но оба алгоритма так-

же не возвращают в качестве результата *оптимальное* приближение заданного формата.

Будем использовать математический аппарат, связанный с понятием 'Гауссовой ширины' [45] множества, не накладывающий практически никаких условий на само множество. Будет показано, что теоретическая или практическая невозможность построения *оптимального* тензорного приближения заданного формата не является помехой для построения оценок на невязку приближения возмущенного тензора, имеющего точный малый ранг. Приведенные оценки будут подкреплены численными экспериментами на искусственных данных, однако потребуют введения предположения на качество 'приближения', и будут опираться на гипотезу, связанную с *оптимальным* приближением тензора белого шума посредством тензора малого канонического ранга.

3.1 Обозначения тензорных операций

Введем несколько обозначений, которые понадобятся для работы с тензорами в этой и последующей главах. Пусть $\mathbf{T} \in \mathbb{R}^{m_1 \times m_2 \times \dots \times m_d}$ - d -мерный тензор, а m_1, m_2, \dots, m_d - его линейные размеры. Многомерные тензоры будем обозначать жирными буквами. Будем обозначать за $T_k \in \mathbb{R}^{m_k \times (m_1 \dots m_{k-1} m_{k+1} \dots m_d)}$ матрицу развертки тензора \mathbf{T} по размерности k . Введем обозначение для произведения тензора по направлению k с матрицей:

$$(\mathbf{T} \times_k Q)_{j_1 \dots j_d} = \sum_l Q_{j_k l} \mathbf{T}_{j_1 \dots j_{k-1} l j_{k+1} \dots j_d}.$$

Для симметризации выкладок будем иногда рассматривать число как тензор произвольной размерности со всеми линейными размерами, равными 1. Тогда, можно, например, использовать выражение вида

$$\begin{aligned} \mathbf{T} &= \mathbf{1} \times_1 \vec{x} \times_2 \vec{y} \times_3 \vec{z}, \\ \vec{x} &\in \mathbb{R}^{m_1}, \vec{y} \in \mathbb{R}^{m_2}, \vec{z} \in \mathbb{R}^{m_3}, \mathbf{T} \in \mathbb{R}^{m_1 \times m_2 \times m_3}. \end{aligned} \quad (3.1)$$

Выражение (3.1) определяет в частном трехмерном случае тензор с разделяющимися переменными, иначе называемый тензором ранга 1. Тензором канонического ранга r называется тензор, имеющий представление вида

$$\mathbf{T} = \sum_{i=1}^r \mathbf{T}_i,$$

где каждый из тензоров T_i допускает разделение переменных, или, иными словами, имеет ранг 1. Говорят, что тензор имеет разложение Таккера с рангами $r_1 \dots r_d$, если для некоторых \mathbf{C}, U_k выполнено

$$\begin{aligned} \mathbf{T} &= \mathbf{C} \times_1 U_1 \times_2 U_2 \cdots \times_d U_d, \\ \mathbf{C} &\in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_d}, U_k \in \mathbb{R}^{m_k \times r_k}, k = 1 \dots d. \end{aligned}$$

При этом тензор \mathbf{C} называют ядром Таккера, а матрицы U_k называют факторами Таккера; не ограничивая общности, можно считать, что они имеют ортогональные системы столбцов. Говорят, что тензор имеет разложение в формате тензорного поезда [22] с рангами $r_1 \dots r_{d-1}$, если для некоторых 'вагонов' \mathbf{W}_k выполнено

$$\mathbf{T}_{j_1 j_2 \dots j_d} = \sum_{r_1, r_2, \dots, r_{d-1}=1}^{r_1, r_2, \dots, r_{d-1}} W_{j_1 r_1}^{(1)} W_{r_1 j_2 r_2}^{(2)} \cdots W_{r_{d-2} j_{d-1} r_{d-1}}^{(d-1)} W_{r_{d-1} j_d}^{(d)}, j_k = 1 \dots m_k, k = 1 \dots d,$$

где первый и последний 'вагоны' $W^{(1)}, W^{(d)}$ имеют вид матриц, а остальные являются трехмерными тензорами.

Канонический формат, формат Таккера и формат тензорного поезда совпадают при рангах, равных единице.

3.2 Аппарат гауссовой ширины

3.2.1 Общие рассуждения

Рассмотрим множество $K \subset \mathbb{R}^n$, удовлетворяющее 'условию конуса'

$$\forall \vec{x} \in K, \alpha \in \mathbb{R} \Rightarrow \alpha \vec{x} \in K.$$

В частности, под этим множеством можно подразумевать множество матриц ранга не выше r , множество тензоров канонического ранга не выше r , или множество тензоров, все ранги формата Таккера или формата тензорного поезда которых не превосходят r , где r - константа.

Пусть $\mathcal{P}_K(\vec{x}) \in K$ обозначает результат работы некоего алгоритма приближения вектора \vec{x} вектором из множества K . Например, в случае, если K - множество матриц ранга не выше 2, $\mathcal{P}_K(\vec{x})$ может обозначать наилучшее приближение матрицы \vec{x} среди всех матриц ранга 2, построенное при помощи сингулярного разложения. Или, если K - множество тензоров канонического ранга не выше 2, $\mathcal{P}_K(\vec{x})$ может обозначать результат работы алгоритма ALS с рангом 2, примененного к тензору \vec{x} (который, вообще говоря, не является оптимальным приближением).

Рассмотрим произвольный вектор $\vec{x} \in K$, и некоторое возмущение \vec{n} , такое что, вообще говоря $\vec{z} = \vec{x} + \vec{n} \notin K$. Исследуем отклонение $\mathcal{P}_K(\vec{z}) = \mathcal{P}_K(\vec{x} + \vec{n})$ от исходного невозмущенного \vec{x} . Для этого потребуется ввести какое-то требование к алгоритму \mathcal{P}_K , которым будет математически закреплен смысл названия „алгоритм приближения“. Будем опираться на следующее предположение:

$$\|\mathcal{P}_K(\vec{z}) - \vec{z}\| \leq \|\vec{x} - \vec{z}\|. \quad (3.2)$$

Иными словами, $\mathcal{P}_K(\vec{z})$ приближает возмущенный вектор \vec{z} не хуже, чем исходный чистый вектор \vec{x} . Это предположение заведомо выполнено всегда, если под \mathcal{P}_K понимается *оптимальное* приближение входного вектора из множества K , но оптимальность \mathcal{P}_K не подразумевается. Предположение (3.2) постулирует что $\mathcal{P}_K(\vec{z})$ дает невязку с \vec{z} лучшую, чем один конкретный вектор \vec{x} , а не чем

всевозможные векторы из множества K . Проведем следующие преобразования:

$$\begin{aligned} \|\mathcal{P}_K(\vec{z}) - \vec{z}\|_2^2 &\leq \|\vec{x} - \vec{z}\|_2^2, \\ \|(\mathcal{P}_K(\vec{z}) - \vec{x}) + (\vec{x} - \vec{z})\|_2^2 &\leq \|\vec{x} - \vec{z}\|_2^2, \\ \|\mathcal{P}_K(\vec{z}) - \vec{x}\|_2^2 &\leq 2(\mathcal{P}_K(\vec{z}) - \vec{x}, \vec{z} - \vec{x})_2, \\ \|\mathcal{P}_K(\vec{z}) - \vec{x}\|_2 &\leq 2\left(\frac{\mathcal{P}_K(\vec{z}) - \vec{x}}{\|\mathcal{P}_K(\vec{z}) - \vec{x}\|_2}, \vec{z} - \vec{x}\right)_2. \end{aligned} \quad (3.3)$$

Имея ввиду, что $\frac{\mathcal{P}_K(\vec{z}) - \vec{x}}{\|\mathcal{P}_K(\vec{z}) - \vec{x}\|_2} \in K - K$ - вектор единичной длины, можно провести оценку вида

$$\|\mathcal{P}_K(\vec{z}) - \vec{x}\|_2 \leq 2\|\vec{z} - \vec{x}\|_{2,K},$$

где величина $\|\cdot\|_{2,K}$ задается формулой

$$\|\vec{y}\|_{2,K} := \sup_{\vec{v} \in K-K, \|\vec{v}\|_2=1} (\vec{y}, \vec{v})_2.$$

Обоснуем введенное обозначение $\|\cdot\|_{2,K}$. Так как точная верхняя грань суммы не превосходит суммы точных верхних граней слагаемых, $\|\cdot\|_{2,K}$ удовлетворяет неравенству треугольника. Очевидно, что $\|\cdot\|_{2,K}$ также допускает вынос скалярного множителя, однако можно привести искусственный пример множества K , для которого существует $x \in K \neq 0$, $\|x\|_{2,K} = 0$. В частности, это выполнено для $K \subset \mathbb{R}^n$, являющимся линейным попространством, для вектора $\vec{x} \perp K$.

Тем не менее, для интересующих нас в этой работе способов задания множества K как множеств матриц и тензоров малого ранга, $\|\cdot\|_{2,K}$ является полноценной нормой. Во-первых, множество $K - K$ в этом случае несет такой же смысл, как и само K , но с удвоением верхней оценки на ранг. Во-вторых, так как любой тензор, содержащий только один ненулевой элемент, имеет ранг 1, то $\|\vec{x}\|_{2,K} = 0$ означает что скалярные произведения \vec{x} со всеми такими тензорами равны нулю, откуда каждый элемент \vec{x} равен нулю.

Отметим, что всегда верно $\|\vec{n}\|_{2,K} \leq \|\vec{n}\|_2$, так как выполнено $\|\vec{n}\|_2 = \sup_{\|\vec{v}\|_2=1} (\vec{n}, \vec{v})$, где точная верхняя грань берется по более широкому множеству, чем в определении $\|\vec{n}\|_{2,K}$. Если возмущение \vec{n} является случайным и поэлементно гауссово распределено, матожидание $\mathbb{E}\|\vec{n}\|_{2,K}$ обычно связывают с понятием *Гауссовой ширины* множества K [45]. В дальнейшем будет показано, что для рассматриваемых K , соответствующих множествам матриц и тензоров мало-

го ранга, и для случайного возмущения \vec{n} с большой вероятностью выполнено $\|\vec{n}\|_{2,K} \ll \|\vec{n}\|_2$. Поясним полученные неравенства

$$\|\vec{z} - \vec{x}\|_2 \leq \|\vec{n}\|_2, \quad (3.4)$$

$$\|\mathcal{P}_K(\vec{z}) - \vec{x}\|_2 \leq 2\|\vec{n}\|_{2,K}. \quad (3.5)$$

схемой, приведенной на рисунке 3.1, где облако символизирует множество K . Векторы \vec{y} и $\mathcal{P}_K(\vec{x} + \vec{n})$ принадлежат множеству K , и расстояние между ними оценивается сравнительно малой величиной $\|\vec{n}\|_{2,K}$, в то время как расстояние между \vec{x} и $\vec{x} + \vec{n}$ характеризуется сравнительно большой величиной $\|\vec{n}\|_2$. Исходное предположение (3.2) о работе алгоритма \mathcal{P}_K соответствует тому, что синяя стрелка на схеме короче красной. При этом достаточно выполнения условия (3.2) на конкретной паре векторов \vec{x}, \vec{n} .

Наконец отметим, что в случае задания множества K как множества матриц или тензоров малого ранга вторая норма векторов \vec{x}, \vec{n} из таким образом заданного множества K соответствует норме Фробениуса соответствующих тензоров(матриц). В дальнейшем будем предполагать прямое соответствие между обозначениями тензоров(матриц) и векторов: будет подразумеваться, что вектора, обозначенные маленькими буквами, соответствуют векторизациям тензоров(матриц), обозначенных большими версиями тех же букв, например $\vec{n} = \text{vec}(\mathbf{N})$, $\|\vec{n}\|_2 = \|\mathbf{N}\|_F$. Таким образом, для анализа устойчивости d -мерных тензорных приближений далее будем оценивать величину

$$\sup_{\text{rank}(\mathbf{V}) \leq r, \|\mathbf{V}\|_F = 1} (\mathbf{N}, \mathbf{V})_F, \quad (3.6)$$

где $\mathbf{N}, \mathbf{V} \in \mathbb{R}^{m_1 \times m_2 \times \dots \times m_d}$, а ранг понимается в матричном смысле для $d = 2$, и может пониматься в различных смыслах в случае тензоров, в частности - в каноническом.

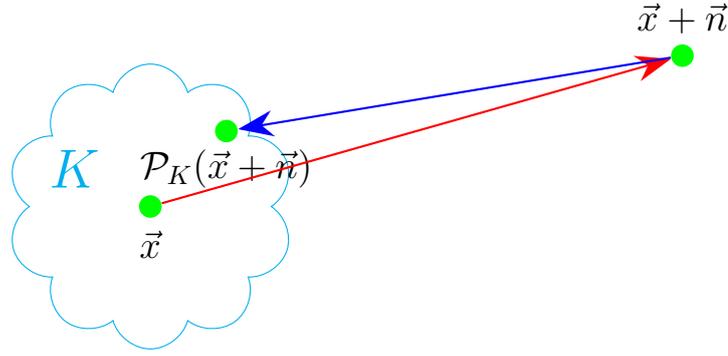


Рисунок 3.1 — Схема анализа устойчивости тензорных приближений

3.2.2 Аналитический вид введенной нормы для множества матриц ограниченного ранга

В качестве примера, рассмотрим K , заданное как множество квадратных матриц ранга не выше r . Тогда будем иметь

$$\begin{aligned}
\|N\|_{2,K} &= \sup_{A \in K-K, \|A\|_F=1} (N, A)_F = \sup_{\text{rank}(A) \leq 2r, \|A\|_F=1} (N, A)_F \\
&= \sup_{U_A^* U_A = V_A^* V_A = I, \|\Sigma_A\|_F=1, \text{rank}(\Sigma_A) \leq 2r} (U_N \Sigma_N V_N^*, U_A \Sigma_A V_A^*)_F \\
&= \sup_{U_A^* U_A = V_A^* V_A = I, \|\Sigma_A\|_F=1, \text{rank}(\Sigma_A) \leq 2r} ((U_A^* U_N) \Sigma_N (V_A^* V_N), \Sigma_A)_F \\
&\leq \sup_{\|\Sigma_A\|_F=1, \text{rank}(\Sigma_A) \leq 2r} (\Sigma_N, \Sigma_A)_F \leq \sqrt{2r} \|N\|_2,
\end{aligned}$$

где $U_A, \Sigma_A, V_A^* \in \mathbb{R}^{m \times m}$ соответствуют полному сингулярному разложению A , и использовано $|(U_A^* U_N)_{ij}| \leq 1$.

Сингулярные числа поэлементно гауссовых матриц N убывают медленно: $\|N\|_2 \approx 2\sqrt{m}$, $\|N\|_F \approx m$ в случае $(N)_{ij}$, распределенных по стандартному гауссовому распределению [46]. Таким образом, в случае белого шума N имеем $\|\vec{n}\|_{2,k} \leq \sqrt{2r} \|N\|_2 \ll \|N\|_F = \|\vec{n}\|_2$. Можно видеть, что даже в случае матриц $X, N \in \mathbb{R}^{m \times m}$ величина $\|\mathcal{P}_r(X+N) - N\|_F$ может быть значительно ниже $\|N\|_F$ и определяться рангом r и величиной $\|N\|_2$.

Для матриц этот результат уже известен; более того, улучшенная версия такого результата уже использовалась при анализе возмущений сингулярных подпространств ранее в Главе 2. В частности, согласно Леммам 2.1.1 ([39]), 2.1.2

в случае матриц величина $\|\mathcal{P}_R(X + N) - N\|_F$ может быть оценена с помощью величин $\|P_{U_Y}N\|_F, \|NP_{V_Y}\|_F$, где $U_Y, V_Y \in \mathbb{R}^{m \times r}$ - сокращенные сингулярные базисы Y . Иными словами, оценка вида $\sqrt{2r}\|N\|_2$ может быть улучшена для случая, когда возмущение N почти или полностью ортогонально сингулярным пространствам X . Тем не менее, анализ общей конструкции $\|\vec{n}\|_{2,K}$ для случая тензоров, описанный в последующих параграфах, позволит получить оценки, представляющие новизну.

3.2.3 Связь введенной величины с задачей приближения тензора шума тензором малого ранга

Покажем, что рассматриваемая величина $\sup_{rank(\mathbf{V}) \leq r, \|\mathbf{V}\|_F = 1} (\mathbf{N}, \mathbf{V})_F$ имеет связь с задачей приближения вида

$$\|\mathbf{N} - \mathbf{T}\|_F \rightarrow \inf, rank(\mathbf{T}) \leq r. \quad (3.7)$$

Представим неизвестное \mathbf{T} в виде произведения тензора единичной нормы $\mathbf{V} = \frac{\mathbf{T}}{\|\mathbf{T}\|_F}$ и неотрицательного вещественного числа $\alpha = \|\mathbf{T}\|_F$. Имеем

$$\begin{aligned} \|\mathbf{N} - \alpha\mathbf{V}\|_F^2 &= \|\mathbf{N}\|_F^2 - 2(\mathbf{N}, \alpha\mathbf{V})_F + \|\alpha\mathbf{V}\|_F^2 \\ &= \|\mathbf{N}\|_F^2 - 2\alpha(\mathbf{N}, \mathbf{V})_F + \alpha^2 \\ &= \|\mathbf{N}\|_F^2 - (\mathbf{N}, \mathbf{V})_F^2 + ((\mathbf{N}, \mathbf{V})_F - \alpha)^2. \end{aligned}$$

Слагаемое $(\mathbf{N}, \mathbf{V})_F^2$ не меняется при изменении знака \mathbf{V} , а слагаемое $((\mathbf{N}, \mathbf{V})_F - \alpha)^2 \geq 0$ минимизируется при $(\mathbf{N}, \mathbf{V})_F \geq 0, \alpha = (\mathbf{N}, \mathbf{V})_F$. Откуда имеем

$$\inf_{rank(\mathbf{T}) \leq R} \|\mathbf{N} - \mathbf{T}\|_F^2 = \|\mathbf{N}\|_F^2 - \sup_{rank(\mathbf{V}) \leq R, \|\mathbf{V}\|_F = 1} (\mathbf{N}, \mathbf{V})_F^2,$$

то есть оптимальные величины из уравнений (3.7), (3.6) имеют прямую аналитическую связь. Эта связь и вывод формул (3.3) не претерпевает изменений и в случае комплексных тензоров, кроме замены скалярного произведения на его вещественную часть для корректности использования супремума и раскрытия

квадрата разности. Задача оптимизации (3.7) не имеет практического применения: так как матрица из случайных гауссовых элементов не может быть „хорошо“ приближена матрицей малого ранга в силу существенного различия ее спектральной и фробениусовой нормы [46], то и тензор с тем же количеством элементов по норме Фробениуса приближается тензором малого канонического ранга не лучше. Однако, согласно (3.5), задача имеет (3.7) теоретическое применение: чем хуже тензор белого шума приближается тензором канонического ранга не выше константы, тем лучше характеристика точности приближения $\|\mathcal{P}_K(\vec{x} + \vec{n}) - \vec{x}\|_2$ в случае множества K , соответствующего множеству тензоров малого канонического ранга. В самом деле, дальнейший анализ величины (3.6) покажет, что тензором ранга 1 набор случайных гауссово-распределенных величин приближается с еще большей погрешностью, чем матрицей ранга 1, при одинаковом числе элементов у тензора и матрицы.

3.3 Гауссова ширина множества тензоров ранга 1 ограниченной нормы

Проведенные ранее рассуждения показывают, что устойчивость канонических тензорных приближений малого ранга к случайному гауссовому возмущению можно охарактеризовать и оценить величиной

$$\|\vec{n}\|_{2,K} = \sup_{\text{rank}(\vec{v}) \leq r, \|\vec{v}\|_2 = 1} (\vec{n}, \vec{v}), \quad (3.8)$$

при различных r (в частности, вдвое больших, чем ранг элементов K).

Предлагается сначала рассмотреть соответствующую точную грань (3.8) по множеству тензоров единичного ранга $\text{rank}(\vec{v}) \leq 1$, $\|\vec{v}\|_2 = 1$. Для удобства записи будем рассматривать трехмерный тензор размера $m_1 \times m_2 \times m_3$. Будем считать, что каждый элемент \vec{n} распределен независимо по нормальному распределению с нулевым матожиданием. Дисперсия распределения может быть различна, но так как $\|\vec{n}\|_{2,K}$ допускает вынесение скалярного множителя, не ограничивая общности можно считать, что дисперсия каждого элемента \vec{n} рав-

на единице. Рассмотрим

$$\|\vec{n}\|_{2,K} = \sup_{\|\vec{a}\|_2=\|\vec{b}\|_2=\|\vec{c}\|_2=1} (\mathbf{N}, \mathbf{1} \times_1 \vec{a}^T \times_2 \vec{b}^T \times_3 \vec{c}^T),$$

предлагается использовать соотношение

$$\begin{aligned} (\mathbf{N}, \mathbf{1} \times_1 \vec{a}^T \times_2 \vec{b}^T \times_3 \vec{c}^T) &= (\mathbf{N}, \mathbf{1} \times_1 \vec{\hat{a}}^T \times_2 \vec{\hat{b}}^T \times_3 \vec{\hat{c}}^T) \\ &+ \|a - \hat{a}\|_2 (\mathbf{N}, \mathbf{1} \times_1 \frac{a^T - \hat{a}^T}{\|a - \hat{a}\|_2} \times_2 \vec{b}^T \times_3 \vec{c}^T) \\ &+ \|b - \hat{b}\|_2 (\mathbf{N}, \mathbf{1} \times_1 \vec{\hat{a}}^T \times_2 \frac{b^T - \hat{b}^T}{\|b - \hat{b}\|_2} \times_3 \vec{c}^T) \\ &+ \|c - \hat{c}\|_2 (\mathbf{N}, \mathbf{1} \times_1 \vec{\hat{a}}^T \times_2 \vec{\hat{b}}^T \times_3 \frac{c^T - \hat{c}^T}{\|c - \hat{c}\|_2}), \end{aligned} \quad (3.9)$$

выполненное для произвольных $\hat{a}, \hat{b}, \hat{c}$. Сведем точную верхнюю грань (3.8) к максимуму по конечному множеству. Будем использовать так называемые ϵ -покрытия единичной сферы. Под ϵ -покрытием единичной сферы в \mathbb{R}^m будем понимать конечное множество векторов \mathcal{N}_ϵ^m , такое что

$$\forall \vec{y} \in \mathbb{R}^m, \|\vec{y}\|_2 = 1 \rightarrow \exists \vec{y}_i \in \mathcal{N}_\epsilon^m : \|\vec{y} - \vec{y}_i\|_2 \leq \epsilon.$$

В литературе (например, в [47], главе 4.2) построены оценки на зависимость достаточного числа элементов в множестве \mathcal{N}_ϵ^m в зависимости от m и ϵ . Для этого рассматривается максимальное количество векторов из единичной сферы, такое что расстояние между любой парой таких векторов не меньше ϵ ; благодаря компактности единичной сферы это число конечно. Так как увеличить это число нельзя, соответствующий набор векторов и представляет собой ϵ -покрытие; вместе с этим, по неравенству треугольника набор шаров с центрами в выбранных элементах покрытия и радиусом $\frac{\epsilon}{2}$ не имеет попарных пересечений. Используя то, что объем единичного шара вычисляется как $C(m)r^m$, где m - размерность, и то, что все шары радиуса $\frac{\epsilon}{2}$ с выбранным набором центров полностью лежат в шаре с центром в нуле радиуса $1 + \frac{\epsilon}{2}$, можно оценить число элементов в покрытии как

$$|\mathcal{N}_\epsilon^m| \leq \frac{(1 + \frac{\epsilon}{2})^m}{(\frac{\epsilon}{2})^m} = (1 + \frac{2}{\epsilon}) \leq (\frac{3}{\epsilon})^m.$$

Возвращаясь к оценке (3.9), и предполагая, что построены конечные ϵ -сети, покрывающие единичные сферы $\|\vec{a}\|_2 = 1, \|\vec{b}\|_2 = 1, \|\vec{c}\|_2 = 1$ и состоящие из векторов единичной длины, можно оценить

$$(1 - d\epsilon) \sup_{\|V\|_F=1, \text{rank}(V)=1} (\mathbf{N}, V)_F \leq \max_{i,j,k} (\mathbf{N}, \mathbf{1} \times_1 \vec{a}_i^T \times_2 \vec{b}_j^T \times_3 \vec{c}_k^T)$$

где i, j, k соответствуют индексам ϵ -сети, а d обозначает размерность тензора (в формулах выше равна 3 для удобства). Имея ввиду, что по свойству нормального распределения величины $(\mathbf{N}, \mathbf{1} \times_1 \vec{a}_i^T \times_2 \vec{b}_j^T \times_3 \vec{c}_k^T)_F$ распределены нормально, и используя, что порядок числа необходимых элементов в ϵ -сетях имеет вид $\approx (\frac{3}{\epsilon})^{m_1}, (\frac{3}{\epsilon})^{m_2}, (\frac{3}{\epsilon})^{m_3}$, можно видеть, используя свойство вероятности объединения событий (хоть и зависимых)

$$\begin{aligned} \mathbb{P} \left((1 - d\epsilon) \sup_{\|V\|_F=1, \text{rank}(V)=1} (\mathbf{N}, V)_F > t \right) &\leq \left(\frac{3}{\epsilon}\right)^{\sum_{s=1}^d m_s} \mathbb{P}(\mathcal{N}(0,1) > t) \\ &\leq \frac{1}{\sqrt{2\pi t}} \left(\frac{3}{\epsilon}\right)^{\sum_{s=1}^d m_s} e^{-\frac{t^2}{2}}, \end{aligned} \quad (3.10)$$

где использована классическая оценка на „хвост“ нормального распределения $\mathbb{P}(\mathcal{N}(0,1) > t) \leq \frac{1}{\sqrt{2\pi t}} e^{-\frac{t^2}{2}}$. Оценка (3.10) имеет вид произведения „быстро растущей“ и „быстро убывающей“ функции; ликвидируем „растущий“ множитель с помощью нетривиального выбора параметра t . Выбирая $\epsilon = \frac{1}{2d} \Rightarrow \frac{3}{\epsilon} = 6d$ и проводя замену $t \rightarrow t + 2\sqrt{\log(6d) \sum_{s=1}^d m_s}$, получим экспоненциальное убывание оценки с ростом t :

$$\mathbb{P} \left(\frac{\sup_{\|V\|_F=1, \text{rank}(V)=1} (\mathbf{N}, V)_F}{2} > 2\sqrt{\log(6d) \sum_{s=1}^d m_s} + t \right) < C e^{-\frac{t^2}{2}}, \quad (3.11)$$

где C не зависит от t . Обобщая рассуждение на случай d -мерных тензоров с одинаковыми размерами m_s , имеем

$$\begin{aligned} \|\mathbf{N}\|_2 &\lesssim m^{d/2}, \\ \sup_{\|V\|_F=1, \text{rank}(V)=1} (\mathbf{N}, V)_F &\lesssim 2\sqrt{\log(d)dm} \end{aligned}$$

с большой вероятностью. Обозначив за $M = m^d$ полное число элементов тензора, имеем

$$\|\mathbf{N}\|_F^2 \sim M, \quad (3.12)$$

$$\sup_{\|\mathbf{V}\|_2=1, \text{rank}(\mathbf{V})=1} (\mathbf{N}, \mathbf{V})^2 \lesssim \log(d) d M^{\frac{1}{d}}. \quad (3.13)$$

Аналог оценки (3.11) может быть получен и для комплексного случая путем о веществления: достаточно использовать комплекснозначные векторы $\vec{a}_i, \vec{b}_j, \vec{c}_k$, покрывающие комплекснозначную единичную сферу; объем d -мерного комплексного единичного шара тогда можно будет оценивать как объем $2d$ -мерного вещественного единичного шара, что приведет к изменению константы $\log 6 \rightarrow \log 12$ в оценке. При этом, чтобы аналогично вещественнозначному случаю, $(\mathbf{N}, \mathbf{1} \times_1 \vec{a}_i \times_2 \vec{b}_j \times_3 \vec{c}_k)_F \sim \mathcal{N}(0, 1)$ и $\|\mathbf{N}\|_F \sim M$, достаточно положить, что вещественная и мнимая часть каждого элемента \mathbf{N} имеет независимое распределение $\mathcal{N}(0, \frac{1}{\sqrt{2}})$.

На основании оценок (3.12, 3.13), ожидается, что устойчивость приближения тензоров малого канонического ранга с фиксированным числом элементов значительно улучшается с ростом его размерности d (т.е., например, среди тензоров с 1024 элементами наиболее устойчив десятимерный тензор с размерами 2,2,2,2...2).

Эта гипотеза была проверена практически: был рассмотрен тензор \vec{x} с элементами вида $\vec{x}_s := e^{i\phi s}$ с некоторым вещественным ϕ и $s = 1 \dots 2^{24}$. Этот тензор можно рассматривать несколькими способами: как двумерный $2^{12} \times 2^{12}$, трехмерный $2^8 \times 2^8 \times 2^8$, и так далее, вплоть до 24-мерного; каждый из них имеет точный канонический ранг 1 в силу разделяемости переменных в экспоненте вида

$$e^{i\phi s} = e^{i\phi(s_1+m_1(s_2+m_2(s_3+\dots)))} = e^{i\phi s_1} e^{i\phi m_1 s_2} e^{i\phi m_1 m_2 s_3} \dots e^{i\phi m_1 m_2 \dots m_{d-1} s_d},$$

где $s_1, s_2 \dots s_d, m_1, m_2 \dots m_d$ - индексы и размеры тензора согласно выбранной размерности. Для всех возможных разложений такого тензора был проведен следующий эксперимент: к тензору был добавлен случайный гауссов шум, а затем применен алгоритм поиска канонического разложения ALS соответствующей размерности. Результаты представлены на рис. 3.2, 3.3, где желтым пока-

зана теоретическая оценка $dM^{\frac{1}{d}}$, соответствующая (3.13) без множителя $\log(d)$, а синим - практическое значение ошибки $\|\mathcal{P}_K(\vec{x} + \vec{n}) - \vec{x}\|_2$. На графиках видно асимптотическое соответствие аналитической гипотезы (3.13) практическим экспериментам с точностью до сравнительно небольшого члена $\log(d)$, однако при большом уровне шума и большой размерности d полученное алгоритмом ALS приближение может полностью потерять аппроксимационные свойства. Экспериментально проверено, что это всегда соответствует невыполнению на практике условия (3.2); таким образом, тензоры канонического ранга 1 наиболее устойчивы к случайному шуму в случае максимального d , однако алгоритм ALS в случае экстремально высокого d и уровня шума на практике может не удовлетворять условию квазиоптимальности (3.2). На рисунке 3.4 проиллюстрирован вывод алгоритма ALS при поступлении на вход тензора $\vec{x}_s := e^{i\phi_s}$ с добавлением шума, десятикратно превышающего фробениусову норму тензора: на графиках представлена скалированная вещественная часть тензоров с 2^{24} элементами в формате двумерного массива $2^{12} \times 2^{12}$. На зашумленных входных данных (рис. 3.4б) искомый тензор точного ранга 1 (рис. 3.4а) практически неразличим, в то время как выход алгоритма ALS все более похож на незашумленный тензор с увеличением размерности, пока выполняется условие (3.2). На этом примере при достижении размерности 8 гипотеза (3.2) нарушается, и вывод алгоритма ALS имеет случайную разреженную „сетчатую“ структуру, не имеющую ничего общего с искомым тензором ранга 1. При этом поведение алгоритма ALS изменяется - при малых размерностях сходимость по критерию неухудшения невязки выше машинного нуля достигается за небольшое число итераций (около 10), при размерности 6 сходимость замедляется и стагнирует большее число итераций (около 50), но затем возобновляется, а при размерностях, на которых условие (3.2) в конечном итоге нарушается, алгоритм ALS проводит большое число итераций до выполнения критерия неухудшения невязки.

На основании описанных практических свойств алгоритма ALS для построения канонического разложения зашумленных тензоров, представимых в виде тензоров разной размерности, и имеющих ранг, не зависящий от размерности тензора (например, тензора, заданного формулой $\vec{x}_s := e^{i\phi_s}$), предлагается использовать иерархическую схему. Схема состоит из следующих шагов:

- Запуск алгоритма ALS на малой размерности (например, 3) - на пониженной размерности более вероятно выполнение условия (3.2).

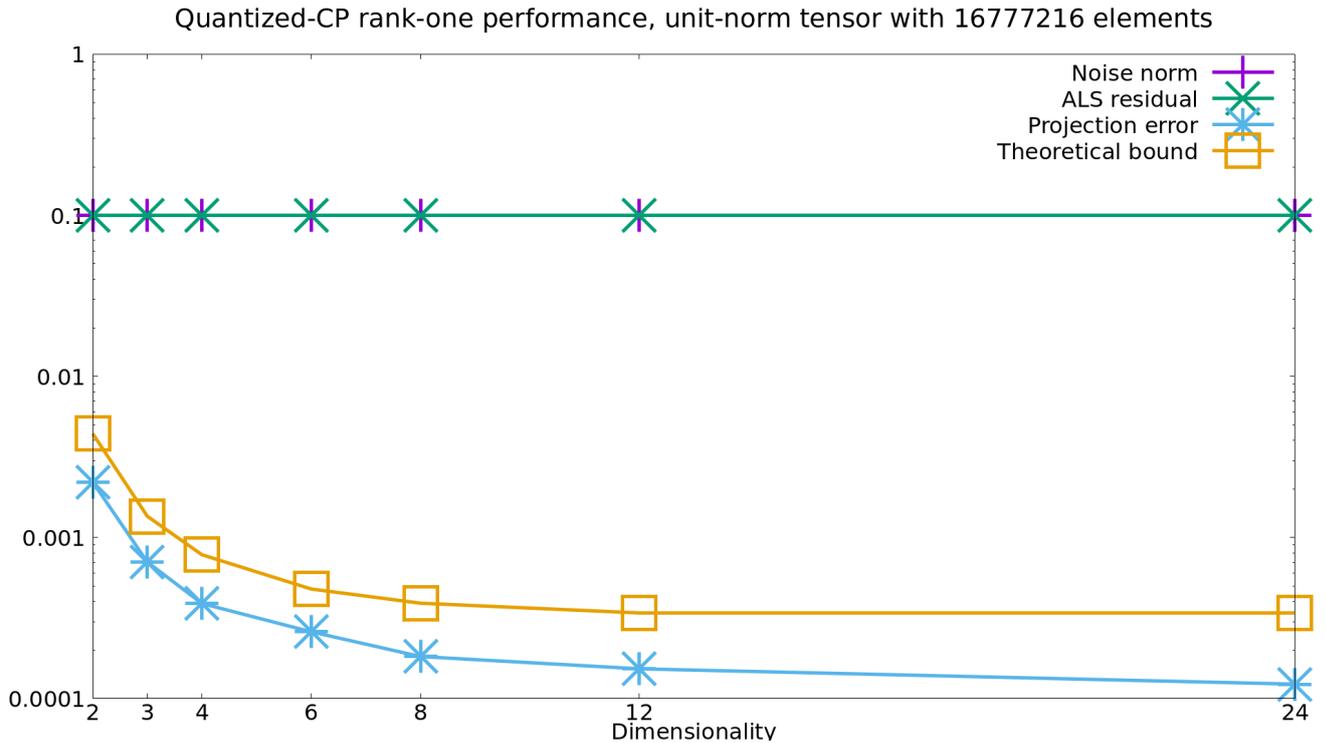


Рисунок 3.2 — Устойчивость тензорного приближения ранга 1 к „малому“ шуму, в 10 раз меньшему тензора по норме Фробениуса

- Запуск алгоритма ALS или использование сингулярного разложения для приближения каждого фактора по отдельности как тензора (или, возможно, матрицы) ранга 1. Процедура может повторяться рекурсивно.

Таким образом, для построения канонического приближения размерности 24 для зашумленного тензора ранга 1 с 2^{24} элементами предлагается сначала найти с помощью алгоритма ALS трехмерное разложение в формате $2^8 \times 2^8 \times 2^8$, затем каждый полученный фактор размера 2^8 с помощью сингулярного разложения приблизить матрицей ранга 1 размера $2^4 \times 2^4$, каждый фактор результата приблизить матрицей $2^2 \times 2^2$ ранга 1, и каждый фактор результата приблизить матрицей 2×2 ранга 1. На практике оказывается, что построенный таким образом алгоритм „справляется“ с более высоким уровнем шума: при постепенном повышении уровня шума до бесконечности условие на результат (3.2) выполняется дольше, а невязка результата с незашумленным тензором ниже, чем для трехмерного разложения ALS.

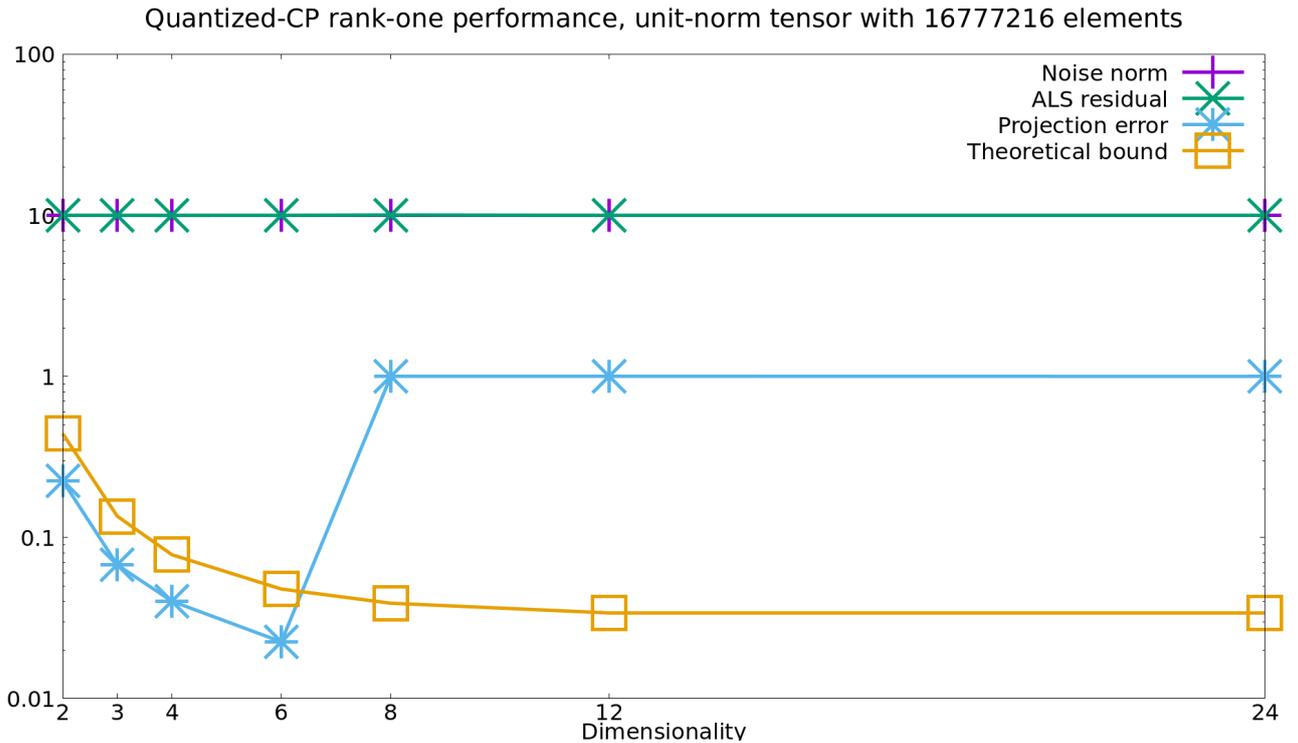


Рисунок 3.3 — Устойчивость тензорного приближения ранга 1 к „большому“ шуму, в 10 раз большему тензора по норме Фробениуса

3.4 Обобщение на случай тензоров большего ранга и различных тензорных форматов

3.4.1 Канонический формат

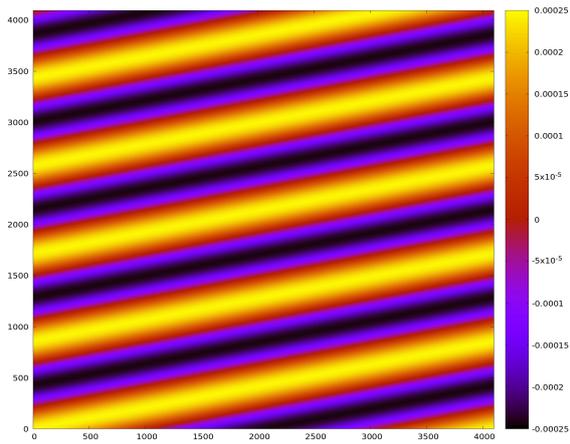
Как было видно ранее, для полноты анализа требуется провести переход от точной верхней грани

$$\sup_{\text{rank}(\mathbf{V}) \leq 1, \|\mathbf{V}\|_F = 1} (\mathbf{N}, \mathbf{V})_F$$

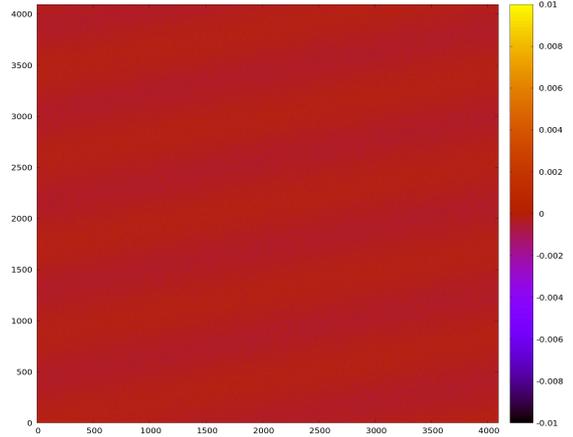
к величине

$$\sup_{\text{rank}(\mathbf{V}) \leq r, \|\mathbf{V}\|_F = 1} (\mathbf{N}, \mathbf{V})_F$$

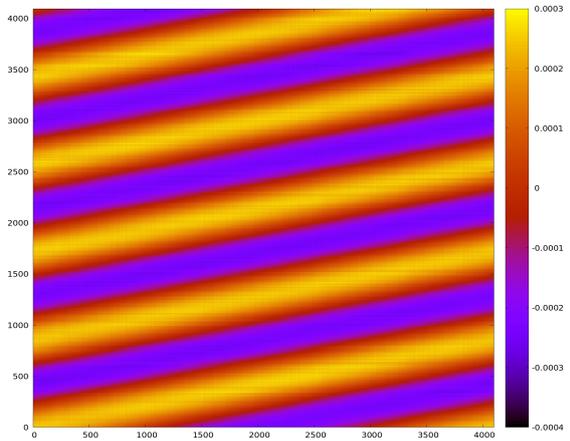
где ранг понимается в каноническом смысле: это позволило бы корректно рассмотреть множество $K \times K$ для тензоров ранга 1, а так же расширить теорию на множество K тензоров канонического ранга, не превосходящего константы.



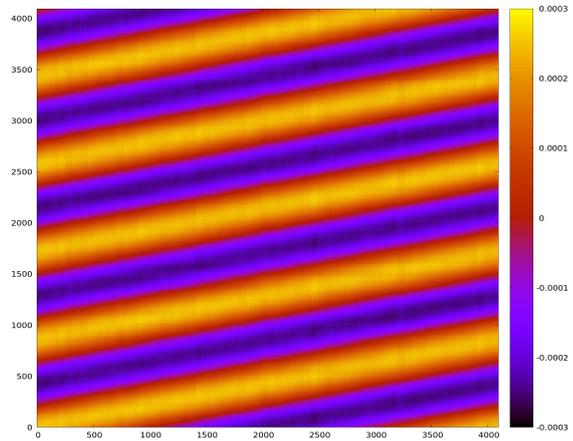
а) Тензор без шума



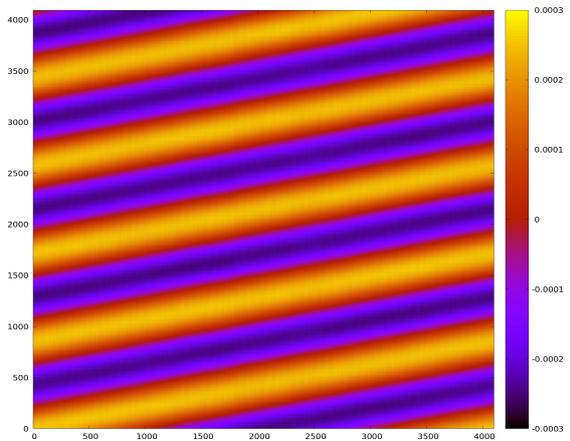
б) Тензор с шумом, превосходящим норму тензора в 10 раз



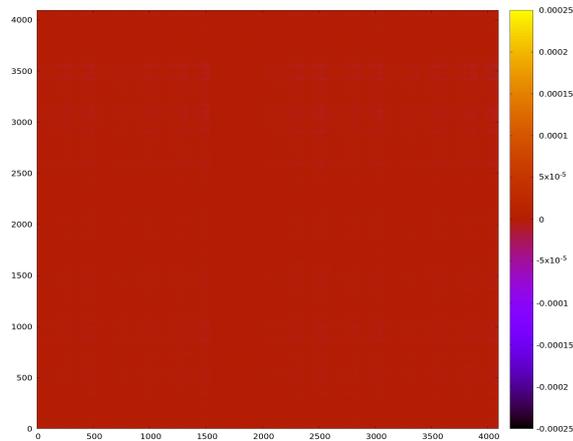
в) Результат работы ALS размерности 3



г) Результат работы ALS размерности 4



д) Результат работы ALS размерности 6



е) Результат работы ALS размерности 8

Рисунок 3.4 — Результаты работы алгоритма ALS с одним и тем же входным тензором, рассматриваемым как тензор различных размерностей

Для этого, предлагается представить тензор $\text{vec}(\mathbf{V}) = \vec{v}$ в виде суммы $\sum_{k=1}^r \alpha_k \vec{v}_k$, где тензоры \vec{v}_k имеют ранг 1 и единичную норму Фробениуса. Тогда,

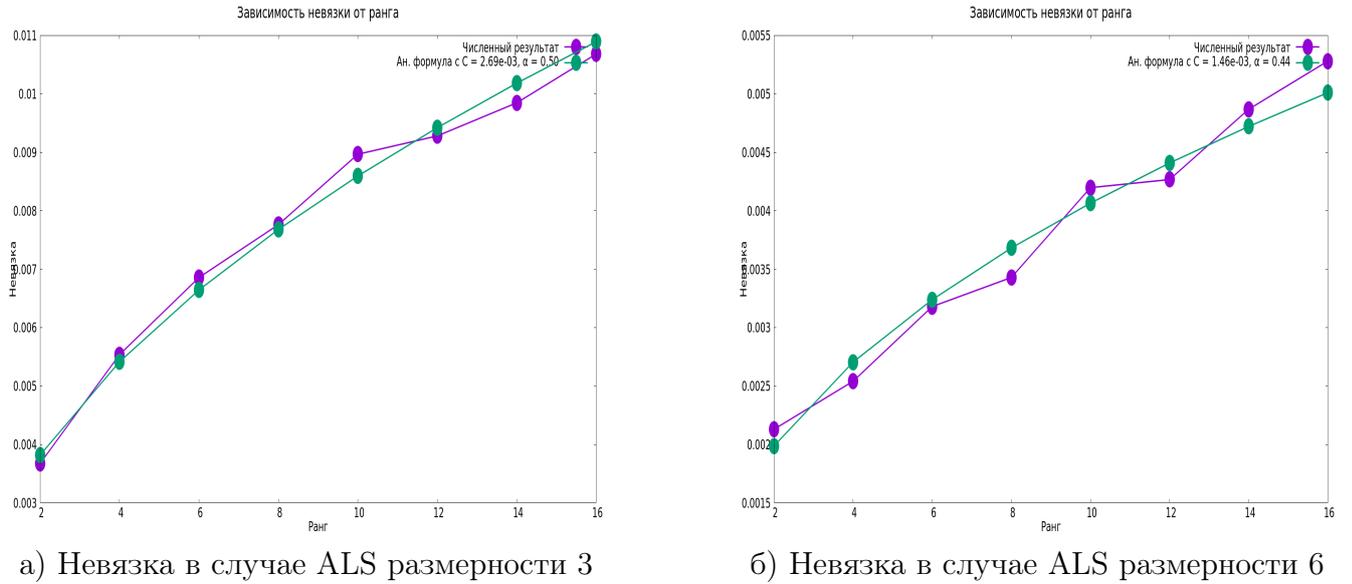
$$\begin{aligned} \sup_{\text{rank}(\vec{v}) \leq r, \|\vec{v}\|_2=1} (\vec{n}, \vec{v}) &= \sup_{\text{rank}(\vec{v}) \leq r, \|\vec{v}\|_2=1} \sum \alpha_k (\vec{n}, \vec{v}_k) \\ &\leq \sup_{\text{rank}(\vec{v}) \leq 1, \|\vec{v}\|_2=1} \sqrt{\sum \alpha_k^2} \sqrt{\sum_{k=1}^r (\vec{n}, \vec{v}_k)}. \end{aligned} \quad (3.14)$$

Если предположить, что рассмотренный супремум может достигаться только на тензорах канонического ранга R с компонентами, близкими к ортогональным, можно считать, что $\sqrt{\sum \alpha_k^2} \leq \text{const}$, и тогда

$$\sup_{\|\vec{v}\|_2=1, \text{rank}(\vec{v})=r} (\vec{n}, \vec{v}) \lesssim \sqrt{r} \sqrt{\log(d) d M^{\frac{1}{d}}}. \quad (3.15)$$

Эта оценка, опираясь на основное общее неравенство (3.5), была проверена эмпирически - для этого была построена последовательность тензоров $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \dots$ такая что $\text{rank}(\mathbf{X}_r) = r$ в каноническом смысле, а фробениусова норма каждого из Y_r равна единице. Затем, к каждому из этих тензоров был добавлен гауссов шум \vec{n} , и проведено исследование зависимости величины $\|\mathcal{P}_K(\vec{x}_r + \vec{n}) - \vec{x}_r\|_2$ от ранга r , где в качестве \mathcal{P}_K использовался алгоритм ALS. Для этой величины строилась аппроксимация вида Cr^α , где неизвестные C, α потенциально могли зависеть от размерности d . Поиск C, α логарифмированием сводился к задаче наименьших квадратов с двумя неизвестными. Результаты приведены на рис. 3.7, где фиолетовым обозначен результат эксперимента, а зеленым его аналитическое приближение в формате Cr^α с C, α , найденными методом наименьших квадратов после логарифмирования. На графиках видно, что вне зависимости от d , рост невязки хорошо характеризуется величиной \sqrt{r} , т.е. найденный показатель степени близок к $\alpha = 0.5$. Более того, экспериментально с высокой точностью невязка в случае \vec{x} со случайным заданием факторов оценивается формулой (без дополнительных констант!)

$$\|\mathcal{P}_K(\vec{x}_r + \vec{n}) - \vec{x}_r\|_2 \approx \sqrt{drm} \frac{\|\vec{n}\|_2}{\sqrt{M}},$$



а) Невязка в случае ALS размерности 3

б) Невязка в случае ALS размерности 6

Рисунок 3.5 — Зависимость устойчивости канонического тензорного приближения от ранга

где как и ранее m - линейный размер тензора, а $M = m^d$ - его общее число элементов.

Тем не менее, использованную гипотезу об ограниченности $\sqrt{\sum \alpha_k^2} \leq \text{const}$ для векторов \vec{v} , на которых значение (\vec{n}, \vec{v}) близко к оптимальному, доказать формально не удалось. С учетом связи задачи оптимизации (3.6) и задачи аппроксимации (3.7) теоретическая проверка этой гипотезы может быть крайне трудной; задачи о канонической аппроксимации и ее рангах не были решены значительное время даже для конкретных тензоров, например, тензоров матричного умножения.

3.4.2 Формат Таккера

Исследование устойчивости тензорных разложений было проведено и для других форматов тензорных разложений. Для этого, согласно проведенному ранее анализу, необходимо снова оценить величину

$$\sup_{\text{rank}(\mathbf{V}) \leq r, \|\mathbf{V}\|_F = 1} (\mathbf{N}, \mathbf{V})_F$$

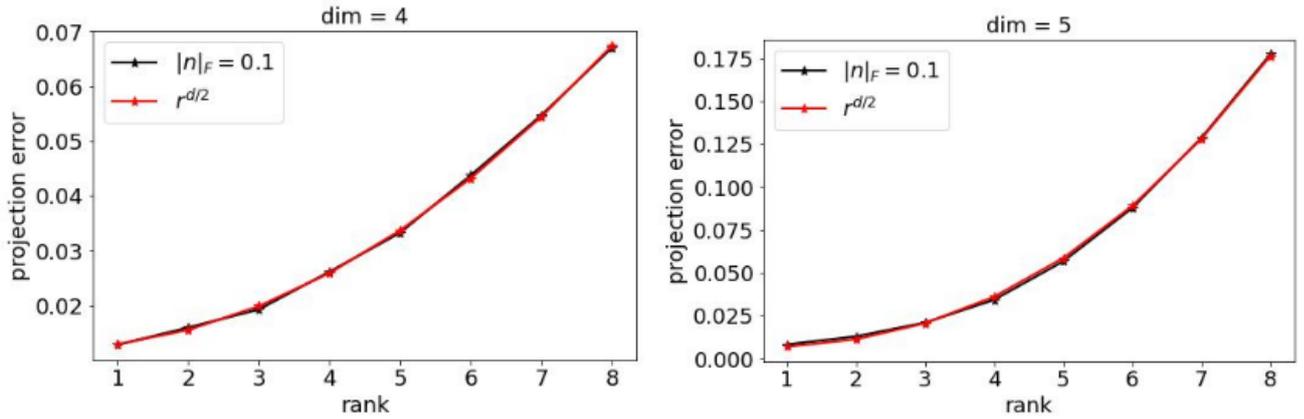


Рисунок 3.6 — Зависимость устойчивости тензорного приближения Таккера от ранга

где $rank(\mathbf{V})$ теперь понимается в смысле Таккера, а также ввести в рассмотрение алгоритм построения приближения в формате Таккера, и проверить удовлетворение им гипотезы (3.2).

В качестве такого алгоритма был использован алгоритм HOSVD [21]. Снова была построена последовательность тензоров $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \dots$ такая что $rank(\mathbf{X}_r) = r$ в смысле Таккера, а фробениусова норма каждого из \mathbf{X}_r равна единице. При этом ядро Таккера задавалось поэлементно случайным, а факторы Таккера задавались в виде случайных ортогональных базисов. Затем, к каждому из этих тензоров был добавлен гауссов шум \vec{n} , и проведено исследование зависимости величины $\|\mathcal{P}_K(\vec{x}_r + \vec{n}) - \vec{x}_r\|_2$ от ранга r , где в качестве \mathcal{P}_K использовался алгоритм HOSVD.

В отличие от случая канонического разложения и алгоритма ALS, в случае использования разложения типа Таккера и алгоритма HOSVD проблем с невыполнением условия (3.2) на практике не обнаружено. Однако, разложение Таккера просто не имеет смысла при наиболее высоких значениях d (на которых не выполнялось (3.2) в случае ALS), так как число элементов ядра при этом становится быстрее размера тензора; разложение Таккера несет нетривиальный смысл только при $r < m = M^{1/d}$.

Исходя из этого, при размере построенного набор тензоров $Y_1, Y_2, Y_3 \dots$ вида 2^{20} были рассмотрены размерности тензора 4 и 5 и ранги вплоть до 16. Экспериментальные результаты представлены на рис. 3.6, где в обоих случаях экспериментальные результаты хорошо приближаются функцией $Cr^{\frac{d}{2}}$.

Такая зависимость невязки от ранга Таккера не требует дополнительного теоретического анализа, так как тензор в формате Таккера с рангами не выше r всегда представим в формате канонического разложения с рангом не выше r^d ; таким образом, полученная полиномиальная оценка напрямую следует из аналогичной оценки для канонического разложения, и не может быть улучшена, согласно практическим экспериментам.

3.4.3 Формат тензорного поезда

Наконец, для исследования устойчивости тензорных приближений в формате тензорного поезда была снова оценена величина

$$\sup_{\text{rank}(\mathbf{V}) \leq r, \|\mathbf{V}\|_F = 1} (\mathbf{N}, \mathbf{V})$$

где $\text{rank}(\mathbf{V})$ понимается в смысле наибольшего из рангов тензорного поезда. В качестве алгоритма приближения рассматривался алгоритм TT-SVD [22]. Снова была построена последовательность тензоров $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \dots$ такая что $\text{rank}(\mathbf{X}_r) = r$ в смысле тензорного поезда, а фробениусова норма каждого из \mathbf{X}_r равна единице. Эти тензоры задавались как скалированный результат применения TT-SVD к поэлементно случайному тензору. Затем, к каждому из этих тензоров был добавлен гауссов шум \vec{n} , и проведено исследование зависимости величины $\|\mathcal{P}_K(\vec{x}_r + \vec{n}) - \vec{x}_r\|_2$ от ранга r , где в качестве \mathcal{P}_K использовался алгоритм TT-SVD.

Как и в случае канонического разложения, для численных результатов было построено приближение в формате Cr^α , где C, α выбирались посредством метода наименьших квадратов после логарифмирования. Экспериментально полученные α близки к единице вне зависимости от размерности d вплоть до максимально возможных.

Данную зависимость невязки от ранга не удастся напрямую теоретически свести к каноническому случаю (как в случае разложения Таккера) с рангом $r^{2\alpha} = r^2$. Известно, что тензор в формате тензорного поезда со всеми рангами, не превышающими r , можно представить в виде тензора канонического фор-

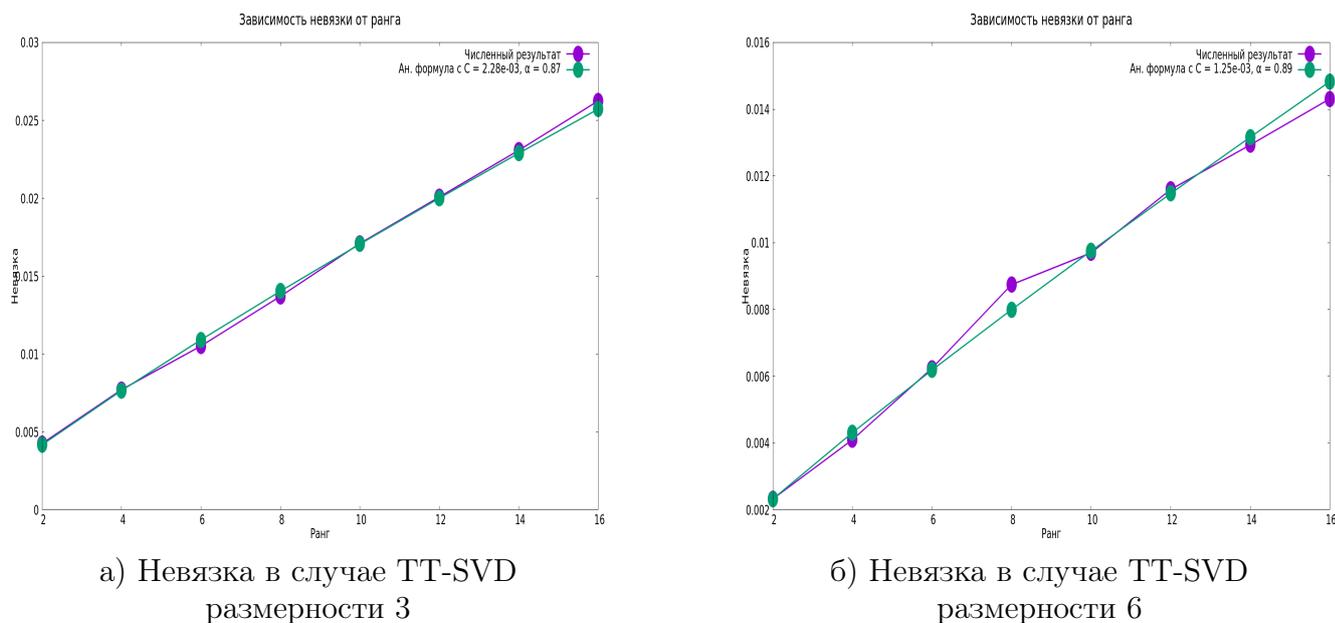


Рисунок 3.7 — Зависимость устойчивости тензорного приближения в формате тензорного поезда от ранга

мата с рангом асимптотического порядка $r^{\log \frac{d}{2}}$ (для этого рекурсивно перемножаются пары вагонов тензорного поезда), что вообще говоря больше, чем r^2 . Пример оператора Лапласа, имеющего ранг тензорного поезда, равный двум, и канонический ранг, равный d , указывает, что оценка $\log \frac{d}{2}$ асимптотически достигается.

На основании проведенных экспериментов с тензорными разложениями в различных форматах таким образом можно видеть, что невязка $\|\mathcal{P}_K(\vec{x}_r + \vec{n}) - \vec{x}_r\|_2$ зависит от ранга во всех случаях так же, как и корень из числа параметров разложения. Для случая тензорного поезда достаточного теоретического обоснования этому пока построено не было.

Глава 4. Восполнение тензоров в формате Таккера

В этом разделе будет рассмотрена модель тензоров в формате Таккера при условии неполных данных. Глава не будет содержать теоретических обоснований, однако в случае модели Таккера будет рассмотрена интересная возможность применения восполнения как алгоритма аппроксимации. Идея заключается в том, что в случае многомерных тензоров малого ранга любой алгоритм, нетривиально использующий все элементы тензора, имеет сложность не ниже произведения линейных размеров тензора, и уже заведомо неприменим в некоторых приложениях, поэтому работоспособность алгоритмов восполнения в условиях неполных данных является преимуществом.

4.1 Обобщение метода проекции градиента на восполнение тензоров в формате Таккера

4.1.1 Сведение к матрицам развертки

Метод проекции градиента для задачи восполнения матриц, рассмотренный в Главе 1, состоит из двух основных операций: вычисления градиента и вычисления проекции на множество матриц малого ранга. Естественным образом распространим функционал восполнения на случай тензоров с помощью

$$\psi_{\mathcal{A}_\Omega}(T) = \frac{1}{2} \|\mathcal{A}_\Omega(T) - B\|_F^2 \rightarrow \inf, \quad \text{rank}(T) \leq R, \quad (4.1)$$

где оператор $\mathcal{A}_\Omega : \mathbb{R}^{m_1 \times m_2 \times \dots \times m_d} \rightarrow \mathbb{R}^{m_1 \times m_2 \times \dots \times m_d}$, как и ранее, приравнивает нулю элементы входного тензора T , не лежащие на наперед заданной маске Ω , и масштабирует остальные для выполнения условия ограниченной изометрии:

$$\mathcal{A}_\Omega(T)_{j_1 j_2 \dots j_d} = \begin{cases} \frac{1}{\sqrt{\rho}} T_{j_1 j_2 \dots j_d}, & (j_1, j_2, \dots, j_d) \in \Omega \\ 0, & (j_1, j_2, \dots, j_d) \notin \Omega \end{cases}$$

Формально обобщим метод проекции градиента на случай тензоров формата Таккера. Операция вычисления градиента не будет иметь изменений по сравнению с матричным случаем; изменения потребуются в операции построения проекции. Точная проекция наперед заданного тензора на множество тензоров, ранги Таккера которых не превосходят константы - всегда существует (в отличие от канонического формата тензоров), но ее поиск затруднителен. Вместо этого, для построения будем опираться на более классический алгоритм HOSVD [21], сводящий построение к малоранговым приближениям матриц развертки тензора.

От обобщения методов выполнения на случай тензоров естественно ожидать понижения требуемой плотности ρ известных элементов по сравнению со случаем матриц, иначе построение алгоритмов выполнения тензоров не несет большого практического смысла, так как не даст преимуществ по сравнению с запуском матричных алгоритмов на тех же данных. Поэтому, в случае $m_1 = m_2 = \dots = m_d = m$, ожидается $\rho = o(\frac{1}{m})$. В связи с этим, матрицы разверток тензора с большой вероятностью будут содержать целые столбцы, в которых не содержится ни одного известного элемента; это отражено на рис. 4.1. С учетом того, что шаг градиента имеет ненулевые значения только в позициях известных элементов, предлагается опираться только на левый столбцовый сингулярный базис шагов градиента. На основании этих рассуждений будем опираться на нерекурсивный вариант алгоритма HOSVD [21], и введем Алгоритм 5. Применение \mathcal{A}_Ω к развертке тензора в описании алгоритма неявно подразумевает перестановочность оператора выполнения и построения развертки.

Можно видеть, что каждый шаг Алгоритма 5 эквивалентен матричному методу проекции градиента, примененному к развертке, с дополнительным проектированием результата на текущие приближения к базисам Таккера вида

$$\mathbf{T} := \mathcal{P}_{k,r_k}(F_k(\mathbf{T}, \mathbf{B})) \times_1 U_1 U_1^* \dots \quad (4.2)$$

$$\dots \times_{k-1} U_{k-1} U_{k-1}^* \times_{k+1} U_{k+1} U_{k+1}^* \dots \times_d U_d U_d^* \quad (4.3)$$

т.е. полному шагу алгоритма SVP для одной развертки тензора, спроектированному дополнительно на текущее приближение для подпространств Таккера вдоль оставшихся размерностей.

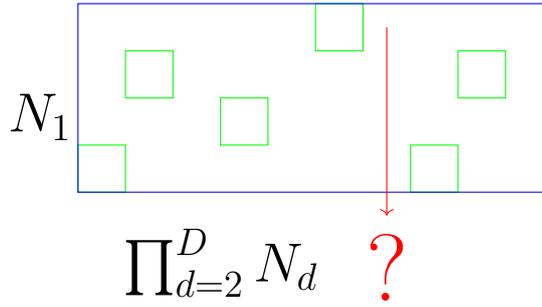


Рисунок 4.1 — Схема распределения известных элементов тензора на развертке

Algorithm 5: Попеременный метод проекции градиента для восполнения тензоров в формате Таккера

Data: Инициализация: $\mathbf{C}_{j_1 j_2 \dots j_d} := 0$; $V_k, k = 1 \dots d$ задаются как случайные матрицы с ортогональными столбцами.

```

1 while  $\|\mathcal{A}_\Omega(\mathbf{T}) - \mathbf{B}\|_F \geq \epsilon$  do
2   for  $k = 1 \dots d$  do
3     Вычисляется матрица шага градиента вдоль размерности  $k$ :
4      $F_k(\mathbf{T}, \mathbf{B}) := T_k - \tau \mathcal{A}_\Omega(T_k)^* (\mathcal{A}_\Omega(T_k) - B_k)$  ;
5     Вычисляется точная или приближенная матричная проекция
6     малого ранга  $\mathcal{P}_{k,r_k}(F_k(\mathbf{T}, \mathbf{B})) = U \Sigma V^*$  ;
7     Обновляется фактор Таккера  $U_k := U$  ;
8     Обновляется ядро Таккера
9      $\mathbf{C} := F_k(\mathbf{T}, \mathbf{B}) \times_1 U_1^* \dots \times_k U_k^* \dots \times_d U_d^*$  ;

```

4.1.2 Снижение сложности метода проекции градиента на развертках

Явное построение тензорных разверток $F_k(\mathbf{T}, \mathbf{B})$ и их точных или приближенных SVD-проекции $\mathcal{P}_{k,r_k}(F_k(\mathbf{T}, \mathbf{B}))$ - слишком затратно с точки зрения и вычислительной сложности, и памяти: практически интересными являются алгоритмы, сложность которых не зависит от произведения всех размеров тензора. В связи с этим, предлагается провести замену (неэквивалентную!) порядка операций (4.2), которая позволила бы значительно снизить сложность алгорит-

ма: рассмотрим вместо (4.2) формулу

$$T := \mathcal{P}_{k,r_k}(F_k(\mathbf{T}, \mathbf{V}) \times_1 U_1 U_1^* \dots \dots \times_{k-1} U_{k-1} U_{k-1}^* \times_{k+1} U_{k+1} U_{k+1}^* \dots \times_d U_d U_d^*) \quad (4.4)$$

Иными словами, поменяем очередность операции линейного проектирования и операции проектирования развертки на множество матриц малого ранга. При этом структура Таккера позволит понизить сложность вычислений: линейная проекция тензора в формате Таккера на $d - 1$ базисов того же тензора равна произведению ядра на оставшийся базис, а разреженную матрицу можно проектировать на $d - 1$ базис Таккера поэлементно, причем проекция одного элемента будет определяться произведением строк базисных матриц для подпространств Таккера. На основе этих рассуждений предлагается Алгоритм 6, в котором величина, находящаяся под оператором проекции \mathcal{P}_{k,r_k} вычисляется под обозначением Y_k . Сложность итерации этого алгоритма характеризуется объемом двух вложенных циклов и оценивается величиной $|\Omega|r^d$ в случае равенства всех рангов Таккера величине r .

Предложенные варианты метода проекции градиента для тензоров в формате Таккера были протестированы на тензорах размерности 4, размера $20 \times 20 \times 20 \times 20$ точных рангов Таккера 4,3,3,2. Тестовые тензоры генерировались как тензоры с поэлементно случайным ядром и случайными ортогональными базисами факторов. В качестве входных данных алгоритму предоставлялось 0.8% элементов искомого тензора (можно сравнить с типичными 5-10 % элементов, необходимыми для восполнения матрицы порядка 400 и ранга 9). Полученные следующие графики функционала (4.1) для Алгоритмов 5, 6 представлены на рис. 4.2: эмпирически наблюдается линейная сходимость обоих вариантов алгоритма.

Интересно заметить, что в случае „двумерного“, то есть матричного, разложения Таккера, итерация Алгоритма 6 сводится к формуле итерационного обновления вида

$$U := U - \tau(\mathcal{A}(UV^*) - B)V,$$

не содержащей операции проектирования (в таком случае число столбцов матрицы Y_k в точности равно r_k , и построение частичного сингулярного разложения сводится к построению ортогонального базиса). Эксперименты подтвер-

Algorithm 6: Вычислительно упрощенный попеременный метод проекции градиента для восполнения тензоров в формате Таккера

Data: Инициализация: $\mathbf{C}_{j_1 j_2 \dots j_d} := 0$; $U_k, k = 1 \dots d$ задаются как случайные матрицы с ортогональными столбцами.

```

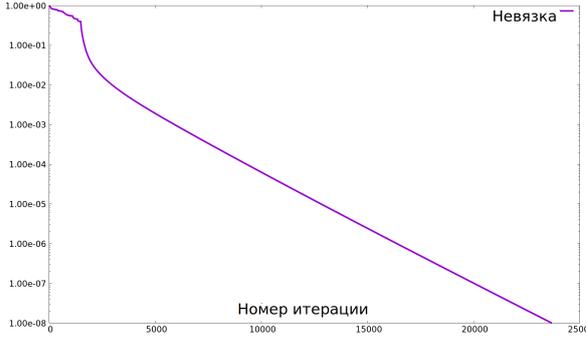
1 while  $\|\mathcal{A}_\Omega(\mathbf{T}) - \mathbf{B}\|_F \geq \epsilon$  do
2   for  $k = 1 \dots d$  do
3     Вычисляется произведение ядра и одного из факторов Таккера
4      $Y_k := (\mathbf{C} \times_k U_k)_k$ ;
5     for  $(j_1, j_2 \dots j_d) \in \Omega$  do
6       Вычисляется один элемент текущего приближения
7        $\kappa := \mathbf{T}_{j_1 j_2 \dots j_d}$  посредством задания  $\kappa$  нулем и суммирования
8       вида
9       for  $s_1 = 1 \dots r_1, s_2 = 1 \dots r_2, \dots, s_d = 1 \dots r_d$  do
10         $\tau := \mathbf{C}_{s_1 s_2 \dots s_d}$ ;
11        for  $l = 1 \dots d$  do
12           $\tau := (U_l)_{j_l s_l} \tau$ ;
13         $\kappa := \kappa + \tau$ ;
14      Вычисляется невязка  $\kappa := \mathbf{B}_{j_1 j_2 \dots j_d} - \kappa$ ;
15      Вычисляется проекция одного элемента градиентного шага;
16       $Y_k := Y_k + (\frac{\kappa}{\rho} \times_1 (U_1^*)_{j_1} \times_2 (U_2^*)_{j_2} \dots$ 
17       $\dots \times_{k-1} (U_{k-1}^*)_{j_{k-1}} \times_k \vec{e}_{j_k} \times_{k+1} (U_{k+1}^*)_{j_{k+1}} \dots \times_d (U_d^*)_{j_d})$ ;
18      Для  $Y_k$  вычисляется точное или приближенное частичное
19      сингулярное разложение с рангом  $r_k$ :  $Y_k \approx U_{r_k} \Sigma_{r_k} V_{r_k}^*$ ;
20      Обновляется один базис и ядро Таккера  $U_k := U_{r_k}, \mathbf{C} := \Sigma_{r_k} V_{r_k}^*$ ;

```

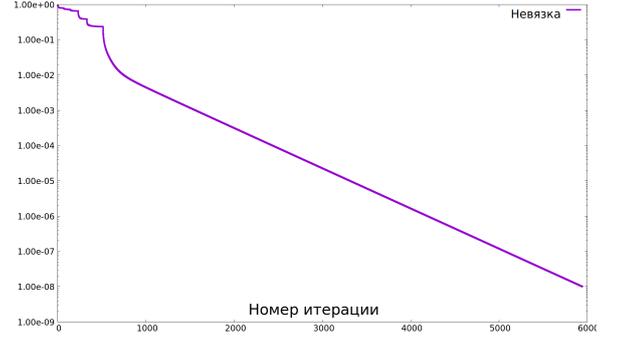
ждают геометрическую сходимость и такого алгоритма в случае равномерного распределения известных элементов матрицы.

4.2 Замена оператора восполнения

Вспомним, что исходная теория матричного метода проекции градиента из Теоремы 1.1.1 [29], а также предложенная в рамках данной работы теория в виде Теорем 1.1.2, построена не для конкретного оператора \mathcal{A}_Ω , задающего задачу восполнения, а для произвольного линейного оператора \mathcal{A} , удовлетворяющего



а) Алгоритм 5 без перестановки операций



б) Алгоритм 6 с перестановкой операций

Рисунок 4.2 — Падение функционала невязки (4.1) с итерациями двух вариантов метода проекции градиента для восполнения тензоров в формате Таккера

условию ограниченной изометрии вида

$$(1 - \delta)\|Y\|_F^2 \leq \|\mathcal{A}(Y)\|_F^2 \leq (1 + \delta)\|Y\|_F^2, \\ \forall Y \in \mathbb{R}^{m_1 \times m_2}, \forall \text{rank}(Y) \leq \text{const}.$$

Введенный ранее оператор восполнения \mathcal{A}_Ω частично удовлетворяет этому условию (удовлетворяет при дополнительном условии некогерентности входных данных), но не является единственно подходящим.

Рассмотрим альтернативный способ задания оператора \mathcal{A} , основанный на теории поэлементно случайных матриц. Опорной точкой такой теории является следующая теорема:

Теорема 4.2.1 (Лемма Джонсона-Линденштраусса, [48]). Пусть заданы $0 < \epsilon < 1$ и размерность M . Пусть для целочисленной величины L выполнено

$$R > \frac{4 \log M}{\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}};$$

Пусть каждый элемент матрицы $J \in \mathbb{R}^{M \times L}$ выбран случайно и независимо по нормальному распределению с математическим ожиданием 0 и стандартным отклонением 1. Тогда для любого наперед заданного вектора $\vec{x} \in \mathbb{R}^M$ выполнено

$$\mathbb{P}\left\{\left|\left\|\frac{1}{\sqrt{R}}J^*\vec{x}\right\|_2^2 - \|\vec{x}\|_2^2\right| > \epsilon\right\} \leq \frac{2}{M^2}.$$

В литературе утверждение также расширено на случай целых подпространств вместо конкретных векторов [49], других распределений элементов J [50] (например, дискретных), а также на случай сильно разреженных матриц J : в [51] показано, что даже $O(\log^2 n)$ случайно расставленных ненулевых элементов в расчете на столбец J достаточно для приближенного сохранения нормы вектора \vec{x} с константной вероятностью если \vec{x} неразрезан; также в [51] на основе преобразования Адамара предложено линейное преобразование, превращающее с большой вероятностью любой вектор в неразрезанный.

Используем такое преобразование с разреженным J в качестве альтернативы оператору восполнения в методе проекции градиента. Пусть $J_k \in \mathbb{R}^{(m_1 \dots m_{k-1} m_{k+1} \dots m_d) \times L}$ имеет s случайно расположенных элементов в каждом столбце, а каждый такой ненулевой элемент, как и ранее, распределен независимо по стандартному нормальному распределению. Пусть $\tau = \frac{M}{sL}$. Тогда вместо ранее рассмотренного шага градиента рассмотрим шаг вида

$$F_k(\mathbf{T}, \mathbf{B}) = T_k - (\tau(\mathbf{T} - \mathbf{B}) \times_{1 \dots k-1, k+1 \dots d} J J^*)_k,$$

и, как и ранее, будем использовать комбинацию проекций вида (4.4). Теперь будем считать что тензор $B \in \mathbb{R}^{m_1 m_2 \dots m_d}$ известен полностью, однако использование всех его элементов при построении приближения в формате Таккера нежелательно в силу большой вычислительной сложности. Структура матрицы J будет гарантировать неполное использование входных данных. Результирующая последовательность действий описана в Алгоритме 7 и обладает геометрической сходимостью эмпирически при $L = 2r + 6$ и $s = 4r^2$ на искусственных тензорах, в точности имеющих малые ранги Таккера со случайными базисами подпространств вдоль каждой размерности. В описании алгоритма Ω обозначает позиции ненулевых элементов J . При таких параметрах и равных размерах m и рангах r сложность итерации алгоритма оценивается величиной $O(mr^d + r^{d+2})$, и таким образом, методы восполнения могут быть эффективно применены к задачам аппроксимации.

Algorithm 7: Аппроксимация тензоров в формате Таккера на основе алгоритмов восполнения

Data: Инициализация: $\mathbf{C}_{j_1 j_2 \dots j_d} := 0$; $U_k, k = 1 \dots d$ задаются как случайные матрицы с ортогональными столбцами; $\gamma := +\infty$

- 1 **while** $\gamma > \epsilon$ **do**
- 2 **for** $k = 1 \dots d$ **do**
- 3 Вычисляется произведение ядра и одного из факторов Таккера
 $Y_k \in \mathbb{R}^{m_k \times (r_1 \dots r_{k-1} r_{k+1} \dots r_d)}$
- 4 $Y_k := (\mathbf{C} \times_k U_k)_k$;
- 5 Вычисляется $Z_k \in \mathbb{R}^{L \times (r_1 \dots r_{k-1} r_{k+1} \dots r_d)}$, $Z_k =$
 $J^* \times_1 U_1 \cdots \times_{k-1} U_{k-1} \times_{k+1} U_{k+1} \cdots \times_d U_d$ с помощью
- 6 $Z_k := 0$;
- 7 **for** $(j_1, j_2 \dots j_d) \in \Omega$ **do**
- 8 $Z_k := Z_k + J_{j_1 j_2 \dots j_d} \times_1 (U_1^*)_{j_1} \times_2 (U_2^*)_{j_2} \cdots \times_{k-1} (U_{k-1}^*)_{j_{k-1}} \times_k$
 $\vec{e}_{j_k} \times_{k+1} (U_{k+1}^*)_{j_{k+1}} \cdots \times_d (U_d^*)_{j_d}$
- 9 Вычисляется $D_k \in \mathbb{R}^{m_k \times L}$, $D_k := B_k J^*$ (матричное произведение разреженной J^* с матрицей развертки B). ;
- 10 Вычисляется невязка $J^*(B_k - T_k)$ и ее норма γ для проверки условия выхода:
- 11 $D_k := D_k - Y_k Z_k^*$, $\gamma := \|D_k\|_F$;
- 12 Выполняется шаг градиента:
- 13 $Y_k := Y_k + \tau D_k Z_k$;
- 14 Выполняется матричное проектирование с рангом R_k :
 $Y_k \approx U_{R_k} \Sigma_{R_k} V_{R_k}^*$;
- 15 Обновляется один базис и ядро Таккера:
- 16 $U_k := U_{R_k}$, $\mathbf{C} := \Sigma_{R_k} V_{R_k}^*$;

Глава 5. Практические приложения построенной теории и алгоритмов

5.1 Автомобильные радары

5.1.1 Структура возникающих массивов данных

Автомобильный радар представляет собой набор отправляющих и принимающих антенн, способных излучать и принимать волновые сигналы на различных частотах и в различные моменты времени. Данные таких радаров естественным образом представимы в виде многомерных комплекснозначных тензоров [19]. Описание модельной физической структуры такого сигнала начнем со случая одной антенны и одного объекта: предполагается, что сигнал, порождаемый антенной, отражается от объекта и возвращается обратно. Простейшая математическая модель электромагнитной волны имеет вид

$$x(t) = e^{i2\pi f(t)t},$$

где $f(t)$ - частота передачи. Будем предполагать использование технологии частотной модуляции („Frequency Modulated Continuous Waveform“, FMCW): в рамках этой технологии вводится временной интервал T , в рамках которого частота передачи меняется линейно от некоей f_0 вплоть до $f_0 + B$, где величину f_0 называют „несущей частотой“, а величину B - шириной полосы. Тогда имеем

$$\begin{aligned} f(t) &= f_0 + B\frac{t}{T} \\ x(t) &= e^{i2\pi t(f_0 + B\frac{t}{T})} = e^{i2\pi(f_0 t + \frac{B}{T}t^2)} \end{aligned}$$

После отражения от объекта сигнал возвращается к антенне радара с некоторой задержкой, определяемой пройденным волной удвоенным расстоянием l от радара до объекта. Для демодуляции возвращенного сигнала посредством вычислительных мощностей радара полученный сигнал умножают на сопряженный

к отправленному:

$$r(t) = x^*(t)x(t - \Delta t) = e^{i2\pi(f_0(t-\Delta t) + \frac{B}{T}(t-\Delta t)^2 - f_0t - \frac{B}{T}t^2)} = \alpha(\Delta t)e^{-i2\pi t(\Delta t \frac{2B}{T})} \quad (5.1)$$

$$\Delta t = \frac{2l}{c}, \alpha(\Delta t) = e^{-i2\pi(f_0\Delta t - \frac{2B}{T}(\Delta t)^2)}.$$

Величина $r(t)$ дискретизируется по времени с шагом s и индексацией j_{time} и определяет массив данных $r(t_{j_{time}}) = r(j_{time}s)$, доступных одноантенному радару.

Теперь рассмотрим случай многоантенного радара, схематично представленного на рис. 5.1. У такого радара имеется несколько отправляющих антенн („ТХ“, синих) и принимающих антенн („RX“, красных). Пусть такие антенны для упрощения расположены равномерно на двух отрезках с началом в начале координат и направленных вдоль осей OX, OY соответственно. Предполагается, что каждая принимающая антенна j_{rx} умеет выделять сигнал, излученный с каждой конкретной отправляющей антенны j_{tx} посредством использования кодирования. Будем использовать модели параллельного падения лучей:

$$\Delta t \approx \frac{2l}{c} + hj_{rx} \cos \theta + hj_{tx} \cos \psi,$$

где $\theta = \angle XOQ$, $\psi = \angle YOQ$, а h - расстояние между двумя последовательными антеннами. Подставим это приближение в вывод (5.1), предполагая, что произведение $h \frac{2B}{T}$ пренебрежимо мало: тогда, дополнительные слагаемые $hj_{rx} \cos \theta$ и $hj_{tx} \cos \psi$ повлияют только на $\alpha(\Delta t)$. Массив данных радара тогда будет приближенно иметь вид

$$r(j_{rx}, j_{tx}, j_{time}) \approx \beta e^{-i2\pi f_0 h \cos \theta j_{rx}} e^{-i2\pi f_0 h \cos \psi j_{tx}} e^{-i2\pi \frac{4Bl s}{Tc} j_{time}}. \quad (5.2)$$

В случае наличия нескольких объектов на сцене с различными наборами параметров l, ϕ, ψ , массив данных $r(j_{rx}, j_{tx}, j_{time})$ представим в виде суммы конечного числа слагаемых, каждое из которых имеет формат (5.2). В более общем виде, тензор данных радара имеет структуру

$$\mathbf{T}(j_1, j_2 \dots j_d) = \sum_{r=1}^R \beta_r e^{i\gamma_{1,r} j_1} e^{i\gamma_{2,r} j_2} \dots e^{i\gamma_{d,r} j_d}, \quad (5.3)$$

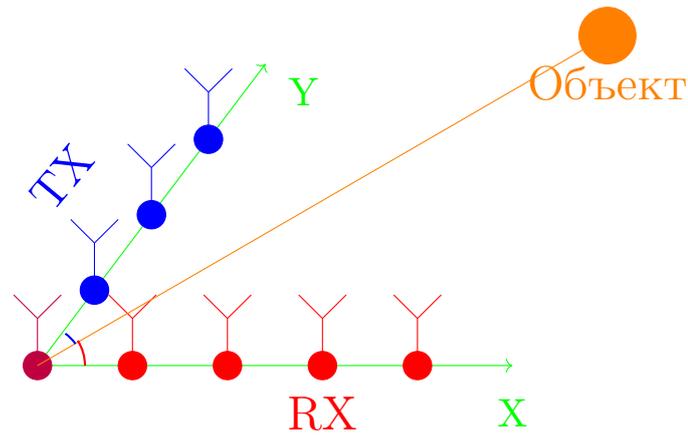


Рисунок 5.1 — Иллюстрация многоантенного радара

где параметры $\gamma_{k,r}$ находятся в прямой зависимости с физическими характеристиками расположения объектов. Использование многократного повторения временных интервалов излучения частотно-модулированного сигнала позволяет добавить к списку оцениваемых параметров объекта и его скорость. При этом в тензоре данных оказывается две размерности, соответствующих времени, но структура (5.3) сохраняется. Тогда тензор данных радара становится четырехмерным.

Структура массива данных (5.3) соответствует модели тензора канонического ранга R , с дополнительным ограничением на вид факторов канонического представления.

Основной математической задачей, связанной с радарными, в частности, автомобильными, является анализ многомерного тензора T формата (5.3), и определение параметров $\gamma_{k,r}$. При этом в приложениях, связанных с радарными, тензор T является существенно зашумленным.

Классическим алгоритмом обработки массивов данных радара является многомерное дискретное преобразование Фурье: по индексам разреженных пиков результата определяется приближение к $\gamma_{k,r}$. Проблемами такого подхода являются:

- Высокая сложность, зависящая как $O(M \log M)$ от общего числа M элементов тензора данных радара
- Фиксированная разрешающая способность: результат такого алгоритма $\hat{\gamma}_{k,r}$ принимает конечный набор значений. Расширенное преобразование Фурье позволяет повысить разрешение, но еще выше поднимает сложность.

5.1.2 Квантизация

В качестве альтернативы преобразованию Фурье предлагается построить для T приближение по модели тензора малого канонического ранга. При этом предлагается опираться на идею теоремы Крускала о единственности: естественно ожидать, что любое построенное приближение канонического формата будет автоматически иметь структуру факторов, близкую к структуре из (5.3), при условии попарного несовпадения параметров $\gamma_{k,r}$. Параметры $\gamma_{k,r}$ тогда предлагается определять путем анализа факторов полученного разложения; с учетом малого размера таких факторов для этого можно использовать большой спектр алгоритмов. Например, можно провести линейную интерполяцию фаз элементов каждого фактора.

При использовании модели тензора малого канонического ранга можно опираться на теорию, построенную в Главе 3, и проведенные в Главе 3 численные эксперименты. Структура (5.3) допускает введение искусственных размерностей, что также известно как „квантизация“: каждую размерность тензора линейного размера m_k можно рассматривать как совокупность s_k искусственных размерностей с длинами $q_1, q_2 \dots q_{s_k}$, где $m_k = q_1 q_2 \dots q_{s_k}$ - разложение m_k на простые множители, а s_k - число таких множителей. При этом смысл малоранговых тензорных моделей изменяется, но благодаря свойствам экспоненты и структуре (5.3) ранг канонического представления данных автомобильного радара не изменяется и остается равным числу объектов на сцене. Опираясь на основной результат Главы 3

$$\|\mathcal{P}_R(T + N) - T\|_F \sim \sqrt{drM^{\frac{1}{d}}} \frac{\|N\|_F}{\sqrt{M}},$$

где M - общее число элементов тензора, а N - тензор случайного белого шума, использование квантизованного канонического разложения дает наилучшие оценки погрешности $\|\mathcal{P}_R(T + N) - T\|_F$ и, соответственно, позволяет с наименьшей погрешностью восстановить расположение объектов на сцене в условиях зашумленных данных, типичных для автомобильных радаров.

5.1.3 Комбинирование моделей

Несмотря на преимущества высокой устойчивости к случайному белому шуму и наличия теории единственности, у квантизованной канонической модели тензора малого ранга есть недостаток: ограниченность методов построения таких разложений и ограниченность теоретических результатов о сходимости таких методов. Так, наиболее широко распространенный алгоритм построения канонического разложения ALS является итерационным, но число его итераций может быть неприемлемым: построение канонического разложения напрямую на основе ALS по сложности будет проигрывать многомерному преобразованию Фурье: при полной квантизации в случае $d = \log_2(M)$ сложность уже одной итерации ALS асимптотически совпадает со сложностью многомерного FFT.

В связи с этим возможно использование промежуточной модели, построение приближения по которой имеет сравнительно низкую сложность и позволяет впоследствии использовать его для построения финального канонического приближения. Примером может служить модель Таккера: в частности, в Главе 4 был предложен Алгоритм 7 быстрого построения приближений в формате Таккера с равными рангами R , сложность которого не зависит от общего числа элементов тензора данных M . Затем, для построения финального канонического разложения достаточно применить ALS к ядру Таккера; сложность такой операции будет зависеть только от размера ядра и иметь вид R^d , и таким образом, сложность всей комбинации алгоритмов будет асимптотически ниже M .

Однако, для конкретного приложения из области радаров, структура Таккера не подходит, так как не допускает полной квантизации, и более эффективной промежуточной моделью оказалось разложение в формате квантизованного тензорного произведения. Такое разложение строилось при помощи крестовой аппроксимации [52].

В Таблице 1 приведены численные результаты погрешности определения физических параметров радара на симуляционных данных. Данные соответствовали 12 отправляющим антеннам и 16 принимающим, и имели две временные размерности длины 512 и 256, по которым определялось расстояние до объекта и его скорость. Расстояние между антеннами h задавалось как половина длины электромагнитной волны, и число объектов (ранг тензора) задава-

Уровень шума в децибелах	Погрешность угла азимута, °	Погрешность угла подъема, °	Погрешность расстояния, м	Погрешность скорости, м/с
-7.7	0.27	0.29	0.092	0.024
-17.7	0.58	0.63	0.097	0.025
-27.7	0.95	1.06	0.105	0.073
-37.7	1.45	1.53	0.119	0.074

Таблица 1 — Численная погрешность определения физических параметров объектов на данных симуляций автомобильного радара

лось равным пяти. Эксперименты проводились на различных уровнях шума, превышающих по норме Фробениуса норму сигнала. Результаты демонстрируют высокую эффективность определения параметров объектов даже при очень больших уровнях шума.

5.2 Распространение сигналов по каналам беспроводной связи MIMO

Другой сферой возможных приложений модели канонического тензора малого ранга служит область беспроводной связи по протоколу MIMO („Multiple Input, Multiple Output“). Протокол предполагает существование базовой станции с большим количеством отправляющих антенн, и набора пользователей с мобильными устройствами, также имеющими несколько принимающих антенн. Отправляющие антенны часто организуются по расположению в виде двумерной равномерной решетки, расположенной в одной плоскости (см. рис 5.2), и передача данных ведется параллельно на определенном наборе частот и в различные моменты времени. Основным набором данных задачи служит многомерный тензор канала беспроводной связи, размерности которого и соответствуют времени, частоте и нескольким размерностям антенн. В силу схожести физических принципов распространения волны, структура этого тензора похожа на (5.3). но может включать в себя и коэффициенты амплитудного затухания (в

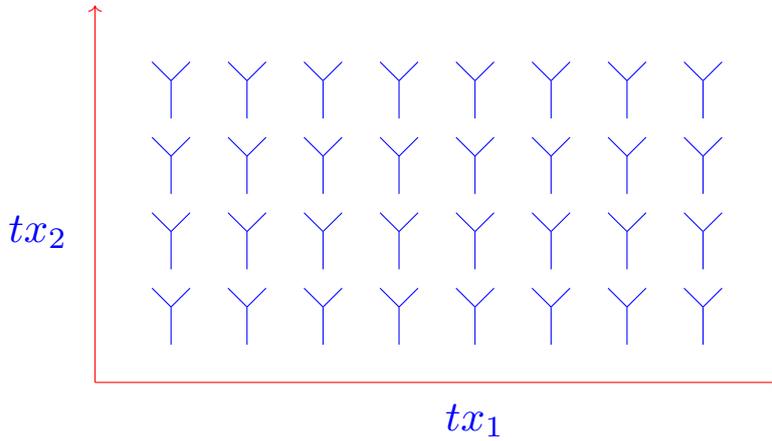


Рисунок 5.2 — Двумерная решетка антенн базовой станции беспроводной связи (частности, по времени): тензор канала имеет приближенный вид

$$\mathbf{T}(j_1, j_2 \dots j_d) = \sum_{r=1}^R \beta_r \kappa_1^{j_1} e^{i\gamma_{1,r} j_1} \kappa_2^{j_2} e^{i\gamma_{2,r} j_2} \dots \kappa_d^{j_d} e^{i\gamma_{d,r} j_d}, \quad (5.4)$$

где величины κ_k вещественны и близки к единице.

Для приложений в области беспроводных сетей имеет актуальность широкий спектр задач, где помогает модель квантизуемого тензора малого канонического ранга. Все эти задачи требуют решения в режиме реального времени, т.е., в течение миллисекунд. Среди этих задач существуют:

- Задача оценки канала: в отличие от приложения в области радаров, тензор канала не является наперед заданным, а измеряется по ходу передачи данных; при этом одно „измерение“ канала соответствует математически вычислению скалярного произведения тензора канала с тензором „пилота“ определенной структуры, а число доступных измерений ограничено.
- Задача сжатия канала: информация о канале должна быть доступна различным участникам сеанса связи, то есть, должна передаваться между базовой станцией и пользователями. Период времени, в который тензор канала можно считать неизменным, очень короткий, поэтому такую передачу необходимо проводить очень часто, и информация о канале требует эффективной параметризации;
- Задача *экстраполяции, предсказания* канала: изменчивость канала беспроводной связи со временем и необходимость предобработки отправляемых с базовой станции сигналов (например, построения матриц предо-

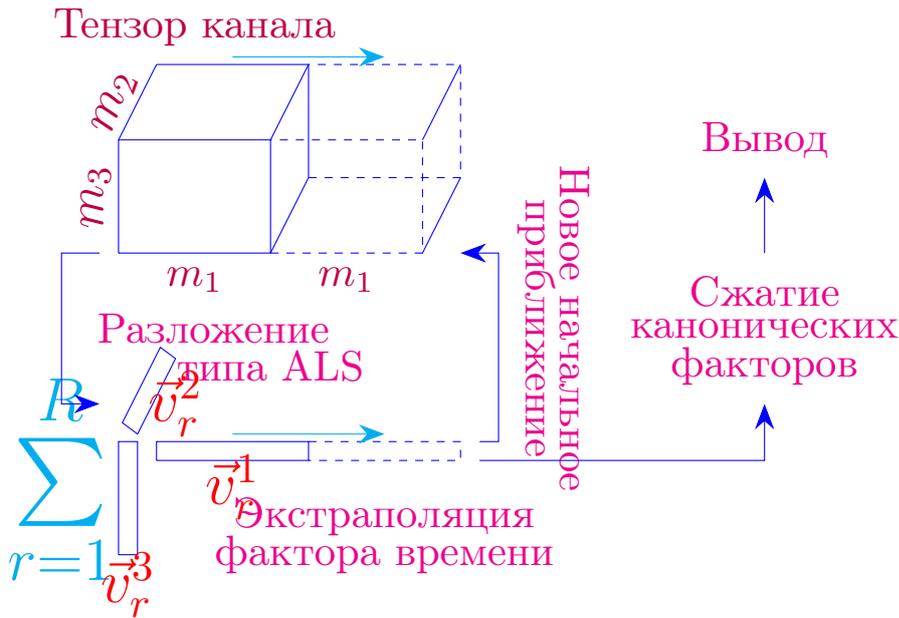


Рисунок 5.3 — Структура алгоритма экстраполяции канала

буставливания) приводит к необходимости знать канал наперед, иначе сигналы на отправляющих антеннах будут сформированы по информации о канале, уже неактуальной в реальном времени. При использовании модели канонического тензора малого ранга в качестве такой процедуры экстраполяции можно использовать экстраполяцию временного фактора тензорного разложения. Исходя из структуры фактора (5.4), достаточно вычислить фазы и логарифмы модуля каждого элемента временного канонического фактора, а затем экстраполировать их с помощью прямой. Этот подход проиллюстрирован на рис. 5.3. С помощью такого подхода удалось провести экстраполяцию канала, сгенерированного по модели 3GPP NLOS в симуляторе Quadriga 2.6.1 вперед на 32 интервала измерения канала по времени с погрешностью, сравнимой с погрешностью стандартного протокола 3GPP Release16.

5.3 Численные методы для интегральных и дифференциальных уравнений

5.3.1 Нелинейное уравнение магнитостатики

Редукция к модели матриц малого ранга имеет применение и в решении нелинейных дифференциальных уравнений. В частности, рассмотрим двумерное уравнение магнитостатики, имеющее приложение к моделированию электродвигателей ([28], [26], [53]).

$$\nabla(\nu(\|\nabla\phi(x,y)\|; x,y)\nabla\phi(x,y)) = \hat{f}(x,y), (x,y) \in \Omega, \quad (5.5)$$

где функция ν моделирует нелинейные магнитные эффекты материалов и имеет вид

$$\begin{cases} \nu(s; x,y) = 4\pi 10^{-7}, (x,y) \notin \Omega_n \\ \nu(s; x,y) = \frac{s^3}{s^4 + 6.25 \times 10^{10}} + 4\pi 10^{-7}, (x,y) \in \Omega_n \end{cases}$$

где область Ω_n соответствует расположению материала с нелинейными свойствами. Применение смешанного метода конечных элементов контрольного объема к задаче приводит к нелинейной системе уравнений вида

$$Ay_\mu + F_\mu(y_\mu) = 0, \quad (5.6)$$

где $A \in \mathbb{R}^M \rightarrow \mathbb{R}^M$ - линейный оператор, а $F_\mu \in \mathbb{R}^M \rightarrow \mathbb{R}^M$ - нелинейный оператор, зависящий от параметров μ , и размер системы M велик. Моделирование электродвигателей предполагает решение такой задачи большое количество раз: рассматриваемое уравнение магнитостатики в приложении является подзадачей в более крупной временной численной схеме с вращающимся ротором двигателя и соответствует одной итерации шага по времени явной схемы ([28], [26], [53]). Одним из параметров задачи, обозначенных μ , таким образом является угол поворота ротора двигателя; этот параметр нетривиально меняет геометрию задачи. Более простым параметром является сила тока, которая сводится к изменению правой части уравнения (5.5).

Покажем, как к нелинейным системам уравнений типа (5.6) применима модель матриц малого ранга. Так, метод POD ('Proper Orthogonal Decomposition', [24], [54]) основан на поиске решения этой системы в линейном подпространстве малого ранга, построенном на основе статистических данных о решениях при различных значениях параметров. Обозначая y_μ решение (5.6), соответствующее параметру μ , метод POD использует следующее предположение:

$$\text{numrank}_\epsilon([y_{\mu_1} y_{\mu_2} \dots y_{\mu_k}]) \approx \text{const}, \forall k \quad (5.7)$$

Здесь, $\text{numrank}_\epsilon(B)$ обозначает численный ранг матрицы B , равный $\max\{i : \sigma_i(B) > \epsilon \sigma_1(B)\}$, где $\sigma_j(B)$ соответствует j -тому сингулярному числу матрицы. Тогда, базис линейного подпространства V ранга r_a выбирается как линейная оболочка сингулярного подпространства матрицы B , построенной по $k > r_a$ (но $k \ll M$) 'типичным' решениям ('снимкам'), соответствующих 'типичным' параметрам.

Тогда, численная схема проектируется на V , и имеем

$$V^T AV x_\mu + V^T F_\mu(V x_\mu) = 0.$$

Сложность вычисления слагаемого $V^T F(V x_\mu)$, по-прежнему зависит от M в силу нелинейности, поэтому далее можно применить метод дискретной эмпирической интерполяции (DEIM, [24]). Приближим F_μ с помощью проектирования PF_μ :

$$V^T AV x_\mu + V^T PF_\mu(V x_\mu) = 0.$$

Проектор P для снижения сложности следует выбирать таким образом, чтобы вычисление Px требовало знания лишь небольшого числа элементов x . Поэтому, проектор $P : \mathbb{R}^M \rightarrow \mathbb{R}^M$ имеет структуру

$$P = U(S^T U)^{-1} S^T, \quad (5.8)$$

где U размера $M \times r_i$ (r_i - ранг интерполяции, r_a - ранг аппроксимации, $r_i > r_a$), а S содержит только один ненулевой элемент в каждом столбце:

$$S = [e_{i_1} \ e_{i_2} \ \dots \ e_{i_{r_i}}].$$

Структура P предполагает что элементов $y_{i_1} \dots y_{i_{r_i}}$, „выбранных“ столбцами S , достаточно для вычисления Py , и лишь небольшое число элементов $F(Vx)$ необходимо для вычисления $V^T P F(Vx)$. Проектор P также называют интерполяционным, так как он сохраняет значения вектора в „выбранных“ позициях $i_1 \dots i_{r_i}$. Множество таких позиций обозначим I_{int} .

Пусть уравнения численной схемы (5.6) локальны: для каждого j , значения $F(y)_j$ зависят только от небольшого числа элементов y (но, возможно, не одного). Обозначим за $C(j)$ соответствующий шаблон локальности, требуемый для вычисления $F(y)_j$. Пусть $C(I_{int}) = \cup_k C(i_k)$. Тогда введем диагональную $D_{C(I_{int})}$, такую что

$$\begin{cases} (D_{C(I_{int})})_{jj} = 1, j \in C(I_{int}) \\ (D_{C(I_{int})})_{jj} = 0, j \notin C(I_{int}) \end{cases}$$

Откуда имеем $V^T P F(Vx) = V^T P F(D_{C(I_{int})} Vx)$, поэтому малое число элементов Vx достаточно для вычисления $V^T P F(Vx)$. Базис U предлагается выбирать, аналогично V , статистически на основании „типичных“ значений $F(Vx)$. Имеем систему

$$V^T A V x_\mu + V^T U (S^T U)^{-1} S^T F_\mu (D_{C(I_{int})} V x_\mu) = 0.$$

Эта система по-прежнему нелинейна, но содержит гораздо меньше неизвестных, чем (5.6). Ее предлагается решать итерационно стандартными методами, например, методом Ньютона с доверительными интервалами [55]. При этом матрицы $V^T A V$ и $V^T U (S^T U)^{-1}$ могут быть предрасчитаны один раз, а сложность каждой итерации метода Ньютона никак не будет зависеть от M .

Специфика конкретной численной схемы, связанной с задачей (5.5) требует аккуратной обработки вращений. Движущиеся объекты могут нарушить предположение о малом ранге (5.7). Например, два решения уравнения, соответствующие различным параметрам сдвига, могут графически выглядеть, как указано на рис. 5.4.

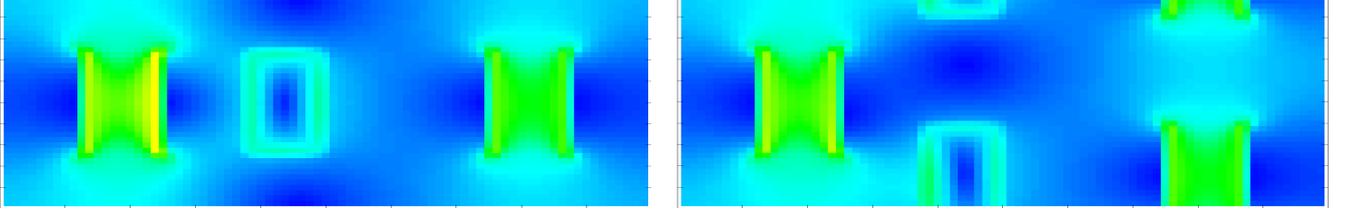


Рисунок 5.4 — Визуализация решений уравнения (5.5) при различных параметрах сдвига ротора

Параметр сдвига является определяющим в моделировании электродвигателей, поэтому для его обработки предлагается использовать перестановку базисов V, U под каждый вариант параметра сдвига; при этом предполагается, что сдвиг дискретен и соответствует по величине узлам численной сетки. Математически, это описывается умножением базисных матриц на матрицу перестановки Z

$$V \rightarrow ZV, U \rightarrow ZU, S \rightarrow ZS, \quad (5.9)$$

которая приводит к следующей трансформации системы:

$$\begin{aligned} V^T AV x_\mu + V^T U (S^T U)^{-1} S^T F_\mu (D_{C(I_{int})} V x_\mu) &= 0 \rightarrow \\ \rightarrow V^T Z^T AZ V x_\mu + V^T U (S^T U)^{-1} S^T Z^T F_\mu (Z D_{C(I_{int})} V x_\mu) &= 0 \end{aligned}$$

Из этих трансформаций видно:

- Вспомогательная матрица $V^T U (S^T U)^{-1}$ присутствует в той же форме (не зависит от перестановки), и не требует дополнительных подсчетов;
- $S^T F_\mu (D_{C(I_{int})} y)$ преобразуется в $S^T Z^T F_\mu (Z D_{C(I_{int})} y)$, $y = Vx$, что соответствует изменению порядка вершин, участвующих в интерполяции F . Наборы элементов вектора y , требуемые для вычисления величин $(F_\mu(y))_j$ и $(Z^T F_\mu(Zy))_j$ - отличаются для индексов j , соответствующих границе сдвига. Так как таких вершин гораздо меньше, чем M , предлагается добавить соответствующие им индексы в множество $C(I_{int})$, что требует сохранения небольшого количества дополнительных строк V , соответствующего одномерному размеру сетки, т.е. имеющему порядок $O(\sqrt{M})$ в двумерном случае.
- Вспомогательная матрицы $V^T AV$ меняется на $V^T Z^T AZV$, что требует возможности построения операции такого пересчета малой сложности.

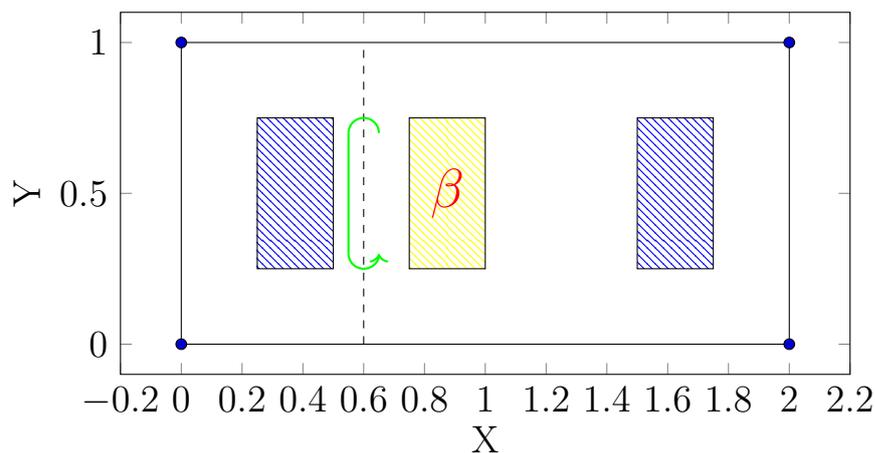


Рисунок 5.5 — Схема модельной части двигателя; заштрихованные синим области соответствуют областям материалов с нелинейными свойствами, а желтая область соответствует позиции ненулевой правой части уравнения.

Эта операция существует для конкретного вида A , соответствующего дискретизации уравнения магнитостатики на основе метода конечных элементов контрольных объемов; в таком случае, матрица A является блочной с диагональными и циркулянтными блоками, и рассмотренная перестановка изменяет A только в позициях вдоль границы сдвига.

Моделирование электродвигателей, как правило, сводится к решению задачи (5.5) в прямоугольнике в полярных координатах, т.е. секторе; такой сектор соответствует одной уникальной части ротора и статора, повторяющейся в двигателе периодическим образом. Для упрощения, в рамках этой работы методы редукции модели на основе матриц малого ранга с обработкой сдвига были применены к решению той же задаче в прямоугольнике в декартовых координатах: см. схему на рис. 5.5.

Тестирование проводилось при следующих параметрах

- Использовались равномерные сетки размера 40×80 или 80×160 ; соответствующее число уравнений дискретизованных систем равно $M = 38480$.
- Рассматривалось два параметра: сила тока (константа правой части) и нелинейный сдвиг. Каждый параметр имел 20 возможных значений.
- 80 (из возможных 400) случайных комбинаций параметров выбиралось для построения „снимков“. иными словами, 80 раз система вида (5.6)

была решена методом Ньютона без редукции с использованием матриц малого ранга

- 2000 случайных экземпляров значений $F(Vx)$ использовалось для построения статистического U . Вектора x выбирались случайно, а базис V соответствовал построенному по явно вычисленным решениям.

Затем, оставшиеся 320 решений были вычислены по рассмотренному методу редукции моделей с помощью матриц малого ранга. Полученные значения ошибки приведены в Таблице 2; в отсутствие аналитических решений у задачи „ошибка“ понимается как расстояние между решением задачи, полученным по полной численной схеме, и решением задачи, полученным по редуцированной модели. При этом, соответствующие времена практического поиска полного и редуциро-

	L_2 relative error	C-norm relative error
Электрическое поле, 40×80 , $r_a = 40$, $r_i = 300$	0.001793	0.003583
Магнитные потоки, 40×80 , $r_a = 40$, $r_i = 300$	0.001581	0.001942
Электрическое поле, 80×160 , $r_a = 40$, $r_i = 600$	0.000624	0.002642
Магнитные потоки, 80×160 , $r_a = 40$, $r_i = 600$	0.000328	0.002374

Таблица 2 — Таблицы экспериментально полученных ошибок редуцированных решений уравнения магнитостатики

ванного решений приведены в Таблице 3. Приближенное решение, таким образом, находилось в 100 – 1000 раз быстрее, чем решение полной численной схемы (5.6). Таким образом, если необходимо найти все 400 решений при всевозможных значениях параметров, сложность такой процедуры с помощью редукции к модели матриц малого ранга была снижена примерно в 5 раз.

	40×80 , $r_a = 40$, $r_i = 300$	80×160 , $r_a = 40$, $r_i = 600$
Решение полной задачи	3.384 сек	41.13 сек
Считывание вспомогательных данных	0.0254 сек	0.0533 сек
Преобразование $V^T AV$	0.000131 сек	0.000286 сек
Итерации с редукцией	0.007770 сек	0.0154 сек

Таблица 3 — Сравнение времени работы программ поиска полного и редуцированного решений уравнения магнитостатики

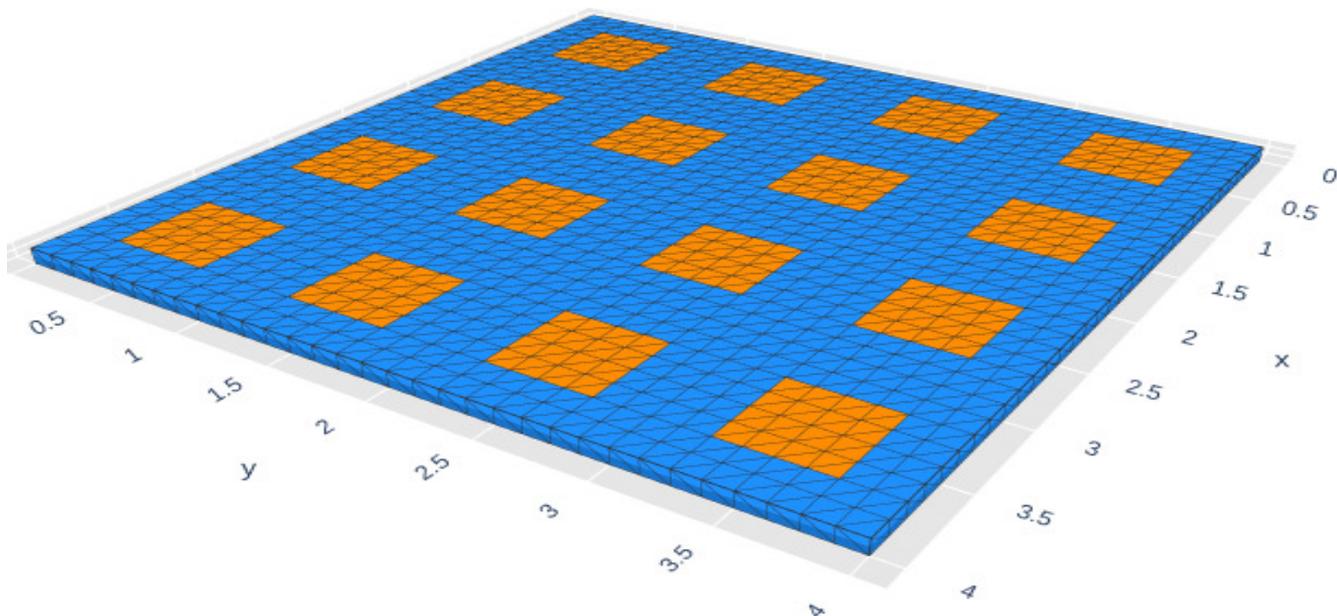


Рисунок 5.6 — Метаповерхность и введенная на ней сетка метода конечных элементов.

5.3.2 Уравнения рассеяния на метаповерхностях

Модель суммы матрицы малого ранга и разреженной матрицы была успешно применена к задаче сжатия большой структурированной плотной матрицы линейной системы, возникающей из конечноэлементной дискретизации интегральных уравнений, связанных с задачей рассеяния на метаповерхностях. Под словом „метаповерхность“ понимается конструкция из проводящих и диэлектрических частей, организованных в форме блоков одинаковой формы: пример приведен на рисунке 5.6.

Рассматриваемая линейная система основана на дискретизации уравнений Максвелла, используя базисные функции Рао-Вилтона-Глиссона. Задача включает в себя два потенциальных оператора, описывающих электрическое и магнитное поле на метаповерхности:

$$(\mathcal{E}(p))(x) = ik \int_{\Gamma} p(y)G_k(x, y)dy - \frac{1}{ik} \nabla_x \int_{\Gamma} \nabla_{\Gamma} \cdot p(y)G_k(x, y)dy,$$

$$(\mathcal{M}(p))(x) = \nabla_x \times \int_{\Gamma} p(y)G_k(x, y)dy,$$

где $G_k(x, y)$ - функция Грина для уравнения Гельмгольца с волновым числом k

$$G_k(x, y) = G_k(r) = \frac{-e^{-ikr}}{4\pi r}, \quad r = \|x - y\|_2.$$

Важным для приложения свойством приведенных уравнений является инвариантность к параллельному переносу, что, вкупе со структурой метаповерхности, приводит к блочно дважды-теплицевой структуре линейной системы численных схем. Каждый блок такой блочной матрицы описывает физическое взаимодействие двух частей-блоков метаповерхности. Таким образом, если метаповерхность дискретизуется посредством $M_1 \times M_2$ двумерной сетки из одинаковых блоков, и n - размер блока, достаточно хранить $(2M_1 - 1)(2M_2 - 1)$ блоков размера $\mathbb{C}^{n \times n}$, чтобы полностью описать рассматриваемую матрицу линейной системы размера $\mathbb{C}^{M_1 M_2 n \times M_1 M_2 n}$.

В рамках данной работы модель суммы матрицы малого ранга и разреженной была применена для дальнейшего сжатия матрицы линейной системы посредством построения модельных приближений к внутренним $n \times n$ блокам дважды-теплицевой структуры; в частности, тем, которые описывают физическое взаимодействие находящихся вдали друг от друга частей метаповерхности.

Эксперименты проводились на дискретизации метаповерхности размера 4×4 , с размером блока $n = 2542$, что соответствует плотной квадратной линейной системе с 40672 уравнениями. В Таблице 4 приведены характеристики полученного сжатия: „эффективность сжатия“ соответствует числу параметров сжатой матрицы по сравнению с хранением всей матрицы в плотном формате. Первая строка соответствует сжатию без потерь с использованием только дважды-теплицевой структуры и симметрии. Последующие две строки соответствуют дальнейшему сжатию внутренних плотных блоков дважды теплицевой структуры по модели суммы матрицы малого ранга и разреженной. Сжатие проводилось посредством двухмасочного алгоритма, а параметры ранга и разреженности подбирались под получение predetermined погрешности приближения всей структурированной матрицы во фробениусовой норме.

Размер блока	Тип сжатия	Погрешность сжатия	Эффективность сжатия
2542	Только дважды-тепиц. и симметрия	0	9.57 %
2542	Со сжатием внутренних блоков	1.0e-4	2.02 %
2542	Со сжатием внутренних блоков	1.0e-3	1.21 %

Таблица 4 — Эффективность сжатия матрицы численной схемы для уравнения рассеяния на метаповерхности с различной погрешностью

Заключение

Основные результаты работы заключаются в следующем.

1. Для метода проекции градиента, применительно к задаче матричного восполнения, построена теория применимости приближенного проектирования. Показана эффективность полученного алгоритма с точки зрения снижения требуемого числа операций и доказана теорема о сохранении геометрического характера сходимости.
2. Исследована возможность получения алгоритмов восполнения матриц, устойчивых к разреженной ошибке в начальных данных. Предложен и программно реализован алгоритм, позволяющий эффективно приближать матрицы в виде суммы матрицы малого ранга и разреженной. Построена теория сходимости этого алгоритма на основе теории возмущений сингулярных подпространств матриц.
3. Построена теория устойчивости тензорных приближений малого ранга с доказательной базой, при выполнении определенных условий на качество приближения. Асимптотическая точность теории подтверждена на численных экспериментах.
4. На основе методов восполнения предложен алгоритм построения тензорных приближений по модели Таккера малой вычислительной сложности.
5. Предложенная теория и разработанные алгоритмы получили широкое применение в области обработки сигналов беспроводной связи.

Список сокращений и условных обозначений

РАН - Российская академия наук

ALS - Alternating Least Squares, метод построения приближений тензоров малого канонического ранга [23]

ASVP - Approximate Singular Value Projection, метод проекции градиента для решения задачи восполнения матриц с приближенным проектированием [29]

DEIM - Discrete Empirical Interpolation, метод интерполяции нелинейных систем [24]

FFT - Fast Fourier Transform, быстрое преобразование Фурье

HOSVD - Higher Order Singular Value Decomposition, классический алгоритм построения приближений в формате Таккера

maxvol - Maximum Volume, алгоритм поиска подматрицы большого объема [6]

MIMO - Multiple Input Multiple Output, протокол беспроводной связи

PDF - Portable Document Format

POD - Proper Orthogonal Decomposition, метод редукции моделей для систем уравнений [54]

RIP - Restricted Isometry Property, свойство ограниченной изометрии

SVD - Singular Value Decomposition, сингулярное разложение матрицы

SVP - Singular Value Projection, метод проекции градиента для решения задачи восполнения матриц [29]

TT - Tensor Train, тензорный поезд

Список литературы

1. *Mirsky L.* Symmetric gauge functions and unitarily invariant norms // The quarterly journal of mathematics. — 1960. — т. 11, № 1. — с. 50—59.
2. *Udell M., Townsend A.* Why are big data matrices approximately low rank? // SIAM Journal on Mathematics of Data Science. — 2019. — т. 1, № 1. — с. 144—160.
3. *Osinsky A.* Rectangular maximum volume and projective volume search algorithms // arXiv preprint arXiv:1809.02334. — 2018.
4. How to find a good submatrix / S. Goreinov [и др.] // Matrix Methods: Theory, Algorithms, Applications, V. Olshevsky and E. Tyrtyshnikov, eds., World Scientific, Hackensack, NY. — 2010. — с. 247—256.
5. *Goreinov S. A., Tyrtyshnikov E. E., Zamaraashkin N. L.* A theory of pseudoskeleton approximations // Linear algebra and its applications. — 1997. — т. 261, № 1—3. — с. 1—21.
6. *Osinsky A., Zamaraashkin N. L.* Pseudo-skeleton approximations with better accuracy estimates // Linear Algebra and its Applications. — 2018. — т. 537. — с. 221—249.
7. *Halko N., Martinsson P.-G., Tropp J. A.* Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions // SIAM review. — 2011. — т. 53, № 2. — с. 217—288.
8. *Boutsidis C., Drineas P., Magdon-Ismail M.* Near-optimal column-based matrix reconstruction // SIAM Journal on Computing. — 2014. — т. 43, № 2. — с. 687—717.
9. *Kang Z., Peng C., Cheng Q.* Top-n recommender system via matrix completion // Proceedings of the AAAI Conference on Artificial Intelligence. т. 30. — 2016.
10. *Ahmed A., Romberg J.* Compressive multiplexing of correlated signals // IEEE Transactions on Information Theory. — 2014. — т. 61, № 1. — с. 479—498.
11. *Davies M. E., Eldar Y. C.* Rank awareness in joint sparse recovery // IEEE Transactions on Information Theory. — 2012. — т. 58, № 2. — с. 1135—1146.

12. Low-Complexity and Basis-Free Channel Estimation for Switch-Based mmWave MIMO Systems via Matrix Completion / R. Hu [и др.] // arXiv preprint arXiv:1609.05693. — 2016.
13. *Argyriou A., Evgeniou T., Pontil M.* Convex multi-task feature learning // Machine learning. — 2008. — т. 73, № 3. — с. 243—272.
14. *Blei D., Carin L., Dunson D.* Probabilistic topic models // IEEE signal processing magazine. — 2010. — т. 27, № 6. — с. 55—65.
15. Surveillance video coding via low-rank and sparse decomposition / C. Chen [и др.] // Proceedings of the 20th ACM international conference on Multimedia. — 2012. — с. 713—716.
16. Video-SAR imaging of dynamic scenes using low-rank and sparse decomposition / M. Moradikia [и др.] // IEEE Transactions on Computational Imaging. — 2021. — т. 7. — с. 384—398.
17. Sparse and low-rank matrix decompositions / V. Chandrasekaran [и др.] // IFAC Proceedings Volumes. — 2009. — т. 42, № 10. — с. 1493—1498.
18. *Almeida A. L. de.* Tensor modeling and signal processing for wireless communication systems : дис. . . . канд. / de Almeida André LF. — Université de Nice Sophia Antipolis, 2007.
19. *Wang S.* Multidimensional Radar Signal Processing Based on Sparse Fourier Transforms : дис. . . . канд. / Wang Shaogang. — Rutgers The State University of New Jersey, School of Graduate Studies, 2019.
20. Optimal resource allocation in coordinated multi-cell systems / E. Björnson, E. Jorswieck [и др.] // Foundations and Trends® in Communications and Information Theory. — 2013. — т. 9, № 2/3. — с. 113—381.
0. *Sadek M., Tarighat A., Sayed A. H.* A leakage-based precoding scheme for downlink multi-user MIMO channels // IEEE transactions on Wireless Communications. — 2007. — т. 6, № 5. — с. 1711—1721.
21. *De Lathauwer L., De Moor B., Vandewalle J.* A multilinear singular value decomposition // SIAM journal on Matrix Analysis and Applications. — 2000. — т. 21, № 4. — с. 1253—1278.

22. *Oseledets I. V., Tyrtysnikov E. E.* Breaking the curse of dimensionality, or how to use SVD in many dimensions // SIAM Journal on Scientific Computing. — 2009. — т. 31, № 5. — с. 3744—3759.
23. *Comon P., Luciani X., De Almeida A. L.* Tensor decompositions, alternating least squares and other tales // Journal of Chemometrics: A Journal of the Chemometrics Society. — 2009. — т. 23, № 7/8. — с. 393—405.
24. *Chaturantabut S., Sorensen D. C.* Nonlinear model reduction via discrete empirical interpolation // SIAM Journal on Scientific Computing. — 2010. — т. 32, № 5. — с. 2737—2764.
25. *Sorensen D. C., Embree M.* A DEIM induced CUR factorization // SIAM Journal on Scientific Computing. — 2016. — т. 38, № 3. — A1454—A1482.
26. Transient simulation of an electrical rotating machine achieved through model order reduction / L. Montier [и др.] // Advanced Modeling and Simulation in Engineering Sciences. — 2016. — т. 3, № 1. — с. 10.
27. Structure Preserving Model Reduction of Low-Frequency Electromagnetic Problem Based on POD and DEIM / L. Montier [и др.] // IEEE Transactions on Magnetics. — 2017. — т. 53, № 6. — с. 1—4.
28. *Henneron T., Clénet S.* Model order reduction applied to the numerical study of electrical motor based on POD method taking into account rotation movement // International Journal of Numerical Modelling: Electronic Networks, Devices and Fields. — 2014. — т. 27, № 3. — с. 485—494.
29. *Meka R., Jain P., Dhillon I. S.* Guaranteed rank minimization via singular value projection // arXiv preprint arXiv:0909.5457. — 2009.
30. *Recht B.* A Simpler Approach to Matrix Completion. // Journal of Machine Learning Research. — 2011. — т. 12, № 12.
31. *Tanner J., Wei K.* Normalized iterative hard thresholding for matrix completion // SIAM Journal on Scientific Computing. — 2013. — т. 35, № 5. — S104—S125.
32. *Klopp O.* Matrix completion by singular value thresholding: sharp bounds // Electronic journal of statistics. — 2015. — т. 9, № 2. — с. 2348—2369.
33. *Drineas P., Mahoney M. W.* Lectures on randomized numerical linear algebra // The Mathematics of Data. — 2018. — т. 25, № 1.

34. *Goreinov S. A., Tyrtyshnikov E. E., Zammarashkin N. L.* A theory of pseudoskeleton approximations // Linear algebra and its applications. — 1997. — т. 261, № 1–3. — с. 1–21.
35. *Osinsky A., Zammarashkin N. L.* Pseudo-skeleton approximations with better accuracy estimates // Linear Algebra and its Applications. — 2018. — т. 537. — с. 221–249.
36. *Zammarashkin N., Osinsky A.* New accuracy estimates for pseudoskeleton approximations of matrices // Doklady Mathematics. т. 94. — Springer. 2016. — с. 643–645.
37. *Davis C., Kahan W. M.* Some new bounds on perturbation of subspaces // Bulletin of the American Mathematical Society. — 1969. — т. 75, № 4. — с. 863–868.
38. *Galántai A.* Subspaces, angles and pairs of orthogonal projections // Linear and Multilinear Algebra. — 2008. — т. 56, № 3. — с. 227–260.
39. *Wedin P.-Å.* Perturbation bounds in connection with singular value decomposition // BIT Numerical Mathematics. — 1972. — т. 12, № 1. — с. 99–111.
40. *Stewart G. W.* Perturbation theory for the singular value decomposition : tex. отч. — 1998.
41. *Lebedeva O., Osinsky A., Petrov S.* Low-Rank Approximation Algorithms for Matrix Completion with Random Sampling // Computational Mathematics and Mathematical Physics. — 2021. — т. 61, № 5. — с. 799–815.
42. *Becker S., Cevher V., Kyriillidis A.* Randomized low-memory singular value projection // arXiv preprint arXiv:1303.0167. — 2013.
43. *Arratia R., Gordon L.* Tutorial on large deviations for the binomial distribution // Bulletin of mathematical biology. — 1989. — т. 51, № 1. — с. 125–131.
44. *Petrov S., Zammarashkin N.* Matrix completion with sparse measurement errors // Calcolo. — 2023. — т. 60, № 1. — с. 9.
45. *Plan Y., Vershynin R., Yudovina E.* High-dimensional estimation with geometric constraints // Information and Inference: A Journal of the IMA. — 2017. — т. 6, № 1. — с. 1–40.

46. *Rudelson M., Vershynin R.* Non-asymptotic theory of random matrices: extreme singular values // Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures. — World Scientific. 2010. — c. 1576–1602.
47. *Vershynin R.* High-dimensional probability: An introduction with applications in data science. т. 47. — Cambridge university press, 2018.
48. *Dasgupta S., Gupta A.* An elementary proof of a theorem of Johnson and Lindenstrauss // Random Structures & Algorithms. — 2003. — т. 22, № 1. — c. 60–65.
49. *Zouzias A.* Randomized primitives for linear algebra and applications. — University of Toronto, 2013.
50. *Achlioptas D.* Database-friendly random projections: Johnson-Lindenstrauss with binary coins // Journal of computer and System Sciences. — 2003. — т. 66, № 4. — c. 671–687.
51. *Ailon N., Chazelle B.* The fast Johnson–Lindenstrauss transform and approximate nearest neighbors // SIAM Journal on computing. — 2009. — т. 39, № 1. — c. 302–322.
52. *Oseledets I., Tyrtshnikov E.* TT-cross approximation for multidimensional arrays // Linear Algebra and its Applications. — 2010. — т. 432, № 1. — c. 70–88.
53. *Henneron T., Clenet S.* Model order reduction of non-linear magnetostatic problems based on POD and DEI methods // IEEE Transactions on Magnetics. — 2014. — т. 50, № 2. — c. 33–36.
54. Certified reduced basis methods for parametrized partial differential equations / J. S. Hesthaven, G. Rozza, B. Stamm [и др.]. — Springer, 2016.
55. *Dennis Jr J. E., Schnabel R. B.* Numerical methods for unconstrained optimization and nonlinear equations. т. 16. — Siam, 1996.

Публикации автора по теме диссертации

Научные статьи, опубликованные в журналах Scopus, WoS, RSCI, а также в изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности

1. Petrov S., Zamarashkin N. Matrix completion with sparse measurement errors //Calcolo. – 2023. – Т. 60. – №. 1. – С. 9. (WoS IF 2.097, Q1; Scopus SJR 0.79, Q1 в 2022г.)
2. Petrov S. Model Order Reduction Algorithms in the Design of Electric Machines //International Conference on Large-Scale Scientific Computing. – Springer, Cham, 2019. – С. 140-147. (Scopus SJR 0.427, Q2 в 2019г.)
3. Lebedeva O. S., Osinsky A. I., Petrov S. V. Low-Rank Approximation Algorithms for Matrix Completion with Random Sampling //Computational Mathematics and Mathematical Physics. – 2021. – Т. 61. – №. 5. – С. 799-815 (WoS IF 0.769, Q2; Scopus SJR 0.503, Q3 в 2021 г.)

Список рисунков

1.1	Иллюстрация входных данных задачи восполнения матрицы . . .	14
1.2	Графики итоговой относительной невязки $\ \mathcal{A}_\Omega(X_t - X_*)\ _F / \ \mathcal{A}_\Omega(X_*)\ _F$ восполнения матриц с сингулярными числами вида $\sigma_i = 1$	39
1.3	Графики итоговой относительной невязки $\ \mathcal{A}_\Omega(X_k - X_*)\ _F / \ \mathcal{A}_\Omega(X_*)\ _F$ восполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{i}$	40
1.4	Графики итоговой относительной невязки $\ \mathcal{A}_\Omega(X_k - X_*)\ _F / \ \mathcal{A}_\Omega(X_*)\ _F$ восполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{i^2}$	41
1.5	Графики итоговой относительной невязки $\ \mathcal{A}_\Omega(X_k - X_*)\ _F / \ \mathcal{A}_\Omega(X_*)\ _F$ восполнения матриц с сингулярными числами вида $\sigma_i = \frac{1}{2^i}$	42
2.1	Поведение методов проекции градиента для восполнения матрицы с разреженной ошибкой данных в обычном сценарии. . .	70
2.2	Поведение методов проекции градиента для восполнения матрицы с разреженной ошибкой данных в сглаженном сценарии. . .	71
2.3	Поведение одномасочного варианта метода проекции градиента для восполнения матрицы с разреженной ошибкой данных в сценарии плохой обусловленности.	71
2.4	Поведение двухмасочного варианта метода проекции градиента для восполнения матрицы с разреженной ошибкой данных в сценарии плохой обусловленности.	72
2.5	Поведение методов проекции градиента для восполнения матрицы с разреженной ошибкой данных в сценарии большого ранга.	72
3.1	Схема анализа устойчивости тензорных приближений	79
3.2	Устойчивость тензорного приближения ранга 1 к „малому“ шуму, в 10 раз меньшему тензора по норме Фробениуса	86

3.3	Устойчивость тензорного приближения ранга 1 к „большому“ шуму, в 10 раз большему тензору по норме Фробениуса	87
3.4	Результаты работы алгоритма ALS с одним и тем же входным тензором, рассматриваемым как тензор различных размерностей	88
3.5	Зависимость устойчивости канонического тензорного приближения от ранга	90
3.6	Зависимость устойчивости тензорного приближения Таккера от ранга	91
3.7	Зависимость устойчивости тензорного приближения в формате тензорного поезда от ранга	93
4.1	Схема распределения известных элементов тензора на развертке .	96
4.2	Падение функционала невязки (4.1) с итерациями двух вариантов метода проекции градиента для восполнения тензоров в формате Таккера	99
5.1	Иллюстрация многоантенного радара	104
5.2	Двумерная решетка антенн базовой станции беспроводной связи .	108
5.3	Структура алгоритма экстраполяции канала	109
5.4	Визуализация решений уравнения (5.5) при различных параметрах сдвига ротора	113
5.5	Схема модельной части двигателя; заштрихованные синим области соответствуют областям материалов с нелинейными свойствами, а желтая область соответствует позиции ненулевой правой части уравнения.	114
5.6	Метаповерхность и введенная на ней сетка метода конечных элементов.	116

Список таблиц

1	Численная погрешность определения физических параметров объектов на данных симуляций автомобильного радара	107
2	Таблицы экспериментально полученных ошибок редуцированных решений уравнения магнитостатики	115
3	Сравнение времени работы программ поиска полного и редуцированного решений уравнения магнитостатики	115
4	Эффективность сжатия матрицы численной схемы для уравнения рассеяния на металлоповерхности с различной погрешностью	118