ОТЗЫВ официального оппонента
о диссертации на соискание ученой степени
доктора физико-математических наук Осинского Александра Игоревича
на тему: «Существование и построение близких к оптимальным
столбцовых и крестовых аппроксимаций матриц»
по специальности 1.1.6 Вычислительная математика

Актуальность темы диссертации.

Малоранговые аппроксимации матриц играют ключевую роль в различных областях математики, поскольку позволяют задавать большие наборы данных с помощью небольшого числа переменных. Так, на их основе можно строить тензорные поезда для приближения многомерных данных и уменьшать сложность операций умножения на вектор в итеративных алгоритмах. Особое значение при этом имеют столбцовые и крестовые аппроксимации, использующие небольшое число элементов исходной матрицы. На их основе можно находить набор наиболее важных параметров, что часто используется в задачах редукции размерности и, в частности, в рекомендательных системах. Кроме того, такие аппроксимации не требуют знания всех элементов исходной матрицы для их построения, что позволяет существенно снизить вычислительную сложность соответствующих алгоритмов.

Диссертационная работа Осинского А. И. посвящена теоретическому исследованию точности крестовых аппроксимаций и построению эффективных на практике алгоритмов их построения. При этом важно заметить, что из-за невозможности изучить все элементы матрицы, подобные алгоритмы, как уже существующие, так и новые предложенные, оказываются жадными, и не могут гарантировать высокой точности аппроксимации для всех матриц. В связи с этим оказывается необходим вероятностный подход. Таким образом, кроме доказательства существования аппроксимаций высокой точности, важным вопросом оказывается возможность их достижения с помощью быстрых алгоритмов.

Новизна диссертационной работы.

Такой вывод можно сделать только рассмотрев вероятностную модель, и доказав эффективность алгоритмов на ней с большой вероятностью. Именно это и было сделано в диссертации. Важно отметить, что в отличие от существующих оценок для рандомизированных алгоритмов, здесь сами алгоритмы являются детерминистическими, а случайными считаются сами матрицы.

Степень обоснованности положений, выносимых на защиту, научных выводов и рекомендаций, сформулированных в диссертации, их достоверность.

Достоверность работы следует из строгости математических рассуждений и доказательств и обоснованности использованных численных методов. Численное исследование методов аппроксимации, в том числе в приложении к задачам восстановления матриц и построения неотрицательных аппроксимаций, согласуется с полученными теоретическими оценками. Результаты диссертации опубликованы в 15 работах, индексируемых WoS и Scopus, и доложены на российских и международных конференциях.

Краткое содержание диссертации.

Диссертационная работа состоит из введения, семи глав, заключения и приложения.

Первая глава посвящена рассмотрению крестовых и столбцовых аппроксимаций в целом: какие бывают основные виды, как они связаны с LU и QR разложениями, какими свойствами “хорошие” аппроксимации должны обладать. Рассмотрена связь с задачей одновременной аппроксимации. Приведен вид наилучших столбцовых и крестовых аппроксимаций в спектральной норме и норме Фробениуса. Последняя часть главы посвящена свойствам подматриц большого объема и проективного объема. Даны определения объема, проективного объема и критерия их локальной максимальности. Доказано, что по столбцам, составляющим подматрицу большого объема, можно разложить остальные столбцы, используя в среднем небольшие коэффициенты.

Во второй главе приведены все основные теоретические оценки точности для крестовых и столбцовых аппроксимаций. Глава разбита на разделы, посвященные различным нормам – спектральной, норме Фробениуса и норме Чебышева. Для каждой нормы и каждого типа крестовой и столбцовой аппроксимаций доказаны как верхние оценки – на существование аппроксимаций высокой точности соответствующего вида, так и нижние – на то, достижение какой точности гарантировать нельзя. Для аппроксимаций по норме Чебышева рассмотрено обобщение на тензоры произвольной размерности, а также выведены нижние оценки точности аппроксимаций единичной матрицы. В конце полученные оценки для каждого вида аппроксимации сравниваются с наилучшими из полученных другими авторами.

В третьей главе рассматриваются вероятностные оценки точности. Показано, что использование подматриц локально максимального объема и проективного объема не может гарантировать высокой точности по спектральной норме или норме Фробениуса. Для обоснования использования такого критерия в алгоритмах автором рассматривается вероятностная модель, где сингулярные числа фиксируются, а сингулярные векторы выбираются случайно и равномерно, на основе меры Хаара. Для полученного ансамбля матриц, называемого randsvd, далее строятся столбцовые и крестовые аппроксимации на основе подматриц локально максимального объема и проективного объема и анализируется их точность. При этом получено два типа результатов – для матожидания погрешности по норме Фробениуса и для вероятности того, что погрешность по норме Фробениуса существенно выше средней. В последнем случае также рассматривается два типа результатов: для случая, когда сингулярные числа приближаемой матрицы убывают быстро, и когда сингулярные числа погрешности почти равны, что можно интерпретировать, например, как белый шум.

В четвертой главе рассматриваются различные алгоритмы поиска подматриц с небольшой нормой псевдообратных. Это оказывается особенно важно в связи с тем, что вероятностные оценки содержали не объем напрямую, а именно норму псевдообратной к найденной подматрице в определенном подпространстве. Норма псевдообратных также явно присутствовала в нижних оценках точности, с напоминания которых и начинается глава. Далее автор переходит уже к описанию алгоритмов. Сначала рассматриваются жадные алгоритмы набора столбцов. Затем на их основе строятся алгоритмы замены строк и столбцов для поиска подматриц локально максимального объема. Глава разбита на несколько частей, посвященных как поиску в фиксированных строках и столбцах, так и во всей матрице. Для каждого алгоритма доказывается вычислительная сложность и число шагов, требуемое для достижения близкого к максимальному объема. В конце главы рассматриваются алгоритмы, не связанные с поиском максимального объема, но также приводящие к ограничениям на нормы псевдообратной к подматрице, составленной из части столбцов исходной матрицы. Также на основе поиска подматриц локально максимального объема выводится алгоритм приближенного поиска минимального охватывающего эллипсоида набора векторов.

В пятой главе анализируется эффективность алгоритма последовательных замен строк и столбцов для поиска локально максимального объема. Известно, что такой алгоритм не может гарантировать близости объема найденной подматрицы к максимальному, а потому снова рассматривается вероятностная модель. На этот раз случайными задаются сингулярные векторы некоторого приближения исходной матрицы, после чего шум выбирается произвольным образом (то есть предполагается наихудшим). Показано, что если шум достаточно мал, стандартный алгоритм последовательных замен (maxvol) позволяет достичь высокой точности крестовой аппроксимации. Во второй части главы предлагается способ последовательного увеличения ранга аппроксимации для достижения заданной точности. Классический адаптивный крестовый метод модифицируется для того, чтобы максимизировать вероятность на каждом шаге найти элемент, близкий к максимальному. Выбор модификации при этом основан на только что полученных вероятностных оценках и предполагает увеличение числа шагов и выбора нескольких стартовых (тестовых) столбцов.

В шестой главе проводятся численные эксперименты на случайных матрицах. Хотя матрицы также генерируются из randsvd ансамбля, в отличие от доказанных в третьей главе оценок, здесь напрямую тестируются быстрые алгоритмы, основанные на поиске подматриц локально максимального объема в фиксированных строках и столбцах. Численные эксперименты показывают, что даже такие упрощенные алгоритмы позволяют достичь точности, близкой к наилучшей. Это наблюдается как в среднем, так и на худших сгенерированных случаях. Также показано, что даже если требовать от алгоритмов достижения точного локально максимального объема, число замен на практике оказывается ограниченным. Также заметно, что аппроксимации на основе подматриц большого проективного приводят не только к наименьшей средней погрешности, но также позволяют избежать худших случаев, показывая существенно более быстро убывающий хвост распределения погрешностей, чем, например, для алгоритма maxvol.

В седьмой главе рассмотрены применения алгоритмов крестовой аппроксимации, основанных на подматрицах локально максимального объема и проективного объема. Первое – вычисление правой части при решении дифференциальных уравнений типа Смолуховского. Здесь крестовый метод позволяет построить и использовать аппроксимацию, не вычисляя целиком матрицу ядра из правой части. Далее рассмотрен метод восстановления матриц, основанный на проекции градиента. Здесь крестовый метод используется в качестве проекции на множество матриц фиксированного ранга. Доказано, что, при высокой относительной точности проектора по норме Фробениуса, наблюдается геометрическая сходимость. В третьем примере рассматривается построение неотрицательных аппроксимаций методом переменных проекций, где крестовый метод также используется в качестве альтернативы сингулярному разложению, позволяя быстро проектировать на многообразие подматриц фиксированного ранга. В приложениях особенно подчеркивается низкая вычислительная сложность построения аппроксимаций, а также их способность достичь высокой точности по норме Фробениуса, требуемой в приложениях, что и позволило использовать крестовый метод в качестве приближенного проектора.

В приложении к диссертации выписан псевдокод для всех основных рассматриваемых алгоритмов.

Автореферат в полной мере передает содержание диссертации.

Замечания по диссертационной работе.

1. Диссертация содержит опечатки. Например:

   Стр. 25, 4 строка снизу: должно быть m <= n

   Стр. 29, формула (1.9): после первого равенства вместо $\tilde{A}$ должно быть $\tilde{Q}$

   Стр. 32 в доказательстве леммы 1.4 вместо $\tilde{A}$ должна быть A

   Стр. 36 в (2.9) вместо n должно быть k

Стр. 43, первая строка: в формуле для τ пропущено ε

Стр. 49, в последней формуле вместо P_r должно быть Z

Стр. 68, формула (2.68): точнее min(M,N) вместо N, согласно доказательству

Стр. 101, в начале главы следует определить $\tilde{A}$

При взятии min и max в некоторых местах не указана размерность или смысл переменной.

2. Некоторые обозначения не введены в тексте, хотя часть присутствует в списке обозначений. В частности, речь идет об использовании символа + для обозначения псевдообратной матрицы (чаще используется крест). Или обозначении П как ортогонального проектора.

3. Крестовые аппроксимации – довольно широко рассматриваемая тема с большим количеством оценок в точности. При этом среди источников литературы очень мало статей, посвященных CUR аппроксимациям. В результате неясно, насколько полученные результаты действительно превосходят другие известные работы по теме. Например, было бы интересно сравнение результатов диссертации с работой Rudelson, M., & Vershynin, R. (2005). Sampling from large matrices: An approach through geometric functional analysis. *J. ACM, 54*, 21.

Вместе с тем, указанные замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В. Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 1.1.6 Вычислительная математика (по физико-математическим наукам), а также критериям, определенным пп. 2.1-2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова, а также оформлена согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Московского государственного университета имени М.В. Ломоносова.

Таким образом, соискатель Осинский Александр Игоревич заслуживает присуждения ученой степени доктора физико-математических наук по специальности 1.1.6 Вычислительная математика.

Официальный оппонент:
доктор физико-математических наук, доцент,
профессор базовой кафедры Института проблем передачи информации имени А.А. Харкевича РАН
факультета компьютерных наук
ФГАОУ ВО Национальный исследовательский университет «Высшая школа экономики»

НАУМОВ Алексей Александрович

16.04.2025 г.

Контактные данные:

тел.: +7 (495) 772-95-90, 27358, e-mail: anaumov@hse.ru
Специальность, по которой официальным оппонентом защищена диссертация:
05.13.17 – теоретические основы информатики

Адрес места работы:

109028, Россия, г. Москва, Покровский бульвар, д. 11
Федеральное государственное автономное образовательное учреждение высшего образования Национальный исследовательский университет «Высшая школа экономики»
Базовая кафедра Института проблем передачи информации имени А.А. Харкевича РАН
Тел.: +7 (495) 772-95-90, 27358, e-mail: anaumov@hse.ru

Подпись сотрудника Федерального государственного автономного образовательного учреждения высшего образования Национальный исследовательский университет «Высшая школа экономики» А.А. Наумова удостоверяю: