



Российский университет
дружбы народов (РУДН)

ул. Миклухо-Маклая, д. 6, Москва, Россия, 117198
ОГРН 1027739189323; ОКПО 02066463; ИНН 7728073720

Телефон: +7495 434 53 00, факс: +7495 433 15 11
www.rudn.ru; rudn@rudn.ru

4 марта 20 19
№ 0240-19-22/41

УТВЕРЖДАЮ

Первый проректор – проректор
по научной работе

ФГАОУ ВО «Российский университет
дружбы народов» (РУДН)



Н.С. Кирабаев

2019 г.

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

федерального государственного автономного образовательного учреждения высшего образования «Российский университет дружбы народов» (РУДН) на диссертационную работу Азаряна Алекса Артуровича «Быстрые алгоритмы моделирования многомерных линейных регрессионных зависимостей на основе метода наименьших модулей», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.18 — математическое моделирование, численные методы и комплексы программ

Актуальность диссертационной работы

В задачах мониторинга и диагностики многомерных систем достаточно распространено использование регрессионных моделей, и в связи с этим в та-

ких приложениях алгоритмы численного оценивания коэффициентов моделей за ограниченное время и с приемлемой точностью представляют значительный интерес. Так как при оценивании этих коэффициентов во многих случаях приходится сталкиваться со стохастической неоднородностью экспериментальных данных, то в этих условиях их оценивание требуется выполнять с помощью более устойчивых методов чем метод наименьших квадратов. К их числу относят метод наименьших модулей (МНМ) и обобщенный метод наименьших модулей (ОМНМ). Однако в настоящее время вопрос обеспечения высокой вычислительной эффективности моделирования регрессионных зависимостей на основе указанных методов решен не в полной мере. Использование известных точных алгоритмы на основе идей линейного программирования ограничено малыми размерностями исходных данных и числа наблюдений. Приближенные алгоритмы для достижения приемлемой точности требуют значительных вычислительных затрат. Поэтому представляется актуальной разработка единого подхода к вычислительно эффективному моделированию многомерных линейных регрессионных зависимостей в условиях стохастической неоднородности на основе МНМ и ОМНМ, не имеющих ограничений на порядок моделей и объем экспериментальных данных и проведение исследований для его теоретического обоснования.

Общая характеристика и содержание диссертации

Диссертация состоит из введения, 4 глав, заключения, списка литературы и приложений. Полный объем составляет 148 страниц, включая 21 рисунок, 27 таблиц, список литературы из 135 наименований и 3 приложения.

Во Введении обосновывается актуальность и научная новизна работы, ставятся цели работы и формулируются основные задачи для достижения поставленных целей, приводятся методы исследования, содержится информация по апробации работы, описана структура диссертации и ее краткое содержание.

Первая глава посвящена обзору устойчивых методов моделирования линейных регрессионных зависимостей. В первом разделе рассматривается проблема устойчивости моделирования линейных зависимостей в условиях стохастической неоднородности. При оценивании параметров линейной модели, использование классических процедур на основе МНК, ориентированных на выполнение основных предпосылок математической статистики может привести к грубым ошибкам. Особенно это проявляется, если в экспериментальных данных имеются аномальные наблюдения в виде выбросов, а также при нестационарности случайных ошибок. В этой ситуации используют устойчивые (робастные и непараметрические) методы оценивания. В их основе лежит метод наименьших модулей (МНМ). Используемая при оценивании параметров регрессии функция потерь либо является модулем, либо функцией от модуля. Наиболее актуальным является обеспечение устойчивости оценок параметров моделей временных рядов, поскольку они гораздо более чувствительны к стохастической неоднородности данных, ввиду зависимости наблюдений. В результате при расчете коэффициентов моделей временных рядов, неустойчивыми к выбросам оказываются и МНМ-оценки. Недостатком МНМ также является то, что, он рассчитан на симметричный характер выбросов. С целью устранения этих недостатков используют оценки с функциями потерь, имеющими горизонтальную асимптоту. Одним из таких методов является обобщенный метод наименьших модулей (ОМНМ).

Во втором разделе описывается метод наименьших модулей, который представляет собой минимизацию суммы модулей невязок. Целевая функция $Q(a)$ является выпуклой, непрерывной и кусочно-линейной, но у нее имеется большое число изломов в виде отрезков прямых, причем они вытянуты, скрещены, многие из них почти параллельны и находятся на близком расстоянии друг от друга, образуя пучок. Минимум у $Q(a)$ находится внутри этого пучка. Стенки изломов представляют собой выпуклые линейные гиперплоскости, которые при при-

ближении к минимуму целевой функции постепенно становятся почти параллельными. Это к затруднениям при реализации разных алгоритмов нахождения точного решения задачи на основе идей линейного программирования. Другим методом для нахождения точного решения задачи является алгоритм полного перебора, который осуществляет перебор всех особых точек, в которых не существует производная функции $Q(a)$ ни по одному из возможных направлений. Но для этого алгоритма с ростом объема выборки и количества переменных наблюдается экспоненциальный рост вычислительных затрат. Приближенным методом решения задачи является итерационный метод вариационно-взвешенных квадратических приближений, для которого в точке решения задачи m невязок равны нулю. Для решения этой проблемы используется прием регуляризации. В случае регуляризации, вопрос о сходимости алгоритма вариационно-взвешенных-квадратических приближений остается открытым. Для приближенного решения задачи также применяют методы поиска безусловного экстремума нулевого порядка. В этом случае при повышении точности наблюдается быстрый рост вычислительных затрат.

В третьем разделе описывается ОМНМ. Как и в случае МНМ, для нахождения точного решения задачи можно применить переборный алгоритм. Но с ростом объема выборки и количества переменных наблюдается экспоненциальный рост вычислительных затрат. Итерационный алгоритм для нахождения решения задачи основанный на идеях линейного программирования не гарантирует сходимость к точному решению. Повышение точности оценивания за счёт осреднения многократного повторного формирования подвыборок приводит к росту вычислительных затрат.

Во второй главе описан и исследован новый подход к вычислительно эффективному построению многомерных линейных регрессионных моделей в условиях стохастической неоднородности на основе МНМ и ОМНМ. Доказана его сходимость для МНМ. Описаны эффективные алгоритмы оценивания ли-

нейных моделей на основе предложенного подхода и проведен анализ их вычислительной трудоемкости.

В первом разделе, для предложения нового подхода к эффективному построению линейных моделей на основе метода наименьших модулей и доказательства его сходимости, вводятся обозначения и определения. Во втором разделе описываются алгоритмы спуска по узловым прямым для метода наименьших модулей. В третьем разделе описывается алгоритм спуска по узловым прямым для обобщенного метода наименьших модулей. В четвертом разделе рассмотрены вопросы оценивания параметров нелинейных регрессионных моделей с помощью ОМНМ, сформулированы условия, при которых возможно получение решения. Доказано, что если функция регрессии является линейной по параметрам или имеет свободный член, то тогда ОМНМ-оценки будут находиться в узловых точках.

В третьей главе представлено описание программного комплекса для устойчивого построения линейных многомерных регрессионных моделей на основе МНМ и ОМНМ, а также для проведения вычислительных экспериментов с целью исследования эффективности предложенных алгоритмов оценивания многомерных линейных регрессионных моделей. В качестве платформы для реализации алгоритмов используется язык программирования R.

В первом разделе представлено описание архитектуры программного комплекса, дано подробное описание всех модулей, входящих в его состав. Во втором разделе, на основе разработанного программного комплекса проведен сравнительный анализ по быстродействию решения задачи алгоритмом спуска по узловым прямым с известными алгоритмами. Результаты сравнены с точными алгоритмами на основе полного перебора узловых точек и решения эквивалентной задачи линейного программирования. В третьем разделе проводится сравнительный анализ алгоритма спуска по узловым прямым с известными точ-

ными и приближенными методами реализации обобщенного метода наименьших модулей.

В четвертой главе рассмотрены примеры реализации и практического использования разработанного программного комплекса. В первом разделе на модельных примерах выполнено сравнение разработанных алгоритмов на основе спуска по узловым прямым с известными алгоритмами. Во втором разделе приведены четыре примера применения разработанного программного комплекса при решении практически важных задач в технике и экономике.

В Заключении приведены итоги проведенного исследования, рекомендации и перспективы дальнейшей разработки темы.

Правильность оформления диссертации и автореферата, соответствие автореферата диссертации её содержанию

Диссертация и автореферат оформлены в соответствии с принятыми для научных квалификационных работ нормами и требованиями. Автореферат адекватно и в полной мере отражает основные научные результаты и положения, сформулированные в тексте диссертации. Автореферат содержит краткое изложение материалов диссертационной работы по разделам и полностью соответствует содержанию самой диссертационной работы. В автореферате выделены все решаемые в каждом разделе задачи и представлены полученные результаты.

Новизна исследования и полученных результатов диссертационной работы

В диссертационной работе А. А. Азаряна разработан и обоснован новый подход к вычислительно эффективному математическому моделированию многомерных линейных регрессионных зависимостей в условиях стохастической неоднородности. Разработаны и протестированы алгоритмы для моделирования многомерных линейных регрессионных зависимостей методом наименьших модулей и обобщенным методом наименьших модулей основанные на спуске

по узловым прямым. Выполнена реализация разработанных методов и алгоритмов в виде комплекса программ. С помощью данного комплекса исследованы эффективности предложенных алгоритмов, проведены вычислительные эксперименты, а также решены практически важные задачи в технике и экономике.

Обоснованность и достоверность выводов

Постановки задач и полученные результаты обладают достаточной математической строгостью. Согласованность результатов вычислительных экспериментов с модельными примерами, решением большого количества тестовых задач достаточно убедительно выполнены на основе метода статистических испытаний Монте-Карло. Приведенные примеры практического применения разработанного комплекса программ убедительно показывают вычислительную эффективность предложенного подхода к моделированию регрессионных зависимостей на основе спуска по узловым прямым. Все выносимые на защиту результаты опубликованы.

Теоретическая и практическая значимость

Теоретическая значимость диссертационного исследования состоит в том, что полученные результаты развивают теорию математического моделирования многомерных линейных регрессионных зависимостей на основе метода наименьших модулей. Предложенная идея спуска по узловым прямым и реализующие ее алгоритмы моделирования обладают высокой вычислительной эффективностью.

Результаты диссертационной работы могут быть использованы в приложениях, в которых важно обеспечить высокое быстродействие и устойчивость моделирования многомерных линейных регрессионных зависимостей к грубым ошибкам и выбросам в экспериментальных данных, например, в динамических задачах мониторинга и диагностики многомерных систем.

Рекомендации по использованию результатов диссертационной работы

Результаты диссертационной работы позволяют применять их для:

- мониторинга и диагностики технических систем, в условиях стохастической неоднородности экспериментальных данных (задачи вибрационной диагностики и текущего контроля состояния нагруженных технических объектов и др.);

- моделирования социально-экономических явлений и систем в виде многомерных линейных моделей регрессии, а также авторегрессии и распределённого лага;

- решения технико-экономических задач, требующих использования многомерных данных и достаточно больших объемов выборок (оптимизация периода эксплуатации высоконагруженной техники на основе анализа средних удельных затрат, оценивание относительной производительности центральных процессоров и др.).

Разработка новых программных систем с качественным интерфейсом пользователя и программных модулей для проведения вычислительных экспериментов с использованием современных средств и подходов программирования позволяет использовать результаты диссертации в учебном процессе.

Соответствие паспорту специальности

Содержание диссертационной работы соответствует требованиям паспорта специальности 05.13.18 – «Математическое моделирование, численные методы и комплексы программ». Область исследований охватывает области исследований № 1-4 паспорта специальности:

- область исследований №1 «Разработка новых математических методов моделирования объектов и явлений»;
- область исследований №2 «Развитие качественных и приближенных аналитических методов исследования математических моделей»;

- область исследований №3 «Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий»;
- область исследований №4 «Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента».

Результат проверки в системе «Антиплагиат»

Результат проверки диссертации Азаряна Александра Артуровича в системе «Антиплагиат» показал 64,12 % оригинального текста, при этом из 35,88 % текстуальных совпадений с другими источниками 23,02 % составляют совпадения с опубликованными статьями самого А. А. Азаряна и в соавторстве с научным руководителем; 0,68 % составляет цитирование стандартов и руководящих документов, технической документации и спецификаций. Таким образом, после исключения данных источников из рассмотрения оригинальный текст автора составил 87,82 %. Оставшиеся 12,18 % текстуальных совпадений с источниками других авторов присутствуют в 130 источниках, не являются критичными и представляют собой общенаучные выражения, а также выражения и высказывания в рамках предметной области исследования, соответствующей тематике диссертации; общепринятые в рассматриваемой предметной области наименования, термины, словосочетания, используемые при постановке задач, формулировке и доказательстве лемм, теорем; совпадения по списку использованных источников и литературы; совпадения в наименованиях публикаций и их авторов, наименованиях научных конференций.

Работа выполнена достаточно грамотно, материал изложен последовательно и логично, однако в ходе ознакомления с диссертацией был сделан ряд замечаний, которые приведены ниже.

Замечания

1. На стр. 19 указаны условия Гаусса-Маркова для использования метода наименьших квадратов при моделировании линейных регрессионных зависимостей. Было бы полезно далее также привести аналогичные условия для использования метода наименьших модулей, тем более, что эти условия известны. Например, формально для существования узловых точек условие 5 выглядит более общим как $Rg(\mathbf{X}) = m \leq n$.
2. В тексте допущена неточность. На странице 25 оценки коэффициентов a_j регрессии в формуле $e_i = y_i - (a_1 + a_2 x_{i2} + a_3 x_{i3} + \dots + a_m x_{im})$ должны быть записаны с крышечкой (\hat{a}_j).
3. Из контекста диссертации понятно, какие методы решения задачи (1.7) считают точными, а какие – приближенными. Однако из-за наличия вычислительных погрешностей все методы при их численной реализации становятся приближенными. Поэтому следовало бы в тексте четко сформулировать данный вопрос.
4. Желательно было бы рассмотреть иные приближенные алгоритмы решения задачи (1.7), например метод внутренних точек и метод отжига.
5. Программный комплекс был реализован на языке программирования R. Проводился ли сравнительный анализ вычислительной эффективности использования языка R по сравнению с другими популярными и распространенными языками программирования.

Указанные замечания не влияют на положительную оценку диссертационной работы.

Заключение

Диссертационная работа А. А. Азаряна «Быстрые алгоритмы моделирования многомерных линейных регрессионных зависимостей на основе метода наименьших модулей» является законченной научной работой, выполненной на актуальную, практически и теоретически значимую тему. Все основные результаты являются новыми и достоверными, они дополняют и обобщают результаты отечественных и зарубежных исследований. Полученные результаты соответствуют поставленной цели. Диссертация обладает внутренним единством, научные и практические результаты которой отражены в 18 публикациях, в том числе в 5 публикациях в изданиях из перечня ВАК, 1 монографии. Получены 3 свидетельства Роспатента о государственной регистрации программ для ЭВМ. Содержание и оформление диссертационной работы и автореферата отвечают предъявленным требованиям. Текст диссертации представляет собой самостоятельную научно-квалификационную работу, не содержит заимствованного материала без ссылки на автора и (или) источник заимствования. Диссертационное исследование не содержит результатов научных работ, выполненных в соавторстве, без ссылок на соавторов.

На основании сказанного считаем, что диссертационная работа А. А. Азаряна «Быстрые алгоритмы моделирования многомерных линейных регрессионных зависимостей на основе метода наименьших модулей» соответствует всем требованиям п. 9 действующего «Положения о присуждении ученых степеней», предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 05.13.18 — «Математическое моделирование, численные методы и комплексы программ», а её автор Азарян Алексан Артурович заслуживает присуждения ему учёной степени кандидата физико-математических наук по специальности 05.13.18 — «Математическое моделирование, численные методы и комплексы программ».

Отзыв обсужден и одобрен на расширенном научном семинаре кафедры прикладной информатики и теории вероятностей 12.02.2019, протокол № 0200-19-04/08.

Профессор кафедры прикладной информатики и теории вероятностей, доктор физико-математических наук, доцент, специальность 05.13.18 — «Математическое моделирование, численные методы и комплексы программ»

e-mail: kulyabov-ds@rudn.ru



Кулябов Дмитрий Сергеевич

Профессор кафедры прикладной информатики и теории вероятностей, доктор физико-математических наук, профессор, специальность 05.13.16 — «Применение компьютеров, математического моделирования и математических методов в научных исследованиях»

e-mail: sevastianov-la@rudn.ru



Севастьянов Леонид Антонович

Декан факультета физико-математических и естественных наук, доктор химических наук по специальности 02.00.03 – Органическая химия, профессор



Воскресенский Леонид Геннадьевич

Федеральное государственное автономное образовательное учреждение высшего образования «Российский университет дружбы народов»

Адрес: 117198, Российская Федерация, г. Москва, ул. Миклухо-Маклая, д. 6

Сайт: <http://www.rudn.ru>

Телефон: +7 (495) 434-53-00

Факс: +7(495) 433-95-88

E-mail: rector@rudn.ru