

Федеральное государственное автономное образовательное учреждение
высшего образования «Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа прикладной математики и информатики
Кафедра вычислительных технологий и моделирования
в геофизике и биоматематике

На правах рукописи



Катруца Александр Михайлович

Оптимизация численных методов методами машинного обучения

Специальность 01.01.07 — Вычислительная математика

АВТОРЕФЕРАТ

*диссертации на соискание учёной степени
кандидата физико-математических наук*

Москва 2019

Работа выполнена на кафедре вычислительных технологий и моделирования в геофизике и биоматематике Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)».

Научный руководитель: *Оселедец Иван Валерьевич,*
доктор физико-математических наук, доцент
Сколковского института науки и технологий

Ведущая организация: *Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук»*

Защита состоится «26» декабря 2019 г. в 14:00 на заседании диссертационного совета ФПМИ.01.01.07.001 по адресу: 141701, Московская область, г. Долгопрудный, Институтский переулок, д. 9

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)»

<https://mipt.ru/education/post-graduate/soiskateli-fizikomatematicheskienauki.php>.

Работа представлена «16» октября 2019 г. в Аттестационную комиссию федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт» (национальный исследовательский университет) для рассмотрения советом по защите диссертаций на соискание учёной степени кандидата наук в соответствии с п. 3.1 ст. 4 Федерального закона «О науке и государственной научно-технической политике».

Общая характеристика работы

Актуальность темы исследования. Диссертационная работа посвящена разработке методов оптимизации значений параметров численных методов. Многие численные методы зависят от параметров, которые необходимо задавать перед запуском метода для решения конкретной задачи. Примерами подобных методов и соответствующими параметрами являются многосеточный метод¹, где параметрами являются операторы проекции и интерполяции между соседними уровнями иерархии сеток, а также параметр сглаживателя², предобусловленный метод сопряжённых градиентов³ с предобусловливателем неполным разложением Холецкого с релаксацией⁴. Данный предобусловливатель зависит от параметра релаксации, который необходимо задать перед использованием предобусловливателя. Также в диссертации рассматриваются другие численные методы и их параметризации. Зачастую выбор значений параметров численных методов существенно влияет на скорость их сходимости^{5,6}, поэтому построение алгоритмов получения оптимальных значений параметров для численных методов является актуальной задачей. Под оптимальными значениями параметров будем понимать такие значения, которые обеспечивают наилучшую скорость сходимости численного метода.

У существующих подходов для определения значений параметров численных методов есть следующие недостатки. Во-первых, они обладают высокой вычислительной сложностью, так как зачастую основаны на переборе возможных значений параметра по сетке с большим числом точек⁷. Во-вторых, для некоторых подходящих значений параметров (таких что метод сходится) отсутствуют гарантии оптимальности, то есть неизвестно можно ли улучшить данные значения параметров или нет. Также разработано

¹Trottenberg U., Oosterlee C. W., Schuller A. Multigrid. Academic press, 2000.

²Ruge J. W., Stüben K. Algebraic multigrid // Multigrid methods. 1987. Т. 3, № 13. С. 73—130.

³Axelsson O. Iterative solution methods. Cambridge : Cambridge University Press, 1994.

⁴Axelsson O., Lindskog G. On the eigenvalue distribution of a class of preconditioning methods // Numerische Mathematik. 1986. Т. 48, № 5. С. 479—498.

⁵Stüben K. A review of algebraic multigrid // Journal of Computational and Applied Mathematics. 2001. Т. 128, № 1. С. 281—309.

⁶Saad Y. Iterative Methods for Sparse Linear Systems. 2d-e изд. SIAM, 2003. Available from <http://www-users.cs.umn.edu/~saad/books.html>.

⁷Eshof J. van den, Hochbruck M. Preconditioning Lanczos approximations to the matrix exponential // SIAM Journal on Scientific Computing. 2006. Т. 27, № 4. С. 1438—1457.

множество алгоритмов для выбора значений параметров в случае решения задач определённого типа, однако эти алгоритмы не могут быть использованы для того же численного метода, используемого для решения задач другого типа⁸. Поэтому необходимо разработать универсальные методы выбора значений параметров численных методов, которые будут гарантировать оптимальность найденных значений.

Цель диссертационной работы.

1. Разработка новых подходов к ускорению алгоритмов решения систем линейных уравнений с разреженными матрицами
2. Разработка методов для решения проблемы мультиколлинеарности при решении задачи регрессии
3. Разработка новых быстрых методов приближённого вычисления меры похожести вершин в графе
4. Реализация алгоритмов для ускорения численных методов на основе подходов машинного обучения и проведение вычислительных экспериментов для обоснования эффективности предложенных алгоритмов

Научная новизна. Предложены новые методы для поиска оптимальных значений параметров численных методов. Предложенные методы основаны на подходах машинного обучения и позволяют сделать вывод об оптимальности найденных значений параметров. Основная идея, лежащая в основе полученных результатов, заключается в записи численного метода как суперпозиции дифференцируемых функций, введения некоторого функционала качества и его оптимизация по отношению к параметрам численного метода. В работе исследованы свойства и методы решения полученных задач оптимизации для различных численных методов. Получены результаты о сложности вычисления стохастических градиентов в случае гладких целевых функций, а также приведены экспериментальные результаты по использованию безградиентных методов в случае недифференцируемых функционалов качества. Также были исследованы выпуклые релаксации задач дискретной оптимизации, возникающие при решении проблемы мультиколлинеарности. Получен метод малоранговой коррекции для приближённого вы-

⁸Stolk C. C., Ahmed M., Bhowmik S. K. A multigrid method for the Helmholtz equation with optimized coarse grid corrections // SIAM Journal on Scientific Computing. 2014. Т. 36, № 6. A2819—A2841.

числения величин SimRank'a между вершинами графа. Предложенный метод показал высокую эффективность по сравнению с другими подходами к решению этой задачи.

Практическая значимость работы. Предложенные методы позволяют автоматически искать значения параметров численных методов, доставляющие их наилучшую скорость сходимости. Общий подход к оптимизации значений параметров рассмотренных численных методов может быть использован для исследования других методов, которые зависят от некоторого набора параметров. Реализации предложенных методов позволяют использовать их для различных входных данных и получать для них оптимальные значения параметров. Например, для многосеточного метода и метода сопряжённых градиентов можно получить оптимальные значения параметров для каждой матрицы линейной системы, которая может быть решена этими методами, и в дальнейшем использовать их для любых правых частей. Для метода вычисления похожести между вершинами графа достаточно один раз вычислить факторы малорангового разложения корректирующего члена, чтобы в дальнейшем использовать их для вычисления величины похожести SimRank между запрашиваемыми вершинами.

Теоретическая значимость работы В диссертационной работе представлена стохастическая оценка спектрального радиуса матрицы итераций для итерационного метода решения системы линейных уравнений. Эта оценка используется для оптимизации значений параметров многосеточного метода. Также исследована сложность вычисления градиентов этой оценки по параметрам многосеточного метода. Для метода сопряжённых градиентов введена новая концепция сходимости в среднем, которая является обобщением оценки, полученной для линейного итерационного метода. На основании этой концепции сходимости предложен алгоритм оптимизации предобусловливателей в методе сопряжённых градиентов. Далее приведён анализ решения проблемы мультиколлинеарности и установлен глобальный характер найденного решения для поставленной задачи квадратичной оптимизации. В заключении, для метода приближённого вычисления матрицы SimRank'a доказана теорема о представлении этой матрицы в виде ряда. Эта теорема позволяет ввести малоранговую поправку к первому члену этого ряда и использовать такое представление для приближённого вычисления SimRank'a.

Основные положения, выносимые на защиту. Основными результатами настоящей работы являются методы оптимизации значений параметров численных методов для решения различных практически важных задач. В частности рассматривается задача решения линейных систем большой размерности с разреженными матрицами, задача регрессии и задача приближённого вычисления меры похожести между вершинами графа. На защиту выносятся следующие положения:

1. Метод проверки оптимальности значений параметров многосеточного метода и их оптимизации
2. Метод оптимизации предобусловливателей в методе сопряжённых градиентов
3. Проанализирована проблема мультиколлинеарности в задаче регрессии и предложен новый метод для её решения
4. Метод приближённого вычисления меры похожести между вершинами графа SimRank

Апробация работы. Результаты, полученные в настоящей диссертации, имеют математическое обоснование, а также были экспериментально проверены на модельных и реальных задачах. Достоверность результатов также подтверждается публикациями в рецензируемых журналах, индексируемых Scopus и Web of Science. Результаты диссертации докладывались и обсуждались на следующих научных конференциях:

- Международная конференция «20th Conference of the International Federation of Operational Research Societies», 2014, Барселона
- 58-я научная конференция МФТИ, 2015, Москва
- 60-я научная конференция МФТИ, 2017, Москва
- 61-я научная конференция МФТИ, 2018, Москва
- Международная конференция «19th Copper Mountain Conference On Multigrid Methods», 2019, Copper Mountain

Публикации. Основные результаты настоящего диссертационного исследования опубликованы в следующих работах:

1. Работы, опубликованные в изданиях, входящих в перечень рецензируемых научных изданий, индексируемых Web of Science:

- (a) Katrutsa A., Strijov V. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria // Expert Systems with Applications. — 2017. — Т. 76. — С. 1-11.
- (b) Oseledets I. V., Ovchinnikov G. V., Katrutsa A. M. Fast, memory-efficient low-rank approximation of SimRank // Journal of Complex Networks. — 2016. — Т. 5. — №. 1. — С. 111-126.
- (c) Katrutsa A. M., Strijov V. V. Stress test procedure for feature selection algorithms // Chemometrics and Intelligent Laboratory Systems. — 2015. — Т. 142. — С. 172-183.
- (d) Katrutsa A., Daulbaev T., Oseledets I. Black-box learning of multigrid parameters // Journal of Computational and Applied Mathematics. — 2019. — <https://doi.org/10.1016/j.cam.2019.112524>.

2. Работы, опубликованные в прочих изданиях:

- (a) Оселедец И. В., Бочев М. А., Катруца А. М., Овчинников Г. В. Как оптимизировать предобусловливатели в методе сопряжённых градиентов: стохастический подход // Препринты Института прикладной математики им. М. В. Келдыша РАН. — 2018. — №. 164. — С. 1-26.
- (b) Катруца А. М. Оптимизация параметров двухсеточного метода с помощью глубоких нейронных сетей // Труды 60-й научной конференции МФТИ, 2017
- (c) Катруца А. М. Как оптимизировать предобуславливатели в методе сопряжённых градиентов: стохастический подход // Труды 61-й научной конференции МФТИ, 2018

Личный вклад автора. Диссертационное исследование является самостоятельным законченным трудом автора. В работе Лично автором была предложена процедура вычисления градиентов по параметрам многосеточного метода в работе [1d]. Также лично автором в работе [1a] было предложено использовать подход квадратичного программирования для анализа мультиколлинеарности. Постановка задачи в работах [1d,2a,1b] была выполнена И.В. Оселедцем. Реализация алгоритмов, а также подготовка численных экспериментов в работах [1a-1d,2a] были выполнены автором самостоятельно. В

работе [1b] автором совместно с Г. В. Овчинниковым была получена теорема о разложении матрицы SimRank'а в ряд, вклад авторов равнозначен.

Структура работы. Диссертация состоит из оглавления, введения, пяти глав, заключения, списка иллюстраций, списка таблиц, перечня основных обозначений и списка литературы из 125 наименований. Полный объем диссертации составляет 115 страниц с 17 рисунками и 14 таблицами.

Содержание работы

Во **введении** приводится обоснование актуальности темы настоящего исследования, формулируются цели и задачи, а также указывается научная новизна диссертации, её теоретическая и практическая значимость.

В **первой главе** приведена постановка задачи обучения с учителем — одна из основных задач машинного обучения. Эта задача формулируется следующим образом. Пусть дан некоторый набор пар $(\mathbf{x}_i, y_i), i = 1, \dots, N_s$, то есть каждому объекту $\mathbf{x}_i \in \mathcal{X}$ ставится в соответствие некоторая метка $y_i \in \mathcal{Y}$. Множества $\mathcal{X}$ и $\mathcal{Y}$ соответственно обозначают множество объектов и множество меток. Тогда задача обучения с учителем состоит в том, чтобы по данному набору из N_s пар объектов и меток найти некоторую функцию $f^* : \mathcal{X} \rightarrow \mathcal{Y}$, такую что для любого объекта из множества $\mathcal{X}$ она возвращала бы правильную метку $y_i \in \mathcal{Y}$. В машинном обучении функцию $f : \mathcal{X} \rightarrow \mathcal{Y}$ также называют моделью. Для упрощения поиска оптимальной модели f^* вводят предположение о том, что искомая модель лежит в некотором параметрическом семействе $\mathcal{F}$. Тогда для поиска f^* необходимо ввести некоторую функцию потерь $\ell : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$, такую что

$$\hat{f}^* = \arg \min_{f \in \mathcal{F}} \frac{1}{N_s} \sum_{i=1}^{N_s} \ell(f(\mathbf{x}_i), y_i). \quad (1)$$

Также учитывая, что семейство $\mathcal{F}$ параметрическое, задачу (1) можно переписать в виде

$$\hat{\mathbf{w}}^* = \arg \min_{\mathbf{w} \in \mathcal{W}} \frac{1}{N_s} \sum_{i=1}^{N_s} \ell(f(\mathbf{w} | \mathbf{x}_i), y_i), \quad (2)$$

где $\mathcal{W}$ — множество параметров, соответствующих семейству функций $\mathcal{F}$. Запись $f(\mathbf{w} | \mathbf{x}_i)$ означает, что функция f зависит от $\mathbf{w}$ как от параметра при данном объекте $\mathbf{x}_i$. Задачу (2) также называют задачей минимизации эмпирического риска⁹.

Далее показана эквивалентность между решением задачи нахождения оптимальных значений параметров численного метода и решением задачи обучения с учителем. В частности, предположим, что численный метод зависит от параметров $\mathbf{v}$ и результат его выполнения через $k_{\max}$ итераций можно представить как значение некоторой функции вида $g(\mathbf{v} | x_0, k_{\max}, \mathcal{P})$, где

⁹Vapnik V. The nature of statistical learning theory. Springer science & business media, 2013.

x_0 — начальное приближение, $k_{\max}$ — максимальное число итераций, $\mathcal{P}$ — это множество всех параметров, которые задают задачу, решаемую численным методом. Каждому начальному приближению соответствует приближение x_k , полученное после выполнения k итераций численного метода. В то же время для данной задачи $\mathcal{P}$ существует точное решение x^* , которое должно быть приближено с помощью x_k после достаточно большого числа итераций. Одним из основным свойств численных методов, является скорость их сходимости, то есть то насколько быстро приближение x_k сходится к точному решению x^* после выполнения k итераций¹⁰. Скорость сходимости численных методов зачастую зависит от параметров $\mathbf{v}$, используемых для решения задачи, описанной множеством $\mathcal{P}$. Таким образом, если мы введём некоторую функцию потерь $\mathcal{L}$, отражающую скорость сходимости численного метода для данного начального приближения x_0 , то сможем записать следующую задачу

$$\mathbf{v}^* = \arg \min_{\mathbf{v} \in \mathcal{V}} \frac{1}{N_s} \sum_{i=1}^{N_s} \mathcal{L}(g(\mathbf{v} | x_0^{(i)}, k_{\max}, \mathcal{P}), x^*), \quad (3)$$

где $\mathcal{V}$ — пространство параметров численного метода, $x_0^{(i)}$ — i -ое начальное приближение. В последующих главах для каждого рассмотренного численного метода будем указывать вид функции $\mathcal{L}$, её свойства, а также свойства функции g , которые напрямую влияют на выбор способа решения задачи (3). Также будет рассмотрен вопрос о способе задания набора начальных приближений $\{x_0^{(1)}, \dots, x_0^{(N_s)}\}$. Таким образом, показана равносильность между решением задачи обучения с учителем и задачи определения значений параметров численных методов. Далее в главе обсуждаются способы решения задачи оптимизации эмпирической скорости сходимости, выраженной функцией $\mathcal{L}$, в зависимости от свойств множества $\mathcal{V}$ и функции g . В последующих главах диссертации будут представлены конкретные примеры функций g , множеств $\mathcal{V}$ и, в целом, задач типа (3) для конкретных численных методов.

Вторая глава посвящена разработке метода оптимизации параметров многосеточного метода для решения систем линейных уравнений с разреженными матрицами. То есть решается задача вида

$$Au = f,$$

¹⁰ Tyrtyshnikov E. E. A brief introduction to numerical analysis. Springer Science & Business Media, 2012.

где A — разреженная матрица большой размерности, а f — вектор правой части. Идея двухсеточного метода заключается в проекции на грубую сетку вектора невязок, полученного на мелкой сетке, решении корректирующего уравнения со спроецированной на эту сетку матрицей A_c и последующей интерполяции решения обратно на мелкую сетку. Эту идею легко обобщить на случай иерархии сеток, в котором для решения уравнения на грубой сетке будет рекурсивно вызываться та же самая функция только уже со спроецированной матрицей и с вектором невязок в правой части. Одну итерацию двухсеточного метода можно записать в виде:

$$u^{(k+1)} = Cu^{(k)} + b, \quad (4)$$

где C — матрица итерации, которая имеет вид

$$C = (M_2^{-1}K_2)^{s_2}(I + P(RAP)^{-1}RA)(M_1^{-1}K_1)^{s_1}, \quad (5)$$

где I — единичная матрица, а вектор b вычисляется как

$$b = ((M_2^{-1}K_2)^{s_2}P(RAP)^{-1}R(s_1AM_1^{-1} - I) + s_2M_2^{-1})f.$$

Матрица итерации C зависит от операторов R , P и демпфирующего фактора ω : $C = C(R, P, \omega)$ (далее будем использовать обозначение C вместо $C(R, P, \omega)$). Эти параметры определяют то, как именно осуществляется переход от мелкой сетки к грубой и наоборот, а также как именно на каждой из сеток сглаживается решение. Матрицы M_1 , M_2 , K_1 и K_2 задают сглаживатели и зависят от параметра ω . Отметим, что умножение матрицы C на любой вектор y подходящей размерности эквивалентно выполнению одной итерации двухсеточного метода для $u^{(k)} = y$ и $f \equiv 0$.

Далее в главе ставится задача оптимизации стохастической оценки спектрального радиуса

$$\rho(C) = \max_{i=1, \dots, n} |\lambda_i(C)| \rightarrow \min_{R, P, \omega}, \quad (6)$$

полученная с помощью стохастической оценки следа матрицы и формулы Гельфанда¹¹

$$\rho(C) = \lim_{k \rightarrow \infty} \sqrt[k]{\|C^k\|}, \quad (7)$$

¹¹Kozyakin V. On accuracy of approximation of the spectral radius by the Gelfand formula // Linear Algebra and its Applications. 2009. Т. 431, № 11. С. 2134—2141.

где $\|\cdot\|$ некоторая матричная норма. А именно предлагается следующая аппроксимация

$$\rho(C) \approx \sqrt[k]{\|C^K\|_F}, \quad (8)$$

где K — положительная константа и $\|\cdot\|_F$ — норма Фробениуса. Введённая аппроксимация ограничивает сверху спектральный радиус

$$\gamma^{(1+\ln K)/K} \|C^K\|_F^{1/K} \leq \rho(C) \leq \|C^K\|_F^{1/K}, \quad (9)$$

для некоторой константы $\gamma \in (0, 1)$ и положительной константы K . Точность оценки возрастает с ростом K . Таким образом, мы фиксируем некоторое значение для константы K и минимизируем верхнюю границу в (9). Такая задача минимизации эквивалентна следующей задаче:

$$F_K = \|C^K\|_F^2 \rightarrow \min_{R, P, \omega}. \quad (10)$$

Построим несмещённую стохастическую оценки целевой функции (10), для вычисления которой необходимо только умножать матрицу C на вектор:

$$\|C^K\|_F^2 = \mathbb{E}_z \|C^K z\|_2^2, \quad (11)$$

где $z = [z_i]$ — случайный вектор, элементы которого сгенерированы из распределения Радемахера:

$$\mathbb{P}(z_i = \pm 1) = \frac{1}{2}. \quad (12)$$

Поэтому мы минимизируем матожидание:

$$\mathbb{E}_z \|C^K z\|_2^2 \rightarrow \min_{R, P, \omega}. \quad (13)$$

Для решения (13) используем стандартные методы, основанные на стохастической оценке градиента целевой функции, равной градиенту стохастической оценки самой целевой функции. Эта оценка целевой функции может быть представлена в виде:

$$\hat{F}_K = \frac{1}{N} \sum_{i=1}^N \|C^K z^i\|_2^2, \quad (14)$$

где N — количество векторов z , используемых для усреднения, а z^i — i -ый случайный вектор с элементами из распределения Радемахера (12).

Следующая лемма является прямым следствием стохастической оценки следа¹² и показывает, что оценка (14) является несмещённой.

¹²*Hutchinson M. F. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines // Communications in Statistics-Simulation and Computation. 1990. Т. 19, № 2. С. 433—450.*

Лемма 1. Стохастическая оценка (14) целевой функции (10) имеет следующее матожидание и дисперсию:

$$\mathbb{E}_{z^i} \left(\frac{1}{N} \sum_{i=1}^N \|C^K z^i\|_2^2 \right) = \|C^K\|_F^2$$

$$\text{Var}_{z^i} \left(\frac{1}{N} \sum_{i=1}^N \|C^K z^i\|_2^2 \right) = \frac{2}{N} \left(\left\| (C^K)^\top C^K \right\|_F^2 - \sum_{i=1}^n \left((C^K)^\top C^K \right)_{ii}^2 \right).$$

Таким образом, мы получили несмещённую оценку (14) целевой функции. Сравнение задач минимизации оценки (14) и задачи (3) показывает, что поставленная задача минимизации оценки спектрального радиуса формально совпадает с поставленной в главе 1 задачей оптимизации численного метода. В частности, моделью в данном случае является функция g , которая эквивалентна применению многосеточного метода K раз к начальному приближению z_i , набор векторов z_i образует обучающую выборку. Метки y_i для векторов z_i соответственно нулевые векторы, поскольку умножение матрицы C на вектор эквивалентно одной итерации многосеточного метода с нулевой правой частью, а решением такой системы уравнений является нулевой вектор. Таким образом, функцией $\mathcal{L}$ является квадрат евклидовой нормы разности между меткой для вектора z_i , полученной с помощью модели g , и правильной меткой для этого вектора, которая равна нулевому вектору согласно написанному выше. В итоге получаем, что задачи эквивалентны.

Далее в главе обсуждаются способы решения поставленной задачи минимизации на стохастических градиентных методах. Градиенты целевой функции по параметрам могут быть автоматически вычислены с помощью графа вычислений¹³, пример которого для одной итерации двухсеточного метода приведён на Рис. 1. В разделе 2.1.3. приводится пример вычисления градиентов в для операции вычисления проекции матрицы системы на грубую сетку и оценивается вычислительная сложность этой операции.

В заключении данной главы приводятся результаты вычислительных экспериментов, в которых рассматривались двумерное уравнения Пуассона, одномерно уравнение Гельмгольца и двумерное сингулярно возмущённое уравнение конвекции–диффузии. Во всех случаях было получено меньшее

¹³Automatic differentiation in machine learning: a survey / A. G. Baydin [и др.] // Journal of machine learning research. 2018. Т. 18, № 153.

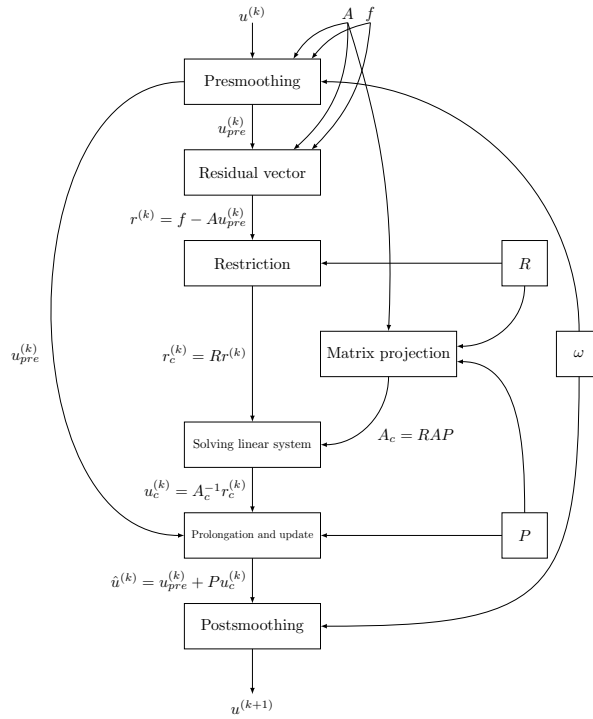


Рис. 1: Одна итерация двухсеточного метода, представленная в виде графа вычислений

значение спектрального радиуса матрицы итераций по сравнению с геометрическим многосеточным методом с линейной и билинейной интерполяцией (Linear), а также по сравнению с алгебраическим многосеточным методом (AMG), который адаптивно задаёт операторы проекции и интерполяции зависимости от вида сглаживателя. Везде далее предложенный метод оптимизации параметров с помощью решения оптимизационной задачи (13) называется DMG. Результаты собраны в Таблице 1 для уравнения Пуассона и в Таблицах 2, 3 для одномерного уравнения Гельмгольца в случае различных диапазонов частот.

Таблица 1: Спектральные радиусы ρ для DMG, AMG и линейных параметров при решении двумерного уравнения Пуассона

(a) Два уровня				(b) Четыре уровня			
Размер сетки	Линейный	AMG	DMG	Размер сетки	Линейный	AMG	DMG
15 × 15	0.06172	0.06172	0.02739	15 × 15	0.19004	0.10991	0.02699
31 × 31	0.06172	0.06172	0.03551	31 × 31	0.19563	0.12374	0.03492
63 × 63	0.06172	0.06172	0.03741	63 × 63	0.19705	0.16366	0.04747
127 × 127	0.06172	0.06172	0.05274	127 × 127	0.19741	0.26354	0.0736

Таблица 2: Спектральный радиус ρ для случая низких частот, размер сетки $n = 127$, четыре уровня, инициализация линейными операторами

k	Линейный	AMG	DMG
5	0.086543	0.082556	0.039514
10	0.271969	0.082073	0.048242
15	0.466403	0.097001	0.057631
20	4.927016	0.262049	0.055113

Таблица 3: Спектральный радиус ρ для случая высоких частот, размер сетки $n = 1023$, три уровня, инициализация гомотопией

k	Линейный	AMG	DMG
100	0.554602	0.082526	0.079775
200	3.883023	0.082154	0.088978
300	10.995468	0.214390	0.083191
400	17.112969	0.758239	0.121774
500	4428.612730	3.541510	0.204888
600	84.168512	31.749316	0.300613

Для сингулярно возмущённого уравнения конвекции–диффузии получена следующая зависимость спектрального радиуса для различных классов многосеточных методов, см. Рис. 2.

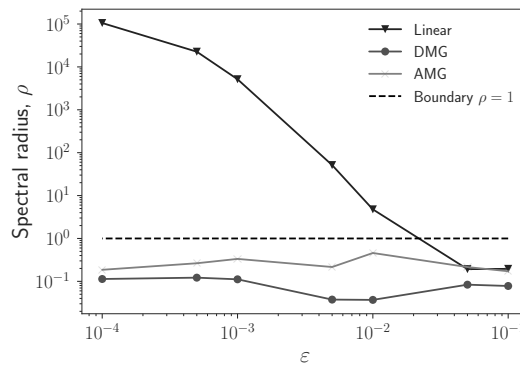


Рис. 2: Спектральный радиус матрицы итераций при решении сингулярно возмущённого уравнения конвекции–диффузии, сетка 127×127 , пять уровней

В **третьей главе** рассматривается задача оптимизация значения параметра предобусловливателя в методе сопряжённых градиентов для решения системы линейных уравнений $Ax = b$, где $A \in \mathbb{R}^{m \times m}$. Для постановки такой задачи требуется ввести специальный функционал качества который, по аналогии с изложенным в главе 2, оценивал бы сходимость итерационного метода. В этой главе показано, что в отличие от линейных итерационных методов вида (4), оценка сходимости итерационного метода в среднем для нелинейных итерационных методов отличается от оценки сходимости в худшем случае. Это приводит нас к оценке сходимости нелинейного итерационного метода в среднем, которую можно вычислить следующим образом. Как k итераций стационарного линейного метода (4) эквивалентно выполнению k умножений матрицы итераций на вектор, так и k итераций нестационарного итерационного метода можно представить как действие некоторого нелинейного преобразования $\mathcal{G}$, определяющего метод, k раз. Таким образом, можно записать

$$x_* - x_k = \mathcal{G}[x_* - x_{k-1}] = \mathcal{G}^k[x_* - x_0], \quad (15)$$

где $\mathcal{G}$ — это отображение, эквивалентное применению одной итерации нелинейного метода, а $\mathcal{G}^k$ обозначает применение этого отображения k раз. Тогда практический способ наблюдения сходимости метода — это выполнение k итераций для некоторого числа случайных начальных приближений $x_0^{(i)}$. Действительно, определим стохастический функционал сходимости как

$$F_s \equiv \frac{1}{n} \sum_{i=1}^n \|x_* - x_k^{(i)}\|_*, \quad (16)$$

где n — число случайных начальных приближений и $x_k^{(i)}$ — приближение решения на k -ой итерации с начальным приближением $x_0^{(i)}$. Возникает вопрос: как вычислять функционал F_s , если точное решение x_* , как правило, неизвестно на практике. Для метода сопряжённых градиентов эту трудность легко обойти, если вычислять F_s для A -нормы. Можно также воспользоваться следующим подходом, работающим для любых итерационных методов: при вычислении F_s считать точное решение x_* равным нулю, что соответствует решению линейной системы с нулевой правой частью $b = 0$. Везде далее в данной главе F_s вычислялся для $x_* = 0$ и евклидовой нормы. Заметим, что

мы можем записать

$$F_s = \frac{1}{n} \sum_{i=1}^n \|x_* - x_k^{(i)}\|_* = \frac{1}{n} \sum_{i=1}^n \|\mathcal{G}^k[x_* - x_0^{(i)}]\|_* \approx \mathbb{E}_{x_0} \|\mathcal{G}^k[x_* - x_0]\|_*.$$

Возможное применение данного подхода, рассматриваемое в данной работе, заключается в оптимизации параметра предобусловливателя в методе сопряжённых градиентов для получения более быстрой сходимости. Очевидно, что такой оптимизационный процесс, основанный на многократных пробных запусках итерационного метода, является вычислительно затратным (один пробный запуск означает выполнение nk итераций метода). Такие затраты оправданы, только если полученный оптимальный предобусловливатель будет использован для проведения большого числа итераций, например, при решении различных систем с одинаковой матрицей и различными правыми частями.

Обратим внимание на то, что задача минимизации стохастического функционала F_s эквивалентна задаче обучения с учителем (3) из главы 1. В текущей главе роль функции $\mathcal{L}$ играет евклидова норма, модель — это отображение $\mathcal{G}^k$, реализующее k итераций предобусловленного метода сопряжённых градиентов и потому зависящее от параметра предобусловливателя, а обучающая выборка состоит из n векторов с элементами из стандартного нормального распределения. Поскольку нам необходимо найти параметр предобусловливателя, который не зависит от правой части, то будем использовать нулевую правую часть и нулевое точное решение соответственно. Начальные приближения $x_0^{(i)}$ генерировались из стандартного нормального распределения по аналогии с главой 2. Таким образом, в данном случае снова получили задачу аналогичную задаче обучения с учителем.

Далее в разделе 3.2 предлагается использовать безградиентный метод оптимизации параметра релаксации в предобусловливателе (неполном разложении Холецкого с релаксацией). Также исследуется сложность такой процедуры оптимизации и её зависимость от параметров n и K . После чего в разделе 3.3 приведена тестовая задача, для которой выполнен поиск оптимального значения параметра предобусловливателя. Она имеет вид

$$\begin{aligned} -(D_1 u_x)_x - (D_2 u_y)_y &= g(x, y), \quad (x, y) \in \Omega = [0, 1] \times [0, 1], \\ u(x, y)|_{\partial\Omega} &= 0, \end{aligned} \tag{17}$$

где индексы $\cdot_x, \cdot_y$ обозначают частные производные по отношению к x и y , соответственно. Мы рассматриваем два случая: в первом случае коэффициенты $D_{1,2}$ берутся тождественно равными единице во всей области Ω , а во втором — разрывными:

$$D_1 = \begin{cases} 1000, & \text{если } (x, y) \in [\frac{1}{4}, \frac{3}{4}] \times [\frac{1}{4}, \frac{3}{4}], \\ 1, & \text{иначе,} \end{cases} \quad D_2 = \frac{1}{2}D_1.$$

Функция правой части $g(x, y)$ выбирается так, чтобы значения функции

$$u(x, y) = \sin(\pi x) \sin(\pi y) \quad (18)$$

на конечно-разностной сетке являлись элементами вектора точного решения дискретизированной краевой задачи. В экспериментах показано, что предложенный метод определения оптимального значения параметра предобусловливателя позволяют лучше кластеризовать спектр предобусловленной матрицы. Это свойство приводит к более быстрой сходимости предобусловленного метода сопряжённых градиентов по сравнению с использованием в качестве функции ошибки числа обусловленности предобусловленной матрицы. На Рис. 3 показано сравнение предложенного функционала качества на основе сходимости в среднем и классического функционала качества, основанному на числе обусловленности. Сравнение проводится как по скорости сходимости предобусловленного метода сопряжённых градиентов, так и по виду спектра предобусловленной матрицы. Таким образом, предложенный метод оптимизации значения параметра предобуславливателя в методе сопряжённых градиентов даёт параметр, приводящий к более быстрой сходимости метода сопряжённых градиентов за счёт лучшей кластеризации спектра, чем использование классической оценки на основе числа обусловленности.

В **четвёртой главе** рассматривается задача регрессии при наличии мультикоррелирующих столбцов в матрице плана. Пусть $\mathbf{X} = [\chi_1, \dots, \chi_n] \in \mathbb{R}^{m \times n}$ матрица плана, где $\chi_j \in \mathbb{R}^m$ обозначает j -ый столбец. Обозначим за $\mathcal{J} = \{1, \dots, n\}$ множество индексов столбцов и пусть $\mathcal{A} \subseteq \mathcal{J}$ некоторое индексное подмножество. Пусть также $\mathbf{y} \in \mathbb{R}^m$ целевой вектор. Задача регрессии заключается в поиске вектора параметров $\mathbf{w}^* \in \mathbb{R}^n$ такого что

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbb{R}^n} S(\mathbf{w}, \mathcal{A} | \mathbf{X}, \mathbf{y}, \mathbf{f}), \quad (19)$$

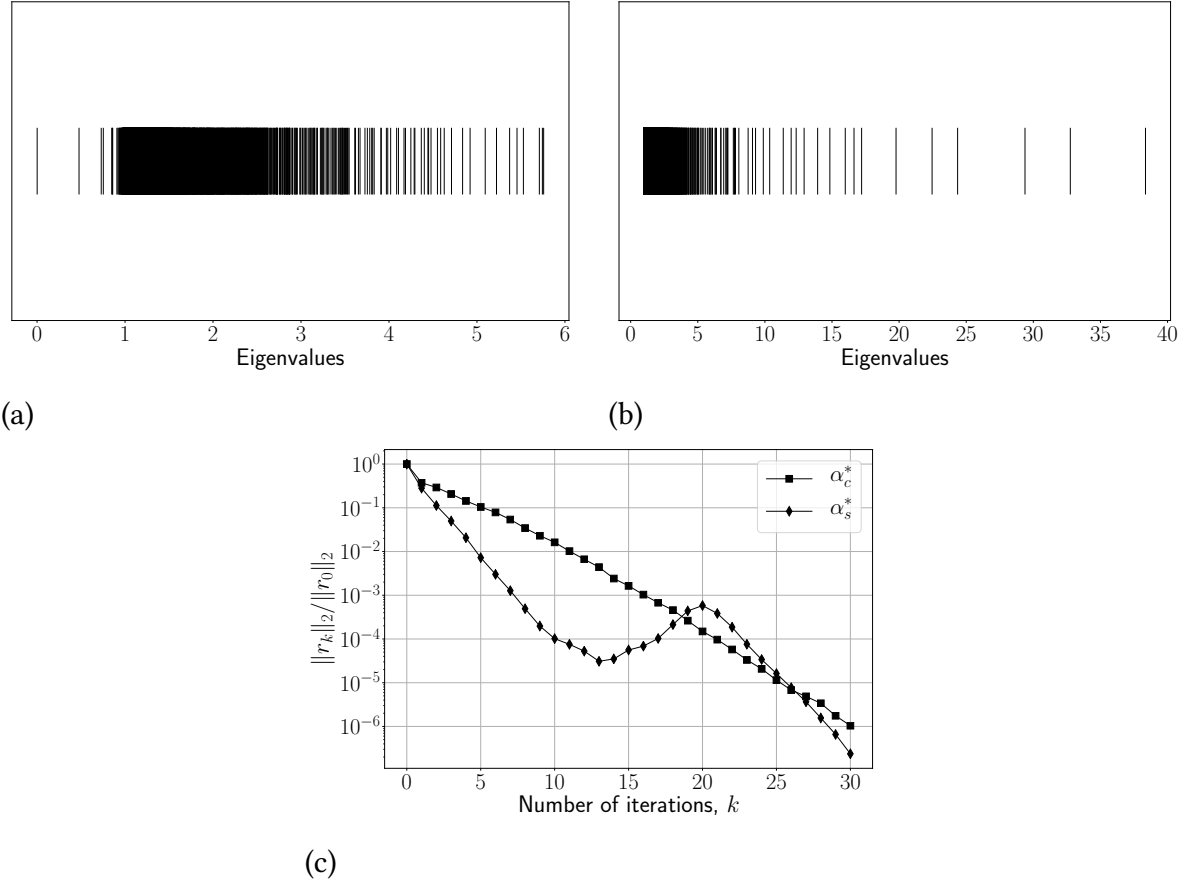


Рис. 3: Тестовый пример с параметром $m = 2500$ и непостоянными коэффициентами $D_{1,2}$. Рис. 3а, 3б: собственные значения предобусловленной матрицы A для оптимального значения параметра α_s^* (найденного для F_s , рис. 3а) и α_c^* (найденного для F_c , рис. 3б). Рис. 3с: сходимость относительной нормы невязки для метода сопряжённых градиентов с предобусловливателем $\text{RIC}_\alpha(0)$ для α_s^* и α_c^* .

где S функция ошибки, которая оценивает качество вектора параметров $\mathbf{w}$ и соответствующего множества индексов столбцов, при данной матрице плана, целевом векторе $\mathbf{y}$ и модели $\mathbf{f}$. Модель $\mathbf{f}$ приближает целевой вектор $\mathbf{y}$.

В данной главе анализируются только линейная функция вида

$$\mathbf{f}(\mathbf{X}, \mathcal{A}, \mathbf{w}) = \mathbf{X}_{\mathcal{A}} \mathbf{w},$$

где $\mathbf{X}_{\mathcal{A}}$ усечённая матрица в плана, состоящая только из столбцов из индексного множества $\mathcal{A}$, и квадратичная функция ошибки

$$S(\mathbf{w}, \mathcal{A} | \mathbf{X}, \mathbf{y}, \mathbf{f}) = \|\mathbf{f}(\mathbf{X}, \mathcal{A}, \mathbf{w}) - \mathbf{y}\|_2^2. \quad (20)$$

Предположим, что столбцы матрицы плана зашумлены, нерелевантны целевому вектору или мультиколлинеарны. Это влечёт за собой дополни-

тельные ошибки в оценке оптимального вектора параметров $\mathbf{w}^*$ и повышает неустойчивость этого вектора.

Методы снижения размерности могут быть использованы для удаления подобных столбцов из матрицы плана $\mathbf{X}$. Такие методы повышают устойчивость оптимального вектора $\mathbf{w}^*$ и скорость работы метода решения задачи наименьших квадратов. Задача выбора столбцов матрицы состоит в следующем

$$\mathcal{A}^* = \arg \min_{\mathcal{A} \subseteq \mathcal{J}} Q(\mathcal{A}|\mathbf{X}, \mathbf{y}), \quad (21)$$

где $Q : \mathcal{A} \rightarrow \mathbb{R}$ критерий качества, который оценивает качество выбранного индексного подмножества $\mathcal{A} \subseteq \mathcal{J}$. Задача (21) не требует одновременно с выбором множества $\mathcal{A}^*$ получать оценку оптимального вектора параметров $\mathbf{w}^*$. Для решения задачи (21) можно использовать взаимное расположение между столбцами $\chi_j, j \in \mathcal{J}$ и целевым вектором $\mathbf{y}$.

Пусть $\mathbf{a} \in \mathbb{B}^n = \{0, 1\}^n$ индикаторный вектор, такой что $a_j = 1$ тогда и только тогда, когда $j \in \mathcal{A}$. Тогда задача (21) может быть переписана в виде

$$\mathbf{a}^* = \arg \min_{\mathbf{a} \in \mathbb{B}^n} Q(\mathbf{a}|\mathbf{X}, \mathbf{y}), \quad (22)$$

где $Q : \mathbb{B}^n \rightarrow \mathbb{R}$ другая форма записи критерия Q , область определения которого — бинарные векторы из множества $\mathbb{B}^n$. Вектор $\mathbf{a}^*$ и индексное множество $\mathcal{A}^*$ связаны следующим образом

$$a_j^* = 1 \Leftrightarrow j \in \mathcal{A}^*, j \in \mathcal{J}. \quad (23)$$

Далее даются формальные определения явлению мультиколлинеарности и приводятся важные частные случаи этого определения. Пусть столбец χ_j и целевой вектор нормированы:

$$\|\mathbf{y}\|_2 = 1 \text{ and } \|\chi_j\|_2 = 1, j \in \mathcal{J}. \quad (24)$$

Рассмотрим множество столбцов $\mathcal{A} \subseteq \mathcal{J}$.

Определение 1. *Столбцы с индексами из множества $\mathcal{A}$ называются мультиколлинеарными, если найдётся такой индекс j , коэффициенты λ_k , индекс $k \in \mathcal{A} \setminus j$ и достаточно малое положительное число $\delta > 0$, что*

$$\left\| \chi_j - \sum_{k \in \mathcal{A} \setminus j} \lambda_k \chi_k \right\|_2^2 < \delta. \quad (25)$$

Чем меньше δ , тем выше степень мультиколлинеарности.

В частности, примерами мультикоррелирующих столбцов матрицы плана являются следующие конфигурации.

Определение 2. *Столбцы с индексами i, j называются коррелирующими, если существует достаточно малое положительное число $\delta_{ij} > 0$ такое, что*

$$\|\chi_i - \chi_j\|_2^2 < \delta_{ij}. \quad (26)$$

Из определения следует, что $\delta_{ij} = \delta_{ji}$. Неравенства (25) и (26) идентичны при условии $\lambda_k = 0, k \neq j$ и $\lambda_k = 1, k = j$.

Определение 3. *Столбец χ_j называется коррелирующим с целевым вектором y , если найдётся достаточно малое положительное число $\delta_j > 0$ такое, что*

$$\|y - \chi_j\|_2^2 < \delta_j.$$

После введения всех необходимых определений, в разделе 4.2 приводится постановка задачи квадратичного программирования для решения проблемы мультиколлинеарности. В работе¹⁴, было показано, что ни один из рассмотренных методов выбора столбцов в матрице (LARS, Lasso, Ridge, Stepwise и генетический алгоритм) не решают одновременно задачу (19) и не дают устойчивый, точный и неизбыточный набор столбцов. Поэтому предлагается использовать подход, основанный на постановке задачи квадратичной оптимизации.

Основная идея предлагаемого подхода заключается в минимизации числа похожих столбцов в матрице плана и максимизации числа столбцов, релевантных целевому вектору. Для формализации этой идеи мы представим критерий Q из задачи (22) в виде квадратичной функции

$$Q(\mathbf{a}) = \mathbf{a}^\top \mathbf{Q} \mathbf{a} - \mathbf{b}^\top \mathbf{a}, \quad (27)$$

где $\mathbf{Q} \in \mathbb{R}^{n \times n}$ матрица попарных похожестей между столбцами матрицы, а вектор $\mathbf{b} \in \mathbb{R}^n$ состоит из мер релевантности каждого столбца матрицы плана $\mathbf{X}$ целевому вектору y .

Для вычисления матрицы $\mathbf{Q}$ и вектора $\mathbf{b}$ введём функции Sim и Rel:

$$\begin{aligned} \text{Sim}: \mathcal{J} \times \mathcal{J} &\rightarrow [0, 1], \\ \text{Rel}: \mathcal{J} &\rightarrow [0, 1]. \end{aligned} \quad (28)$$

¹⁴Katrutsa A., Strijov V. Stress test procedure for feature selection algorithms // Chemometrics and Intelligent Laboratory Systems. 2015. Т. 142. С. 172–183.

Это проблемно-ориентированные функции, определяемые пользователем до запуска процедуры удаления столбцов матрицы плана для каждой конкретной задачи. Они показывают, каким образом определяется похожесть между столбцами матрицы плана (Sim) и релевантность целевому вектору (Rel). Чтобы подчеркнуть зависимость метода удаления столбцов на основе квадратичной оптимизации от функций похожести и релевантности, введём следующее определение.

Определение 4. Пусть $QP(Sim, Rel)$ метод выбора столбцов матрицы плана, который решает задачу оптимизации вида

$$\mathbf{a}^* = \arg \min_{\mathbf{a} \in \mathbb{B}^n} \mathbf{a}^\top \mathbf{Q} \mathbf{a} - \mathbf{b}^\top \mathbf{a}, \quad (29)$$

где матрица $\mathbf{Q}$ вычисляется с помощью функции Sim:

$$\mathbf{Q} = [q_{ij}] = Sim(\chi_i, \chi_j), \quad (30)$$

а вектор $\mathbf{b}$ вычисляется с помощью функции Rel:

$$\mathbf{b} = [b_i] = Rel(\chi_i). \quad (31)$$

Примерами таких функций являются коэффициент корреляции Пирсона и взаимная информация. Однако ни одна из этих функций не показывают напрямую релевантность столбца матрицы плана целевому вектору. Для учёта релевантности столбцов матрицы мы предлагаем использовать нормализованную значимость столбца, которую можно оценить с помощью стандартного t-теста в предположении линейности модели. Для отбора релевантных столбцов мы ставим следующую задачу проверки гипотез для j -го столбца:

$$\begin{aligned} H_0 : w_j &= 0, \\ H_1 : w_j &\neq 0. \end{aligned} \quad (32)$$

Значение p -value p_j показывает релевантность j -го столбца целевому вектору. Если $p_j < 0.05$, тогда мы отклоняем нулевую гипотезу и предполагаем, что j -ый элемент вектора параметров не нулевой.

Определение 5. Нормализованная значимость столбца для j -ого столбца матрицы плана, $j \in \mathcal{J}$ определяется как

$$\hat{p}_j = 1 - \frac{p_j}{\sum_{k=1}^n p_k}.$$

Мы предлагаем использовать нормализованную значимость столбца как меру релевантности столбца целевому вектору:

$$b_j = \text{Rel}(\chi_j) = \hat{p}_j. \quad (33)$$

Далее в разделе 4.2.4 рассматривается выпуклая релаксация поставленной задачи (29), так как она является NP-сложной в силу бинарности элементов вектора $\mathbf{a}^*$. Поэтому необходима аппроксимация этой задачи некоторой задачей выпуклой оптимизации, которая может быть эффективно решена.

Пусть функция Sim даёт положительно определённую матрицу $\mathbf{Q}$. Тогда квадратичная форма (27) — выпуклая функция. Для представления задачи в выпуклой форме мы должны заменить невыпуклое множество $\mathbb{B}^n$ на некоторое выпуклое множество. Естественный кандидат на такую замену — это выпуклая оболочка множества $\mathbb{B}^n$:

$$\text{Conv}(\mathbb{B}^n) = [0, 1]^n.$$

Теперь задача (29) аппроксимирована следующей *задачей выпуклой оптимизации*:

$$\begin{aligned} \mathbf{z}^* &= \arg \min_{\mathbf{z} \in [0, 1]^n} \mathbf{z}^\top \mathbf{Q} \mathbf{z} - \mathbf{b}^\top \mathbf{z} \\ \text{s.t. } \|\mathbf{z}\|_1 &\leq 1. \end{aligned} \quad (34)$$

Мы добавили ограничение на вектор $\mathbf{z}$, чтобы показать, что он может быть рассмотрен, как вектор ненормализованных вероятностей попадания каждого столбца в множество $\mathcal{A}^*$.

Для возврата от непрерывного вектора $\mathbf{z}^*$ к булеву вектору $\mathbf{a}^*$ и как следствие к определению множества активных индексов $\mathcal{A}^*$ (см (23)), мы используем *порог значимости* τ .

Определение 6. Значение τ называется *порогом значимости* если $z_j^* > \tau$ тогда и только тогда когда $a_j^* = 1$ и $j \in \mathcal{A}^*$.

Настройка значения τ зависит от конкретной задачи и основана на допустимом уровне ошибки, числе отобранных столбцов матрицы и значениях внешних критериев качества. Для получения наиболее подходящего значения порога значимости для конкретной задачи нам необходимо проверить некоторый диапазон значений τ .

В заключении приводятся критерии качества, по которым оценивается выбранное подмножество столбцов матрицы плана и результаты вычислительных экспериментов. С помощью тестовых выборок, на которых не работают другие рассмотренные методы выбора столбцов матрицы плана, показано, что предложенный метод позволяет выбрать такое подмножество столбцов матрицы плана, что итоговая модель будет одновременно точной, устойчивой и обладать наименьшей сложностью. Отметим, что этого удаётся добиться для всех конфигураций тестовых выборок.

В **пятой главе** рассматривается задача аппроксимации величины похожести между вершинами графа SimRank с помощью быстрого и эффективного по памяти алгоритма. Наивный алгоритм заключается в выполнении итераций метода неподвижной точки вида

$$\begin{aligned} S_0 &= I \\ S_{k+1} &= W^T S_k W - \text{diag}(W^T S_k W) + I, \end{aligned} \quad (35)$$

где S_k — приближение матрицы S с элементами равными значениям SimRank'a для любой пары вершин графа, $W = \sqrt{c}\bar{A}$, а $\bar{A}$ — матрица смежности графа.

Уравнение неподвижной точки для вычисления матрицы S напрямую следует из итерационного процесса (35):

$$S = W^T S W - \text{diag}(W^T S W) + I. \quad (36)$$

Для получения предлагаемого метода доказана следующая теорема.

Теорема 1. *Решение уравнения (36) существует, единственно и может быть записано в виде:*

$$S = \sum_{k=0}^{\infty} \mathcal{L}(\mathcal{D}^k(I)), \quad (37)$$

где оператор $\mathcal{L}(X)$ даёт решение L дискретного уравнения Ляпунова¹⁵:

$$L = W^T L W + X, \quad (38)$$

и оператор $\mathcal{D}(X)$ определяется следующим образом:

$$\mathcal{D}(X) = -\text{diag}(W^T \mathcal{L}(X) W). \quad (39)$$

¹⁵Hammarling S. Numerical solution of the discrete-time, convergent, non-negative definite Lyapunov equation // Systems & Control Letters. 1991. Т. 17, № 2. С. 137–139.

Из этой теоремы следует следующая малоранговая модель

$$\tilde{S} = I + W^\top W - \text{diag}(W^\top W) + UDU^\top, \quad (40)$$

где I — единичная $n \times n$ матрица, U и D — ортогональная $n \times m$ и диагональная $m \times m$ матрица соответственно. Матрицы U и D надо найти для использования модели в вычислении SimRank'a.

Для этого вместо итерационного процесса (35), который сходится к точной матрице SimRank'a S , рассмотрим следующий итерационный процесс:

$$\begin{aligned} \tilde{S}_0 &= I \\ \tilde{S}_{k+1} &= W^\top \tilde{S}_k W - \text{diag}(W^\top \tilde{S}_k W) + I. \end{aligned} \quad (41)$$

Подставляя (40) в (41) с произвольными начальными матрицами U_0 и D_0 получим следующее уравнение на U_{k+1} и D_{k+1} :

$$\begin{aligned} I + W^\top W - \text{diag}(W^\top W) + U_{k+1} D_{k+1} U_{k+1}^\top &= \\ W^\top (I + W^\top W - \text{diag}(W^\top W) + U_k D_k U_k^\top) W - & \\ \text{diag}(W^\top (I + W^\top W - \text{diag}(W^\top W) + U_k D_k U_k^\top) W) + I, & \end{aligned} \quad (42)$$

из которого следует, что

$$U_{k+1} D_{k+1} U_{k+1}^\top = M_k,$$

где $M_k = W^\top (W^\top W - \text{diag}(W^\top W) + U_k D_k U_k^\top) W - \text{diag}(W^\top (W^\top W - \text{diag}(W^\top W) + U_k D_k U_k^\top) W)$. Вычисление матриц U_{k+1} и D_{k+1} основано на спектральном разложении матрицы M_k :

$$M_k = \tilde{U} \tilde{D} \tilde{U}^\top,$$

где $\tilde{U}$ — это матрица, столбцы которой являются собственными векторами матрицы M_k , а матрица $\tilde{D}$ диагональная, и на диагонали стоят собственные значения матрицы M_k , упорядоченные по убыванию. Используя заранее выбранный ранг r , мы формируем матрицы U_{k+1} и D_{k+1} :

$$U_{k+1} = \tilde{U}_r, \quad D_{k+1} = \tilde{D}_r,$$

где $\tilde{U}_r$ — это прямоугольная матрица, образованная первыми r столбцами матрицы $\tilde{U}$, а $\tilde{D}_r$ — диагональная матрица из первых r строк и r столбцов матрицы $\tilde{D}$. Таким образом, получен алгоритм вычисления U и D , однако для его ускорения мы используем вероятностное спектральное разложение.

На примере графа, полученного из статей Simple English Wikipedia и ссылок на них, покажем ближайшие статьи к заданным в первой строке в Таблице 4.

Таблица 4: Близкие концепты, полученные с помощью предложенного метода аппроксимации значений SimRank'a

GNU	Earth	Liquid
Richard Matthew Stallman	Ceres(planet)	Hematological
Linux operating system	Global warming controversy	Blude
Encyclopedia of Computer Terms	Effects of global warming	Human Blood
January 05	Si	Blood-forming
Sleep (operating system)	Doldrums	Hemic
History of computing	Planetary wind belt	Oxygenated blood
All-in-one PC	Horse latitude	Bloed

Основные результаты работы

Основным результатом являются новые методы оптимизации значений параметров численных методов с помощью методов машинного обучения.

- Предложен метод оптимизации значений параметров многосеточного метода для заданных шаблонов разреженности. Получена стохастическая оценка спектрального радиуса матрицы итераций и поставлена задача оптимизации значений параметров многосеточного метода. Показана сложность вычисления градиента целевой функции по этим параметрам. Вычислительные эксперименты показали, что предложенный метод даёт локально оптимальные значения параметров в отличие от методов из семейства алгебраических многосеточных методов.
- Предложен новый способ оценки сходимости нелинейных итерационных методов решения линейной системы уравнений. Он сводится к вычислению усреднённой ошибки аппроксимации для нескольких начальных приближений. Это даёт оценку сходимости в среднем, в отличие от классической оценки в худшем случае. Предложенная оценка скорости сходимости применена к задаче оптимизации параметров предобусловливателя в методе сопряжённых градиентов.

- Проанализирована проблема мультиколлинеарности при решении задачи регрессии и предложен подход к её решению, основанный на постановке задачи квадратичного программирования. Полученный метод показывает одновременно высокую точность и устойчивость при минимальном количестве отобранных столбцов матрицы плана.
- Разработан метод приближённого вычисления величин похожести между вершинами графа SimRank, основанный на малоранговой модели. Доказана теорема о представлении матрицы из этих величин в виде ряда. Полученный метод обладает существенно более низкой вычислительной сложностью и требует меньше памяти по сравнению с другими подходами к аппроксимации SimRank'a.

Публикации автора по теме диссертации

1. Katrutsa A., Strijov V. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria // Expert Systems with Applications. — 2017. — Т. 76. — С. 1-11.
2. Oseledets I. V., Ovchinnikov G. V., Katrutsa A. M. Fast, memory-efficient low-rank approximation of SimRank // Journal of Complex Networks. — 2016. — Т. 5. — №. 1. — С. 111-126.
3. Katrutsa A. M., Strijov V. V. Stress test procedure for feature selection algorithms // Chemometrics and Intelligent Laboratory Systems. — 2015. — Т. 142. — С. 172-183.
4. Katrutsa A., Daulbaev T., Oseledets I. Black-box learning of multigrid parameters // Journal of Computational and Applied Mathematics. — 2019. — <https://doi.org/10.1016/j.cam.2019.112524>.
5. Оселедец И. В., Бочев М. А., Катруца А. М., Овчинников Г. В. Как оптимизировать предобусловливатели в методе сопряжённых градиентов: стохастический подход // Препринты Института прикладной математики им. М. В. Келдыша РАН. — 2018. — №. 164. — С. 1-26.
6. Катруца А. М. Оптимизация параметров двухсеточного метода с помощью глубоких нейронных сетей // Труды 60-й научной конференции МФТИ, 2017
7. Катруца А. М. Как оптимизировать предобуславливатели в методе сопряжённых градиентов: стохастический подход // Труды 61-й научной конференции МФТИ, 2018