МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное

учреждение высшего образования

«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

На правах рукописи

Гуртовая Ольга Владимировна

Методы онлайн оптимизации квадратичной функции потерь, основанные на использовании случайных признаков Фурье

Специальность 1.2.2. — «Математическое моделирование, численные методы и комплексы программ»

АВТОРЕФЕРАТ

диссертации на соискание учёной степени кандидата

физико-математических наук

Ростов-на-Дону — 2025

Работа выполнена в Федеральном государственном автономном образовательном учреждении высшего образования «ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Научный руководитель:	**Рохлин Дмитрий Борисович**, доктор физико-математических наук, доцент, заведующий кафедрой методов оптимизации и машинного обучения ФГАОУ ВО «Южный федеральный университет»,
Официальные оппоненты:	**Бутакова Мария Александровна**, доктор технических наук, главный научный сотрудник Ростовского филиала акционерного общества «Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте»; АО «НИИАС»
	Гисин Владимир Борисович кандидат физико-математических наук, профессор кафедры математики и анализа данных Федерального государственного бюджетного образовательного учреждения высшего образования «Финансовый университет при Правительстве Российской Федерации»

Защита диссертации состоится «18» декабря 2025 года в 17 часов на заседании диссертационного совета ЮФУ801.01.09 на базе Южного федерального университета по адресу: 344090, г. Ростов-на-Дону, ул. Мильчакова 8А, ауд. 211.

С диссертацией можно ознакомиться в Зональной научной библиотеке Южного федерального университета по адресу: г. Ростов-на-Дону, ул. Зорге, 21-ж и на сайте `https://hub.sfedu.ru/diss/show/1346889/`

Автореферат разослан «31» октября 2025 года.

Ученый секретарь
диссертационного совета
д.ф.-м.н. Говорухин В.Н.

1 Общая характеристика работы

Актуальность темы. Современные задачи в таких областях, как обработка финансовых временных рядов, анализ потоковых данных в интернете вещей, адаптивные системы рекомендаций, робототехника и оптимальное управление, требуют построения математических моделей, способных адаптироваться к изменяющимся условиям в реальном времени. Ключевая сложность заключается в том, что часто бывает неизвестна сама природа данных, а также конкретные механизмы и факторы, вызывающие их изменение, которые могут иметь состязательный (adversarial) характер. В связи с этим актуальна задача математического моделирования и анализа таких процессов. В этом контексте онлайн оптимизация выделяется как одно из наиболее перспективных направлений. В отличие от классической (офлайн) оптимизации, где всё множество данных доступно заранее, онлайн алгоритмы обрабатывают информацию последовательно, постоянно адаптируя свои решения на основе как новых данных, так и предыдущего опыта. Это позволяет моделям динамически перестраиваться и поддерживать высокую эффективность работы без необходимости полного переобучения на всех накопленных данных.

Основным инструментом, исследуемым в диссертации, является метод случайных признаков Фурье, разработанный A. Rahimi и B. Recht [1]. Данный подход позволяет эффективно работать с нелинейными зависимостями путём преобразования данных в пространство случайных признаков, где модель становится линейной относительно своих коэффициентов. Аналогично преобразованию Фурье, ядро (мера схожести признаков) аппроксимируется с помощью набора простых тригонометрических функций. Сначала генерируется набор случайных векторов-частот ω_i из распределения, связанного с ядром преобразованием Фурье, и случайных сдвигов фазы b_i, равномерно распределённых на отрезке $[0, 2\pi]$. Затем для каждой входной точки данных (вектора признаков) x вычисляется её проекция на случайные направления ω_i, к которой прибавляются сдвиги b_i. Новыми признаками будут $\cos(\langle \omega_i, x\rangle + b_i)$. Такой подход можно рассматривать как построение моделей типа «чёрного ящика», способных аппроксимировать широкий класс зависимостей без необходимости точного знания внутренней структуры системы.

Разработанные в работе численные методы онлайн оптимизации, основанные на этой идее, демонстрируют свою универсальность для построения математических моделей в самых разных предметных областях. Процесс моделирования можно описать следующим образом: данные поступают последовательно (онлайн), на каждом шаге по имеющимся признакам $x_t \in \mathbb{R}^d$ (входные параметры системы) строится их образ в пространстве случайных признаков Фурье $\Phi(x_t) \in \mathbb{R}^m$, и на его основе строится прогноз целевой переменной $\hat{y}_t = \langle w_t, \Phi(x_t)\rangle$, где $w_t \in \mathbb{R}^m$ — вектор весов модели. Затем становится известно истинное значение целевой

[1]**Rahimi A.** Random features for large-scale kernel machines / A. Rahimi, B. Recht // Advances in Neural Information Processing Systems 20 (NIPS 2007). — 2007. — P. 1177–1184.

переменной $y_t \in \mathbb{R}$, и вычисляется квадратичная функция потерь

$$l_t(w_t) = (\hat{y}_t - y_t)^2 = (\langle w_t, \Phi(x_t) \rangle - y_t)^2,$$

характеризующая точность модели. Задача состоит в построении такой последовательности векторов весов w_t, которая минимизирует суммарные потери за всё время работы. Эффективность алгоритма оценивается величиной сожаления (regret)

$$R_T(u) = \sum_{t=1}^{T} l_t(w_t) - \sum_{t=1}^{T} l_t(u),$$

представляющего собой разность между накопленными потерями нашего алгоритма и потерями «эксперта» u. Этот «эксперт» может выбирать свою стратегию ретроспективно, обладая полным знанием всех данных. Таким образом, $R_T(u)$ измеряет, насколько мы «сожалеем», что не следовали стратегии u.

Проиллюстрируем широту применения данного подхода в разных задачах математического моделирования, рассмотренных в диссертации.

Моделирование стационарных марковских процессов (глава 1), которые являются фундаментальным инструментом для описания систем, развивающихся во времени. Задача состояла в том, чтобы, наблюдая только текущее состояние системы (входные признаки), построить прогноз не просто следующего значения, а некоторой сложной функции от её будущего состояния через несколько шагов (целевая переменная). В качестве конкретного примера для проверки теоретических выводов была использована модель векторной авторегрессии (VAR). Такая модель позволяет описывать взаимосвязанную динамику нескольких временных рядов (например, курсов валют и цен на акции), где поведение каждого ряда в настоящем зависит от его собственных прошлых значений и прошлых значений других рядов.

Моделирование и анализ сложных физических, экономических и социальных явлений на основе имеющихся табличных данных (глава 2). В частности:

- Моделирование аэроакустического шума, генерируемого при обтекании профиля крыла. Цель: установить нелинейную зависимость между набором входных признаков (частота, угол атаки, толщина профиля и др.) и целевой переменной — уровнем звукового давления.
- Моделирование рынка недвижимости. Цель: установить сложную, многофакторную зависимость между входными признаками, описывающими объект и его местоположение (площадь, количество комнат, характеристики района и т.д.), и целевой переменной — стоимостью дома.
- Моделирование социально-экономического поведения на примере данных об оставлении

чаевых. Цель: описать зависимость между набором входных факторов (общая сумма счёта, день недели, пол плательщика и др.) и целевой переменной — размером оставленных чаевых.

Как будет показано в третьей главе диссертации, разработанные для таких задач численные методы онлайн оптимизации обладают достаточной общностью, позволяющей применять их к принципиально иному классу проблем. Это демонстрируется на примере применения методов онлайн оптимизации для решения задачи моделирования внешнего воздействия в рамках задачи вариационного исчисления с квадратичным функционалом качества. Современные подходы к решению таких задач в онлайн режиме открывают новые возможности для адаптивного управления динамическими системами, обработки нестационарных сигналов и оптимального планирования в условиях неопределённости.

Предложенный в работе метод решения вариационных задач с неизвестным внешним воздействием интегрирует технику тригонометрической аппроксимации с аппаратом онлайн оптимизации. В отличие от стохастических подходов на основе случайных признаков Фурье, данный метод использует детерминированную тригонометрическую аппроксимацию для задач поиска оптимальной траектории. Теоретическая значимость результатов подтверждается строгими оценками ошибок аппроксимации и границами сожаления.

Основой данного исследования является теория онлайн оптимизации, историческое развитие которой восходит к фундаментальным работам J. Hannan [2] и D. Blackwell[3] по теории минимизации сожаления, заложившим основы для последующего формализма предсказания с экспертами. Значительный прорыв был достигнут в работах N. Littlestone [4] и В. Вовка[5], разработавших первые практически применимые алгоритмы агрегирования прогнозов.

Современный этап развития теории онлайн оптимизации опирается на результаты, изложенные в классических монографиях N. Cesa-Bianchi и G. Lugosi[6], а также S. Shalev-Shwartz[7]. Дальнейшее развитие и систематизация этого направления представлены в современных трудах E. Hazan [8] и F. Orabona[9]. Важными достижениями в этой области стали разработка адаптив-

[2] **Hannan J.** Approximation to Bayes risk in repeated play / J. Hannan // Contributions to the Theory of Games. — 1957. — Vol. 3. — P. 97–139.

[3] **Blackwell D.** An analog of the minimax theorem for vector payoffs / D. Blackwell // Pacific Journal of Mathematics. — 1956. — Vol. 6, no. 1. — P. 1–8. — DOI: 10.2140/pjm.1956.6.1.

[4] **Littlestone N.** Mistake bounds and logarithmic linear-threshold learning algorithms / N. Littlestone // University of California, Santa Cruz, Technical Report. — 1989. — 43 p.

[5] **Vovk V. G.** Aggregating strategies / V. G. Vovk // Proceedings of the Third Annual Workshop on Computational Learning Theory (COLT 1990). — 1990. — P. 371–383.

[6] **Cesa-Bianchi N.** Prediction, learning, and games / N. Cesa-Bianchi, G. Lugosi. — Cambridge: Cambridge University Press, 2006. — 394 p.

[7] **Shalev-Shwartz S.** Online Learning and Online Convex Optimization / S. Shalev-Shwartz // Foundations and Trends® in Machine Learning. — 2011. — Vol. 4, no. 2. — P. 107–194. — DOI: 10.1561/2200000018.

[8] **Hazan E.** Introduction to online convex optimization / E. Hazan // Foundations and Trends® in Optimization. — 2016. — Vol. 2, no. 3–4. — P. 157–325. — DOI: 10.1561/2400000013.

[9] **Orabona F.** A modern introduction to online learning / F. Orabona // arXiv preprint. — 2019. — arXiv:1912.13213. — URL: https://arxiv.org/abs/1912.13213 (дата обращения: 15.01.2025).

ных алгоритмов оптимизации, таких как AdaGrad [10], и исследование методов ядерного последовательного обучения[11].

Теоретические основы анализа алгоритмов для зависимых данных были заложены в работе A. Agarwal и J. C. Duchi[12], а задачи регрессионного анализа для временных рядов получили развитие в исследовании O. Anava, E. Hazan и S. Mannor[13].

Также не менее важны разработки в области многоядерного обучения[14], которые предоставляют мощные инструменты для комбинирования информации из различных источников, что позволяет создавать более гибкие и точные модели. В контексте возрастающей сложности моделей и данных актуальным становится и автоматизированное машинное обучение (например, платформа AutoGluon[15]), которое стремится автоматизировать процесс выбора моделей, настройки гиперпараметров и создания признаков. Эти направления — многоядерное обучение и AutoML — отражают две ключевые тенденции современного машинного обучения: необходимость эффективной интеграции разнородных данных и автоматизации сложных этапов анализа. В связи с этим, во второй главе диссертационной работы предлагаются новые онлайн алгоритмы многоядерного обучения VAW^2 и S-VAW^2.

Цели работы. Целью диссертационной работы является разработка математических моделей и численных методов онлайн оптимизации, основанных на использовании случайных признаков Фурье для сведения задач непараметрической регрессии к конечномерному случаю. В качестве варианта данного подхода исследуется его детерминированный аналог — аппроксимация тригонометрическими полиномами — для решения задач вариационного исчисления. Ключевой задачей для всех рассматриваемых методов является вывод оценок сожаления.

Для достижения указанной цели необходимо решить следующие задачи.

1. Разработать методику применения численного метода Вовка–Азури–Вармута (VAW) к случайным признакам Фурье для моделирования данных с марковской зависимостью с теоретическим обоснованием и экспериментальной проверкой на моделях AR(1) и VAR(1).

2. Разработать новый численный метод VAW^2, основанный на многоядерном подходе и построении ансамбля моделей, для моделирования различных регрессионных зависимо-

[10] **Duchi J.** Adaptive subgradient methods for online learning and stochastic optimization / J. Duchi, E. Hazan, Y. Singer // Journal of Machine Learning Research. — 2011. — Vol. 12, no. 7. — P. 2121–2159.

[11] **Kivinen J.** Online learning with kernels / J. Kivinen, A. J. Smola, R. C. Williamson // IEEE Transactions on Signal Processing. — 2004. — Vol. 52, no. 8. — P. 2165–2176. — DOI: 10.1109/TSP.2004.830991.

[12] **Agarwal A.** The generalization ability of online algorithms for dependent data / A. Agarwal, J. C. Duchi // IEEE Transactions on Information Theory. — 2012. — Vol. 59, no. 1. — P. 573–587. — DOI: 10.1109/TIT.2012.2214202.

[13] **Anava O.** Online learning for time series prediction / O. Anava, E. Hazan, S. Mannor // Proceedings of the 28th Conference on Learning Theory (COLT 2015). — 2015. — P. 172–184.

[14] **Lanckriet G. R. G.** Learning the kernel matrix with semidefinite programming / G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, M. I. Jordan // Journal of Machine Learning Research. — 2004. — Vol. 5. — P. 27–72. — URL: https://www.jmlr.org/papers/volume5/lanckriet04a/lanckriet04a.pdf (дата обращения: 15.01.2025).

[15] **Erickson N.** Autogluon-tabular: Robust and accurate automl for structured data / N. Erickson [et al.] // arXiv preprint. — 2020. — arXiv:2003.06505. — URL: https://arxiv.org/abs/2003.06505 (дата обращения: 15.01.2025).

стей. Установить для него оценку сожаления. Провести сравнительный анализ предложенного численного метода с известными из литературы родственными алгоритмами на ряде эталонных наборов данных.

3. Разработать трёхуровневый численный метод S-VAW^2, объединяющий иерархическое агрегирование и стратегии масштабирования данных. Провести сравнительный анализ с ведущим AutoML-фреймворком.

4. Сформулировать и решить задачу моделирования внешнего воздействия динамической системы с квадратичным функционалом качества, включая её сведение к конечномерной задаче и вывод границ статического и динамического сожалений.

5. Реализовать комплекс программ на Python для численного моделирования и сравнительного анализа предложенных онлайн алгоритмов.

Объект исследования — алгоритмы онлайн оптимизации в гильбертовых пространствах для задач, где данные поступают последовательно и могут иметь как стохастическую, так и состязательную (adversarial) природу.

Предмет исследования — методология сведения бесконечномерных задач онлайн оптимизации к конечномерным и анализ качества соответствующих алгоритмов. В частности, исследуются:

- применение случайных признаков Фурье для аппроксимации ядерных методов в задачах онлайн регрессии;
- применение аппроксимации тригонометрическими полиномами в задачах вариационного исчисления с неизвестным внешним воздействием;
- теоретические оценки границ сожаления как ключевая метрика качества разработанных подходов.

Методы исследования. Работа опирается на методы функционального анализа (в частности, теорию гильбертовых пространств с воспроизводящим ядром), теории вероятностей (включая теорию марковских цепей), методы выпуклого анализа (включая субградиентные методы онлайн оптимизации).

Научная новизна исследования. Результаты, выносимые на защиту, являются новыми. Автором совместно с научным руководителем проводилась постановка задач, обсуждались полученные основные результаты и формулировки выводов.

Теоретическое значение исследования заключается в развитии теории онлайн оптимизации для решения задач в гильбертовых пространствах. В работе установлены новые теоретические оценки сожаления для предложенного класса многоядерных алгоритмов, а также для алгоритмов, решающих задачи непараметрической регрессии с марковскими данными и вариационного исчисления с неизвестным внешним воздействием.

Практическая значимость исследования определяется разработкой и экспериментальной апробацией эффективных численных методов VAW^2 и S-VAW^2, работающих в онлайн режиме. Данные алгоритмы реализуют гибкий многоядерный подход к онлайн регрессии на основе вычислительно эффективной техники случайных признаков Фурье. Проведённые эксперименты на широком наборе реальных данных показали, что предложенный подход демонстрирует результаты, сопоставимые с ведущим AutoML-фреймворком AutoGluon-Tabular в задачах регрессии. Все разработанные алгоритмы были реализованы в виде программного модуля на языке Python, что подтверждает их практическую применимость и создаёт основу для их дальнейшего использования в прикладных системах.

Основные положения, выносимые на защиту.
В области математического моделирования:

1. Разработан подход к моделированию условных математических ожиданий для стационарных марковских процессов по одной наблюдаемой траектории, основанный на аппроксимации случайными признаками Фурье.

2. Предложен метод математического моделирования нелинейных регрессионных зависимостей, когда данные обладают состязательным свойством, на основе иерархических ансамблей моделей со случайными признаками Фурье.

3. Сформулирована и исследована задача математического моделирования внешнего воздействия неизвестной природы в вариационных задачах с квадратичным функционалом качества в режиме онлайн.

В области численных методов:

4. Разработан численный метод для решения задач онлайн регрессии на марковских данных, основанный на комбинации алгоритма Вовка–Азури– Вармута (VAW) со случайными признаками Фурье. Для данного метода получены теоретические оценки сожаления для квадратичной функции потерь, которые обосновывают его эффективность в условиях зависимых данных. Экспериментально подтверждена конкурентоспособность предложенного подхода на моделях авторегрессии первого порядка и векторной авторегрессии.

5. Предложен новый двухуровневый численный метод многоядерного обучения VAW^2, обладающий значительно более низкой вычислительной сложностью (порядка $O(Nm^2)$ на итерацию) по сравнению с прямым применением алгоритма VAW к конкатенированным векторам признаков ($O(N^2m^2)$). Для метода VAW^2 установлена оценка сожаления порядка $O(T^{1/2}\ln T)$, что подтверждает его теоретическую эффективность. Здесь N — число ядер, m — количество случайных признаков для каждого ядра; T — длина выборки. Вычислительные эксперименты показали, что алгоритм VAW^2 превосходит в качестве предсказаний известные из литературы родственные численные методы на ряде наборов данных.

6. Разработан трёхуровневый численный метод (S-VAW^2), интегрирующий иерархическое агрегирование и различные стратегии масштабирования данных. Экспериментально показано, что данный метод является вычислительно более эффективной альтернативой AutoML-системе AutoGluonTabular для решения задач регрессии на широком наборе эталонных данных.

7. Для задачи моделирования внешнего воздействия динамической системы с квадратичным функционалом качества теоретически обосновано сведение исходной бесконечномерной задачи к конечномерной и выведены границы статического и динамического сожалений. Проделаны вычислительные эксперименты, подтвердившие теоретические результаты.

В области программного обеспечения:

8. Разработан комплекс программ на языке Python, позволяющий проводить численное моделирование и сравнительный анализ предложенных онлайн алгоритмов, основанных на использовании случайных признаков Фурье: алгоритм VAW; двухуровневый алгоритм VAW^2; трёхуровневый алгоритм S-VAW^2. На основе данного комплекса проведены вычислительные эксперименты, подтвердившие теоретические выводы и продемонстрировавшие эффективность разработанных методов.

Степень достоверности результатов. Достоверность результатов, полученных в диссертации, обеспечивается строгостью приведённых доказательств, а также имеющимися публикациями в рецензируемых изданиях и выступлениями на конференциях. Все численные эксперименты, проводимые в рамках диссертационной работы, находятся в открытом доступе.

Апробация результатов. Результаты настоящего исследования были представлены на следующих конференциях.

- Всероссийская научно-практическая конференция «Математика, информатика, компьютерные науки, математическое моделирование, образование (МИКМО-2024)» (Симферополь, 2024);

- XIX Владикавказская молодёжная математическая школа (Владикавказ, 2024);
- Санкт-Петербургская молодёжная конференция по теории вероятностей и математической физике (Санкт-Петербург, 2024);
- Международная научная конференция «Порядковый анализ и смежные вопросы математического моделирования, XVIII: Теория операторов и дифференциальные уравнения» (РСО-А, Дзинага, 2025);
- Международная научная конференция «Современные методы и проблемы теории операторов и гармонического анализа и их приложения - 2025 (ОТНА-2025)» (Ростов-на-Дону, 2025).

Публикации и личный вклад автора. Основные результаты диссертационного исследования изложены в 7 научных публикациях, из них 4 в сборниках трудов конференций. Статья [2] опубликована в журнале, входящем в международную базу данных Scopus; статья [3] — в журнале, входящем в базы данных Scopus и Web of Science; получено свидетельство о государственной регистрации программы для ЭВМ [1]. Тезисы докладов [4], [5], [6], [7] опубликованы в материалах конференций. Кроме того, статья[16] принята к публикации в научном издании, входящем в Перечень ВАК. Имеется также препринт[17].

Статьи [3], [2] опубликованы в соавторстве с научным руководителем. Д.Б. Рохлину принадлежат постановки задачи, указание методов исследования и общее руководство. Автору диссертации принадлежат доказательства теорем и численные эксперименты.

Основные положения, выносимые на защиту, являются личным вкладом автора.

Структура и объём диссертации. Диссертационная работа состоит из введения, трёх глав, заключения, приложений и списка литературы, содержащего 115 наименований. Полный объём диссертации составляет 142 страницы (в том числе приложений 21 стр.).

Благодарности. Автор искренне благодарит своего научного руководителя Дмитрия Борисовича Рохлина за наставничество и поддержку.

Диссертационная работа выполнена при поддержке Регионального научно-образовательного математического центра ЮФУ, соглашение Минобрнауки России № 075-02-2025-1720.

[16] **Гуртовая, О. В.** О тройном алгоритме VAW с масштабированием для многоядерной онлайн линейной регрессии. / О. В. Гуртовая // Программная инженерия. - 2025 (принята к публикации).

[17] **Rokhlin, D. B.** Random feature-based double Vovk-Azouny-Warmuth algorithm for online multi-kernel learning / D. B. Rokhlin, O. V. Gurtovaya // arXiv preprint. — 2025. — arXiv:2503.20087 [cs.LG]. — URL: `https://doi.org/10.48550/arXiv.2503.20087`.

Первая глава диссертации посвящена задаче оценивания условных математических ожиданий в моделях с марковской зависимостью и состоит из 5 параграфов. В качестве основного инструмента используется численный метод Вовка-Азури-Вармута (VAW) с ограничениями, который применяется для построения аппроксимации условного математического ожидания некоторой ограниченной функции.

В параграфах 1.1 и 1.2 вводятся постановка задачи и общий численный метод VAW. Рассматривается задача восстановления условного математического ожидания $h(x) = \mathsf{E}(g(X_0, \ldots, X_s) \mid X_0 = x)$ по единственной траектории стационарного марковского процесса X_t, обладающего свойством β-перемешивания. При этом функция $g : \mathbb{R}^{(s+1)d} \to \mathbb{R}$ является борелевской и равномерно ограниченной некоторой константой $\overline{g}$.

Определение 1. Коэффициент β-перемешивания стационарного марковского процесса X задаётся следующей формулой:

$$\beta_X(n) = \int_{\mathbb{R}^d} \pi(dx) d_{\mathrm{TV}}(\pi, P_n(\cdot; x)),$$

где $d_{\mathrm{TV}}(\mathsf{P}, \widetilde{\mathsf{P}}) = \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mathsf{P}(A) - \widetilde{\mathsf{P}}(A)| = \frac{1}{2}\mathsf{E}_{\mathsf{Q}}|\xi - \widetilde{\xi}|$ — расстояние по вариации. Здесь ξ и $\widetilde{\xi}$ — плотности мер P и $\widetilde{\mathsf{P}}$ относительно доминирующей меры Q, $P_n(\cdot; x)$ — переходное ядро за n шагов.

Определение 2. Процесс X обладает свойством геометрического β-перемешивания, если

$$\beta_X(n) \leq C\rho^n,$$

где $C > 0$, $0 < \rho < 1$ — константы.

Для аппроксимации искомой функции используется алгоритм VAW, который строит коэффициенты линейной регрессии $w \in [-c, c]^m$ по некоторому набору нелинейных признаков $\Phi(x) = (\phi_1(x), \ldots, \phi_m(x))$. На шаге t коэффициенты w_t вычисляются следующим образом:

$$w_t = \operatorname*{argmin}_{w \in [-c,c]^m} \left\{ \frac{1}{2} \sum_{j=1}^{t-1} ((w, \Phi(X_j)) - Y_j)^2 + \frac{1}{2} \langle w, \Phi(X_t) \rangle^2 + \frac{\epsilon}{2} \|w\|_2^2 \right\}. \tag{1}$$

Пусть

$$\mathcal{E}(v, w) = \mathsf{E}\left[\frac{1}{2}\left(\langle v, \Phi(X_0)\rangle - Y_0\right)^2 - \frac{1}{2}\left(\langle w, \Phi(X_0)\rangle - Y_0\right)^2\right].$$

Ключевой результат параграфа 1.2 формулируется в виде теоремы 1.2 и основан на результатах работы A. Agarwal и J. C. Duchi[12]. В данной работе был приведён пример, показавший, что малого сожаления для алгоритма VAW с ограничениями недостаточно в случае зависимых данных. Дополнительно требуется некоторое свойство устойчивости алгоритма, которое и лежит в основе доказательства следующей теоремы.

Теорема 1.2. *Пусть весовые коэффициенты w_t генерируются алгоритмом VAW (1) с ограничениями $w \in [-c, c]^m$. Предположим, что процесс X_t обладает свойством геометрического β-перемешивания с коэффициентом $\beta_X(n) \leq \beta_X(0)e^{-\gamma n}$. Тогда*

$$\mathsf{E}\mathcal{E}(\bar{w}_n, w) = \hat{O}\left(\frac{c^2m^3 + \overline{g}^2m}{n}(s + 1/\gamma)\right),$$

где n — длина траектории, m — число признаков, c — параметр ограничения, g — граница функции потерь.

Данная оценка обладает полиномиальной зависимостью от числа признаков m (порядка m^3), что делает предложенный подход непрактичным для моделей с высокой размерностью признакового пространства.

Для преодоления этого недостатка в параграфе 1.3 исследуется подход на основе случайных признаков Фурье. Этот метод позволяет свести исходную задачу нелинейной регрессии к линейной регрессии в конечномерном пространстве, обеспечивая при этом эффективную аппроксимацию функций из гильбертова пространства с воспроизводящим ядром (RKHS).

Следуя работе A. Rahimi и B. Recht[1], введём гильбертово пространство

$$\mathcal{H}(\phi, p) = \left\{x \mapsto f(x) = \int \alpha(\theta)\phi(x;\theta)\, d\theta : \int \frac{\alpha^2(\theta)}{p(\theta)}\, d\theta < \infty\right\},$$

со скалярным произведением $\langle f, g\rangle_{\mathcal{H}} = \int \frac{\alpha(\theta)\beta(\theta)}{p(\theta)}\, d\theta$, где $g(x) = \int \beta(\theta)\phi(x;\theta)\, d\theta$. Для признаков $\phi(x;\theta) = \cos(\langle\omega, x\rangle + b)$, где $\theta = (\omega, b)$, и соответствующей плотности вероятности $p(\omega, b)$, являющейся произведением гауссовской плотности и плотности равномерного распределения, доказывается, что $H(\phi, p)$ является RKHS с гауссовским ядром. Объединяя этот результат с выводами параграфа 1.2, доказывается основная теорема (теорема 1.3).

Теорема 1.3. *Пусть $\Phi(x)$ — вектор случайных признаков Фурье размера m и вектор $\bar{w}_n$ построен с помощью алгоритма VAW с ограничением $c = \|f\|_p/m$ для некоторой функции $f \in \mathcal{H}(\phi, p)$. Если процесс X_t является геометрически β-перемешивающим и число признаков $m \propto \sqrt{n}$, то*

$$\begin{aligned}\mathsf{E}\frac{1}{2}\int(\langle\Phi(x), \bar{w}_n\rangle - h(x))^2\pi(dx) \leq \tilde{O}\left(\frac{\|f\|_p^2 + g^2}{\sqrt{n}}(s + 1/\gamma)\right)&\\ + \int(f(x) - h(x))^2\pi(dx).&\end{aligned}$$

Эта оценка не зависит полиномиально от размерности m, а определяется гладкостью целевой функции, что является значительным теоретическим улучшением и делает подход практически применимым.

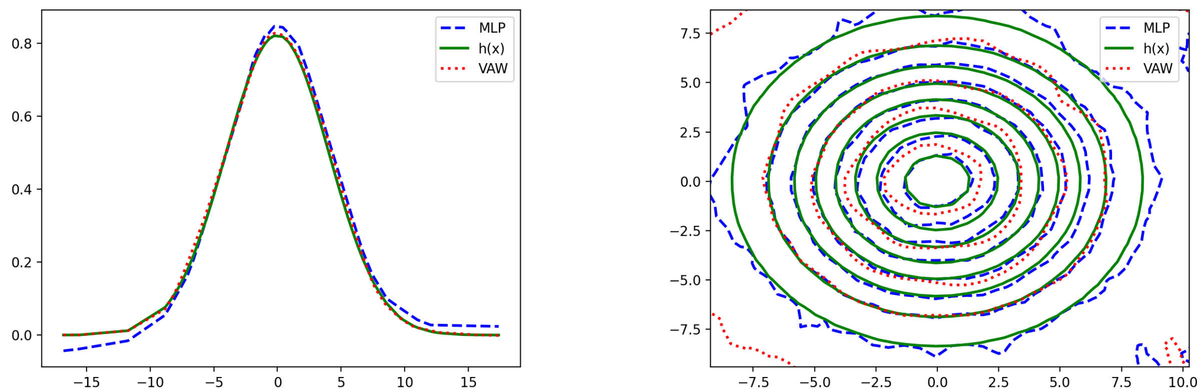
В параграфе 1.4 для апробации теоретических результатов рассматривается модель векторной авторегрессии (VAR)

$$X_t = \kappa Q X_{t-1} + R\xi_t, \quad X_0 = x,$$

где $\xi_t \sim N(0, I_d)$ — н.о.р., $\kappa \in (0,1)$, Q, R — $d \times d$-матрицы, и Q ортогональна. Для этого процесса доказывается выполнение необходимых условий (геометрическое β-перемешивание). Для случая $g(x_s) = e^{-\beta\|x_s\|^2/2}$ удаётся в явном виде вычислить целевую функцию $h(x)$ и показать, что $h(x) \in \mathcal{H}(\phi, p)$, что обнуляет ошибку аппроксимации в теореме 1.3. Это позволяет получить явную оценку скорости сходимости:

$$\mathsf{E}\frac{1}{2}\int (\langle \Phi(x), \bar{w}_n\rangle - h(x))^2 \pi(dx) \le \tilde{O}\left(\frac{\sigma^{2d}}{\beta\kappa^{2s}\sqrt{n}}(s + 1/\gamma)\right).$$

Результаты численных экспериментов для VAR при $d = 1$ и $d = 2$ подтверждают теоретические выводы и демонстрируют конкурентоспособность предложенного численного метода по сравнению со стандартными нейросетевыми моделями (MLPRegressor), особенно на длинных выборках. На рис. 1 представлены график (для $d = 1$) и линии уровня (для $d = 2$) функции $h(x)$, восстановленной алгоритмами VAW и MLPRegressor при объёме выборки $n = 100000$, шаге $s = 5$ и количестве признаков $m = 80$ (для $d = 1$) и $m = 400$ (для $d = 2$).



(a) График h и его аппроксимаций ($d = 1$) (b) Линии уровня h и их аппроксимации ($d = 2$)

Рис. 1: Визуализация функции h

Комплекс программ для воспроизведения экспериментов находится в открытом доступе на платформе GitHub [18].

В заключительном параграфе 1.5 подводятся итоги и формулируются выводы по всем результатам, полученным в главе.

[18] **Gurtovaya, O. V.** Vovk-Azoury-Warmuth algorithm and random Fourier features for a regression problem with Markovian data [Электронный ресурс] / O. V. Gurtovaya — 2025. — URL: `https://github.com/O-Gurt/VAW_Markov` (дата обращения: 21.08.2025).

Вторая глава диссертации посвящена разработке и анализу вычислительно эффективных методов многоядерной онлайн оптимизации. Рассматривается задача регрессии с квадратичной функцией потерь в пространствах RKHS на данных, обладающих состязательным свойством.

В параграфе 2.1 ставится задача построения ансамбля моделей на основе многоядерного подхода для решения проблемы выбора оптимального ядра в задачах нелинейной регрессии.

Рассматриваются N гильбертовых пространств $\mathcal{H}_j$, каждое из которых является воспроизводящим (RKHS) со своим ядром κ_j. Вводится пространство

$$\mathcal{H} = \mathcal{H}_1 + \cdots + \mathcal{H}_N = \left\{ f = \sum_{j=1}^{N} f_j \mid f_j \in \mathcal{H}_j \right\},$$

которое, как известно, также является RKHS. Пусть $\Phi_\theta(x)$ представляет собой конкатенированный вектор случайных признаков:

$$\Phi_\theta(x) = (\Phi_{\theta_1}(x), \ldots, \Phi_{\theta_N}(x)), \quad \Phi_{\theta_j}(x) = (\phi_j(x, \theta_{jk}))_{k=1}^{m}. \tag{2}$$

Такой подход позволяет объединить случайные признаки Φ_{θ_j}, ассоциированные с каждым индивидуальным ядром κ_j, в единый Nm-мерный вектор. Таким образом, алгоритм VAW обрабатывает информацию, полученную из множества ядер, в унифицированной форме.

Для такого подхода в параграфе 2.2 выводится следующая оценка сожаления.

Теорема 2.1. *Пусть $w_t = (w_{t,1}, \ldots, w_{t,Nm}) \in \mathbb{R}^{Nm}$ — вектор весов, полученный при помощи алгоритмов VAW с использованием последовательности данных $(\Phi_\theta(x_t), y_t)$. Тогда*

$$\begin{aligned}\frac{1}{2}\mathsf{E}_\theta \sum_{t=1}^{T} (\langle w_t, \Phi_\theta(x_t)\rangle - y_t)^2 \le \frac{1}{2}\sum_{t=1}^{T} (f(x_t) - y_t)^2 + \left(\frac{\lambda}{2} + NT\right) \frac{\|f\|_{\mathcal{H}}^2}{m} \\ + \frac{NmY^2}{2} \ln\left(1 + \frac{2T}{\lambda}\right)\end{aligned} \tag{3}$$

Данная оценка справедлива для любой функции $f \in \mathcal{H}$. Более того, при $T \to +\infty$

$$\begin{aligned}\frac{1}{2}\mathsf{E}_\theta \sum_{t=1}^{T} (\langle w_t, \Phi_\theta(x_t)\rangle - y_t)^2 \le \frac{1}{2} \inf_{f \in B_R(\mathcal{H})} \sum_{t=1}^{T} (f(x_t) - y_t)^2 \\ + O\left(N(R^2 + Y^2 \ln T)\sqrt{T}\right),\end{aligned} \tag{4}$$

если $m \propto \sqrt{T}$.

Главный недостаток такого подхода — его высокая вычислительная сложность (порядка $O(N^2 m^2)$ на каждую итерацию), делающая его неприменимым для большого числа ядер.

Для преодоления этого недостатка в главе разработан новый двухуровневый численный метод VAW^2. При таком подходе алгоритм VAW применяется для построения экспертных стратегий по случайным признакам, сгенерированным для каждого ядра на первом уровне, а затем для объединения их предсказаний на втором уровне. Было показано, что для этого алгоритма оценка сожаления сопоставима с базовым подходом.

Теорема 2.2. *Пусть $w_{t,j} \in \mathbb{R}^m$ — вектор весов, генерируемый алгоритмами VAW, применёнными к последовательности $(\Phi_{\theta_j}(x_t), y_t)$ для каждого эксперта j. Вектор $\alpha_t \in \mathbb{R}^N$ также генерируется алгоритмом VAW, но уже применённым к вектору экспертных предсказаний z_t:*

$$z_t = (\langle w_{t,1}, \Phi_{\theta_1}(x_t)\rangle, \ldots, \langle w_{t,N}, \Phi_{\theta_N}(x_t)\rangle).$$

Тогда

$$\begin{aligned}\frac{1}{2}\mathsf{E}_\theta \sum_{t=1}^{T}(\langle \alpha_t, z_t\rangle - y_t)^2 \le \frac{1}{2}\sum_{t=1}^{T}(y_t - f_j(x_t))^2 + \frac{\lambda}{2} \\ + \left(\frac{\lambda}{2} + T\right)\frac{\|f_j\|^2_{\mathcal{H}_j}}{m} + \frac{mY^2}{2}\ln\left(1 + \frac{2T}{\lambda}\right) \\ + \frac{NY^2}{2}\ln\left(1 + \frac{NY^2}{\lambda}\left(2T(T+1) + 2mT\ln\left(1+\frac{2T}{\lambda}\right)\right)\right)\end{aligned} \tag{5}$$

Это справедливо для любой функции $f_j \in \mathcal{H}_j$ и любого $j = 1, \ldots, N$. При $T \to +\infty$

$$\begin{aligned}\frac{1}{2}\mathsf{E}_\theta \sum_{t=1}^{T}(\langle \alpha_t, z_t\rangle - y_t)^2 \le \frac{1}{2}\min_{1\le j\le N}\inf_{f_j \in B_R(\mathcal{H}_j)}\sum_{t=1}^{T}(y_t - f_j(x_t))^2 \\ + O\left((R^2 + Y^2\ln T)\sqrt{T}\right), \quad \text{если} \quad m \propto \sqrt{T}.\end{aligned} \tag{6}$$

Помимо сильных теоретических гарантий алгоритм VAW^2 обладает низкой вычислительной сложностью (порядка $O(Nm^2)$), что является его ключевым практическим преимуществом.

Предположим, что верхняя граница Y для значений y_t известна. Возьмём усечённую версию прогнозов экспертов

$$\overline{z}_t = \min(Y, \max(z_t, -Y)), \quad z_{t,j} = \langle w_{t,j}, \Phi_{\theta_j}(x_t)\rangle, \tag{7}$$

где операции max и min применяются к каждой компоненте вектора. Было показано, что при использовании мета-алгоритма EWA для агрегирования прогнозов экспертов оценка (5) может быть улучшена:

$$\frac{1}{2}\mathsf{E}_\theta \sum_{t=1}^{T}(\langle \alpha_t, \overline{z}_t\rangle - y_t)^2 \le \frac{1}{2}\sum_{t=1}^{T}(f_j(x_t) - y_t)^2 + 4Y^2\ln N$$

$$+\left(\frac{\lambda}{2}+T\right)\frac{\|f_j\|^2_{\mathcal{H}_j}}{m}+\frac{mY^2}{2}\ln\left(1+\frac{2T}{\lambda}\right). \qquad (8)$$

Параграф 2.3 содержит два подраздела. В разделе 2.3.1 проведено численное сравнение предложенных численных методов с современными методами онлайн многоядерной регрессии, включая Raker[19] и OMKL-GF[20]. В сравнении также участвовали и другие алгоритмы, в которых в качестве экспертов выступал алгоритм VAW, применённый к разным ядрам, а в качестве мета-алгоритмов были использованы: агрегирующий алгоритм Вовка[21], ML-Prod[22], ML-Poly[22], BOA[22]. Эксперименты выполнены на пяти наборах данных, включая четыре реальных набора: Airfoil, Naval, Bias, Concrete. Численный метод VAW^2 продемонстрировал значительное превосходство по метрике среднеквадратичной ошибки (MSE) на всех реальных наборах данных. Анализ графика (рис. 2) подтвердил, что алгоритм VAW^2 обеспечивает монотонное уменьшение среднеквадратичной ошибки на всех наборах данных, что свидетельствует о его высокоэффективной и стабильной работе на протяжении всего процесса обучения. Комплекс программ для воспроизведения экспериментов находится в открытом доступе на платформе GitHub[23].

В разделе 2.3.2 приводится трёхуровневый численный метод S-VAW^2 (Scaled Vovk-Azoury-Warmuth), объединяющий в себе многоядерное моделирование с использованием случайных признаков Фурье и стратегии предварительной обработки данных. Численный метод S-VAW^2 реализует подход, при котором исходные данные подвергаются различным видам масштабирования, и для каждой модификации независимо запускается алгоритм VAW^2. После этого S-VAW^2 агрегирует полученные предсказания экспертных прогнозов при помощи метода VAW, что позволяет ему показывать более стабильные и точные предсказания, а также справляться с разнородностью в данных. В реализации алгоритма участвовало восемь комбинаций стратегий масштабирования, включая методы `StandardScaler`, `MinMaxScaler` и `RobustScaler` для признаков и `MinMaxScaler` для целевых переменных. Во всех комбинациях дополнительно участвовал вариант без применения масштабирования к данным. Алгоритм S-VAW^2 продемонстрировал сопоставимую MSE с одним из ведущих AutoML-фреймворков, AutoGluon-Tabular, превзойдя его на большинстве (10 из 13) исследованных датасетов (см. таблицу 1).

Комплекс программ для воспроизведения экспериментов находится в открытом доступе на

[19]**Shen Y.** Random feature-based online multi-kernel learning in environments with unknown dynamics / Y. Shen, T. Chen, G. B. Giannakis // Journal of Machine Learning Research. 2019. Vol. 20, no. 22. P. 1–36.

[20]**Ghari P. M.** Graph-aided online multi-kernel learning / P. M. Ghari, Y. Shen // Journal of Machine Learning Research. 2023. Vol. 24, no. 21. P. 1–44.

[21]**Cesa-Bianchi N.** Prediction, learning, and games / N. Cesa-Bianchi, G. Lugosi. — Cambridge: Cambridge University Press, 2006. — 394 p.

[22]**Gaillard P.** A second-order bound with excess losses / P. Gaillard, G. Stoltz, T. Van Erven // Proceedings of the 27th Conference on Learning Theory (COLT 2014). — 2014. — P. 176–196.

[23]**Gurtovaya, O. V.** Random feature-based double Vovk-Azoury-Warmuth algorithm for online multi-kernel learning [Электронный ресурс] / O. V. Gurtovaya — 2025. — URL: `https://github.com/O-Gurt/VAW2` (дата обращения: 13.08.2025).

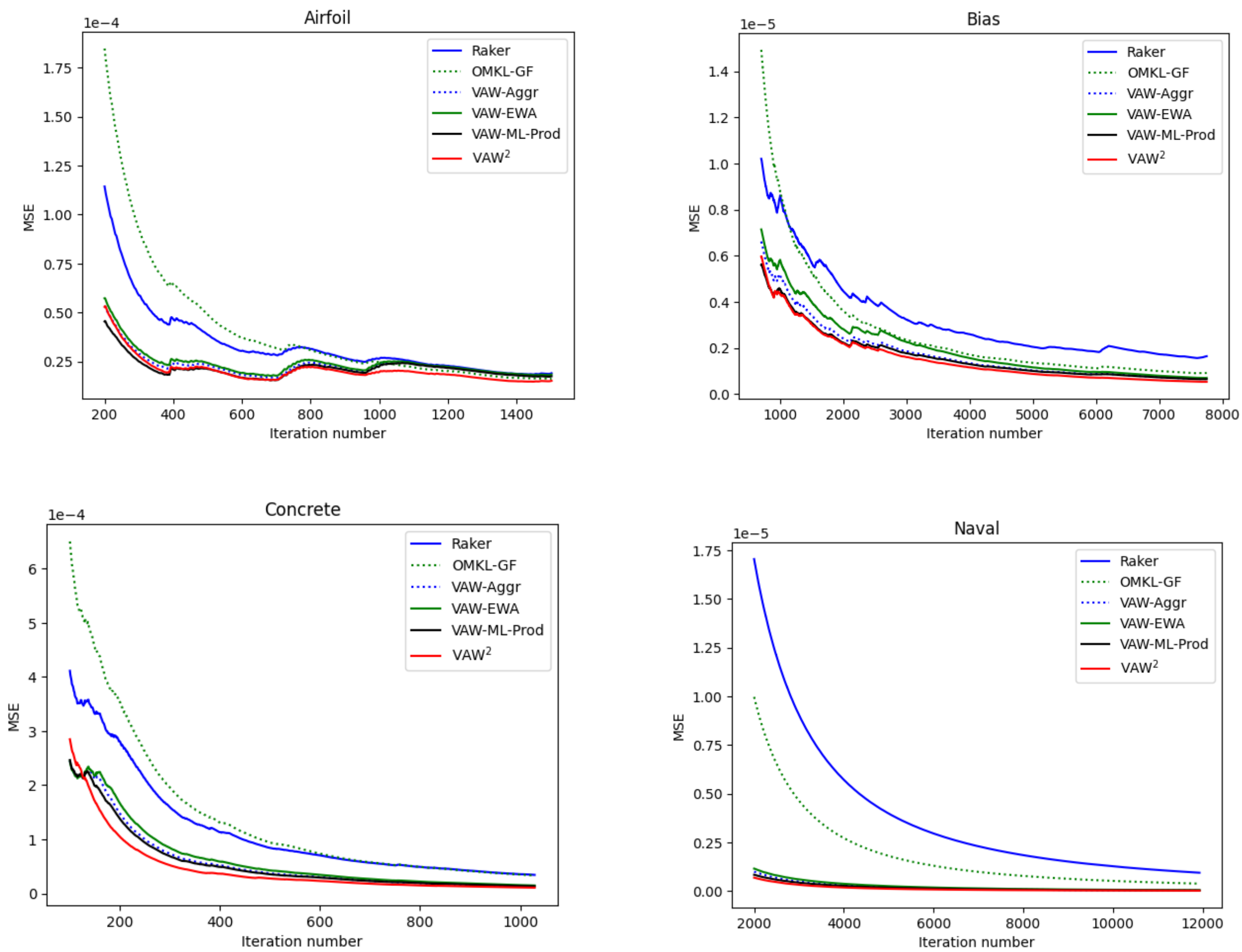


Рис. 2: Итеративное изменение MSE для многоядерных алгоритмов.

платформе GitHub[24].

В параграфе 2.4 подводятся итоги и формулируются выводы по результатам, представленным во второй главе.

Третья глава диссертации посвящена применению методов онлайн оптимизации для решения задачи моделирования внешнего воздействия в рамках задачи вариационного исчисления с квадратичным функционалом качества.

В параграфе 3.1 ставится постановка задачи. Рассмотрим многошаговую игру между игроком и противником. На шаге n игрок выбирает функцию управления $u_n(t)$, где $t \in [0,1]$ и $\int_0^1 u_n(t)\,dt = 0$, и начальную точку $a_n \in \mathbb{R}$. Затем противник выбирает функцию $h_n(t)$, $t \in [0,1]$, и игрок несёт потери

[24] **Gurtovaya, O. V.** Triple Vovk-Azuri-Warmuth Algorithm as an Automated Machine Learning Approach [Электронный ресурс] / O. V. Gurtovaya — 2025. — URL: `https://github.com/O-Gurt/TVAW` (дата обращения: 13.08.2025).

	S-VAW^2	AG-TABULAR
AIRFOIL	**13.686**	20.160
CONCRETE	**20.977**	163.799
BIAS	**0.878**	1.177
NAVAL	0.0006	**3.461E-14**
MERCEDES	**48.759**	50.617
HOUSE-PRICES	1.356×10^9	$\mathbf{8.568 \times 10^8}$
URBAN TRAFFIC	**3.781**	21.824
DIAMONDS	8.063×10^5	$\mathbf{4.654 \times 10^5}$
TIPS	**0.803**	1.436
BOSTON	**6.538**	22.914
FETCH CALIFORNIA	**0.318**	0.362
ENERGY EFFICIENCY	**0.415**	1.088
MTCARS	**6.944**	8.083

Таблица 1: Результаты применения алгоритмов S-VAW^2 и AG-Tabular к различным задачам регрессии. Жирным шрифтом выделеные наименьшие MSE для каждого набора данных.

$$J_n(x_n) = \int_0^1 \left(\frac{1}{2}\dot{x}_n^2(t) + \frac{q^2}{2}x_n^2(t) - h_n(t)x_n(t) \right) dt, \tag{9}$$

где $\dot{x}_n(t) = u_n(t)$ и $x_n(0) = a_n$. Здесь $q > 0$ — некоторая константа. Эквивалентным образом, будем считать, что игрок непосредственно выбирает траекторию x_n с периодическим условием на концах рассматриваемого отрезка $x_n(0) = x_n(1)$.

Для решения задачи используется аппроксимация в ортонормированном базисе тригонометрических полиномов. Приближённое решение ищется в следующем виде

$$w \cdot \Psi_\alpha = \sum_{j=0}^{2m} w_j \sqrt{\alpha_j} \psi_j, \quad w \in \mathcal{W}_{2m} = \prod_{j=0}^{2m} \left[-\sqrt{\alpha_j}, \sqrt{\alpha_j} \right],$$

где

$$\alpha_j = \frac{C_0 b}{\sqrt{\sum_{k=0}^{2} (4\pi^2)^k \lceil j/2 \rceil^{2k}}}$$

с некоторой константой C_0.

Данная задача рассматривается в терминах онлайн игры между игроком и противником. На каждом шаге n игры игрок выбирает вектор весов w_n, а противник — функцию h_n. Цель игрока — минимизировать динамическое сожаление $R_N(w_1^N, y_1^N) = \sum_{n=1}^N (J_n(w_n \cdot \Psi_\alpha) - J_n(y_n))$ по отношению к оптимальной последовательности сравнения

$$y_n = \arg\min \left\{ \int_0^1 \left(\frac{1}{2}\dot{x}^2(t) + \frac{q^2}{2}x^2(t) - g_n(t)x(t) \right) dt : x \in H_p^1 \right\} \tag{10}$$

для некоторой функции $g_n \in L^2$ с условием $\|g_n\|_{L^2} \leq b$. Здесь H_p^1 обозначает пространство Соболева 1-периодических функций.

Введём функцию потерь

$$f_n(w) = J_n(w \cdot \Psi_\alpha). \tag{11}$$

Тогда конечномерное сожаление принимает следующий вид:

$$\overline{R}_N(w_1^N, u_1^N) = \sum_{n=1}^{N}(f_n(w_n) - f_n(u_n)), \quad u_n \in \mathcal{W}_{2m}.$$

Ключевой результат третьей главы содержится в лемме 3.1, изложенной в параграфе 3.2, которая демонстрирует, что для оценки истинного сожаления $R_N(w_1^N, y_1^N)$ достаточно оценить «конечномерное» сожаление $\overline{R}_N(w_1^N, u_1^N)$. Для обеспечения необходимой точности аппроксимации требуется выбор достаточного числа m «признаков» ψ_j, определяющих размерность конечномерного пространства.

Лемма 3.1. *Пусть $w_n \in \mathcal{W}_{2m}$ — последовательность, полученная с помощью некоторого алгоритма онлайн оптимизации для функций, заданных выражением (11). Пусть y_n удовлетворяет условию (10), и пусть*

$$\hat{v}_n = (\alpha_0^{-1/2}\hat{y}_{n,0}, \ldots, \alpha_{2m}^{-1/2}\hat{y}_{n,2m}).$$

Тогда

$$R_N(w_1^N, y_1^N) \leq \overline{R}_N(w_1^N, \hat{v}_1^N) + C\frac{N}{m}b^2. \tag{12}$$

Главный аспект, обуславливающий эффективность сведения к конечномерному случаю, заключен в лемме 3.2. Данная лемма устанавливает, что такие величины, как диаметр множества допустимых решений D_{2m}, параметр сильной выпуклости μ_{2m}, константа Липшица функции потерь $G_{n,2m}$ и константа гладкости $L_{n,2m}$ равномерно ограничены по размерности конечномерного приближения m и по номеру шага итерации n, что гарантирует стабильность и применимость стандартных алгоритмов онлайн оптимизации вне зависимости от выбранной размерности аппроксимации и длительности процесса.

Лемма 3.2. *Диаметр $\mathcal{W}_{2m}$, константа Липшица, константа гладкости и параметр сильной выпуклости функции f_n удовлетворяют следующим оценкам, равномерным по m и n:*

$$D_{2m} \leq D := Cb^{1/2}, \quad \mu_{2m} \geq \mu := Cb, \tag{13}$$

$$G_{n,2m} \leq G =: Cb^{3/2}, \quad L_{n,2m} \leq L =: Cb \tag{14}$$

В параграфе 3.3 выводятся оценки статического и динамического сожалений для ряда онлайн алгоритмов, включая онлайн градиентный спуск[25], его оптимистичную версию[26], алго-

[25] **Zinkevich M.** Online convex programming and generalized infinitesimal gradient ascent / M. Zinkevich // Proceedings of the 20th International Conference on Machine Learning (ICML 2003). — 2003. — P. 928–935.

[26] **Chen S.** Optimistic online mirror descent for bridging stochastic and adversarial online convex optimization / S. Chen, Y.-J. Zhang, W.-W. Tu, P. Zhao, L. Zhang // Journal of Machine Learning Research. — 2024. — Vol. 25, no. 178.

ритм Ader[27] и алгоритм OMDHedge[26].

С учётом сильной выпуклости функции потерь для алгоритма OGD получена оценка статического сожаления:

$$R_N(w_1^N, y) \le Cb^2 \left(1 + \ln N + \frac{N}{m}\right), \quad \|y\|_{H^2} \le C_0 b.$$

Здесь $\|y\|_{H^s} = \sqrt{\sum_{k=0}^{s} |x|_{H^k}^2}$ при $s = 2$. Для оптимистичного онлайн градиентного спуска также получена оценка статического сожаления для сильно выпуклой функции потерь, которая может быть лучше в случае медленно меняющейся функции $h_n(t)$:

$$R_N^s(w_1^N, y) \le Cb^2 \frac{N}{m} + \widetilde{O}\left(b^2 + \Sigma_{\max}^2\right), \quad \Sigma_{\max}^2 = \max_{1 \le n \le N} \|h_n - h_{n-1}\|_{L^2}^2, \quad h_0 := 0.$$

Для оценки динамического сожаления используются алгоритмы Ader и OMDHedge. Для них установлены соответствующие оценки

$$R_N(w_1^N, y_1^N) = \widetilde{O}\left(b^{3/2}\sqrt{N(b + P_N(g_1^N))} + b^2 \frac{N}{m}\right)$$

и

$$R_N(w_1^N, y_1^N) \le Cb^2 \frac{N}{m} + \widetilde{O}\left(b^2 + bP_N(g_1^N) + \sqrt{b + P_N(g_1^N)}\Sigma_N b\right),$$

где $\|g_n\|_{L^2} \le b$ и $P_N(g_1^N) = \sum_{n=2}^{N} \|g_n - g_{n-1}\|_{L^2}$ —L^2-длина последовательности $(g_n)_{n=1}^N$.

В таблице 2 приведены результаты численного моделирования работы алгоритмов при различных сценариях поведения внешнего воздействия. Комплекс программ для воспроизведения экспериментов находится в открытом доступе на платформе GitHub [28]. В качестве метрики эффективности алгоритмов использовалось усреднённое по числу итераций N сожаление

$$\frac{1}{N}\overline{R}_N(w_1^N, \hat{x}_1^N) = \frac{1}{N}\sum_{n=1}^{N}\left(J_n(w_n \cdot \Psi_\alpha) - J_n(\hat{x}_n \cdot \Psi_\alpha)\right), \tag{15}$$

вычисленное относительно оптимальной последовательности сравнения $\hat{x}_1^N = (\hat{x}_1, \ldots, \hat{x}_N)$.

Анализ результатов (табл. 2) показал, что ни один численный метод не является универсально лучшим. При этом алгоритмы OGD и его оптимистичная версия (OptOGD) исследовались в двух модификациях: с учётом сильной выпуклости (v2) и без неё (v1). Наименьшее сожаление в стабильных и предсказуемых сценариях (постоянное, медленно меняющееся и пе-

[27] **Zhang L.** Adaptive online learning in dynamic environments / L. Zhang, S. Lu, Z.-H. Zhou // Advances in Neural Information Processing Systems 31 (NeurIPS 2018). — 2018. — P. 1323–1333.

[28] **Gurtovaya, O. V.** Online learning in a one-dimensional periodic quadratic variational problem with an adversarial external force [Электронный ресурс] / O. V. Gurtovaya — 2025. — URL: `https://github.com/O-Gurt/PQVP_1d` (дата обращения: 13.08.2025).

реключающееся воздействие) демонстрирует Ader. В условиях стохастического воздействия наиболее эффективен OGD v2. Алгоритм OMDHedge показал превосходство в самом сложном сценарии (медленно меняющееся воздействие с шумом), подтвердив устойчивость к стохастическим помехам. Наибольшее сожаление, особенно при сильных стохастических компонентах, наблюдалось у OGD v1 и его оптимистичных вариантов (OptOGD v1/v2).

	$\psi = 1.5$	$\psi \sim$ N(0, 0.03)	$\psi \sim$ U(2, 3)	$\psi = \{0.5, 1.5\}$	$\psi \sim U(2,3)$ $+N(0,0.03)$
OGD v1	$5.958 \cdot 10^2$	$7.011 \cdot 10^2$	$5.954 \cdot 10^2$	$5.954 \cdot 10^2$	$5.983 \cdot 10^2$
OGD v2	$2.674 \cdot 10^1$	$1.656 \cdot 10^2$	$8.178 \cdot 10^0$	$1.636 \cdot 10^1$	$8.436 \cdot 10^0$
Ader	$2.110 \cdot 10^{-1}$	$1.696 \cdot 10^2$	$1.980 \cdot 10^{-1}$	$2.400 \cdot 10^{-1}$	$1.228 \cdot 10^0$
OptOGD v1	$1.014 \cdot 10^2$	$1.760 \cdot 10^3$	$5.022 \cdot 10^2$	$5.250 \cdot 10^2$	$2.063 \cdot 10^2$
OptOGD v2	$3.367 \cdot 10^0$	$1.689 \cdot 10^2$	$1.151 \cdot 10^1$	$1.970 \cdot 10^1$	$1.178 \cdot 10^1$
OMDHedge	$6.850 \cdot 10^{-1}$	$1.813 \cdot 10^2$	$5.400 \cdot 10^{-1}$	$1.100 \cdot 10^0$	$9.800 \cdot 10^{-1}$

Таблица 2: Сравнение эффективности алгоритмов онлайн оптимизации при разном внешнем воздействии.

В параграфе 3.4 подводятся итоги и формулируются выводы по результатам, представленным в третьей главе.

Заключение содержит основные результаты диссертационного исследования:

1. Разработана методика применения численного метода Вовка–Азури–Вармута (VAW) к случайным признакам Фурье для моделирования данных с марковской зависимостью с теоретическим обоснованием и экспериментальной проверкой на моделях AR(1) и VAR(1).

2. Предложен новый численный метод VAW^2 для моделирования различных регрессионных зависимостей. Получена оценка сожаления порядка $O(T^{1/2} \ln T)$. Алгоритм VAW^2 обладает значительно более низкой вычислительной сложностью по сравнению с алгоритмом VAW, применённым к конкатенированным векторам случайных признаков. В ходе численных экспериментов установлено, что алгоритм VAW^2 превосходит известные из литературы родственные алгоритмы на ряде эталонных наборов данных.

3. Разработан трёхуровневый численный метод S-VAW^2, объединяющий иерархическое агрегирование и стратегии масштабирования данных. Сравнительный анализ показал, что его точность сопоставима с ведущим AutoML-фреймворком, при этом он является вычислительно более эффективной альтернативой.

4. Сформулирована и решена задача моделирования внешнего воздействия динамической системы с квадратичным функционалом качества, включая её сведение к конечномерной задаче и вывод границ статического и динамического сожалений.

5. Реализован комплекс программ на Python для численного моделирования и сравнительного анализа предложенных онлайн алгоритмов: алгоритм VAW с использованием случайных признаков Фурье; двухуровневый алгоритм VAW^2; трёхуровневый алгоритм S-VAW^2. Проведённые эксперименты подтвердили теоретические результаты и продемонстрировали практическую эффективность методов.

В **приложении А** приводится описание программного комплекса, позволяющего проводить численное моделирование и сравнительный анализ предложенных онлайн алгоритмов.

В **приложении Б** находится свидетельство о государственной регистрации программы для ЭВМ.

Публикации автора по теме диссертации

[1] Гуртовая, О. В. Свидетельство о государственной регистрации программы для ЭВМ. Программная реализация трёхуровневого алгоритма Вовка-Азури-Вармута / О. В. Гуртовая. — № 2025680750; заявл. 21.07.2025 ; опубл. 08.08.2025 (Рос. Федерация).

Статьи в научных изданиях, входящих в Scopus

[2] Rokhlin, D. B. Online learning in a one-dimensional periodic quadratic variational problem with an adversarial external force / D. B. Rokhlin, O. V. Gurtovaya // Journal of Mathematical Sciences. — 2024. — P. 1–12. — Publisher: Springer.

Статьи в научных изданиях, входящих в Scopus, Web of Science

[3] Rokhlin, D. B. Vovk-Azoury-Warmuth algorithm and random Fourier features for a regression problem with Markovian data / D. B. Rokhlin, O. V. Gurtovaya // Lobachevskii Journal of Mathematics. — 2024. — Vol. 45, no. 12. — P. 6186–6200. — Publisher: Springer.

Публикации в сборниках трудов конференций

[4] Рохлин, Д. Б. Алгоритм Вовка-Азури-Вармута для аппроксимации условного математического ожидания марковского процесса по одной траектории / Д. Б. Рохлин, О. В. Гуртовая // В сборнике: Всероссийская научно-практическая конференция «Математика, Информатика, Компьютерные науки, Моделирование, Образование». Сборник научных трудов Всероссийской научно-практической конференции МИКМО-2024. - Симферополь, 2024. -С. 37-43. - Режим доступа: https://micme.ru/files/theses/МИКМО-сборка-2024.pdf (дата обращения: 13.08.2025).

[5] Гуртовая, О. В. Алгоритм Вовка-Азури-Вармута для аппроксимации условного математического ожидания марковского процесса по одной траектории / О. В. Гуртовая, Д. Б. Рохлин // В сборнике: XIX Владикавказская молодёжная математическая школа. - Владикавказ, 2024. -С. 48-50. - Режим доступа: `https://smath.ru/upload/iblock/001/Sborka_Tez_VMMSH_2024.pdf` (дата обращения: 13.08.2025).

[6] Гуртовая, О. В. Об аппроксимации решения периодической одномерной квадратичной задачи вариационного исчисления в режиме онлайн с неизвестным внешним воздействием. / О. В. Гуртовая // St. Petersburg Youth Meeting on Probability and Mathematical Physics. - Санкт-Петербург, 2024. -С. 6-7. - Режим доступа: `https://indico.eimi.ru/event/1672/attachments/427/892/abstracts_2024.pdf` (дата обращения: 13.08.2025).

[7] Гуртовая, О. В. Мультиядерная онлайн оптимизация: двойной алгоритм Вовка-Азури-Вармута, основанный на использовании случайных признаков Фурье. / О. В. Гуртовая, Д. Б. Рохлин // Международная научная конференция «Порядковый анализ и смежные вопросы математического моделирования, XVIII: Теория операторов и дифференциальные уравнения». - РСО-А, Дзинага, 2025. -С. 80-82. - Режим доступа: https://smath.ru/upload/iblock/003/_Sborka_MF_2025.pdf (дата обращения 13.08.2025).