

УДК 004.8

DOI: 10.63430/TIA2025020018

## Повышение устойчивости физически-информированных нейронных сетей в задаче конвекции\*

Ч.А. Цгоев<sup>1</sup>, М.А. Братенков<sup>1</sup>, Д.И. Сахаров<sup>1</sup>, В.А. Травников<sup>1</sup>,  
А.В. Серёдкин<sup>1,2</sup>, В.А. Калинин<sup>1</sup>, Д.В. Фомичев<sup>3,4</sup>, Р.И. Мулляджанов<sup>1,2</sup>

<sup>1</sup>Новосибирский государственный университет

<sup>2</sup>Институт теплофизики им. С.А. Кутателадзе СО РАН, Новосибирск

<sup>3</sup>Университет «Сириус», Краснодарский край,  
федеральная территория «Сириус»

<sup>4</sup>Госкорпорация «Росатом», Москва

E-mails: c.tsгоеv@g.nsu.ru

Физически-информированные нейронные сети (PINN) представляют собой инновационный метод решения разнообразных задач в областях математики, физики и инженерии. Указанный метод объединяет в себе концепции нейронных сетей и физических уравнений с целью моделирования и анализа различных физических процессов. В частности, PINN может применяться для решения дифференциальных уравнений, включая одномерное уравнение конвекции. Исследование показало, что стандартная реализация PINN эффективно решает одномерное уравнение конвекции при относительно небольших значениях скорости конвекции, однако теряет устойчивость при увеличении данного параметра. В настоящей работе представлен обзор существующих подходов к решению одномерного уравнения конвекции с использованием PINN, а также демонстрируется, как различные методы могут улучшить качество моделей. Результаты сравнения методов свидетельствуют о превосходстве подхода, основанного на динамическом изменении точек коллокации в зависимости от невязки на текущем шаге обучения, по сравнению с другими подходами.

**Ключевые слова:** физически-информированное машинное обучение, нейронные сети, задача конвекции.

### Введение

Активное развитие методов глубокого обучения приводит к созданию новых инструментов в науке и инженерии, которые могут упростить математическое моделирование сложных физических систем. Нейронные сети рассматриваются как вспомогательный, а иногда и альтернативный подход к решению уравнений в частных производных.

---

\* Результаты получены при финансовой поддержке исследования/проекта/работы, реализуемого в рамках государственной программы федеральной территории «Сириус» «Научно-технологическое развитие федеральной территории «Сириус» (соглашение № 18-03 от 10.09.2024).

Первоначальные попытки применения нейросетевых методов основывались на классическом обучении с учителем, где обучающие данные создавались с помощью численного решателя, а нейросеть обучалась на парах «пример – ответ». Такие методы в контексте решения дифференциальных уравнений часто требуют значительных усилий из-за необходимости генерации разнообразной и качественной обучающей выборки. Кроме того, эти подходы страдают от недостатка точности и обобщающей способности. Для получения качественного описания моделируемого процесса необходимо не только создать подходящую обучающую выборку, но и внедрить физическое знание в процесс обучения нейросети.

Одной из концепций, позволяющих интегрировать физическое знание в архитектуру нейронной сети, является применение физически-информированных нейронных сетей (PINN) [1–3]. Этот метод включает законы физики (уравнения в частных производных, граничные условия, начальные условия и т.д.) непосредственно в процесс обучения через функцию потерь модели. Указанный подход использует не только синтетические или экспериментальные данные, но и физические законы, управляющие системой, что позволяет достичь более точного и надежного приближения.

Использование PINN, помимо практических результатов [4–7], показывающих состоятельность подхода, имеет также (в случае некоторых классов задач) и теоретическое обоснование, гарантирующее сходимость нейронной сети к искомому решению уравнений в частных производных при увеличении числа нейронов [8]. Одно из ключевых преимуществ PINN заключается в возможности решения уравнения без построения «классической» регулярной/нерегулярной расчетной сетки даже при решении задач с довольно сложной геометрией. Это достигается благодаря использованию алгоритма автоматического дифференцирования для подсчета входящих в дифференциальное уравнение производных и, следовательно, расчета итоговой невязки, а также всех компонент для градиентного спуска. Изначально PINN разрабатывался для работы с классическими уравнениями в частных производных, однако он легко адаптируется для решения интегродифференциальных уравнений, дробных дифференциальных уравнений [9] и стохастических дифференциальных уравнений [10].

Несмотря на свои преимущества, физически-информированные нейронные сети сталкиваются с рядом проблем, из которых наиболее значимые касаются их обучения и применения к определенным видам задач. Одна из основных трудностей заключается в отсутствии четких рекомендаций по оптимизации архитектуры, параметров обучения сети, а также методов обучения. Это делает процесс настройки PINN сложным и не интуитивным. Также не существует универсального подхода к улучшению точности аппроксимации решений, аналогичного уточнению сетки в классических численных методах или изменению сложности модели в традиционных нейросетевых подходах.

Кроме того, физически-информированные нейросети могут демонстрировать плохую обучаемость и сходимость при решении жестких или плохо обусловленных задач [11]. Эти проблемы усложняют применение метода и требуют дополнительных исследований для разработки более эффективных подходов к обучению и адаптации моделей PINN к различным типам задач. В частности, возникают следующие проблемы: проблема тривиального или нулевого решения, которая заключается в том, что физически-информированная нейросеть может сходиться к некорректному нулевому решению или плавно в него переходить в некоторой области. Это происходит, когда сеть не может адекватно моделировать сложное поведение системы, переносить информацию от границ во внутреннюю часть области и выбирает простейший вариант, равный нулю, либо

плавно затухает к нему [12]. Дополнительно можно выделить еще один момент — трудность корректного выбора обучающего множества для сложных задач [13]. Несмотря на относительную простоту генерации точек коллокации, определение оптимального размера выборки представляет собой сложную подзадачу, для которой пока что не существует каких-то конкретных алгоритмов. Стоит также отметить, что невозможно заранее гарантировать точность итоговой аппроксимации и стабильность обучения при фиксированной выбранной архитектуре [11]. Такие особенности показывают, что обучение и настройка PINN требуют дополнительных, в том числе численных, исследований и разработки методик для повышения их универсальности и эффективности в широком спектре приложений.

В настоящей статье рассматривается применение PINN для решения задачи конвекции — одной из самых проблемных модельных задач в рамках физически-информированного подхода. Ранее было показано [12], что при увеличении жесткости посредством варьирования скорости конвекции  $\beta$  обучение PINN значительно страдало, приводя к уже упомянутому эффекту затухания к нулю. Предложенные на сегодняшний день решения данной проблемы не всегда оказываются эффективными и требуют достаточно сложной настройки. Как правило, они сильно зависят от заданной геометрии задачи, могут быть трудозатратными и увеличивать итоговое время обучение PINN в несколько раз.

В рамках представленной работы исследуются подходы, направленные на повышение устойчивости физически-информированных нейронных сетей при решении задачи конвекции. Рассматриваются и анализируются методы, ранее не применявшиеся для этой проблемы: лагранжева регуляризация [14, 15], позволяющая задавать баланс внутри области и на границе, адаптивное распределение точек коллокации [16], которое динамически изменяет обучающее множество в зависимости от локализации областей с большой невязкой, а также поэтапное обучение с постепенным увеличением скорости конвекции. Проводится сравнительный анализ эффективности указанных подходов с предложением рекомендаций по их применению и определяются ограничения для каждого из них.

### Основы физически-информированных нейронных сетей

Семейство PINN образует класс нейронных сетей, интегрирующих физические законы в свою структуру и в основном направленных на решение дифференциальных уравнений и их систем. Основная цель PINN заключается в улучшении предсказательной мощности и точности нейросетевых моделей в научных и инженерных задачах за счет внедрения знаний о физических процессах в цикл обучения. Включение физических законов происходит через добавление специальных членов в функцию потерь, которые обеспечивают соответствие решения физическим уравнениям или заданным экспериментальным наблюдениям. Этот подход позволяет нейронной сети не просто аппроксимировать данные, но и соблюдать фундаментальные законы природы, что особенно ценно в задачах, где важна физическая точность модели.

PINN, как и любая другая нейронная сеть, функционирует как аппроксиматор решений:

$$NN(x, t, \theta) = \sigma_L(W_L \sigma_{L-1}(\dots \sigma_1(W_1[x, t]^T + b_1) \dots + b_{L-1}) + b_L),$$

состоящий из обучаемых параметров: весов  $W$  и смещений  $b$ , обозначаемых совокупно через  $\theta$ , а также из функций активации  $\sigma$ . Входные данные, как правило, представляют собой пространственные и/или временные координаты  $(x, t)$ . На сегодняшний день

не существует каких-либо ограничений на архитектурную часть PINN. Чаще здесь используются вполне стандартные модели, такие как полносвязные сети. Основная отличительная особенность физически-информированных нейронных сетей, как уже отмечалось, заключается в использовании специфических функций потерь, которые включают термины, отражающие уровень нарушения установленных физических законов. Чаще всего функция потерь для PINN представляет собой сумму невязок дифференциального уравнения, а также ошибок в начальных и граничных условиях:

$$L(\theta) = L_E(\theta) + \sum_i \alpha_i L_{B_i}(\theta) + \sum_j \beta_j L_{IC_j}(\theta) + \gamma L_{ad}(\theta),$$

где  $L(\theta)$  — общая функция потерь, по которой происходит шаг оптимизации весов нейронной сети,  $L_E(\theta)$  — функция потерь, связанная с невязкой дифференциального уравнения или системы,  $L_B(\theta)$  — функция потерь, связанная с граничными условиями,  $L_{IC}(\theta)$  — функция потерь, связанная с начальными/входными условиями,  $L_{ad}(\theta)$  — дополнительная функция потерь, которая может учитывать отклонения от имеющихся экспериментальных данных или других априорных условий. При этом  $\alpha$ ,  $\beta$  и  $\gamma$  представляют собой веса, которые позволяют регулировать вклад соответствующих компонентов функции потерь, а через  $\theta$  обозначены обучаемые веса нейронной сети.

Особенность функции потерь  $L(\theta)$  заключается в необходимости вычисления производных различных порядков аппроксимируемой функции по входным параметрам в процессе обучения. В методе PINN критически важно обеспечить непрерывную дифференцируемость всех слоев нейронной сети, что делает возможным эффективное применение автоматического дифференцирования [1]. Автоматическое дифференцирование позволяет с высокой точностью вычислять производные целевых функций (выходов нейронной сети) в заданном наборе точек, что является фундаментальным элементом метода PINN. Это, с одной стороны, позволяет отказаться от традиционных методов построения расчетных сеток. С другой стороны, благодаря обобщающей способности нейронных сетей обученную модель можно использовать на любом дискретном наборе точек, что приводит к возможности разработки гибридного подхода, объединяющего PINN с классическими численными методами.

По сути, PINN решает оптимизационную задачу минимизации  $L(\theta)$ , зависящей от весов нейронной сети. На вход сети подаются точки коллокации, на основе которых модель генерирует прогнозы значений целевых функций. С использованием техники автоматического дифференцирования вычисляются необходимые частные производные. Эти производные затем используются для вычисления невязок, а также ошибок в начальных, входных и граничных условиях, которые затем интегрируются в функцию потерь. Методом обратного распространения ошибки [17] определяются градиенты весов сети, и на последнем шаге цикла осуществляется шаг оптимизации, обновляющий веса модели. Критерием остановки может служить, например, заранее заданное количество шагов.

Как уже указывалось, одним из ключевых преимуществ данного подхода является отсутствие необходимости использования традиционных методов построения расчетных сеток для аппроксимации решений. Это существенно упрощает процесс численного решения дифференциальных уравнений. Дополнительно следует отметить, что PINN обладает уникальной способностью принимать в качестве входных данных не только временные и пространственные координаты, но и различные параметры задачи, такие как коэффициенты дифференциальных уравнений, включая теплопроводность,

число Рейнольдса и другие. Такие возможности позволяют PINN моделировать и решать целый класс задач с различными параметрами, не требуя пропорционального увеличения объема обучения нейронной сети. Благодаря этому нейросеть обучается на обобщенной функции, которая способна учитывать изменения параметров задачи. В результате одна и та же обученная модель может эффективно применяться для широкого диапазона значений параметров, обеспечивая решение различных задач в рамках одной архитектуры. Это значительно повышает гибкость и эффективность метода, поскольку исключает необходимость повторного обучения сети при изменении параметров задачи и позволяет решать множество связанных задач одновременно, что существенно снижает вычислительные затраты и время обучения.

## Постановка задачи и основные инструменты

### Задача конвекции

Рассматривается одномерная нестационарная тестовая задача конвекции в безразмерном виде, представляющая трудности в обучении для физически-информированных нейронных сетей [11]. Математическая постановка выглядит следующим образом:

$$\begin{aligned} u_t + \beta u_x &= 0, \quad t \in [0, t_{\max}], \quad x \in [0, 2\pi], \\ u(t, 0) &= u(t, 2\pi), \quad t \in [0, t_{\max}], \\ u(0, x) &= \sin(x), \quad x \in [0, 2\pi], \end{aligned}$$

где  $u$  — искомая функция,  $u_t$  и  $u_x$  — частные производные  $u$  по  $t$  и  $x$ ,  $\beta$  — параметр задачи, определяющий скорость конвекции в пространстве,  $t_{\max}$  определяет временной интервал, в пределах которого ищется решение задачи. В рамках данной постановки классическая физически-информированная нейронная сеть испытывает проблемы, связанные с доминированием нулевого решения по мере увеличения параметра  $\beta$  (см. рис. 1). При этом аналитическое решение поставленной задачи представляется следующим образом:

$$u = \sin(x - \beta t).$$

### Лагранжева регуляризация

Лагранжева регуляризации основана на методе множителей Лагранжа. Идея подхода состоит в модификации функции потерь за счет добавления обучаемых весовых коэффициентов для каждой точки и последующего их изменения путем решения задачи двойной оптимизации:

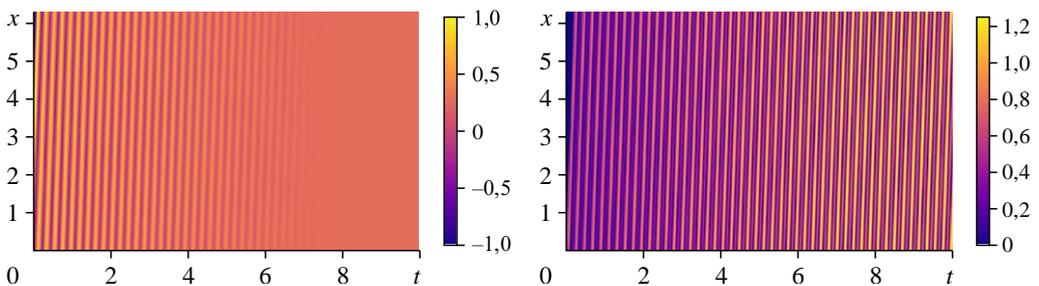


Рис. 1. Пространственные распределения предсказаний  $NN(t, x, \theta)$  обученного PINN (слева) и итоговой абсолютной ошибки  $|NN(t, x, \theta) - u_{\text{sol}}(x, t)|$  (справа) для задачи конвекции с  $\beta = 30$ .

$$\begin{aligned} & \min_{\theta} \max_{\lambda_F, \lambda_B, \lambda_C} L(\theta; \lambda_D, \lambda_B, \lambda_C) = \\ & = \min_{\theta} \max_{\lambda_F, \lambda_B, \lambda_C} (L_D(\theta) + L_F(\theta; \lambda_F) + L_B(\theta; \lambda_B) + L_C(\theta; \lambda_C)), \\ & L_F(\theta, \lambda_F) = \sum_{i=1}^{N_{\Omega_1}} \lambda_F^{(i)} \varphi(F(x^{(i)}; \theta)), \\ & L_B(\theta, \lambda_B) = \sum_{i=1}^{N_{\Omega_2}} \lambda_B^{(i)} \varphi(B(x^{(i)}; \theta)), \\ & L_C(\theta, \lambda_C) = \sum_{i=1}^{N_{\Omega_3}} \lambda_C^{(i)} \varphi(C(x^{(i)}; \theta)), \end{aligned}$$

где  $L_D$  — это функции потерь «основного» дифференциального уравнения,  $L_F, L_B, L_C$  — функции потерь начальных, граничных и иных дополнительных условий/уравнений,  $\lambda_i$  — множители Лагранжа,  $\Omega_i$  — область, где выполняется соответствующее условие,  $\varphi$  — функция с неотрицательной областью значений,  $F, B, C$  — начальные, граничные и дополнительные условия.

На первом шаге к ошибкам в каждой точке области для начальных и граничных условий применяется максимизация методом градиентного подъема по множителям Лагранжа. На данном шаге ошибка в каждой точке коллокации умножается на дополнительный весовой коэффициент. Таким образом, для точек с большой ошибкой будет назначен более высокий весовой коэффициент, что позволит модели на следующем шаге уделять больше внимания именно этим точкам. Это предотвращает ситуацию, когда несколько точек с высокими ошибками могут быть почти игнорированы на фоне средней ошибки по всему набору данных. На втором шаге модель минимизирует сумму всех невязок. Такой подход помогает сгладить поверхность ошибок, убирая тем самым множество локальных минимумов. В качестве функции  $\varphi$  в эксперименте ниже использовалась квадратичная функция.

### *Адаптивное распределение точек коллокации*

В настоящей работе представлена адаптивная стратегия обновления точек коллокации, опирающаяся на анализ распределения поля невязок на каждой итерации вычислительного процесса [16]. Основной целью является улучшение точности модели за счет оптимального выбора точек для обучения, которое базируется на распределении ошибок предсказания.

На определенной итерации вычисляется аппроксимация поля невязок  $\varepsilon(x) = |L(x; \hat{u}(x))|$ , демонстрирующая области с наибольшими ошибками предсказания нейросети. С помощью этого поля создается вероятностное распределение, которое служит основой для выбора обучающих точек:

$$p \propto \frac{\varepsilon^k(x)}{E[\varepsilon^k(x)]} + c.$$

Параметр  $k$  управляет концентрацией точек: при больших значениях  $k$  точки сосредотачиваются в областях с высокой невязкой, при меньших значениях они распределяются более равномерно по всей области. В то время как параметр  $c$  служит для смещения и масштабирования вероятностного распределения.

Описанная адаптивная стратегия предполагает постепенное добавление точек в обучающую выборку на основе вероятностного распределения. При этом точки с наимень-

шими значениями невязок исключаются, что дает возможность поддерживать постоянный размер обучающей выборки, изменяя только ее распределение в области. Важным аспектом предложенного метода является динамическое перераспределение обучающей выборки, что предположительно обеспечивает постоянное улучшение качества предсказаний нейросети. Такой подход позволяет модели сосредотачиваться на наиболее сложных для предсказания областях, что потенциально может привести к улучшению ее обобщающей способности.

### ***Постепенный переход к жесткой постановке***

Как будет показано ниже, метод физически-информированных нейронных сетей демонстрирует высокую эффективность при решении задачи конвекции при малых значениях скорости конвекции. В таком случае модель не сталкивается с проблемой нулевого решения. Однако по мере увеличения значения  $\beta$  указанная проблема начинает проявляться все более явно. Один из способов ее преодоления заключается в постепенном увеличении значения этого параметра. В данной работе рассматриваются два подхода к решению этой проблемы.

В первом подходе используется одна нейросетевая модель, которая аппроксимирует целевую функцию. Процесс обучения разбивается на несколько этапов. На каждом этапе модель обучается до достижения установившегося состояния с фиксированным значением  $\beta$ . После этого, на следующем этапе, значение параметра увеличивается, а обучение продолжается, начиная с уже обученной на предыдущем этапе модели. Таким образом, каждый этап (за исключением первого) представляет собой переобучение модели относительно предыдущего этапа для нового значения параметра. Второй подход предполагает непрерывное равномерное увеличение значения скорости конвекции в процессе обучения без разделения на этапы. Он позволяет плавно адаптировать модель к изменениям параметра, что может способствовать более стабильному и эффективному обучению.

## **Экспериментальная часть и анализ результатов**

### ***Конфигурации экспериментов***

В проведенных экспериментах в качестве модели использовалась полносвязная нейронная сеть, состоявшая из четырех слоев, каждый из которых содержал по 128 нейронов. Размер нейронной сети определялся на основе методических расчетов, в рамках которых дополнительно подавалась информация о точном решении в PINN. В ходе расчетов было установлено, что мощности выбранной архитектуры достаточно, чтобы аппроксимировать имеющиеся данные. При этом размерность входного слоя составляла 2, а выходного — 1. Количество эпох варьировалось от 100000 до 300000, оно выбиралось из соображений сходимости. Внутри области использовалось от 25000 до 200000 точек коллокации, а на каждой границе — от 2000 до 16000 точек коллокации, при соотношении 25:2. Все точки генерировались согласно равномерному распределению. В качестве функции активации на всех слоях рассматривался  $\sin(x)$ . Для оптимизации применялся экспоненциальный алгоритм уменьшения шага обучения [18] с уменьшением шага на 0,003 % на каждой эпохе. В качестве входных данных были выбраны точки коллокации в пространственно-временном пространстве  $(x, t)$ , а выходные данные модели представляли собой предсказанные значения целевой функции  $u$  в этих точках  $(x, t)$ .

В проведенных экспериментах использовался программный стек, включающий PyTorch и собственноручно разработанную библиотеку, реализующую технологию

физически-информированных нейронных сетей. Время выполнения типового расчета варьировалось в зависимости от количества точек коллокации и числа эпох обучения. Например, при использовании 25000 внутренних точек и 2000 точек на границах время обучения составляло примерно 0,01 секунды за эпоху. В более ресурсоемком эксперименте с 200000 внутренних точек и 16000 точек на границах время обучения увеличивалось до 0,09 секунды за эпоху. Все расчеты выполнялись на системе с видеокартой NVIDIA 4060 Ti 16 ГБ, процессором Intel Core i9-13900K и 32 ГБ оперативной памяти.

### Проблема тривиального решения

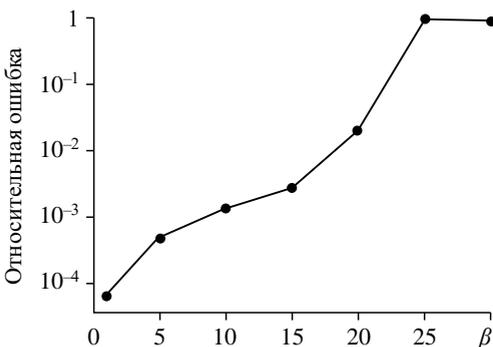
Проиллюстрируем справедливость утверждения, что при относительно большом значении параметра  $\beta$  классический PINN без должной настройки не справляется с аппроксимацией искомого решения. В частности, в рамках эксперимента рассмотрим следующую конфигурацию: пусть  $\beta = 30$ ;  $t_{\max} = 10$ , обучение длится около 100000 итераций; генерируется около 25000 точек внутри области. Как и в работе [16], расчеты продемонстрировали (рис. 1), что классический PINN плохо справляется с рассматриваемой задачей. Так, модель не смогла распространить начальные данные на всю область и подверглась проблеме тривиального решения. По распределениям ошибки (рис. 1b) заметен резкий перепад значений по оси  $t$ . Это свидетельствует о том, что модель начала сходиться к локальному минимуму, прекратив распространение начальных данных. Выявленный эффект, в частности, объясняется тем, что нулевое решение удовлетворяет уравнению во внутренней области. Однако в стандартной реализации обучения PINN и его архитектуре не предусмотрены механизмы для выхода из такого локального минимума, при котором граничные условия не соблюдаются, хотя значения во внутренней области аппроксимируются достаточно точно.

Нулевое решение начинает проявляться только при достижении определенного критического значения скорости конвекции. В ходе численного эксперимента параметры варьировались в диапазоне значений  $\{1, 5, 10, 15, 20, 25, 30\}$ . Как показано на рис. 2, нейросеть успешно аппроксимирует решение задачи при значениях  $\beta < 25$ . Однако с уве-

личением  $\beta$  наблюдается рост относительной ошибки  $\varepsilon = \frac{\|NN(t, x, \theta) - u_{\text{sol}}(x, t)\|_{L_2}}{\|u_{\text{sol}}(x, t)\|_{L_2}}$ ,

где  $L_2$  — евклидова норма,  $NN(t, x, \theta)$  представляет собой предсказания обученной нейронной сети, а  $u_{\text{sol}}(x, t)$  — имеющееся аналитическое решение. В частности, в данной серии расчетов получено, что при превышении порога ошибки в  $10^{-1}$  имеет место проблема тривиального решения.

На рис. 3 представлены пространственные распределения предсказаний нейросети



после 100 тыс. итераций обучения для задач с  $\beta \in \{1, 10, 25, 30\}$ . Классический PINN может качественно описать решение вплоть до значения  $\beta = 25$ , где возникает проблема тривиального решения, и относительная ошибка критически возрастает. Проблема нулевого решения в рассматриваемых моделях может проявляться

Рис. 2. Зависимость относительной ошибки  $\varepsilon$  от значений параметра  $\beta$ .

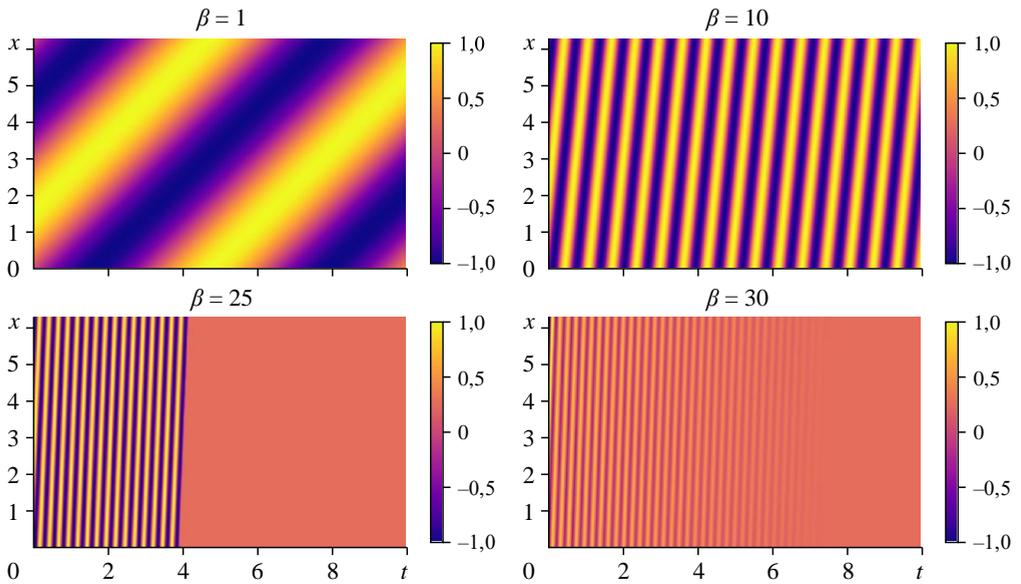


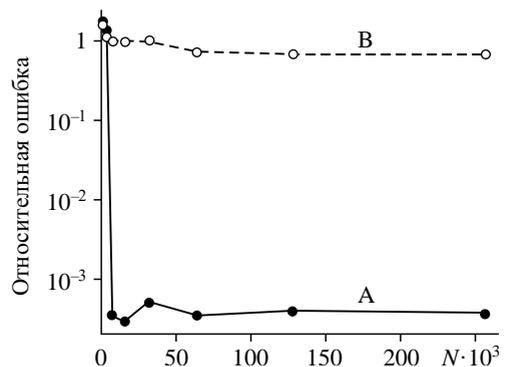
Рис. 3. Пространственные распределения предсказаний обученной на задаче конвекции нейросети  $NN(t, x, \theta)$  при различных значениях параметра  $\beta$ .

в разных формах. В некоторых случаях, как, например, при значении  $\beta = 25$ , данная проблема проявляется относительно резко, так как модель достаточно резко приходит к нулевому решению. В других случаях, например при  $\beta = 30$ , нулевое решение достигается более плавно, что свидетельствует о постепенной утрате модели способности обучаться на данном этапе. Однако независимо от характера проявления этой проблемы ни в одном из рассмотренных сценариев не удастся дообучить модель таким образом, чтобы она могла эффективно распространять решение на всю область задачи. Это указывает на фундаментальные ограничения рассматриваемой методики, препятствующие полноценному распространению решения на заданной области.

#### Точность модели на последовательности вложенных сеток

При решении дифференциальных уравнений с использованием классических численных методов обычно предполагается сходимость метода на последовательности вложенных сеток. Однако, как показано в данном исследовании, подобные эффекты сходимости отсутствуют при использовании PINN. Для начала рассмотрим нежесткую задачу с  $\beta = 5$ , где нейросеть адекватно аппроксимирует искомую функцию. Так, на рис. 4 представлены относительные ошибки предсказаний PINN, полученные на последовательности вложенных сеток. Результаты показывают, что использование статического набора с небольшим числом точек может привести к некачественным решениям.

Рис. 4. Зависимость относительной ошибки от размера обучающей выборки  $N$  в задаче конвекции с  $\beta = 5$  (А) и  $\beta = 30$  (В).



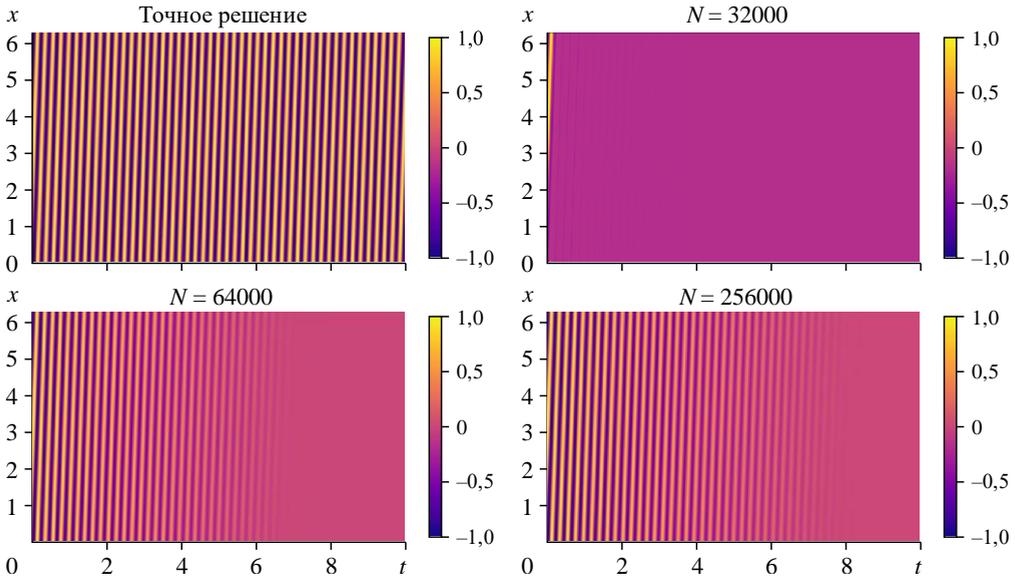


Рис. 5. Пространственные распределения аналитического решения  $u = \sin(x - \beta \cdot t)$  и предсказаний обученной на задаче конвекции с  $\beta = 30$  нейросети  $NN(t, x, \theta)$  для различного размера обучающей выборки  $N$ .

При этом по достижении определенного размера обучающей выборки дальнейшее ее увеличение не приводит к уменьшению ошибки. Таким образом, способы повышения точности и методы анализа сходимости, применяемые в классических численных методах, не подходят для PINN. Следовательно, для повышения точности предсказаний PINN требуется разработка новых подходов, в частности, модифицирующих процесс обучения рассматриваемого класса нейросетей.

Продемонстрируем теперь, что при решении жесткой задачи увеличение числа точек обучения также не позволяет преодолеть проблему тривиального решения. На рис. 4 приведены графики зависимости относительных ошибок предсказания на последовательности вложенных сеток от параметров последних и от эпохи обучения для моделей с фиксированной архитектурой при решении задачи с  $\beta = 30$ . Согласно рис. 4, начиная с 64 тыс. точек постепенное увеличение их количества не приводит к существенному улучшению качества решения, при этом ошибка остается относительно большой.

На рис. 5 представлены пространственные распределения предсказаний нейросети для различных значений  $N$ . Несмотря на возникновение проблемы тривиального решения, использование статической обучающей выборки с числом точек  $\geq 64$  тыс. позволяет модели продолжать решение от начального условия до  $t = 6$  с затуханием. Согласно полученным данным, при обучении на малом числе точек модель не способна продолжить решение даже до  $t = 1$ . Таким образом, хотя увеличение числа точек и позволяет продлить решение, оно не способствует полному преодолению проблемы тривиального решения.

### Анализ влияния лагранжевой регуляризации

В рамках PINN лагранжева регуляризация может считаться одной из наиболее перспективных стратегий адаптивной настройки весовых коэффициентов в функции потерь. Как уже было показано, этот подход позволяет динамически изменять вклад различных

компонентов функции потерь в процессе обучения модели, обеспечивая баланс между сходимостью и точностью решения. Однако несмотря на потенциал этой методики ее применение в определенных задачах может быть ограничено спецификой проблемы.

При исследовании влияния лагранжевой регуляризации для жесткой задачи с  $\beta = 30$  конфигурация эксперимента включала следующие параметры: общее количество эпох составило 300000, обучающая выборка была статической, количество точек коллокации внутри области составляло 50000, а на границах — 4000 точек. Оптимальное число точек было выбрано на основе предыдущих расчетов с использованием последовательности сеток для обеспечения достаточной точности модели.

В экспериментах по варьированию шага оптимизации в рамках лагранжевой регуляризации было обнаружено, что модели отличаются по характеру сходимости к тривиальному решению. Это указывает на сложности в достижении стабильной сходимости. Даже при использовании оптимального шага обучения, определенного на основе вычислительных экспериментов, наиболее успешная модель не продемонстрировала существенного улучшения качества решения по сравнению с «классически» обученной моделью PINN. Более того, ни одна из моделей не смогла эффективно распространить начальные данные на всю область, что свидетельствует о недостаточной адаптивности данного подхода в контексте рассматриваемой задачи.

Таким образом, на основании проведенного анализа можно заключить, что лагранжева регуляризация не является подходящей стратегией для решения того набора проблем, который встречается в задаче конвекции. Эта методика, эффективная в других контекстах, в рассматриваемых условиях сталкивается с серьезными ограничениями, что требует поиска альтернативных методов оптимизации.

### *Анализ влияния адаптивного распределения точек коллокации*

Адаптивные сетки используются в классических численных алгоритмах для повышения точности и эффективности решения задач с переменной степенью сложности, таких как задачи с разрывным решением или сложной геометрией областей. В контексте физически-информированных нейросетей адаптивные коллокационные точки могут применяться с той же целью, что и в классических численных методах: для дообучения нейросети на сложных участках с резкими градиентами или сложной геометрией. Это позволяет улучшить точность модели на указанных критических зонах, однако помимо этого в настоящей работе предлагается использовать адаптивные точки для самого процесса обучения. В этом случае обучающая выборка динамически изменяется в ходе обучения, что потенциально позволяет нейросети постепенно фокусироваться на наиболее проблемных участках, обеспечивая более эффективный перенос граничных условий и улучшая общую обучаемость модели на всей области решения.

В проведенном эксперименте были заданы следующие условия: общее количество эпох составило 300000, точки коллокации пересчитывались каждые 100 эпох, в пределах области обновлялось 1000 точек коллокации, в то время как на каждой границе обновлялось по 50 точек. Параметр  $k$ , управляющий концентрацией точек коллокации в зависимости от значений невязки, был установлен равным 1. Для четырех различных запусков эксперимента количество точек коллокации составляло 25000, 50000, 100000 и 200000 соответственно.

Как видно из рис. 6, даже при использовании относительно небольшого количества точек коллокации модель с адаптивной стратегией демонстрирует способность намного

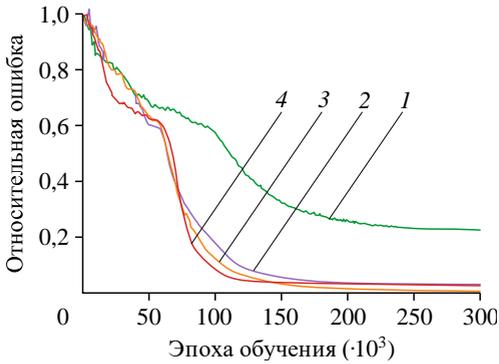


Рис. 6. Зависимость относительной ошибки  $\varepsilon$  от количества эпох при использовании адаптивного подбора точек коллокации. Кривые — модели, обученные на задаче конвекции с  $\beta = 30$  при разном размере адаптивного обучающего множества  $N \in \{25000 (1), 50000 (2), 100000 (3), 200000 (4)\}$ .

и, как показано при обучении на 100000 и 200000 точек, может даже немного ухудшить качество модели. Наилучшее решение в этом случае продемонстрировало относительную ошибку, равную 0,006. Рассмотрим, в частности, поведение невязок и точек коллокации по ходу обучения для модели, обучаемой на 100000 точек.

Как показано на рис. 7, наибольшая невязка наблюдается на границе между областью распространения решения и областью тривиального решения. Вследствие этого предлагаемая адаптивная стратегия на начальном этапе перебрасывает точки коллокации из области нулевого решения в область распространяемого решения, что позволяет модели избегать обучения в областях, куда решение еще не распространилось. По мере смещения границы между распространяемым решением и нулевой областью обучающий набор точек динамически адаптируется, следуя за этой границей. В конце обучения точки коллокации распределяются более равномерно, концентрируясь в локальных областях с повышенной невязкой, что способствует улучшению качества решения в этих зонах.

Таким образом, проведенное исследование показало, что предлагаемая адаптивная стратегия эффективно справляется с изучаемой проблемой, обеспечивая высокое качество решения даже в условиях сложной динамики распространения. Полученный результат демонстрирует новую, ранее не предусматриваемую цель применения адаптивных точек коллокации, которая заключается не только в улучшении точности в областях с резкими градиентами, но и в эффективном распространении решения (обучение нейросети) в регионы, где оно первоначально плохо распространяется. Теоретически этот подход может быть полезен и в других задачах, где наблюдаются аналогичные сложности с распространением решения.

#### *Анализ влияния постепенного перехода к жесткой постановке*

В задачах, связанных с моделированием конвекции посредством физически-информированных нейронных сетей, одной из ключевых стратегий оптимизации является метод поэтапного увеличения скорости конвекции [11]. Эта стратегия была предложена с тем, чтобы преодолеть проблему нулевого решения. За счет постепенного повышения интенсивности конвективных процессов в модели данный подход позволяет достичь лучшей сходимости и способствует повышению точности расчетов, так как модель изначально обучается на простых примерах задачи, а затем переходит к более сложным случаям. В рамках представленной работы было решено проверить эффективность предложенного подхода, а также сравнить его с другими рассматриваемыми методами, оценивая как их результативность, так и простоту внедрения в различных численных моделях.

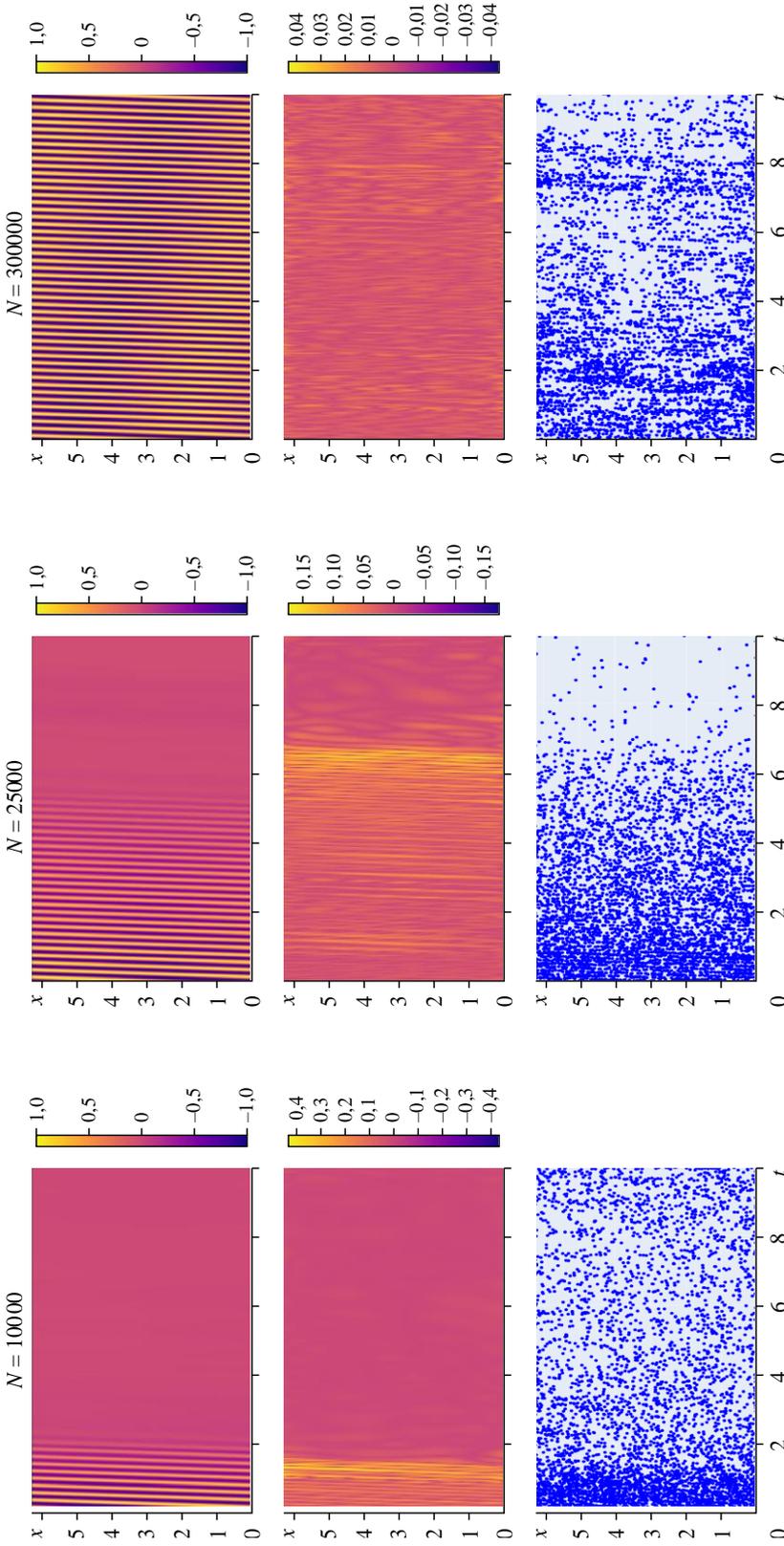


Рис. 7. Пространственные распределения предсказаний нейросети  $NN(t, x, \theta)$  (верхний ряд), распределения итоговых невязок  $L(x; NN(t, x, \theta))$  (средний ряд), а также распределение адаптивного обучающего множества (нижний ряд) для задачи конвекции с  $\beta = 30$  при разных значениях эпохи обучения  $N$ .

В рамках эксперимента рассматривалось 300000 эпох с применением статичного обучающего множества. Для уменьшения шага обучения использовался экспоненциальный алгоритм, при котором шаг обучения уменьшался на 0,003 % в каждой эпохе. Исключение составили этапы, идущие после первого: на них в течение первых 2000 эпох шаг обучения увеличивался с  $10^{-6}$  до  $10^{-3}$ , что позволяло начинать обучение с очень малых шагов, постепенно увеличивая их до оптимального значения. Параметр  $\beta$  менялся различными способами. Для начала рассмотрим «наивное» последовательное увеличение рассматриваемого параметра по следующей схеме: 1, 5, 10, 15, 20, 25, 30.

Согласно рис. 8, 9, подход «резкого» увеличения значения параметра не приводит к качественному улучшению предсказаний модели. Информация, накопленная на предыдущих этапах обучения, не используется новой моделью эффективно для достижения более точного решения. В результате модель продолжает сталкиваться с проблемой тривиального решения, что указывает на недостаточную эффективность данного подхода.

Исследуем более сложную стратегию изменения скорости конвекции, которая позволила бы улучшить качество предсказаний модели и минимизировать влияние проблемы нулевого решения. В рамках этой стратегии параметр  $\beta$  равномерно увеличивается с 1 до 20 в течение первых 100000 эпох обучения. После достижения значения  $\beta = 20$  скорость роста параметра замедляется, поскольку при значениях  $\beta > 20$  модель становится подверженной проблеме нулевого решения. Таким образом, в следующие 100000 эпох параметр  $\beta$  равномерно увеличивается до 30. На заключительном этапе, в оставшиеся 100000 эпох, модель дообучается с фиксированным значением  $\beta = 30$ , что позволяет стабилизировать обучение и добиться более точных результатов.

Согласно рис. 10, 11, второй подход позволил успешно избежать появления нулевого решения. Модель смогла при больших значениях  $\beta$  обучиться до низкого значения ошибки, которая составила примерно 0,0004. Дальнейшие эксперименты показали, что чем более жесткой становится задача при увеличении  $\beta$ , тем медленнее следует увеличивать данный параметр. Например, при проведении экспериментов с равномерным увеличением параметра  $\beta$  на всем фиксированном промежутке избавиться от эффекта нулевого решения не удалось. Этот результат подчеркивает важность контроля скорости изменения  $\beta$  для успешного обучения модели и минимизации ошибок.

Можно предположить, что оптимальная стратегия, вероятно, будет находиться где-то между резким увеличением параметра  $\beta$  и его постепенным изменением в ходе обучения.

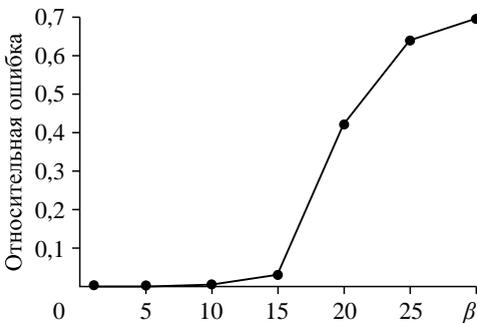


Рис. 8. Зависимость относительной ошибки  $\varepsilon$  от значений коэффициента  $\beta$  при последовательном его изменении в ходе обучения нейросети.

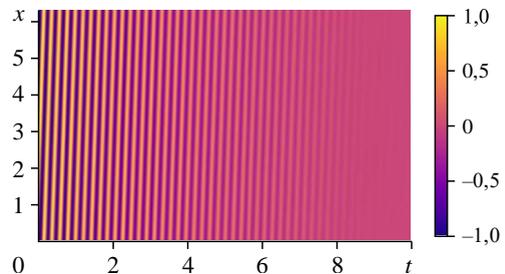


Рис. 9. Пространственное распределение предсказаний нейросети  $NN(t, x, \theta)$ , обученной на задаче конвекции с  $\beta = 30$  при последовательном изменении  $\beta$  в ходе обучения.

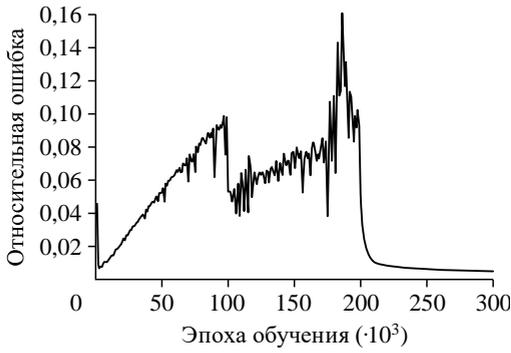


Рис. 10. Зависимость относительной ошибки  $\varepsilon$  от эпохи обучения для задачи конвекции с принятым кусочно-равномерным увеличением по ходу обучения нейросети параметра  $\beta$  вплоть до  $\beta = 30$ .

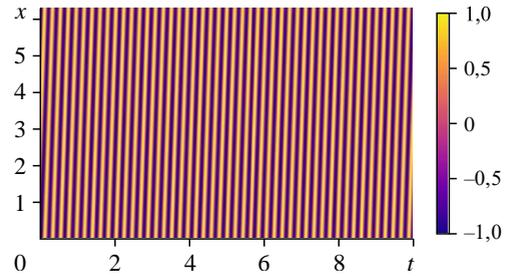


Рис. 11. Пространственное распределение предсказаний нейросети для задачи конвекции с  $\beta = 30$  при принятом кусочно-равномерном увеличении параметра  $\beta$  по ходу обучения нейросети.

Поиск такой оптимальной стратегии представляет собой достаточно сложную, а возможно, и нереализуемую задачу, учитывая множество факторов, влияющих на обучение модели. Кроме того, даже поиск не оптимальной, но просто эффективной стратегии с точки зрения борьбы с нулевым решением также оказывается сложной вычислительной задачей, особенно при рассмотрении достаточно больших значений параметра  $\beta$ . При этом постепенное увеличение параметра  $\beta$  в ходе одного цикла обучения дало результат, сопоставимый с использованием адаптивных подходов выбора обучающего множества. С учетом простоты реализации и настройки адаптивных стратегий в контексте физически-информированных нейросетей этот последний подход представляется наиболее привлекательным. Он сочетает в себе эффективность и простоту, что делает его предпочтительным выбором для решения задач, связанных с избеганием подобного нулевого решения и улучшением качества предсказаний модели.

### Выводы

В настоящей работе проведен анализ применения физически-информированных нейронных сетей к решению задачи конвекции, которая представляет значительные трудности для методов глубокого обучения. Исследование подтвердило, что хотя PINN обладает рядом преимуществ, таких как интеграция физических законов, в сравнении с рядом других подходов глубокого обучения, а также имеет преимущество, выраженное в отсутствии необходимости классического построения расчетных сеток, данный метод сталкивается с серьезными проблемами при решении жестких задач, таких как задачи конвекции с высокими значениями  $\beta$ .

Ключевыми трудностями являются проблемы нулевого решения и низкой обучаемости модели при увеличении жесткости задачи. Рассмотренные в работе подходы, включающие исследование решения на последовательности вложенных сеток, лагранжеву регуляризацию, адаптивный подбор обучающих точек и поэтапное увеличение скорости конвекции, продемонстрировали различную степень эффективности в преодолении указанных проблем. В частности, ранее не исследованные в рамках поставленной задачи адаптивные стратегии подбора обучающего множества и постепенное увеличение скорости конвекции показали наиболее стабильные результаты, обеспечивая лучшее распространение решения и снижение ошибок аппроксимации.

Особое внимание следует уделить неожиданно эффективному применению адаптивных стратегий построения обучающего множества. Эти стратегии не просто дали возможность повысить точность в областях с резкими градиентами, но продемонстрировали способность значительно улучшать обучаемость модели в целом. Они позволили сосредоточиться на критических участках, предотвращая сходимость к тривиальному нулевому решению и улучшая распространение решения по области. Важно отметить, что, в отличие от стратегии с поэтапным увеличением скорости конвекции, адаптивные коллокационные точки представляют собой относительно простой в настройке и использовании инструмент.

Несмотря на достигнутые результаты, дальнейшие исследования должны быть направлены на разработку более универсальных и надежных методов оптимизации обучения PINN. В частности, требуется более глубокое изучение механизмов адаптации сеток в различных типах задач и их влияния на общую сходимость модели. Также к важным направлениям будущих работ следует отнести исследование комбинированных методов, объединяющих рассмотренные адаптивные стратегии построения обучающего множества с другими подходами, такими как поэтапное увеличение параметров задачи, для достижения максимальной эффективности.

### Список литературы

1. Raissi M., Perdikaris P., Karniadakis G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations // *J. Computat. Physics*. 2019. Vol. 378. P. 686–707.
2. Cuomo S., Di Cola V.S., Giampaolo F., Rozza G., Raissi M., Piccialli F. Scientific machine learning through physics-informed neural networks: Where we are and what's next // *J. Sci. Computing*. 2022. Vol. 902, No. 3. 88 p.
3. Karniadakis G.E., Kevrekidis I.G., Lu L., Perdikaris P., Wang S., Yang L. Physics-informed machine learning // *Nature Reviews Physics*. 2021. Vol. 3, No 6. P. 422–440.
4. Cai S., Mao Z., Wang Z., Yin M., Karniadakis G.E. Physics-informed neural networks (PINNs) for fluid mechanics: A review // *Acta Mechanica Sinica*. 2021. Vol. 37, No. 12. P. 1727–1738.
5. Zhu L., Jiang X., Lefauve A., Kerswell R.R., Linden P.F. New insights into experimental stratified flows obtained through physics-informed neural networks // *J. Fluid Mechanics*. 2024. Vol. 981. P. R1–R1-13.
6. Jiang X., Wang D., Fan Q., Zhang M., Lu C., Lau A.P.T. Physics-informed neural network for nonlinear dynamics in fiber optics // *Laser & Photonics Reviews*. 2022. Vol. 16, No. 9. P. 2100483-1–2100483-15.
7. Цгоев Ч.А., Сахаров Д.И., Братенков М.А., Травников В.А., Серёдкин А.В., Калинин В.А., Фомичев Д.В., Мулляджанов Р.И. Моделирование обтекания тела в двумерном канале на основе физически информированных нейронных сетей // *Прикл. механика и технич. физика*. (Принята к печати.)
8. Sirignano J., Spiliopoulos K. DGM: A deep learning algorithm for solving partial differential equations // *J. Computational Physics*. 2018. Vol. 375. P. 1339–1364.
9. Pang G., Lu L., Karniadakis G.E. PINNs: Fractional physics-informed neural networks // *SIAM J. Sci. Computing*. 2019. Vol. 41, No. 4. P. A2603–A2626.
10. Zhang D., Lu L., Guo L., Karniadakis G.E. Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems // *J. Computational Physics*. 2019. Vol. 397. P. 108850-1–108850-34.
11. Krishnapriyan A., Gholami A., Zhe S., Kirby R., Mahoney M.W. Characterizing possible failure modes in physics-informed neural networks // *Advances in Neural Information Processing Systems*. 2021. Vol. 34. P. 26548–26560.
12. Leiteritz R., Pflüger D. How to avoid trivial solutions in physics-informed neural networks // *arXiv preprint arXiv:2112.05620*. Cornell University, 2021. P. 1–17.
13. Chiu P.H., Wong J.C., Ooi C., Dao M.H. CAN-PINN: A fast physics-informed neural network based on coupled-automatic-numerical differentiation method // *Computer Methods in Applied Mechanics and Eng*. 2022. Vol. 395. P. 114909-1–114909-25.
14. Basir S. Investigating and mitigating failure modes in physics-informed neural networks (pinns) // *arXiv preprint arXiv:2209.09988*. 2022. P. 1–32.

15. **Basir S., Senocak I.** An adaptive augmented lagrangian method for training physics and equality constrained artificial neural networks // arXiv preprint arXiv:2306.04904. Cornell University, 2023. P. 1–24.
16. **Hanna J.M., Aguado J.V., Comas-Cardona S., Askri R., Borzacchiello D.** Residual-based adaptivity for two-phase flow simulation in porous media using physics-informed neural networks // *Computer Methods in Applied Mechanics and Eng.* 2022. Vol. 396. P. 115100-1–115100-16.
17. **Rumelhart D.E., Hinton G.E., Williams R.J.** Learning representations by back-propagating errors // *Nature.* 1986. Vol. 323, No. 6088. P. 533–536.
18. **Li Z., Arora S.** An exponential learning rate schedule for deep learning // arXiv preprint arXiv:1910.07454. Cornell University, 2019. P. 1–29.

*Статья поступила в редакцию 9 сентября 2024 г.,  
принята к публикации 5 декабря 2024 г.*