

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ  
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)

---

ФИЗТЕХ-ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ  
КАФЕДРА ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Топологический анализ математических  
доказательств, генерируемых большими  
языковыми моделями на формальном языке Lean4

Выпускная квалификационная работа на степень бакалавра  
студентки Б05-112 группы ФПМИ Линич А.А.

Научный руководитель  
к.ф.-м.н. Баранников С.А.

Москва 2024

# Оглавление

Введение	3
1 Использование Lean4	4
2 TDA	6
2.1 Баркоды в топологическом анализе данных . . . . .	6
2.2 Применение баркодов . . . . .	7
Литература	8

# Введение

# Глава 1

## Использование Lean4

В последние десятилетия формальные языки и системы доказательства стали важным инструментом в области математики и компьютерных наук. Одним из таких современных инструментов является Lean 4 — система автоматизированного доказательства и язык программирования, разработанный для поддержки формализации математических теорий и создания надежного программного обеспечения. В этом введении мы рассмотрим причины использования формальных языков, их структуру, а также современное движение по формализации математики.

Формальные языки и системы доказательства, такие как Lean 4, используются для повышения строгости и точности математических доказательств. Традиционные математические доказательства, представленные в естественном языке, могут содержать двусмысленности и пропуски деталей, что иногда приводит к ошибкам. Формальные системы, напротив, позволяют выражать математические идеи с абсолютной точностью, сводя вероятность ошибок к минимуму. Они обеспечивают строгое следование правилам логики и помогают автоматизировать процесс проверок и доказательств, значительно ускоряя и упрощая верификацию сложных математических заявлений.

Lean 4 и подобные системы строятся на основе логических формализмов и теории типов, которые задают строгие правила для построения и манипуляции математическими объектами и высказываниями. В основе Lean лежит мощный механизм зависимых типов, позволяющий выразительно представлять и обрабатывать математические конструкции. Это делает Lean 4 не только инструментом для доказательства теорем, но и полноценным языком программирования, где корректность программ может быть формально проверена.

Сегодня движение по формализации математики набирает обороты. Проекты, такие как «Формальная верификация доказательств», привлекают внимание как академических кругов, так и индустрии. Одной из главных причин этого является стремление к созданию надежных математических баз данных и библиотек, которые могут быть использованы в дальнейшем исследовании и разработке технологий. Формали-

зация математики также способствует образовательным целям, позволяя студентам и исследователям изучать и развивать математические теории в интерактивной и проверяемой среде.

Таким образом, Lean 4 и другие системы формальных доказательств представляют собой мощные инструменты для математики и компьютерных наук, способствуя повышению надежности и автоматизации математических исследований и разработки программного обеспечения. В следующих разделах данной главы будут подробно рассмотрены архитектура Lean 4, его применение в различных областях и перспективы развития формализации математики.

# Глава 2

## TDA

Топологический анализ данных (TDA) — это относительно новая область в анализе данных, которая использует методы топологии для изучения структуры и формы данных. Одним из ключевых инструментов TDA является концепция "баркодов" которые предоставляют компактное представление топологических свойств данных. В этой главе мы рассмотрим основы топологического анализа данных, как строятся баркоды, и их применение в различных областях.

Топология, как раздел математики, изучает свойства и структуры, которые инвариантны при непрерывных деформациях. В контексте данных, топологические методы используются для выявления устойчивых и значимых структур, таких как кластеры, циклы и дыры, которые могут присутствовать в многомерных данных. Топологический анализ данных предлагает инструменты для извлечения таких структур, что позволяет лучше понять внутренние свойства данных.

Среди методов TDA особое внимание уделяется гомологии, которая изучает топологические пространства через алгебраические инварианты. Гомология позволяет количественно оценить наличие различных топологических особенностей, таких как компоненты связности, циклы и полости.

### 2.1 Баркоды в топологическом анализе данных

Баркоды являются визуальным и алгебраическим представлением топологических характеристик данных. Они возникли в контексте персистентной гомологии, которая изучает, как топологические свойства данных изменяются при изменении масштаба.

Основная идея заключается в том, чтобы изучить данные на различных уровнях детализации, создавая так называемую фильтрацию. В процессе фильтрации данные преобразуются в серию упрощённых представлений, называемых комплексами. На каждом уровне этой фильтрации вычисляются гомологические группы, и отслеживается "рождение" и "смерть" топологических характеристик, таких как компоненты

связности или циклы.

Результатом этого анализа является баркод — набор горизонтальных отрезков на плоскости, где каждая линия соответствует топологической характеристике. Длина линии отражает "продолжительность жизни" этой характеристики в процессе фильтрации. Длинные линии обычно указывают на значимые топологические особенности, которые могут представлять интерес, в то время как короткие линии могут считаться шумом.

## 2.2 Применение баркодов

Баркоды нашли применение в самых различных областях, где анализ структуры данных играет важную роль. В биоинформатике они используются для анализа структур белков, изучения геномных данных и выявления эволюционных связей. В обработке изображений баркоды помогают в распознавании и классификации объектов на изображениях. В машиностроении и материаловедении они применяются для анализа свойств материалов и оптимизации их структуры.

Одним из преимуществ баркодов является их устойчивость к шуму и деформациям, что делает их особенно полезными для анализа реальных данных, которые часто содержат ошибки и неточности. Они также обеспечивают интуитивно понятный способ визуализации сложных многомерных данных, облегчая интерпретацию и принятие решений.

## Литература