Министерство науки и высшего образования РФ

Национальный исследовательский
Нижегородский государственный университет им. Н.И. Лобачевского



МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И СУПЕРКОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ

ТРУДЫ XXV МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ

Нижний Новгород, 17–19 ноября 2025 г.

Нижний Новгород
Издательство Нижегородского госуниверситета
2025

УДК 004.942+519.876.5+519.6
ББК 22.18я43+22.19я43
М34

М34 **Математическое моделирование и суперкомпьютерные технологии.** Труды XXV Международной конференции (Н. Новгород, 17–19 ноября 2025 г.) / Под ред. проф. Д.В. Баландина. – Нижний Новгород: Изд-во Нижегородского госуниверситета, 2025. – 221 с.

Отв. за выпуск К.А. Баркалов

ISBN 978-5-908049-32-0

Сборник материалов Двадцать пятой Международной конференции «Математическое моделирование и суперкомпьютерные технологии», состоявшейся 17–19 ноября 2025 г. на базе Нижегородского государственного университета им. Н.И. Лобачевского, содержит тезисы докладов и короткие статьи, посвященные математическому моделированию сложных процессов и явлений, численным методам их исследования, а также проблемам разработки методов суперкомпьютерных вычислений для решения актуальных задач в различных областях науки, промышленности и образования.

Подробную информацию о конференции можно найти в сети Интернет по адресу https://mmst.unn.ru.

Поддержка конференции

приоритет Программа академического лидерства Приоритет 2030

Научно-образовательный математический центр
«Математика технологий будущего»

ISBN 978-5-908049-32-0

УДК 004.942+519.876.5+519.6
ББК 22.18я43+22.19я43

© Авторы статей, 2025
© ННГУ им. Н.И. Лобачевского, 2025

ПРЕДИСЛОВИЕ

17–19 ноября 2025 г. в рамках международного конгресса «Суперкомпьютерные дни в России» на базе Нижегородского государственного университета им. Н.И. Лобачевского при поддержке программы академического лидерства «Приоритет 2030» и научно-образовательного ма-тематического центра «Математика технологий будущего» проведена XXV Международная конференция и молодежная школа «Математическое моделирование и суперкомпьютерные технологии».

Тематика конференции охватывала основные направления в математического моделирования и суперкомпьютерных технологий:

- математическое моделирование динамики систем и процессов управления;
- теория динамических систем и бифуркаций;
- колебательные процессы в динамических сетях;
- модели и методы поддержки принятия решений;
- модели и методы искусственного интеллекта;
- математическое модели в биологии и медицине;
- математическое моделирование природных процессов;
- алгебра, геометрия и дискретная математика;
- технологии параллельных и распределенных вычислений;
- применение суперкомпьютерных технологий для решения вычислительно-сложных задач.

На молодежной школе в формате пленарных докладов было рассказано о текущем состоянии дел в области разработки и адаптации системного и прикладного ПО для процессоров архитектуры RISC-V. В рамках научной секции представлены доклады участников школы, прошедшие рецензирование и рекомендованные программным комитетом.

Основными задачами проведения мероприятия являлись представление актуальных результатов в области математического моделирования сложных процессов и явлений, обсуждение различных аспектов организации суперкомпьютерных вычислений, расширение контактов между специалистами для решения ресурсоемких прикладных задач, обмен опытом научно-образовательной деятельности при подготовке специалистов в области математического моделирования и параллельных вычислений.

Все перечисленные темы нашли отражение в настоящем сборнике тезисов докладов и коротких статей, представленных на конференцию (публикуются частично в авторской редакции).

THE IMPACT OF NEWS SENTIMENT ON THE QUALITY OF PREDICTIVE ANALYTICS USING A LONG-SHORT TERM MEMORY (LSTM) MODEL

E.A. Kubrakova

Moscow Institute of Physics and Technology

This study investigates the impact of integrating news sentiment analysis into predictive models for stock price forecasting on the Russian market, a domain significantly under-researched compared to its foreign counterparts. While the influence of investor sentiment is well-documented for developed markets, its efficacy in the context of the high-volatility Russian environment remains an open question. We address this gap by proposing a hybrid modeling approach that augments traditional technical indicators (TA) with aggregated news sentiment scores (SA). Our methodology employs advanced neural network architectures, specifically LSTM and a hybrid LSTM/GRU model, to process sequential data and capture complex, non-linear relationships. The models were rigorously tested on a portfolio of six liquid stocks from the iMOEX index during a period of extreme market turbulence (February–March 2022), providing a robust stress-test for predictive accuracy. The empirical results demonstrate a statistically significant enhancement in forecasting performance when sentiment features are included. Across all evaluated metrics (RMSE, MAE, MAPE), models utilizing the hybrid TA+SA approach consistently outperformed those based solely on technical analysis. For instance, the LSTM model for AFLT showed a 60% reduction in RMSE (from 3.90 to 1.56), while the hybrid LSTM/GRU model for SBER achieved a lower error (RMSE 9.10 vs. 10.16). The improvement was most pronounced for stocks sensitive to macroeconomic and public sentiment shifts, such as in the financial and transportation sectors. The findings compellingly indicate that news sentiment is a critical and economically meaningful feature that significantly improves the accuracy of next-day price predictions for Russian stocks, with hybrid architectures proving particularly effective in synthesizing quantitative and qualitative data. This research confirms the vital role of investor sentiment in Russian market dynamics and provides a validated framework for its application in predictive modeling.

Keywords: Investor Sentiment, News Sentiment Analysis, Russian Stock Market, Predictive Modeling, Neural Networks, Technical Analysis, Machine Learning in Finance.

1. Introduction

A significant number of studies demonstrate the substantial impact of investor sentiment on the stock market [1-9]. The assessment of investors' emotional behavior is an important aspect of stock trading. Investment sentiment influences not only the selection of stock assets but also the dynamics of investment decision-making, trading volume, and risk appetite. According to studies, the prevailing daily sentiment in financial news affects buying and selling in the stock market [10-13]; thus, it makes sense to assess the emotional tone not only of individual news items about issuers but also of an aggregated measure for the day. Several studies have shown that if financial news is collected and analyzed as part of an investment strategy, it can lead to profits both in the short term and long term [12-13]. J.G. Vivek et al. [14] demonstrated that stock prices predicted based on historical data and sentiment significantly correlate with actual stock prices of specific companies. K. Bozic and D. Seese [15] proposed a system for quantitatively assessing textual sentiment based on predictors using neural networks and proved a statistically significant relationship between the textual sentiment of published news and future daily returns. A high degree of correlation between investor sentiment and stock prices was also demonstrated in a study by V. Sehgal and C. Song [16] as well as in the work of X. Liang [17], an aggregated measure (the volume of published news about the stock market) was used, without classifying sentiment by issuers, to train a neural network and predict stock price changes, resulting in a statistically significant relationship between the indicators. Most studies on the relationship between

sentiment and price fluctuations in stock market quotations mentioned above have focused exclusively on foreign markets, while such research is practically absent for the Russian market. One of the pioneering studies in this area is the work of researchers from HSE [18], which examines the influence of investor sentiment in social networks on the stock market characteristics of the Russian market, demonstrating nonlinear effects, and proposes an original HYPE sentiment indicator that is compared in terms of explanatory power for stock market characteristics with a wide range of proxy variables. However, this area of research requires more detailed and comprehensive analysis, both in assessing the impact of sentiment on market dynamics as a whole and in the possibility of using news sentiment as a feature in training predictive models for the Russian stock market.

The aim of this report is to provide information on the main indicators of investment sentiment under conditions of extreme financial market volatility and information restriction in order to improve forecasts accuracy as for trends and price dynamics in the Russian stock market and to demonstrate the advantage of hybrid approach using aggregated news sentiment as one of the significant features.

The main indicators of financial sentiment in foreign financial markets are:

- Macroeconomic indices (e.g. Consumer Sentiment Index, Business Confidence Index, etc.), reflecting confidence in the economic situation and readiness to take investment decisions;
- Trader sentiment indicators (e.g. Fear and Greed Index, Volatility Index (VIX), the Bullish/Bearish sentiment surveys, Bull-Bear ratio, Commitment of Traders (COT), ratios of long and short positions, open interest indicators, trading volume etc.;
- Media analysis using data from social media, press releases, and news headlines to assess the overall sentiment toward specific assets or markets.

These indicators are widely recognized factors influencing the financial, and particularly the stock market. They help traders and analysts assess the overall market sentiment and make more informed investment decisions. However, each of these indicators has its strengths and weaknesses. For example, the VIX shows a significant negative correlation with the S&P 500 index, making it a good predictor of trends in the short term for stocks within this index; however, the degree of influence on other assets is less obvious.

The use of technical indicators, such as moving averages or the stochastic oscillator, can be adapted to any period and is suitable for both long-term and short-term traders (moving averages help in decomposing price charts). On the other hand, these technical indicators are calculated using only historical price data. Chart indicators, like any other technical analysis tools, ignore changes in fundamental variables that can affect the future value of a security, such as new competitors, increases or decreases in demand for goods in the industry, and changes in the company's management structure. Moreover, there are no universal rules applicable to all markets, as each asset has its unique price history and volatility levels that are critical for model training. Finally, moving averages do not necessarily reflect the cyclical nature of stock activity. During periods of high market volatility, moving averages are unlikely to detect any significant trends.

2. Methodology

Preliminary analysis of the mentioned above indicators for parameterization in machine learning models reveals several aspects that may affect model effectiveness.

1. Some indicators (such as the consumer sentiment index, business activity index, fear and greed index etc.) are aggregated metrics and require decomposition for use as features in machine learning models.

2. Not all indicators can be directly parameterized in machine learning models, as there is a strong correlation between them. E.g., in case of such features as “open”, “low”, “high”, “close” strong correlation necessitates dimensionality reduction to eliminate "noise" in the data and enhance model efficiency or trader sentiment indicators (bullish or bearish), which are assessed based on both open interest and trading volumes, leading to a strong correlation between these features and posing the question of whether to prioritize one of the two indicators (see fig.1), especially since there may be a time lag between them that affects the model's forecast.

3. In Russian practice, information is sometimes provided by the exchange and the regulator in a form that differs from that accepted abroad. For instance, there is no direct equivalent to the Commit-

ment of Traders (COT) report in the Russian market, but a somewhat comparable source can be considered the data on trader positions published by the Moscow Exchange (MOEX), as well as information on margin requirements and analytical reviews/reports from financial companies and brokers, which may include data on market sentiment and the positions of large players.

4. Fundamental indicators of a company and macroeconomic indicators are more significant features in long-term investment analysis, while in the short term, they are not statistically significant. Hence, this information can be extracted from analysts' and companies' reports and included in sentiment score.

5. In addition to technical indicators, textual information from press releases, news headlines, social media, and analytical reviews has a significant emotional impact on investors. The degree of this influence depends on the type of sentiment, the industry in question, the type of investor (retail investors are more susceptible to strong emotional reactions), and the duration of this influence (the cumulative effect or the "snowball" effect).

We apply filtering methods (correlation assessment, significance assessment), regularization and principal component analysis (PCA) to reduce the dimensionality of feature space along with using aggregated news sentiment to enhance models' forecasting accuracy.

The dataset is split into training and test sets in a 70/30 ratio. The split boundary was set at September 1, 2021, so that the evaluated metrics (RMSE, MAE, MAPE) would be calculated on the test set during the period of highest market turbulence to assess the algorithm's stability. The closing price prediction is made for the portfolio of 6 common stocks from iMOEX index for the next day. The experiment was conducted on LSTM and LSTM with additional GRU layers.

3. Model's architechture

The architecture of the best model uses recurrent layers to process time series and sequential data. It contains:

1. Input Layerm which accepts three-dimensional data typical for sequences (batches, timesteps, features).

2. Stack of LSTM layers that process the input sequence. These layers are configured to return the full sequence of outputs for each timestep to be passed on to the next recurrent layer.

3. Following the LSTM stack is a GRU (Gated Recurrent Unit) layer. This layer returns the full output sequence (`output_shape: (None, 5, 128)`), meaning it produces an output for each of the 5 timesteps, and has 128 hidden units. Number of trainable parameters: 53,760.

4. The next GRU layer returns only the output vector from the last timestep (`output_shape: (None, 64)`), which contains compressed information from the entire processed sequence. It has 64 hidden units. Number of trainable parameters: 37,248.

5. The resulting vector is fed into a fully-connected (Dense) hidden layer with 32 neurons for further feature transformation. Number of parameters: 2,080.

The final fully-connected layer has 1 neuron, which is typical for regression tasks (predicting a single numerical value). Number of parameters: 33.

For forecasting the closing price, **Sequential**-class neural networks **without hidden layers** were used to ensure the model's **complete interpretability**. The LSTM model consisted of **three deep layers**, with the **return_sequences parameter set to True**. The following hyperparameters were applied:

LR (Learning Rate) = 0.001

BATCH_SIZE = 128

N_EPOCH (Number of Epochs) = 50

4. Results

Table 1 presents the metric results for the main instruments of the iMOEX index based on predictions made during the market crash (February–March 2022) and subsequent recovery. This period marks the peak of market volatility in the Russian stock market from 2014 to 2024.

Table 1. Comparison of metrics for neural network algorithms LSTM and LSTM/GRU on the test set including 2022 (next-day prediction)

Ticker	Algorithm	RMSE on test set	MAE (RUR) on test set	MAPE (%) on test set
SBER	LSTM based on TA*	14.48	10.28	5.73
	LSTM (TA + SA)**	9.98	5.10	2.39
	LSTM/GRU based on TA	10.16	5.36	2.38
	LSTM/GRU (TA + SA)	9.10	4.59	2.19
LKOH	LSTM based on TA	183.89	91.39	1.64
	LSTM (TA + SA)	153.53	76.94	1.36
	LSTM/GRU based on TA	202.71	126.69	2.16
	LSTM/GRU (TA + SA)	178.85	105.89	1.83
MOEX	LSTM based on TA	4.89	3.15	2.15
	LSTM (TA + SA)	4.54	2.82	1.90
	LSTM/GRU based on TA	3.61	2.37	1.74
	LSTM/GRU (TA + SA)	4.44	2.84	1.97
GMKN	LSTM based on TA	5.02	3.69	2.33
	LSTM (TA + SA)	4.84	3.22	1.98
	LSTM/GRU based on TA	4.42	2.90	1.81
	LSTM/GRU (TA + SA)	4.18	2.83	1.74
MTSS	LSTM based on TA	6.89	3.82	1.48
	LSTM (TA + SA)	5.92	2.66	1.03
	LSTM/GRU based on TA	8.19	4.68	1.80
	LSTM/GRU (TA + SA)	7.09	3.70	1.42
AFLT	LSTM based on TA	3.90	2.99	8.01
	LSTM (TA + SA) .	1.56	1.00	2.49
	LSTM/GRU based in TA	3.66	2.92	7.85
	LSTM/GRU (TA + SA)	1.98	1.03	2.54

*TA stands for prediction based on technical analysis

**TA + SA stands for predition based on hybrid approach (technical indicators + sentiment analysis)

As the results show, adding sentiment to features based on technical indicators reduces the model's prediction error for the next day, regardless of the architecture (LSTM or GRU) used, particularly for instruments with high liquidity and market volumes (SBER).

The empirical analysis conducted in this study yields statistically significant conclusions regarding the efficacy of integrating sentiment analysis (SA) with technical indicators (TA) for stock price forecasting on the Russian market. The key findings are as follows:

1. The incorporation of news sentiment data consistently resulted in a material improvement in the forecasting performance of both LSTM and hybrid LSTM/GRU models across all evaluated error metrics (RMSE, MAE, MAPE). For instance, the LSTM model for **AFLT** demonstrated a drastic reduction in error upon the inclusion of SA, with RMSE decreasing from **3.90 to 1.56** (a **60%** improvement) and MAE from **2.99 to 1.00**. This pattern is unequivocal for pure LSTM architectures and is prevalent in hybrid models, underscoring sentiment as a critical predictive factor.

2. The magnitude of improvement was observed to be contingent on the specific ticker, suggesting a sector-dependent sensitivity to news sentiment. The most pronounced gains were noted for stocks inherently exposed to macroeconomic and public sentiment shifts, such as the financial sector (**SBER**) and transportation (**AFLT**). For **SBER**, the hybrid LSTM/GRU (TA+SA) model achieved an RMSE of **9.10**, compared to **10.16** for its TA-only counterpart. Conversely, for commodities-driven companies (**LKOH**, **GMKN**), while absolute error reduction was substantial (e.g., LSTM for **LKOH**: RMSE improved from **183.89 to 153.53**), the percentage gain was more moderate, indicating that sentiment acts as a powerful complementary factor rather than the primary driver.

3. **Superiority of Hybrid Architectures with Sentiment:** The hybrid LSTM/GRU model leveraging combined TA+SA features frequently delivered superior or among the best results, minimizing error metrics across the board. For example, for **GMKN**, the LSTM/GRU (TA+SA) model yielded the lowest RMSE (**4.18**) and MAPE (**1.74%**) of all tested configurations for that asset. This suggests that

the architectural synergy of LSTM and GRU is particularly adept at capturing the complex, non-linear interactions between quantitative technical data and qualitative sentiment signals.

4. The reduction in error is consistent across all three metrics, confirming not only a lower absolute deviation (MAE) but also a improved overall error magnitude (RMSE) and relative forecasting precision (MAPE). The reduction in MAPE is particularly noteworthy; for the **LSTM model on SBER**, the error fell from **5.73% to 2.39%**, effectively halving the average percentage forecast error and enhancing the model's practical utility for financial decision-making.

5. Conclusion

In summary, this research provides compelling evidence that investor sentiment, quantified through news headline analysis, is a statistically significant and economically meaningful feature for equity price forecasting on the Russian market. The integration of sentiment data with traditional technical indicators leads to a robust and significant enhancement in model accuracy, with hybrid deep learning architectures (LSTM/GRU) processing these combined datasets achieving optimal performance. These findings confirm the hypothesis that investor sentiment, translated into market behavior, is a critical determinant of stock price dynamics in the Russian context, thereby filling a identified gap in the literature.

In addition to the careful selection of features for the machine learning model based on principal component analysis and factor analysis aimed at reducing the dimensionality of the feature matrix, lowering degrees of freedom to stabilize the model, and other optimization factors, we use a hybrid approach based not only on technical indicators but also on sentiment analysis of news in the Russian stock market, including fundamental information about companies, since it is an important factor in enhancing the predictive accuracy of the model while maintaining an optimal number of features due to the possibility to extract significant financial information from textual data, either explicitly or implicitly, without the need to increase the number of analyzed features.

References

1. Seok, S., Cho, H., Ryu, D., Firm-specific investor sentiment and daily stock returns, The North American Journal of Economics and Finance, 2019, vol. 50, 100857. DOI: https://doi.org/10.1016/j.najef.2018.10.005.
2. Smales, L., The importance of fear: Investor sentiment and stock market returns, Applied Economics, 2017, vol. 49, pp. 3395-3421. DOI: https://doi.org/10.1080/00036846.2016.1259754.
3. Fang, H., Chung, C., Lu, Y., et al., The impacts of investors' sentiments on stock returns using fintech approaches. International Review of Financial Analysis, 2021, vol. 77, 101858. DOI: https://doi.org/10.1016/j.irfa.2021.101858.
4. Kadilli, A., Predictability of stock returns of financial companies and the role of investor sentiment: A multi-country analysis, Journal of Financial Stability, vol. 21, pp. 26-45. DOI: https://doi.org/10.1016/j.jfs.2015.09.004.
5. Seok, I., Cho, H., Ryu, D., Stock market's responses to intraday investor sentiment. The North American Journal of Economics and Finance, 2021, vol. 58, 101516. DOI: https://doi.org/10.1016/j.najef.2021.101516.
6. Maitra, D., Dash, S., Sentiment and stock market volatility revisited: A time–frequency domain approach, Journal of Behavioral and Experimental Finance, 2017, vol. 15, pp. 74-91. DOI: https://doi.org/10.1016/j.jbef.2017.07.009.
7. Jiang, S., Jin, X., Effects of investor sentiment on stock return volatility: A spatio-temporal dynamic panel model, Economic Modelling, 2021, vol. 97, pp. 298-306. DOI: https://doi.org/10.1016/j.econmod.2020.04.002.
8. Kim, K., Ryu, D., Term structure of sentiment effect on investor trading behaviour, Finance Research Letters, 2021. DOI: https://doi.org/10.1016/j.frl.2021.102005.
9. Teplova T.V. et al. Sentiment chastnykh investorov v obiasnenii razlichij v birzhevykh kharakterisitikakh akcii rossijskogo rynka // Journal Novoi ekonomicheskoi associacii, 2022, № 1 (53). C. 53–84. DOI: 10.31737/2221-2264-2022-53-1-3 (in Russ.).

10. de Oliveira Carosia, A., Coelho G., da Silva A., Investment strategies applied to the Brazilian stock market: A methodology based on sentiment analysis with deep learning, expert systems with applications, 2021, vol. 184, 115470. DOI: https://doi.org/10.1016/j.eswa.2021.115470.
11. Parveen, S., Satti, Z., Subhan, Q., et al., Exploring market overreaction, investors' sentiments and investment decisions in an emerging stock market, Borsa Istanbul Review, 2020, vol. 20, pp. 224-235. DOI: https://doi.org/10.1016/j.bir.2020.02.002.
12. Danso, A., Lartey, T., Amankwah-Amoah, J., et al., Market sentiment and firm investment decisionmaking, International Review of Financial Analysis, 2019, vol. 66, 101369. DOI: https://doi.org/10.1016/j.irfa.2019.06.008.
13. Lerner, J., Keltner, D., Beyond valence: Toward a model of emotion-specific influences on judgement and choice, Cognition & Emotion, 2000, vol. 14, pp. 473-493. DOI: https://doi.org/10.1080/026999300402763.
14. Vivek J. G., Darshan M.S., Sneha P., Arun S., Vanipriya Ch., "Indian Stock Market Predictor System", in International Journal of Soft Computing and Artificial Intelligence, ISSN: 2321-404X Volume- 1, Issue- 1, pg. 39-44.
15. Bozic C., Seese D., "Neural Networks for Sentiment Detection in Financial", JEL Codes: C45, D83, and G17, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology.
16. Sehgal V., Song Ch., "SOPS: Stock Prediction using Web Sentiment", 2007 IEEE.
17. Liang X. 2005, "Impacts of Internet Stock News on Stock Markets Based on Neural Networks": ISNN 2005, LNCS 3497, pp. 897–903, 2005. © Springer-Verlag Berlin Heidelberg 2005.
18. Teplova, T. V. et al. Sentiment of Private Investors in Explaining Differences in the Market Characteristics of Russian Stocks. Journal of the New Economic Association, 2022, 53(1), 53–84. https://doi.org/10.31737/2221-2264-2022-53-1-3.

О ЗАДАЧЕ РАСПОЗНАВАНИЯ *A*-ФУНКЦИИ В ФУНКЦИОНАЛЬНОЙ СИСТЕМЕ РАЦИОНАЛЬНЫХ ФУНКЦИЙ С РАЦИОНАЛЬНЫМИ КОЭФФИЦЕНТАМИ

***Н.Ф. Алексиадис*[1,2]**

[1]*Московский государственный университет им. М.В. Ломоносова,*
[2]*Национального исследовательского университета «МЭИ»*

Функциональная система (ф.с.) F = (*F*, *O*) представляет собой множество функций *F* с некоторым набором операций *O*, применяемых к функциям *F* и приводящих к получению других функций из *F*. Одной из основных проблем является проблема полноты, состоящая в описании таких подсистем функций, которые являются полными, т.е. из этих функций с помощью операций из *O* можно получить все функции из *F*. С проблемой полноты связана задача об *A*-функциях. *A*-функция в ф.с. F – это аналог функции Шеффера в двузначной логике, т.е. такая функция *f*, что {*f*} является полной системой в ф.с. F. В настоящей работе рассматривается следующая задача: существует ли алгоритм, который для любой заданной рациональной функции с рациональными коэффициентами выясняет, является ли она *A*-функцией или нет; другими словами, в ф.с. рациональных функций с рациональными коэффициентами проблема распознавания *A*-функции алгоритмически разрешима или нет?

Ключевые слова: функциональная система, полная система, проблема полноты, базис, функция Шеффера, алгоритмическая неразрешимость, рациональная функция.

1. Введение

Несмотря на то, что мы используем стандартные обозначения и общеизвестные понятия дискретной математики (в частности, теории функциональных систем), с целью корректного понимания изложенного, все-таки следует уточнить некоторые моменты.

Приведем предварительные сведения из теории функциональных систем, необходимые для дальнейшего изложения.

При изложении материала в основном используется терминология книг [11,12].

Мы будем использовать следующие стандартные обозначения:

N – множество всех натуральных чисел (включая число 0),

Q – множество всех рациональных чисел.

Для удобства полагаем, что $0^0 = 1$.

Говорят, что функция $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ *существенно зависит от переменной* x_i, если существуют такие два набора (из области определения этой функции)

$$c_1, \cdots, c_{i-1}, a, c_{i+1}, \cdots, c_n \text{ и } c_1, \cdots, c_{i-1}, b, c_{i+1}, \cdots, c_n$$

значений переменных $x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n$, что

$$f(c_1, \cdots, c_{i-1}, a, c_{i+1}, \cdots, c_n) \neq f(c_1, \cdots, c_{i-1}, b, c_{i+1}, \cdots, c_n).$$

В этом случае мы говорим, что x_i является *существенной переменной функции*

$$f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n).$$

Если x_i не является существенной переменной $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ то она называется *фиктивной* (*несущественной*) *переменной* функции $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$.

Пусть $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ и $g(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ две произвольные заданные функции и пусть x_i фиктивная переменная функции $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$.

Если для любых $c_1, \cdots, c_{i-1}, c_i, c_{i+1}, \cdots, c_n$ значений переменных $x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n$ имеем

$$f(c_1, \cdots, c_{i-1}, c_i, c_{i+1}, \cdots, c_n) = g(c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n),$$

то говорят, что функция $g(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ получается из функции $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ *удалением* (*изъятием*) *фиктивной переменной* x_i и, наоборот,

функция $f(x_1,\cdots,x_{i-1},x_i,x_{i+1},\cdots,x_n)$ получена из функции $g(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n)$ *добавлением фиктивной переменной* x_i.

Если одна функция может быть получена из другой функции путем добавления или изъятия некоторых фиктивных переменных, то эти две функции являются *равными*.

Замечание 1. В дальнейшем будем считать, что вместе с функцией *f* заданы и все равные ей функции, т.е. функции рассматриваем с точностью до фиктивных переменных.

Замечание 2. Если дана конечная система функций $f_1,\cdots,f_m$ $(m \geq 1)$, то можно считать, что все они зависят от одних и тех же переменных $x_1,\cdots,x_n$, т.е. имеют вид

$$f_1(x_1,\cdots,x_n),\cdots,f_m(x_1,\cdots,x_n).$$

Замечание 3. Если дана функция, отличная от константы, то путем отождествления переменных из нее можно получить равную ей функцию, все переменные которой являются существенными.

Функциональная система (*ф.с.*) представляет собой множество функций с некоторым набором операций, применяемых к этим функциям и приводящих к получению других функций из этого же множества, т.е. *функциональная система* – это пара вида **F** = (*F*, *O*), где *F* – множество функций, а *O* – множество операций над этими функциями, при этом каждая операция из *O* замкнута относительно множества *F*.

Для произвольного подмножества $A \subseteq F$ обозначим через [*A*] множество всех функций из *F*, которые получаются из функций множества *A* с помощью конечного числа применения операций из *O*. Множество [*A*] называется *замыканием множества A*.

Множество *A* (A ⊆ F) называется *замкнутым* в функциональной системе **F**, если [*A*] = *A*.

Замкнутое множество принято называть *замкнутым классом*.

Множество *A* (A ⊆ F) называется *полным* в функциональной системе **F**, если [*A*] = *F*.

Полное множество принято называть *полной системой*.

Система функций *A* (*A* ⊆ *F*) называется *базисом* в ф.с. **F**, если *A* полная система и никакая ее собственная подсистема не является полной в **F**.

Нашей целью является изучение самых минимальных базисов, т.е. базисов, которые состоят из одной единственной функции (если, конечно, такие существуют).

Если базис состоит из одной единственной функции f, то f называется *A-функцией*; другими словами, f является *A*-функцией, если система $\{f\}$ является полной в ф.с. **F**, т.е. из одной единственной функции f с помощью операций из *O* можно получить все функции из *F*.

2. Определение функциональной системы рациональных функций с рациональными коэффициентами

Выражение вида $cx_1^{k_1}\cdots x_n^{k_n}$, где $n,k_1,\cdots,k_n \in N$, а $c \in Q$, называется *мономом с рациональным коэффициентом*, зависящим от n переменных $x_1,\cdots,x_n$; при этом, когда n = 0, тогда заданный моном является просто константой *c*, т.е. мономом с рациональным коэффициентом, зависящим от 0-го числа переменных.

Конечная сумма мономов с рациональными коэффициентами называется *полиномом с рациональными коэффициентами*.

Функция вида

$$f(x_1,\cdots,x_n)\colon \underbrace{Q\times\cdots\times Q}_{n} \to Q,$$

которая задается с помощью полинома с рациональными коэффициентами, называется *полиномиальной функцией с рациональными коэффициентами*.

Следует обратить внимание на то, что аргументы полиномиальных функций с рациональными коэффициентами и сами эти функции принимают только рациональные значения.

Функция вида

$$f(x_1,\cdots,x_n)=\frac{g(x_1,\cdots,x_n)}{h(x_1,\cdots,x_n)},$$

где $g(x_1,\cdots,x_n)$ и $h(x_1,\cdots,x_n)$ – полиномиальные функции с рациональными коэффициентами, называется *рациональной функцией с рациональными коэффициентами*.

Рациональные функции с рациональными коэффициентами будем называть ещё rq-*функциями*.

Следует обратить внимание на то, что аргументы рациональных функций с рациональными коэффициентами и сами эти функции принимают только рациональные значения.

Обозначим через F_{RQ} множество всех рациональных функций с рациональными коэффициентами.

Функциональная система рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ}$ – это пара $\mathbf{F}_{RQ} = (F_{RQ}, O)$, где F_{RQ} – множество всех рациональных функций с рациональными коэффициентами, а O – множество операций суперпозиции над рациональными функциями с рациональными коэффициентами, которое включает в себя:

1. перестановку переменных,
2. переименование переменных (без отождествления),
3. отождествление переменных,
4. введение фиктивной переменной,
5. удаление фиктивной переменной,
6. подстановку одной функции в другую.

Заметим, что определение функциональной системы $\mathbf{F}_{RQ} = (F_{RQ}, O)$ корректное, так как любая суперпозиция функций из F_{RQ} является опять функцией из F_{RQ}.

3. Постановка задачи

Одной из основных проблем в теории функциональных систем является *проблема полноты, состоящая в описании всех подмножеств данного множества F, которые являются полными в* **F** = (*F*, *O*).

С проблемой полноты связана и *задача о базисах, т.е. изучение вопроса о минимальных полных системах* и, в частности, изучение базисов, состоящих из одной функции (если такие базисы существуют), т.е. изучение вопроса об *A*-функциях.

При изучении *A*-функций выделяют два подхода: *алгебраический* и *алгоритмический*.

В первом случае ставится вопрос о получении критериев (т.е. необходимых и достаточных условий) для того, чтобы данная функция являлась *A*-функцией. Во втором случае задача заключается в существовании алгоритма, который распознает все *A*-функции.

В настоящей работе будет рассмотрен алгоритмический вариант задачи об *A*-функциях для функциональной системы $\mathbf{F}_{RQ} = (F_{RQ}, O)$, который звучит так:

Существует ли алгоритм, который для любой заданной rq*-функции выясняет, является ли она A-функцией или нет; другими словами, в ф.с.* $\mathbf{F}_{RQ}$ *проблема распознавания A-функции алгоритмически разрешима или нет?*

Сразу отметим, что аналогичная задача была решена автором для функциональной системы полиномиальных функций с целыми коэффициентами, о чем было сообщено в [1], а алгебраический подход задачи об *A*-функциях был рассмотрен в докладе [2] и далее расширенная версия этого доклада была опубликована в [3].

4. Решение задачи о A-функции (алгоритмический подход) для рациональных функций с рациональными коэффициентами

При исследовании проблемы об *A*-функциях с алгебраической точки зрения для функциональной системы рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ}$ автором ранее была установлена справедливость следующих теорем.

Теорема 1 [3]. В функциональной системе рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ}$ существует *A*-функция, а именно, функция

$$f(x, y, z, t, u, v) = (t - u) \cdot \frac{x}{v} + xz - yz + z + 1$$

является *A*-функцией.

Теорема 2 [3]. В функциональной системе рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ}$ существует счетное число A-функций.

При дальнейшем исследовании проблемы полноты и задачи о базисах с алгоритмической точки зрения для функциональной системы рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ}$ было получено следующее утверждение.

Теорема 3 [4]. В функциональной системе рациональных функций $\mathbf{F}_{RQ}$ с рациональными коэффициентами $\mathbf{F}_{RQ}$ проблема полноты алгоритмически неразрешима.

Теорема 4 [5]. В функциональной системе рациональных функций $\mathbf{F}_{RQ}$ с рациональными коэффициентами $\mathbf{F}_{RQ}$ задача о базисах алгоритмически неразрешима.

Возникает интересный и важный вопрос: несмотря на то, что в функциональной системе $\mathbf{F}_{RQ}$ в общем случае проблема полноты алгоритмически неразрешима и задача о базисах также алгоритмически не разрешима, может быть, в частном случае для систем, состоящих из одной единственной функции, проблема полноты алгоритмически разрешима? Другими словами, проблема распознавания A-функции алгоритмически разрешима или нет?

К сожалению, и в этом случае ответ отрицателен. Справедлива следующая теорема, которая является основным результатом данного доклада.

Теорема 5. В функциональной системе рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ}$ проблема распознавания A-функции алгоритмически неразрешима.

5. Заключение

Функциональные системы являются одним из основных объектов дискретной математики и математической кибернетики и отражают следующие основные особенности реальных и абстрактных управляющих систем: функционирование (в функциональных системах - это функции), правила построения более сложных управляющих систем из заданных и описание функционирования сложных систем по функционированию их компонент (последние два момента отражены в операциях функциональных систем).

В настоящей работе рассматривается функциональная система рациональных функций с рациональными коэффициентами и для этой функциональной системы решается задача об A-функциях (алгоритмический подход), которая играет ключевую роль не только в самой дискретной математике и математической кибернетике, но и во многих других областях математики, например, в теории функций (аппроксимационные теоремы Вейерштрасса и Чебышева), в вычислительной математике и технике (построение и анализ вычислительных чипов и нейронных сетей). Актуальность полученных результатов также состоит в развитии самой теории функциональных систем как в плане охвата новых модельных объектов типа рациональных функций, так и в вычленении позитивных результатов, и в отсечении негативных ситуаций типа алгоритмической неразрешимости задачи об A-функциях.

Целью настоящего доклада, как уже было сказано выше, было изучение задачи об A-функциях с алгоритмической точки зрения в ф.с. $\mathbf{F}_{RQ} = (F_{RQ}, O)$. Полученная на этом пути теорема (5) об алгоритмической неразрешимости задачи об A-функциях в ф.с. $\mathbf{F}_{RQ} = (F_{RQ}, O)$ является основным результатом данной работы.

В заключении отметим, что представленный в настоящем докладе результат [теорема (5)] является логическим продолжением ранее полученных автором результатов относительно функциональной системы рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ}$, т.е. продолжением цикла работ [1-10].

Литература

1. Алексиадис Н. Ф. Об алгоритмической неразрешимости задачи распознавания А-функции в функциональной системе полиномов с целыми коэффициентами // Материалы XIV Международного семинара "Дискретная математика и ее приложения" имени академика О.Б. Лупанова (Москва, МГУ им. М.В. Ломоносова, 2022 г.) М.: ИПМ им. Келдыша, 2022. С. 121-123.

2. Алексиадис Н.Ф. О рациональных А-функциях с рациональными коэффициентами //XIX Международная конференция «Алгебра, теория чисел, дискретная геометрия и многомасштабное моделирование: современные проблемы, приложения и проблемы истории», посвященная двухсотлетию со дня рождения академика П. Л. Чебышева - Тула, 2021. С. 97-101.
3. Алексиадис Н. Ф. Рациональные А-функции с рациональными коэффициентами. 2022. // Чебышевский сборник. - Том 23. Выпуск 3С. 2-9. DOI 10.22405/2226-8383-2022-23-3-2-10.
4. Алексиадис Н.Ф. О проблеме полноты для функциональной системы рациональных функций с рациональными коэффициентами (алгоритмический подход) // Труды XXIII Международной конференции «Математическое моделирование и суперкомпьютерные технологии» – Нижний Новгород, 2023. С. 8-12.
5. Алексиадис Н. Ф. Задача о базисах в функциональной системе рациональных функций с рациональными коэффициентами (алгоритмический подход) //Труды XXIV Международной конференции «Математическое моделирование и суперкомпьютерные технологии» – Нижний Новгород, 2024. С. 3-7.
6. Алексиадис Н.Ф. О проблеме полноты рациональных функций с рациональными коэффициентами // Тезисы Международной конференции "Мальцевские чтения". Новосибирск, 2021. С. 143.
7. Алексиадис Н.Ф. О базисах рациональных функций с рациональными коэффициентами // Материалы XIX Международной конференции “Алгебра, теория чисел, дискретная геометрия и многомасштабное моделирование: современные проблемы, приложения и проблемы истории”, посвященной 130-летию со дня рождения академика И. М. Виноградова - Тула, 2021. С. 77-80.
8. Алексиадис Н.Ф. Об аналоге теоремы Колмогорова о суперпозициях непрерывных функций для функциональных систем полиномиальных и рациональных функций // Материалы XXII Международной конференции «Алгебра, теория чисел, дискретная геометрия и многомасштабное моделирование: современные проблемы, приложения и проблемы истории», посвященной 120-летию со дня рождения академика А. Н. Колмогорова – ТГПУ им. Л. Н. Толстого, 2023. С. 140-144.
9. Aleksiadis N.F. Bases of complete systems of rational functions with rational coefficients. Chebyshevskii Sbornik. 2023; 24(2):5-14.DOI 10.22405/2226-8383-2023-24-2-5-14.
10. Aleksiadis N. Ph., 2024. “An analogue of Kolmogorov’s theorem on superpositions of continuousfunctions for functional systems of polynomial and rational functions”, Chebyshevskii sbornik,vol. 25, no. 2, pp. 29-42. DOI 10.22405/2226-8383-2024-25-2-29-42.
11. Кудрявцев В.Б. Функциональные системы. – М.: Изд--во МГУ, 1982. 157 с.
12. Яблонский С.В. Введение в дискретную математику. М.: Изд--во Наука, 1986. 384 с.

ГИБРИДНЫЙ МЕТОД ВРЕМЕННОГО СУПЕРРАЗРЕШЕНИЯ С НЕЙРОСЕТЕВОЙ ПОСТОБРАБОТКОЙ КАДРОВ В РЕАЛЬНОМ ВРЕМЕНИ

А.А. Бердюгин, В.Е. Турлапов

Нижегородский государственный университет им. Н.И. Лобачевского

При синтезе фотореалистичных изображений высокого разрешения возникает проблема достижения реального времени для 3D сцен реальной сложности. Сегодня с увеличением размеров экранов проблема реального времени становится неразрешимой из-за высокой трудоемкости даже глубоко оптимизированных процедур трассировки лучей (RT) и глобального освещения (GI). Приемлемым ответом на эту проблему стало исполнение синтеза изображения (рендеринга) в более низком разрешении с последующим повышением разрешения до нативного. Мы предлагаем новый подход, работающий в среде операционной системы macOS. Идея заключается в реализации компактной U-Net-подобной модели для улучшения качества изображений в комбинации с Unreal temporal super resolution.

Ключевые слова: компьютерная графика, нейронные сети, гибридный рендеринг, улучшение качества изображения

1. Введение

Свет в реальном мире взаимодействует с окружающим пространством и поверхностями, создавая отражения и каустики. Спрос на высоко реалистичные системы рендеринга в реальном времени, основанные на физически корректных моделях освещения, более распространен, чем когда-либо [1], однако многие процессы настолько вычислительно затратны, что требуют разработки упрощённых моделей и индивидуальных алгоритмов [2]. Для решения данной проблемы было предложено множество методов: адаптивный метод, позволяющий изменять качество графики в зависимости от текущей нагрузки на компьютерную систему [3], метод отражения в пространстве экрана (SSR), использующий в отражениях только те объекты, которые видит виртуальная камера [4], гибридные конвейеры, объединяющие трассировку лучей с растеризацией [5], фовеальный рендеринг, использующий более низкое разрешение на перефирии поля зрения человека. [6]. Ключевым направлением является восстановление изображения с высоким разрешением из его версии с низким разрешением, что в свою очередь включает non-neural temporal upsampling [7-10], а также нейросетевые методы [11-17], в том числе использующие реконструкцию лучей и генерацию промежуточных кадров [18].

Хотя существуют различные подходы, продемонстрировавшие значительное увеличение частоты кадров при умеренных потерях качества изображения, и методы апскейлинга изображения в реальном времени, основанные на нейросетях, но в данный момент не наблюдается их широкого применения в среде операционной системы macOS. В данной статье предложен метод, совмещающий в себе компактную U-Net подобную модель с процедурой с temporal super resolution (далее TSR), встроенной в игровой движок unreal engine.

В ходе разработки мы обнаружили, что ни один из существующих датасетов не удовлетворяет потребность в решении нашей задачи, поэтому был создан новый датасет, который позволяет обучить нейронную сеть исправлять характерные артефакты и потери деталей при применении TSR, а также выполнить сравнение изображений до её применения и после.

Таким образом, мы представляем:

— уникальный датасет для решения проблем метода TSR;
— компактную U-Net-подобную модель, работающую в реальном времени, специально разработанную для устранения недостатков и артефактов TSR;

— гибридный подход, совмещающий в себе компактную U-Net подобную модель, работающую в комбинации с TSR для рендеринга видеоигр в реальном времени в среде операционной системы macOS.

2. Описание метода

2.1. Архитектура модели

U-Net подобная нейросетевая модель состоит из энкодера, bottleneck и декодера, с использованием skip-связей для сохранения деталей. Входные данные представляют собой изображения 1920x1080 отрендеренные с помощью процедуры TSR, работающей в режиме 50%, т.е. изображения размером 1920x1080, получившиеся в результате апскейла с 960x540 до 1920x1080. В режиме обучения для ускорения и облегчения процесса обучения из данных изображений вырезается участок, выбранный случайным образом, размером 256x256. Для валидации используется изображение целиком.

Энкодер содержит три уровня. Первый уровень применяет блок свёрток с 3 входными каналами (RGB) и 96 выходными каналами. Блок свёрток состоит из двух последовательных свёрток с ядром 3x3, дополненных активацией ReLU, которая обнуляет отрицательные значения для усиления нелинейности, и padding=1, добавляющий один пиксель по краям, чтобы сохранить размер изображения. Последующие два уровня используют модуль Down, который увеличивает количество каналов, используя свёртку до 192 и 384 каналов соответственно и уменьшает изображение в 2 раза по ширине и длине соответственно, используя усредняющий пулинг, берущий среднее значение пикселей в окне 2x2.

Bottleneck применяет блок свёрток с 384 входными и 768 выходными каналами, обеспечивая глубокую обработку признаков.

Декодер состоит из трёх уровней, использующих модуль Up. Каждый Up выполняет билинейный upsampling, увеличивающий каждый линейный размер в 2 раза, объединяет данные с соответствующим уровнем энкодера посредством skip-связей (конкатенация каналов энкодера и декодера) и применяет свёртку, уменьшая число каналов (768+384→384, 384+192→192, 192+96→96). Финальная свёртка (96→3) формирует "дельту", которая добавляется к входному изображению, с ограничением значений в диапазоне [0, 1] для получения RGB-выхода.

2.2. Обучение

2.2.1 Функция потерь

Функцией потерь является комбинация L1, MS-SSIM и TV. Данный выбор основывается на статье Zhao и др. [19] исходя из которой видно, что совместная работа L1 и MS-SSIM приводит к наибольшему росту рассматриваемых в статье метрик, в их числе PSNR, SSIM, которые являются наиболее часто встречающимися в подобных работах. Для повышения устойчивости обучения и улучшения визуального качества восстанавливаемых изображений в нашей работе мы используем регуляризацию на основе Total Variation (TV) в составе комбинированной функции потерь. Несмотря на то, что TV не всегда напрямую улучшает численные метрики качества, её основное преимущество заключается в подавлении локальных шумов и артефактов, что способствует получению более гладких и естественных изображений. Аналогичный подход продемонстрирован в работе Yan и др. [20], где TV применяется совместно с L1 для улучшения качества изображений в задаче повышения визуального качества изображений. Таким образом, совокупная функция потерь сочетает три взаимодополняющих критерия – L1 для точности на уровне пикселей, MS-SSIM для сохранения структурных характеристик (яркость, контрастность) и Total Variation (TV) для подавления мелких шумов и артефактов. Формула выглядит следующим образом:

$$L_{total} = (w_{L1} * L1) + \left(w_{MS\text{-}SSIM} * (1 - MS - SSIM)\right) + (w_{TV} * TV), \quad (1)$$

где $w_{L1} = 1$, $w_{MS\text{-}SSIM} = 0{,}5$ $w_{TV} = 0{,}001$.

2.2.2 Датасет

В связи с тем, что на текущий момент отсутствуют датасеты, разработанные для устранения артефактов процедуры TSR в среде игрового движка Unreal Engine был сформирован уникальный синтетический датасет, включающий 463 пары кадров разрешением 1080p, всего 926 кадров. Каждая пара состоит из нативного изображения (1080p) и соответствующего TSR-изображения, отрендеренного с разрешением 540p апскейленного до 1080p. Кадры представляют собой игровую сцену, содержащию ray-traced освещение, снятую под разными ракурсами.

2.2.3 Параметры обучения

Для обучения из изображения на каждой эпохе вырезается случайный участок из каждого изображения тестового набора, выбранный случайным образом, размер патча - 256x256. Для валидации используется полное изображение, не использованное в обучении. Кадры делятся в пропорции 80/10/10 (обучающий/валидационный/тестовый набор).

Модель обучается с использованием оптимизатора Adam[21], который зарекомендовал себя в работе Xiao и др.[11], с начальной скоростью обучения 2e-4 и scheduler StepLR, уменьшающего скорость обучения в 2 раза каждые 20 эпох для точной сходимости. Градиенты ограничиваются нормой 1.0 для стабильности. Во время обучения результаты отслеживаются на валидационном наборе, для оценки используются следующие метрки: PSNR, SSIM, NR-IQA[22]. Размер батча – 6, количество эпох – 430.

3. Результаты

Для оценки эффективности предложенного метода был проведён эксперимент на тестовом наборе, состоящем из 46 пар изображений разрешением 1080p, включающих нативные кадры (ground truth) и кадры, отрендеренные с использованием Unreal Engine Temporal Super-Resolution (TSR) в режиме 50% (540p, апскейленные до 1080p). Обучение проводилось на Apple MacBook Air M1 (16GB 7-core GPU), используя Metal Performance Shaders (MPS), в среде операционной системы macOS и заняло около двух дней. Качество изображений оценивалось с использованием традиционных метрик PSNR, SSIM, однако такие метрики не всегда коррелируют с субъективным качеством, воспринимаемым человеком, поэтому дополнительно была добавлена безреференсная метрика (NR-IQA) на основе модели MANIQA, описанная в работе Yang и др. [22]. Предложенный метод сравнивался с базовым вариантом TSR (без нейросетевой обработки) и нативными изображениями (эталон качества).

Результаты тестирования представлены в таблицах 1-2. Предложенный метод демонстрирует значительное улучшение качества по сравнению с базовым TSR. Среднее значение PSNR для нашего метода составляет 39.18 дБ, что на 3.18 дБ выше, чем у TSR (36.00 дБ). Это указывает на меньшую пиксельную ошибку и более высокую точность восстановления. Среднее значение SSIM достигает 0.9846, что на 0.0082 выше, чем у TSR (0.9764), демонстрируя улучшение структурного сходства, включая яркость, контраст и текстуры. Среднее значение NR-IQA для составляет 0.4705, что на 0.0136 выше, чем у TSR (0.4569), указывая на лучшее воспринимаемое качество изображения без референса. Однако NR-IQA для нативных изображений (0.4796) немного выше, что ожидаемо, так как они являются эталоном качества.

Таблица 1. Сравнение методов по метрикам PSNR, SSIM

Метод	PSNR (дБ)	SSIM
Наш метод	39,18	0,9846
Unreal TSR (50%)	36,00	0,9764

Таблица 2. Сравнение методов по метрике NR-IQA

Метод	NR-IQA
Наш метод	0,4705
Unreal TSR (50%)	0,4569
Нативное изображение	0,4796

Среднее время инференса составляет 6,48 мс на кадр (примерно 154 FPS), что подтверждает возможность применения модели для работы в реальном времени на устройствах MacBook Air M1. Время инференса варьируется от 3.64 мс до 13.54 мс в зависимости от сложности сцены, что подтверждает эффективность компактной архитектуры при минимальных вычислительных затратах. Первый кадр показал аномально высокое время (401.74 мс) из-за инициализации MPS, что исключено из расчёта среднего времени.

Базовый TSR (без нейросетевой обработки) демонстрирует заметные артефакты, что отражается в более низких значениях PSNR (36.00 дБ) и SSIM (0.9764). Предложенный метод устраняет эти артефакты, приближая качество к нативным изображениям, что изображено на рисунке 1.



Метрика	Unreal TSR	Ours	Native
PSNR	35,57	39,16	-
SSIM	0,9750	0,9848	-
NR-IQA	0,537	0,549	0,557

Рис. 1. Сравнение качества изображения

4. Заключение

Предложенный метод нейросетевой постобработки кадров, полученных с помощью Temporal Super-Resolution (TSR) в Unreal Engine, обеспечивает заметное повышение визуального качества по сравнению с базовым TSR. Улучшения подтверждаются как объективными метриками (PSNR, SSIM), так и безреференсной оценкой восприятия (NR-IQA). Модель демонстрирует высокую скорость инференса, что делает её применимой в реальном времени даже на энергоэффективных устройствах. Таким образом, предложенный подход сочетает высокое качество реконструкции с низкими вычислительными затратами и может быть интегрирован в существующие графические конвейеры.

Литература

1. Polychronakis A., Koulieris G. A., Mania K. Emulating foveated path tracing //Proceedings of the 14th ACM SIGGRAPH Conference on Motion, Interaction and Games. – 2021. – С. 1-9.
2. Ritschel T. et al. The state of the art in interactive global illumination //Computer graphics forum. – Oxford, UK : Blackwell Publishing Ltd, 2012. – Т. 31. – №. 1. – С. 160-188.
3. Гонахчян, В.И. Адаптивная стратегия рендеринга динамических трехмерных сцен : диссертация ... кандидата технических наук : 05.13.11 / Гонахчян Вячеслав Игоревич; [Место защиты: ФГБУН Институт системного программирования им. В.П. Иванникова Российской академии наук]. – Москва, 2021. – 150 с.

4. Hofmann N. et al. Hierarchical multi-layer screen-space ray tracing //Proceedings of High Performance Graphics. – 2017. – С. 1-10.
5. Barré-Brisebois C. et al. Hybrid rendering for real-time ray tracing //Ray Tracing Gems: High-Quality and Real-Time Rendering with DXR and Other APIs. – Berkeley, CA : Apress, 2019. – С. 437-473.
6. Bauer D., Wu Q., Ma K. L. Fovolnet: Fast volume rendering using foveated deep neural networks //IEEE transactions on visualization and computer graphics. – 2022. – Т. 29. – №. 1. – С. 515-525.
7. Lei Yang, Diego Nehab, Pedro V Sander, Pitchaya Sitthi- Amorn, Jason Lawrence, and Hugues Hoppe. Amortized supersampling. ACM Transactions on Graphics (TOG), 28(5):1–12, 2009.
8. Brian Karis. High-quality temporal supersampling. Advances in Real-Time Rendering in Games, SIGGRAPH Courses, 1(10.1145):2614028–2615455, 2014.
9. Robert Herzog, Elmar Eisemann, Karol Myszkowski, and H P Seidel. Spatio-temporal upsampling on the gpu. In Pro- ceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games, pages 91–98, 2010.
10. Epic Games. Unreal engine 4.19: Screen percentage with temporal upsample, 2018.
11. Xiao L. et al. Neural supersampling for real-time rendering //ACM Transactions on Graphics (TOG). – 2020. – Т. 39. – №. 4. – С. 142: 1-142: 12.
12. Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1874–1883.
13. Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. 2017. Real-time video super-resolution with spatio- temporal networks and motion compensation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 4778–4787.
14. Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In Proceedings of the 27th acm interna- tional conference on multimedia, pages 2024–2032, 2019.
15. Jie Liu, Jie Tang, and Gangshan Wu. Residual feature dis- tillation network for lightweight image super-resolution. In Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, pages 41–55. Springer, 2020.
16. Hisham Chowdhury, Rense Robert Kawiak, Gabriel Ferreira de Boer, and Lucas Xavier. Intel xess-an ai based super sam- pling solution for real-time rendering.(2022). In Game Developers Conference, 2022.
17. Davide Abati, Amir Ghodrati, Amirhossein Habibian, and Qualcomm AI Research. Efficient video super resolution by gated local self attention. In 2021 British Machine Vision Conference (BMVC), volume 2, page 3, 2021.
18. Andrew Burnes. NVIDIA DLSS 3.5: Enhancing Ray Tracing With AI. -2023. - https://www.nvidia.com/en-us/geforce/news/nvidia-dlss-3-5-ray-reconstruction
19. Zhao H. et al. Loss functions for image restoration with neural networks //IEEE Transactions on computational imaging. – 2016. – Т. 3. – №. 1. – С. 47-57.
20. Yan J. et al. Lepf-net: Light enhancement pixel fusion network for underwater image enhancement //Journal of Marine Science and Engineering. – 2023. – Т. 11. – №. 6. – С. 1195.
21. Kingma D. P. Adam: A method for stochastic optimization //arXiv preprint arXiv:1412.6980. – 2014.
22. Yang S. et al. Maniqa: Multi-dimension attention network for no-reference image quality assessment //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2022. – С. 1191-1200.

ЦИКЛОПНЫЕ СОСТОЯНИЯ В СЕТЯХ НЕИДЕНТИЧНЫХ ОСЦИЛЛЯТОРОВ СО СВЯЗЬЮ ПО ВЫСШИМ МОДАМ[1*]

М.И. Болотов[1], Л.А. Смирнов[1], В.О. Муняев[1], Г.В. Осипов[1], И.В. Белых[2]

[1]Нижегородский государственный университет им. Н.И. Лобачевского,
[2]Университет штата Джорджия, США

Обычно неоднородность разрушает синхронные режимы в больших сетях осцилляторов. В данной работе мы показываем, что неидентичность осцилляторов, напротив, может стабилизировать сложные структуры синхронизации – циклопные и кластерные состояния, – неустойчивые в системах идентичных элементов. Этот эффект объясняется с помощью мезоскопического подхода, связывающего микроскопическую неодродность сети с макроскопической устойчивостью структур. Универсальность механизма подтверждается его реализацией в биологически и физически релевантных моделях Уинфри и Стюарта–Ландау, что открывает перспективы управления сложной коллективной динамикой в неоднородных сетях.

Ключевые слова: беспорядок, неидентичные элементы, фазовый осциллятор, модель Курамото, старшие моды, циклопный режим, кластеры, стабилизация.

1. Введение

Модель Курамото–Сакагучи второго порядка со связью по старшим модам является стандартной моделью для исследования сложных режимов синхронизации, в частности, циклопных состояний. Данный режим представляет собой сосуществование двух когерентных кластеров и одного уединенного осциллятора [1, 2]. Хотя в сетях идентичных осцилляторов стационарные варианты циклопных состояний, как правило, неустойчивы, мы демонстрируем, что введение неоднородности собственных частот обеспечивает их надёжную стабилизацию.

2. Описание модели

Рассматривается сеть из N глобально связанных неидентичных осцилляторов Курамото–Сакагучи второго порядка:

$$m\ddot{\theta}_n + \dot{\theta}_n = \omega_n + \frac{1}{N}\sum_{k=1}^{N}\sum_{q=1}^{2}\varepsilon_q \sin\big(q(\theta_k - \theta_n) - \alpha_q\big), \qquad (1)$$

где θ_n – фаза, m – параметр инерции, ω_n – собственная частота, ε_q и α_q – амплитуда и фазовый сдвиг q-й гармоники в функции связи.

В случае идентичных осцилляторов ($\omega_n \equiv \omega$), когда идеально симметричное циклопное состояние является неустойчивым, введение неоднородности частот (например, равномерное распределение ω_n в интервале $[-\nu, \nu]$) превращает это неустойчивое состояние в устойчивый аттрактор. Механизм стабилизации обусловлен малыми вариациями фаз внутри кластеров, вызванными неоднородностью. Эти вариации создают компенсирующий эффект, который позволяет средней фазе кластера удовлетворить условию баланса по отношению к изолированному осциллятору. Важно, что данный эффект наблюдается в широком диапазоне неоднородности ν и не требует тонкой настройки параметров.

Для объяснения этого явления строится мезоскопическая модель. Вводится линейный анзац для фазы относительно изолированного осциллятора ($\varphi_n = \theta_n - \theta_N$):

[1*] Работа выполнена при поддержке Российского научного фонда (проект № 24-72-00105).

$$\varphi_n(t) \approx \psi_\mu(t) + \chi_\mu(t)\Delta_n, \tag{2}$$

для осцилляторов в кластере μ ($\mu = 1, 2$), где ψ_μ – средняя фаза кластера μ, χ_μ описывает фазовый дрейф внутри кластера, а $\Delta_n = \omega_n - \omega_N$ – расстройка частот.

С помощью проекционного метода Галёркина высокоразмерная микроскопическая динамика сводится к восьмимерной системе относительно коллективных координат $\psi_1, \psi_2, \chi_1, \chi_2$:

$$m\ddot{\psi}_\mu + \dot{\psi}_\mu = \sum_{q=1}^{2} \mathrm{Im}\left[\frac{\varepsilon_q e^{-i\alpha_q}}{N}\left(1 + K S_1^{(q)} e^{iq\psi_1} + K S_2^{(q)} e^{iq\psi_2}\right)\left(\frac{\sigma_\mu^2 S_\mu^{(q)*} - \delta_\mu J_\mu^{(q)*}}{\sigma_\mu^2 - \delta_\mu^2} e^{-iq\psi_\mu} - 1\right)\right], \tag{3}$$

$$m\ddot{\chi}_\mu + \dot{\chi}_\mu = 1 + \frac{1}{\sigma_\mu^2 + \delta_\mu^2} \sum_{q=1}^{2} \mathrm{Im}\left[\frac{\varepsilon_q e^{-i\alpha_q}}{N}\left(1 + K S_1^{(q)} e^{iq\psi_1} + K S_2^{(q)} e^{iq\psi_2}\right)\left(J_\mu^{(q)*} - \delta_\mu S_\mu^{(q)*}\right) e^{-iq\psi_\mu}\right], \tag{4}$$

где K – размер кластера; δ_μ, σ_μ^2, $S_\mu^{(q)}$, и $J_\mu^{(q)}$ – это, соответственно, средняя по кластеру расстройка частоты, её второй момент, q-й параметр порядка для фазовых отклонений и q-й взвешенный по частоте параметр порядка.

3. Результаты

Стационарные решения редуцированной системы с высокой точностью предсказывают существование и устойчивость циклопных состояний в полной сети (1). Анализ собственных значений редуцированной системы (3)–(4) демонстрирует количественное согласие с результатами численного моделирования исходной системы (1).

Установлено, что выявленный механизм стабилизации, индуцированной неоднородностью, имеет универсальный характер и также наблюдается в сетях осцилляторов Уинфри и Стюарта-Ландау.

Таким образом, выявлен контринтуитивный механизм стабилизации, при котором неоднородность частот индуцирует устойчивые циклопные состояния. Полученные уравнения для коллективных координат образуют эффективный мезоскопический подход, который прямо связывает микроскопическую неоднородность параметров с макроскопической устойчивостью сложных кластерных структур в сетях различной природы.

Литература

1. Munyayev V.O., Bolotov M.I., Smirnov L.A., Osipov G.V., Belykh I. Cyclops States in Repulsive Kuramoto Networks: The Role of Higher-Order Coupling // Physical Review Letters. 2023. Vol. 130, No. 10. P. 107201. DOI: 10.1103/PhysRevLett.130.107201.
2. Bolotov M.I., Munyayev V.O., Smirnov L.A., Osipov G.V., Belykh I. Breathing and switching cyclops states in Kuramoto networks with higher-mode coupling // Physical Review E. 2024. Vol. 109, No. 5. P. 054202. DOI: 10.1103/PhysRevE.109.054202.

ОПТИМИЗАЦИЯ АЛГОРИТМОВ УМНОЖЕНИЯ РАЗРЕЖЕННОЙ МАТРИЦЫ НА ВЕКТОР ДЛЯ ПРОЦЕССОРОВ АРХИТЕКТУРЫ RISC-V[1*]

А.А. Воденеева, В.Д. Волокитин, М.А. Загрядсков, К.И. Ковалев, Е.А. Козинов, А.И. Кулик, Д.Д. Литвяков, А.Ю. Пирова, А.В. Устинов, И.Б. Мееров

[1]Нижегородский государственный университет имени Н.И. Лобачевского

В статье описываются промежуточные результаты разработки библиотеки для выполнения алгоритма умножения разреженной матрицы на вектор (SpMV), оптимизированной для процессоров архитектуры RISC-V и поддерживающей несколько форматов хранения разреженной матрицы. Алгоритм SpMV является одним из ключевых алгоритмов линейной алгебры и основным вычислительно трудоемким ядром итерационных решателей СЛАУ с разреженными матрицами, которые, в свою очередь, применяются для решения широкого спектра задач математической физики с использованием суперкомпьютеров. Программная реализация выполнена с использованием векторных инструкций RVV 1.0. Результаты предварительных экспериментов показали, что использование векторизации и подходящего формата хранения позволяет сократить время выполнения операции SpMV в 1.5–19 раз по сравнению с использованием традиционного формата CRS.

Ключевые слова: разреженные матрицы, разреженная алгебра, RISC-V, SpMV.

1. Введение

Известно, что алгоритм умножения матрицы на вектор (SpMV) является основным вычислительно-трудоемким ядром итерационных решателей систем линейных уравнений (СЛАУ) с разреженными матрицами, которые применяются для решения широкого спектра задач математической физики, прикладной механики, химии, а также для высокопроизводительных реализаций алгоритмов на графах. На сегодняшний день при кажущейся простоте алгоритма матрично-векторного умножения до сих пор не решены проблемы с эффективным представлением разреженных матриц, которое позволило бы раскрыть потенциал современных вычислительных систем.

Для традиционных архитектур x86-64, ARM и графических процессоров предложено немало новых форматов, позволяющих увеличить производительность операции SpMV за счет векторизации вычислений и улучшения локальности доступа к данным. Среди новых, современных форматов отметим SELL-C, SELL-C-σ [1, 2], CRS5 [3], CRS2 [4], VNEC [5], показывающие высокую производительность на матрицах из разных прикладных областей, а также форматы LAV [6], CVR [7], VHCC [8], VBFS [9], ориентированные, в первую очередь, на матрицы специальной структуры и безмасштабные графы. В большом числе работ исследованы блочные и гибридные версии традиционных форматов COO, CRS, ELL. Подробный обзор форматов хранения разреженных матриц можно найти, например, в [10].

В настоящее время наиболее распространенные библиотеки разреженной линейной алгебры, оптимизированные для процессоров архитектуры x86 (Intel OneAPI MKL, AOCL-Sparse, SuiteSparse, Trilinos и др.), до сих пор не используют перечисленные новые форматы хранения матриц, ограничиваясь классическими форматами CRS, CSC, COO, DIA, BCRS. Для процессоров архитектуры RISC-V нет общепринятой оптимизированной реализации алгоритмов разреженной линейной алгебры, а эффективность решений, предложенных ранее для процессоров x86, не изучена.

[1*] Работа поддержана в рамках программы стратегического академического лидерства «Приоритет-2030», соглашение Министерства науки и высшего образования РФ №075-15-2025-217 от 04.04.2025.

В данной работе описываются промежуточные результаты работы по разработке библиотеки, предназначенной для выполнения алгоритма умножения разреженной матрицы на вектор, поддерживающей различные форматы хранения больших разреженных матриц и оптимизированной для процессоров архитектуры RISC-V. Данная библиотека может быть полезна для адаптации под процессоры RISC-V инженерных и научных приложений, приложений для работы с графами, а также может внести вклад в изучение вопроса о перспективных наборах команд для вычислений с разреженными матрицами в процессорах RISC-V.

2. Программная реализация

2.1. Методика выбора форматов хранения разреженных матриц

Для сравнения производительности новых форматов хранения разреженных матриц и выбора наиболее предпочтительных из них для реализации на процессорах архитектуры RISC-V, была проведена серия предварительных вычислительных экспериментов на процессоре архитектуры x86-64. В качестве референсной реализации был выбран алгоритм SpMV для формата CRS. Для тестирования были взяты открытые программные реализации алгоритма SpMV для нескольких форматов хранения и проведены замеры времени их работы для однократного выполнения операции SpMV. Идея основывалась на предположении, что те форматы хранения, которые покажут наилучшие результаты на этой архитектуре, будут наиболее перспективны и для архитектуры RISC-V. По итогам тестирования для разработки были выбраны следующие форматы хранения разреженных матриц: SELL-C-σ [2], CSR5 [3], CSR2 [4], VNEC [5] и LAV [6], и соответствующие им алгоритмы матрично-векторного умножения.

2.2. Краткое описание форматов хранения разреженных матриц

Приведем краткое описание форматов хранения разреженных матриц, использованных в программной реализации. Все эти форматы ориентированы на процессоры, поддерживающие векторизацию с использованием SIMD-инструкций.

SELL-C-σ [2] – гибрид ELL и CSR, предложенный для эффективной векторизации матрично-векторного умножения. Идея – сортировка строк по убыванию числа ненулевых элементов в пределах σ строк (локальное выравнивание), разделение матрицы на полосы по C строк и выделение каждой полосе памяти по длине самой длинной строки в этой полосе. Формат реализует отличную векторизацию и кэш-локальность и показывает хорошую производительность при правильно выбранном σ.

CSR5 [3] – модификация формата CSR. Идея формата – разбить матрицу на блоки фиксированного размера $C \times \sigma$, где σ – среднее число ненулевых элементов на строку. CSR5 обеспечивает регулярный доступ внутри блока, вследствие чего достигается лучшая производительность, чем в CSR. Балансировка нагрузки между потоками реализуется автоматически. Алгоритм матрично-векторного умножения для CSR5 хорошо масштабируется на архитектуре x86-64.

CSR2 [4] – формат, являющийся модификацией CRS. В данном формате информация о ненулевых элементах матрицы хранится в трех массивах, содержащих ненулевые элементы, номера столбцов этих элементов и индексы начала строк. При этом массивы ненулевых элементов и номеров их столбцов сгруппированы в блоки, размер которых кратен длине векторного регистра. Данные, принадлежащие одному блоку, лежат рядом. За счет этого в CSR2 реализуется более эффективная векторизация SpMV и более эффективное использование кэша.

VNEC [5] – адаптивный формат, предназначенный для неравномерно разреженных матриц, плотность которых сильно варьируется. Идея VNEC – разбить матрицу на блоки с переменным числом ненулевых элементов, и для каждого блока хранить собственную локальную структуру (CSR, COO или ELL). Дополнительно хранятся указатели на начало каждого блока и данные о том, какой формат в нем используется. VNEC адаптирован к сильно неравномерным матрицам и показывает высокую производительность при смешанном распределении плотности.

LAV [6] – формат, ориентированный на улучшение локальности при сохранении компактности представления данных. Идея формата – разделение входной матрицы на две части.

Первая часть, называемая плотной, содержит большинство ненулевых элементов. Вторая часть, называемая разреженной, включает оставшиеся ненулевые элементы. Опе-рации над плотной и разреженной частями матрицы осуществляют доступ к непересекающимся подмножествам элементов входного вектора. Плотная часть матрицы зависит от длины векторного регистра и размера кэша. Разреженная часть хранится в обычном CRS формате.

2.3. Программная реализация библиотеки

Программная реализация библиотеки выполнена на языке C++. Для каждого формата хранения разреженной матрицы был разработан отдельный шаблонный класс, функция конвертации матрицы из формата CRS в данный формат и функция умножения матрицы на вектор с использованием векторных инструкций RVV 1.0. Поддерживаются вычисления в одинарной и двойной точности. Распараллеливание выполнено с использованием технологии OpenMP. Дополнительно была реализована тестовая система с использованием библиотек Google Test и Google Benchmark, позволяющая проверить корректность вычислений и провести замеры производительности на большом объеме тестовых данных.

3. Результаты вычислительных экспериментов

Для апробации разработанной библиотеки были проведены вычислительные эксперименты сравнения времени однократного выполнения операции SpMV для всех реализованных форматов хранения разреженных матриц. Для экспериментов было выбрано 54 матрицы из коллекции SuiteSparse [11] размером от 200 тысяч до 5 млн строк.

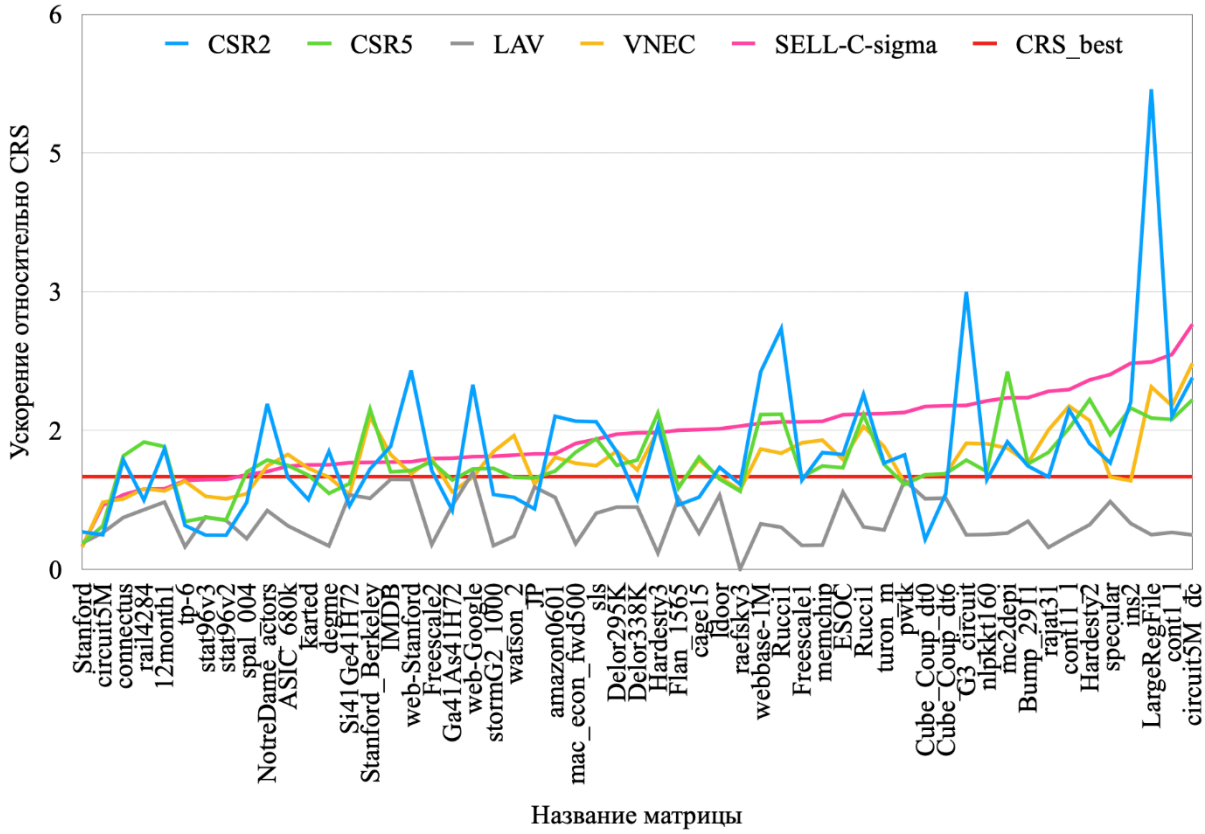


Рис. 1. Ускорение операции SpMV для различных форматов хранения разрежен-ных матриц относительно формата CRS при запуске в 1 поток. По вертикали – величина ускорения, по горизонтали – название тестовой матрицы

Все эксперименты проводились на тестовой системе со следующими техническими характеристиками: плата Banana Pi BPI-F3 с процессором SpacemiT Keystone K1 (8 ядер SpacemiT x60 с частотой 1.6 ГГц и 8-ступенчатым in-order конвейером, профилем RVA22 с поддержкой

256-битного RVV 1.0), 16 GB оперативной памяти и операционной системой Bianbu 1.0.15. Запуски проводились в 1, 2, 4 и 8 потоков, в одинарной и двойной точности. Для сравнения форматов между собой использовалось минимальное из нескольких запусков время выполнения операции SpMV.

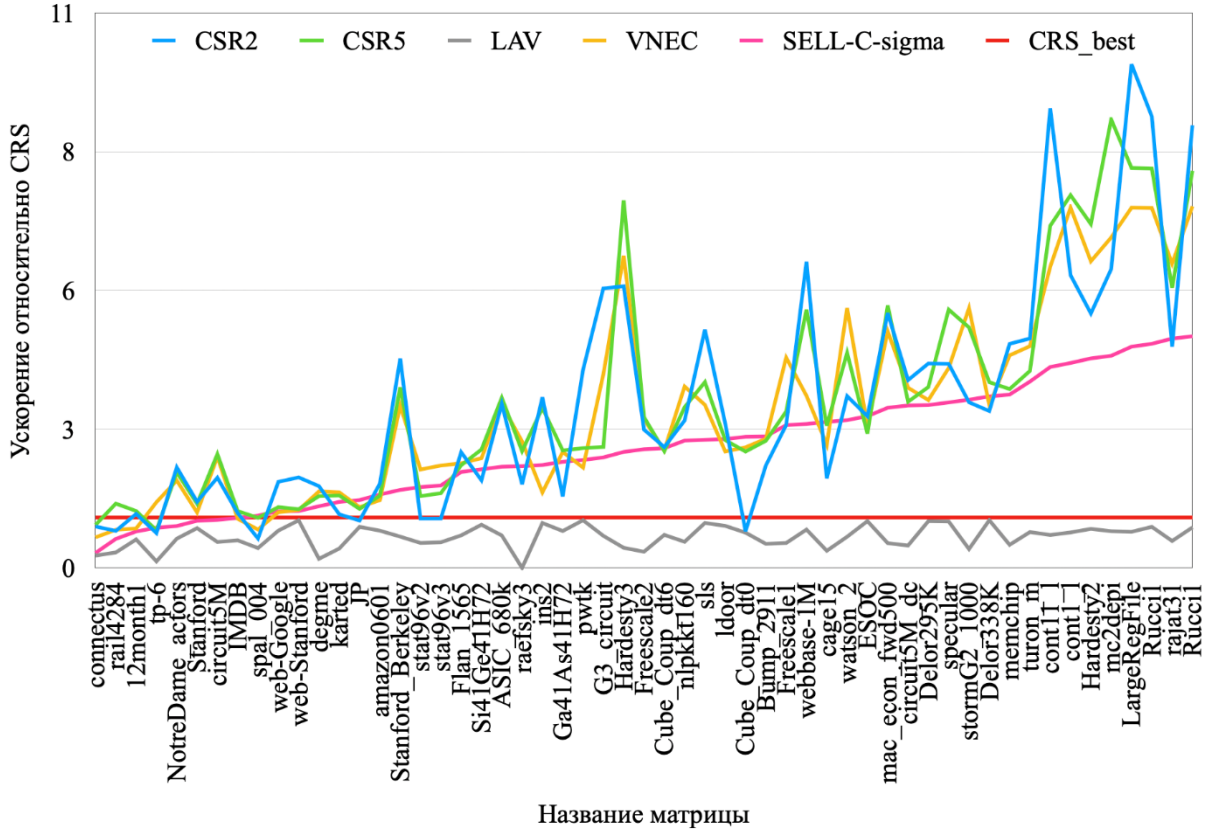


Рис. 2. Ускорение операции SpMV для различных форматов хранения разрежен-ных матриц относительно формата CRS при запуске в 4 потока. По вертикали – величина ускорения, по горизонтали – название тестовой матрицы

Результаты предварительных экспериментов показали, что в параллельной версии для большинства форматов время работы операции SpMV при использовании 4 потоков меньше, чем при использовании 8 потоков, поэтому далее приведены графики времени работы только последовательной версии и параллельной версии в 4 потока.

Результаты экспериментов представлены на рис. 1 и рис. 2. На графиках приведено время работы операции SpMV для каждого из форматов относительно времени работы этой операции для формата CSR, в двойной точности. CRS_best – это наиболее производительная из скалярной и векторной версии формата CSR для каждой матрицы.

В результате экспериментов было получено, что использование предложенных фор-матов позволило ускорить алгоритм матрично-векторного умножения на RISC-V в 1.5–6 раз относительно формата CRS.

Для типа данных double выигрыш новых форматов относительно исходного формата CRS при выполнении операции SpMV встречается почти на всех матрицах в случае работы в 1 поток и на большинстве матриц в случае работы в 4 потока (в 99% и 77% случаев соответственно). Использование CRS2, CRS5, Sell-C-σ и VHCC почти во всех случаях дает ускорение относительно CRS как в последовательном, так и в параллельном запусках. В подавляющем большинстве случаев на всех матрицах для любого числа потоков выигрывает формат SELL-C-Sigma. Тройку лидеров составляют SELL-C-Sigma, CRS5 и CRS2. Количество матриц, на которых их ускорение является наибольшим, составляет 32, 5 и 5 на 1 потоке, и 24, 7 и 5 на 4 потоках соответственно. Остальные форматы дают прирост производительности на определенных типах разреженных матриц, что ожидаемо, так как каждый из них изначально был специализирован под конкретный вид матриц.

4. Заключение

В работе были описаны промежуточные результаты разработки оптимизированной для процессоров архитектуры RISC-V библиотеки форматов хранения разреженных матриц для алгоритма умножения матрицы на вектор. Как видно из результатов экспериментов, достигнутое ускорение на операции матрично-векторного умножения на архитектуре процессора RISC-V относительно стандартного формата хранения разреженных матриц CRS составляет 1.5–6 раз в зависимости от используемого формата хранения. На данном этапе работы наилучшее ускорение было достигнуто при использовании форматов CSR2, CSR5 и SELL-C-$\boldsymbol{\sigma}$.

Ведется дальнейшая работа по улучшению производительности формата LAV и реализации еще нескольких форматов хранения. Проводится анализ типов матриц, на которых несколько форматов стабильно давали замедление или демонстрировали наилучшее ускорение при выполнении операции SpMV. В будущем запланирована интеграция реализуемой библиотеки с одним из итерационных решателей систем линейных уравнений с разреженными матрицами.

Литература

1. Kreutzer M. et al. A unified sparse matrix data format for efficient general sparse matrix-vector multiplication on modern processors with wide SIMD units // SIAM Journal on Scien-tific Computing. – 2014. – Т. 36. – №. 5. – С. C401-C423.
2. Gómez C. et al. Optimizing sparse matrix-vector multiplication in NEC SX-Aurora Vector Engine. – Tech. rep., Technical Report, Barcelona Supercomputing Center, 2020.
3. Liu W., Vinter B. CSR5: An efficient storage format for cross-platform sparse matrix-vector multiplication //Proceedings of the 29th ACM on International Conference on Supercompu-ting. – 2015. – С. 339-350.
4. Bian H. et al. A simple and efficient storage format for SIMD-accelerated SpMV //Cluster Computing. – 2021. – Т. 24. – №. 4. – С. 3431-3448.
5. Wang L. et al. VNEC: A Vectorized Non-Empty Column Format for SpMV on CPUs //2024 IEEE International Parallel and Distributed Processing Symposium (IPDPS). – IEEE, 2024. – С. 14-25.
6. Yesil S. et al. Speeding up SpMV for power-law graph analytics by enhancing locality & vectorization //SC20: International Conference for High Performance Computing, Network-ing, Storage and Analysis. – IEEE, 2020. – С. 1-15.
7. Xie B. et al. CVR: Efficient vectorization of spmv on x86 processors //Proceedings of the 2018 International Symposium on Code Generation and Optimization. – 2018. – С. 149-162.
8. Tang W. T. et al. Optimizing and auto-tuning scale-free sparse matrix-vector multiplication on Intel Xeon Phi //2015 IEEE/ACM International Symposium on Code Generation and Optimization (CGO). – IEEE, 2015. – С. 136-145.
9. Li Y. et al. VBSF: a new storage format for SIMD sparse matrix–vector multiplication on modern processors // Journal of Supercomputing. – 2020. – Т. 76. – №. 3.
10. Gao J. et al. A systematic literature survey of sparse matrix-vector multiplication //arXiv preprint arXiv:2404.06047. – 2024.
11. SuiteSparse Matrix Collection, https://suitesparse-collection-website.herokuapp.com/, дата последнего доступа – 15.10.2025.

МОДЕЛИРОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ МЕТОДА КВАНТОВЫХ ТРАЕКТОРИЙ ДЛЯ ОПТИМИЗАЦИИ ПАРАМЕТРОВ АЛГОРИТМА[1*]

В.Д. Волокитин

Нижегородский государственный университет им. Н.И. Лобачевского

В работе рассматривается задача повышения эффективности метода квантовых траекторий, используемого для моделирования открытых квантовых систем. Стохастический характер моментов квантовых скачков вызывает необходимость синхронизации траекторий и снижает общую производительность вычислений. Для решения этой проблемы предложена модель производительности, описывающая зависимость эффективности алгоритма от параметров системы и свойств эффективного гамильтониана. На основе модели выполняется оценка частоты квантовых скачков и выбирается оптимальное число траекторий на вычислительный узел, что обеспечивает более полное использование ресурсов и уменьшает деградацию производительности. Такой подход позволяет повысить эффективность вычислений без изменения структуры алгоритма и может применяться для автоматической настройки параметров в задачах высокопроизводительного моделирования открытых квантовых систем.

Ключевые слова: модель производительности, метод квантовых траекторий, оценка параметров, высокопроизводительные вычисления, автоматическая настройка параметров

1. Введение

В современном мире использование высокопроизводительного программного обеспечения (ПО) сталкивается с трудностями, связанными с его эффективным запуском на различных аппаратных архитектурах. Разработчики ПО зачастую не могут заранее предусмотреть все сценарии и параметры задач, которые могут возникнуть в процессе эксплуатации их алгоритмов и программ.

Одним из наиболее популярных подходов в настоящее время является построение модели производительности для каждого конкретного алгоритма в составе программы. На основе такой модели можно оперативно подбирать метапараметры алгоритма для решения каждой конкретной задачи, чтобы обеспечить максимально эффективное использование вычислительных ресурсов компьютера.

В работе [1] рассматривается подход итеративного локального поиска оптимальных параметров. В статье [2] предложен метод построения суррогатной модели производительности и подбора параметров с использованием методов машинного обучения. В работе [3] продемонстрирован подход на основе кластеризации задач и применения эффективных конфигураций параметров для каждой группы экземпляров. Этот метод показал высокую эффективность на задачах булевой выполнимости (SAT) и смешанного целочисленного программирования (MIP). В статьях [4, 5] рассматриваются автоматические методы оптимизации гиперпараметров, направленные на повышение эффективности алгоритмов машинного обучения.

В данной работе рассматривается проблема повышения эффективности запуска метода квантовых траекторий. Предлагается модель производительности, которая позволяет выбирать параметры алгоритма таким образом, чтобы обеспечить наиболее полное и рациональное использование вычислительных ресурсов.

[1*] Работа выполнена при поддержке Министерства образования и науки Российской Федерации, проект № FSWR-2023-0034.

2. Метод квантовых траекторий

Ранее в работе [6] рассматривалась модификация метода квантовых траекторий, предназначенная для исследования открытых квантовых систем с кусочно-постоянными операторами. Данный алгоритм позволяет находить решение матричного уравнения Линдблада [7]:

$$\frac{d\rho}{dt} = \mathcal{L}(t)\rho = -i\big[H(t),\ \rho\big] + \mathcal{D}(t)\rho,$$

$$\mathcal{D}(t)\rho = \sum_k \gamma_k \left(A_k(t)\rho A_k^\dagger(t) - \frac{1}{2}\left\{A_k^\dagger(t)A_k(t), \rho\right\}\right)$$

Вместо прямого решения матричного дифференциального уравнения Линдблада, в методе квантовых траекторий интегрируется векторное уравнение с эффективным Гамильтонианом $\tilde{H}(t)$ и случайным процессом квантовых скачков:

$$\frac{d\varphi}{dt} = -i\tilde{H}(t)\varphi,$$

$$\tilde{H}(t) = H(t) - \frac{i}{2}\sum_k^n \left(\gamma_k A_k^\dagger(t)A_k(t)\right)$$

$$\rho = \frac{1}{n}\sum_{j=1}^{n} \frac{\varphi\varphi^*}{\|\varphi\|^2}$$

В работе [8] была предложена модификация метода, в которой различные траектории объединялись в группы, и для них выполнялась единая операция интегрирования с использованием экспоненциального оператора. При таком подходе ключевым вычислительным ядром алгоритма становится не операция умножения матрицы на вектор (GEMV), как ранее, а умножение матрицы на матрицу (GEMM). Это позволяет значительно лучше задействовать доступные вычислительные ресурсы современных архитектур. Было показано, что предложенный метод обеспечивает ускорение вычислений почти в 30 раз по сравнению с исходным алгоритмом.

Предложенный подход относится к семейству методов Монте-Карло и обладает высоким уровнем параллелизма. Он демонстрирует почти идеальную масштабируемость не только в пределах одного вычислительного узла, но и при использовании крупных кластерных систем. При этом метод требует вычисления большого числа квантовых траекторий для получения статистически достоверных результатов, которые могут быть использованы в последующих исследованиях.

3. Модель производительности

Основная проблема и ключевая особенность рассматриваемого метода квантовых траекторий заключается в определении момента скачка траектории, который является случайным во времени. Из-за этой стохастической природы процесса для синхронизации времени интегрирования отдельных траекторий приходится выполнять вычисления на небольших группах, что в определённые моменты работы алгоритма приводит к деградации производительности до уровня обычной операции GEMV. В целом, несмотря на высокую эффективность операции GEMM (около 97–99% на современных процессорах), суммарная эффективность использования ресурсов компьютера в рамках всего метода квантовых траекторий составляет лишь около 65%. На этот показатель влияют, с одной стороны, неизбежные накладные расходы, на которые трудно повлиять изменением метапараметров, а с другой – деградация производительности при переходе к менее эффективному вычислительному ядру. Последнюю проблему можно частично устранить увеличением общего числа квантовых траекторий, обрабатываемых на одном вычислительном узле. Именно этот параметр алгоритма можно рассматривать как метапараметр, подлежащий оптимизации до запуска программы.

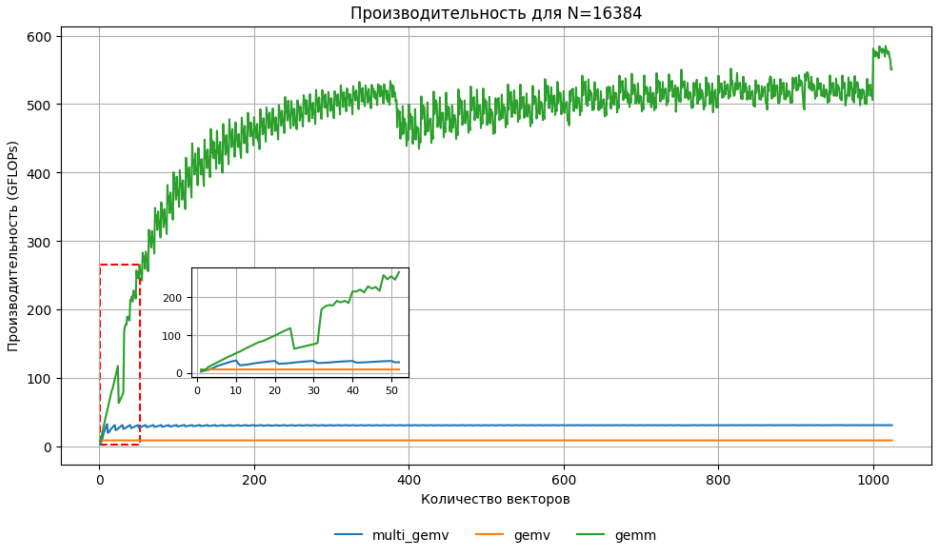


Рис. 1. Производительность разных вычислительных ядер матрично-векторной операции (gemv), матрично-векторной операции с внешним параллелизмом (multi_gemv) и матрично-матричной операции (gemm) в зависимости от количества векторов правого операнда. В случае обращения в DRAM, при размере матрицы N=16384

На рисунках 1 и 2 представлены результаты измерений производительности различных матричных операций в случаях, когда данные располагаются в оперативной памяти или кэше. Эксперименты выполнены на узле (Intel Xeon 4310T, 10 ядер) суперкомпьютера Лобачевский. Как видно из графиков, производительность операции GEMM превосходит различные вариации GEMV в 5–10 раз, даже при сравнительно небольшом числе векторов (квантовых траекторий). Из тех же данных следует, что для достижения максимальной производительности размер малых групп траекторий в алгоритме не должен становиться ниже порядка 400 траекторий.

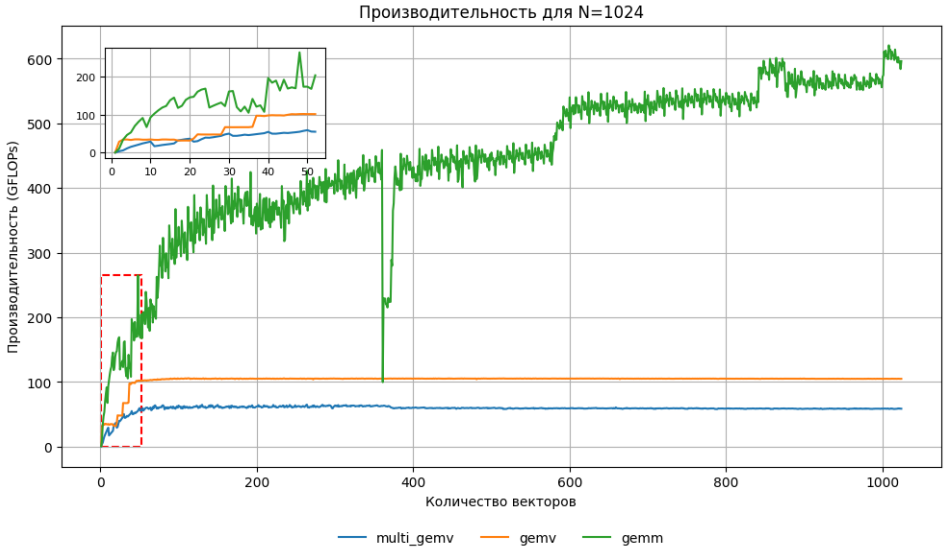


Рис. 2. Производительность разных вычислительных ядер матрично-векторной операции (gemv), матрично-векторной операции с внешним параллелизмом (multi_gemv) и матрично-матричной операции (gemm) в зависимости от количества векторов правого операнда. В случае обращения в кэш, при размере матрицы N=1024

Настоящий доклад посвящён выбору и построению модели производительности метода квантовых траекторий. На основе свойств эффективного гамильтониана будет выполнена оценка частоты возникновения квантовых скачков в алгоритме. Такая оценка позволит определить оптимальное количество траекторий, обрабатываемых на одном вычислительном

узле, что обеспечит не только эффективное распараллеливание вычислений между узлами, но и максимальную загрузку вычислительных ресурсов каждого из них.

Литература

1. Hutter F. et al. ParamILS: an automatic algorithm configuration framework //Journal of artificial intelligence research. – 2009. – Т. 36. – С. 267-306.
2. Hutter F., Hoos H. H., Leyton-Brown K. Sequential model-based optimization for general algorithm configuration //International conference on learning and intelligent optimization. – Berlin, Heidelberg : Springer Berlin Heidelberg, 2011. – С. 507-523.
3. Kadioglu S. et al. ISAC–instance-specific algorithm configuration //ECAI 2010. – Ios Press, 2010. – С. 751-756.
4. Feurer M., Hutter F. Hyperparameter optimization //Automated machine learning: Methods, systems, challenges. – Cham : Springer International Publishing, 2019. – С. 3-33.
5. Bischl B. et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges //Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. – 2023. – Т. 13. – №. 2. – С. e1484.
6. Volokitin V. et al. Computation of the asymptotic states of modulated open quantum systems with a numerically exact realization of the quantum trajectory method //Physical Review E. – 2017. – Т. 96. – №. 5. – С. 053313.
7. Lindblad G. On the generators of quantum dynamical semigroups //Communications in mathematical physics. – 1976. – Т. 48. – №. 2. – С. 119-130.
8. Liniov A. et al. Increasing performance of the quantum trajectory method by grouping trajectories //Russian Supercomputing Days. – Cham : Springer International Publishing, 2017. – С. 136-150.

МЕТОД КОМПЛЕКСНОЙ ВАЛИДАЦИИ ЧИСЛЕННОЙ МОДЕЛИ СОУДАРЕНИЯ КАПЛИ С ПОВЕРХНОСТЬЮ С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ ДИНАМИЧЕСКОГО И КИНЕМАТИЧЕСКОГО ПАРАМЕТРОВ ДВИЖЕНИЯ

Д. С. Жариков[1], М.В. Пискунов[2], Д.С. Коломенский[3]

[1]Московский авиационный институт, Национальный исследовательский университет
[2]Томский политехнический университет, Научно-образовательный центр И.Н. Бутакова
[3]Сколковский институт науки и технологии, Центр системного проектирования

Проведено численное моделирование соударения капли с твёрдой подложкой из сапфирового стекла с помощью ламинарной модели в пакете Ansys Fluent с использованием динамического контактного угла для значений скорости соударения в диапазоне 0.63-2.08 м/с. Валидация модели по максимальному диаметру растекания подтвердила высокую точность подхода с динамическим контактным углом по сравнению со статическим. Для дополнительной верификации модели проведено сравнение рассчитанных полей скорости внутри капли с экспериментальными данными, полученными методом Particle Image Velocimetry. Установлен ключевой вывод: валидация исключительно по кинематике растекания является недостаточной для комплексной оценки адекватности математической модели.

Ключевые слова: численное моделирование, соударение капли, динамический контактный угол, гидродинамика, контактная линия, адаптивная сетка.

1. Введение

Моделирование соударения капли с подложкой представляет собой одну из наиболее сложных и актуальных задач в современной гидродинамике [1]. Этот процесс имеет широкое практическое применение – от разработки систем охлаждения и нанесения покрытий до изучения механизмов образования брызг и эрозии поверхностей. Однако его моделирование осложняется необходимостью учёта множества факторов: поверхностного натяжения, вязкости жидкости, свойств подложки, а также динамики контактной линии [2].

Традиционно для валидации результатов численных моделей используется максимальный диаметр растекания капли - параметр, который легко измерить экспериментально и который напрямую характеризует степень взаимодействия капли с поверхностью. Однако в большинстве исследований основное внимание уделяется только фазе растекания, тогда как поведение капли во время фазы стекания остаётся недостаточно изученным.

В рамках данного исследования были проведены численные расчёты симуляции падения капли на поверхность из сапфирового стекла с использованием ламинарной модели в пакете ANSYS Fluent. При моделировании используется динамический контактный угол.

Особое внимание уделено учёту динамического смачивания, что позволяет более точно описать процесс изменения формы капли при взаимодействии с поверхностью. Для валидации модели использованы не только данные по максимальному диаметру растекания, но и экспериментальные данные по полям скорости внутри капли, полученные методом PIV (Particle Image Velocimetry) [3]. Это позволяет оценить не только внешние характеристики растекания, но и внутренние процессы, происходящие внутри капли.

2. Описание экспериментов

А. Ашихмин и соавторы [4] провели серию экспериментов по соударению капли без примесей и с примесями с подложкой. Результаты измерений скоростей внутренних конвективных потоков, полученные методом PIV (анемометрия по изображениям частиц) в продольном сече-

нии капли водно-глицеринового раствора (60% глицерина, 40% воды), содержащей частицы диоксида титана (концентрация 1 г/л) и растекающейся по поверхности, позволили провести анализ условий, способствующих снижению максимального диаметра растекания. Свойства жидкости представлены в таблице 1.

Таблица 1. Свойства жидкости в экспериментах

Температура, К	Плотность, кг/м^3	Динамическая вязкость, Па*с	Поверхностное натяжение, Н/м
293,15	1154	0.0108	0.06058

Скорость капли в момент соударения составляет в различных экспериментах принимает значения 0.63 м/с, 1.17 м/с, 1.72 м/с и 2.08 м/с, систематическая ошибка измерений составляет 0.1 м/с. Диаметр капли на протяжении всего временного интервала от момента её генерации до момента соударения с поверхностью оставался постоянным и равным 2.9 ± 0.05 мм.

Эксперименты регистрировались с использованием высокоскоростной видеосъемки. Для измерения скоростей жидкости в продольном направлении, параллельном плоскости подложки, применялся метод PIV. Толщина лазерного луча составляла 200 мкм, минимальная высота лазерного луча составляла 200-300 мкм.

3. Моделирование

3.1. Конфигурация модели

Для моделирования соударения капли с подложкой в 2D-осесимметричной постановке использован программный пакет ANSYS Fluent. Для описания течения жидкости использовалась ламинарная модель. Для расчёта многофазного потока использовалась модель Volume of Fluid (VOF) [5], в которой определены две фазы - воздух и жидкость. Поверхностное натяжение задано в соответствии с параметрами жидкости в рассматриваемой работе.

Чтобы локально уточнить решение используется адаптивная сетка с 6-м уровнем адаптации со сгущением на границе раздела фаз. Укрупнение сетки происходит в тех областях, где находится граница раздела фаз и где эта граница имеет высокий градиент. Сетка представлена на рисунке 1.

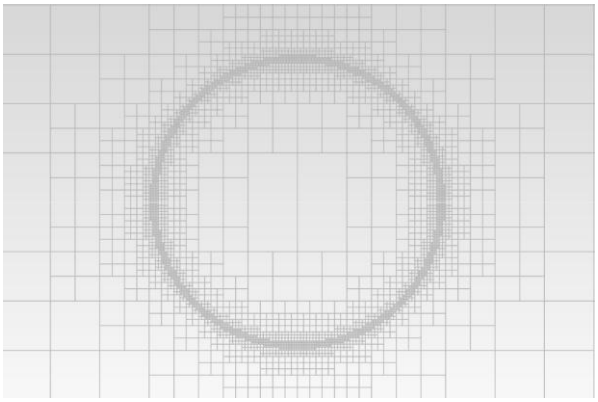


Рис. 1. Визуализация расчетной сетки

Для решения задачи используется алгоритм SIMPLE (Semi-Implicit Method for Pressure Linked Equations) [6], который связывает поля давления и скорости при решении уравнений Навье-Стокса. Градиент давления вычисляется методом наименьших квадратов, а само давление дискретизируется с помощью алгоритма PRESTO! [7]. Уравнения импульса решаются схемой второго порядка (Upwind). Для объемной доли применяется схема Geo-Reconstruct [11].

В данной симуляции используется фиксированный шаг по времени $5 * 10^{-6}$, всего 2500 итераций по времени. Максимальное количество итераций на каждом временном шаге 40. Ста-

тический контактный угол в данном исследовании равен 46.5°. Моделирование динамического краевого угла в фазах растекания и стекания осуществлялось на основе функции Хоффмана [10] и с помощью обобщенного закона Хоффмана-Воинова-Таннера [10].

3.2. Результаты моделирования

В обобщенном законе Хоффмана-Воинова-Таннера присутствует константа, значения которой были подобраны эмпирически таким образом, чтобы в наибольшей степени соответствовать максимальному диаметру растекания в эксперименте. Сравнение моделей по диаметру растекания капли представлено в таблице 2.

Таблица 2. Сравнение результатов по максимальному диаметру растекания. Отличие в процентах от экспериментов

Модель контактного угла	**0.63 м/с**	**1.17 м/с**	**2.08 м/с**
Статический контактный угол	29.60	16.44	10.02
Обобщенный закон Хоффмана-Воинова-Таннера, const=3	15.81	7.64	9.65
Обобщенный закон Хоффмана-Воинова-Таннера, const=5	12.13	5.09	2.63
Обобщенный закон Хоффмана-Воинова-Таннера, const=8.78	5.70	1.62	0.78
Функция Хоффмана	2.88	5.79	3.66

Ламинарная модель со статическим контактным углом показывает значительное расхождение с экспериментальными данными, что свидетельствует о её недостаточной физической адекватности для данного процесса. В то же время использование в ламинарной модели динамического контактного угла позволяет добиться высокой точности: отличие от эксперимента для моделей, использующих динамический контактный угол, не превышает 6%. Это указывает на важность учёта динамики смачивания и изменения контактного угла во времени для корректного моделирования процесса растекания. Общепринятым критерием для верификации моделей соударения капель с поверхностью является максимальный диаметр растекания. Исходя из этого модель, использующая обобщенный закон Хоффмана-Воинова-Таннера, позволяет добиться наилучшего результата.

Для верификации результатов численного моделирования выполнено сравнение рассчитанных полей скорости с экспериментальными данными, полученными методом PIV. Для минимизации систематической погрешности PIV [8, 9], связанной с занижением измеренных значений скорости, с модельных данных исключены ячейки, расположенные за пределами 0.75 радиуса растекания. Кроме того, для подавления высокочастотного шума применён частотный фильтр на основе преобразования Фурье. На рисунке 2 представлено распределение средних значений компонент скорости, параллельных поверхности, в поперечном сечении капли для ламинарной модели с динамическим контактным углом, рассчитанным с помощью функции Хоффмана, а на рисунке 3 расчет с использованием обобщенного закона Хоффмана-Воинова-Таннера.

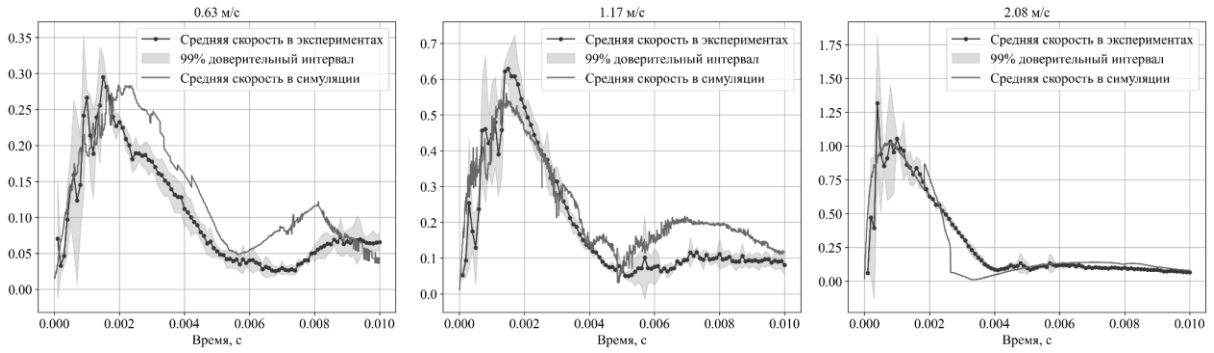


Рис. 2. Сравнение значений средней скорости внутри капли с экспериментальными данными, функция Хоффмана

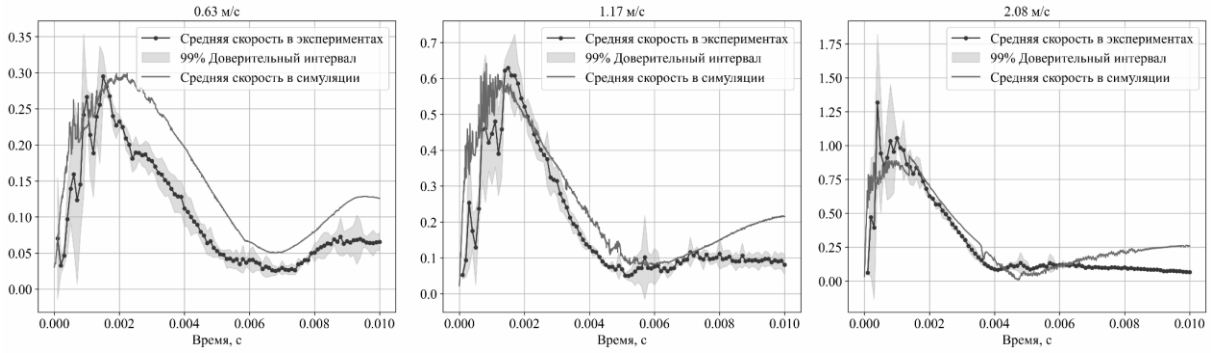


Рис. 3. Сравнение значений средней скорости внутри капли с экспериментальными данными, обобщенный закон Хоффмана-Воинова-Таннера

Сравнительный анализ профилей скорости показывает, что модель, основанная на обобщённом законе (рис. 3), в фазе стекания характеризуется некорректной с физической точки зрения динамикой: на заключительном этапе проявляется монотонное возрастание скорости, что противоречит реальному течению процесса. Напротив, модель с использованием функции Хоффмана (рис. 2) обеспечивает более корректное воспроизведение стекания. Медиана абсолютной ошибки между симуляциями и реальными данными для промежутка времени после 0.008 с. представлена в таблице 3.

Таблица 3. Медиана абсолютной ошибки между симуляцией и реальными данными после 0.008 секунд.

Скорость, м/с	**Обобщенный закон Хоффмана-Воинова-Таннера, const=8.78**	**Функция Хоффмана**
0.63	0.05372	0.13538
1.17	0.00998	0.06338
2.08	0.16088	0.02502

Анализ динамики растекания капли демонстрирует, что для адекватного моделирования процесса, включая кинематическую фазу, фазы активного растекания и релаксации (или стекания), необходимо учитывать не только диаметр растекания, но и распределение скоростей внутри капли. Использование исключительно максимального диаметра растекания в качестве критерия валидации численных результатов и корректности работы модели является методологически некорректным подходом.

Причина заключается в том, что даже при минимальных расхождениях по максимальному диаметру растекания на этапе релаксации (или стекания) могут наблюдаться значительные отклонения, вплоть до качественного изменения физики процесса. В частности, возможны случаи, когда стекание отсутствует вовсе, что свидетельствует о принципиальных различиях в поведении системы. Таким образом, для полноценной валидации модели требуется комплексный

анализ как геометрических параметров (например, диаметра растекания), так и динамических характеристик (например, внутренних скоростей).

4. Заключение

В ходе исследования проведено численное моделирование соударения капли с твёрдой подложкой с использованием ламинарной модели в пакете ANSYS Fluent. Результаты показали, что учёт динамического контактного угла значительно повышает точность моделирования процесса растекания капли. Валидация по максимальному диаметру растекания подтвердила, что модели с динамическим контактным углом обеспечивают лучшее соответствие экспериментальным данным по сравнению с моделями, использующими статический контактный угол.

Однако анализ полей скорости внутри капли, полученных методом PIV, выявил, что даже при высокой точности по максимальному диаметру растекания, динамика стекания может существенно отличаться от реального процесса. Это указывает на необходимость комплексного подхода к валидации численных моделей, включающего не только геометрические параметры, но и динамические характеристики, такие как распределение скоростей внутри капли.

Таким образом, для полноценной оценки адекватности математической модели необходимо учитывать как кинематические параметры растекания, так и внутренние динамические процессы. Это позволит более точно описывать физику процесса и разрабатывать более эффективные модели для практических приложений.

Литература

1. Yarin, A. L. Drop impact dynamics: Splashing, spreading, receding, bouncing…. Annual Review of Fluid Mechanics. 2006, vol. 38, pp. 159–192.
2. Snoeijer, J. H., Andreotti, B. Moving contact lines: Scales, regimes, and dynamical transitions. Annual Review of Fluid Mechanics. 2013, vol. 45, pp. 269–292.
3. Adrian, R. J., Westerweel, J. Particle image velocimetry. Cambridge: Cambridge University Press, 2011.
4. Ashikhmin, A., Khomutov, N., Volkov, R., Piskunov, M., Strizhak, P. Effect of Monodisperse Coal Particles on the Maximum Drop Spreading after Impact on a Solid Wall. Energies. 2023, vol. 16, no. 14, p. 5291.
5. Hirt, C. W., Nichols, B. D. Volume of fluid (VOF) method for the dynamics of free boundaries. Journal of Computational Physics. 1981, vol. 39, no. 1, pp. 201–225.
6. Anderson, D., Tannehill, J. C., Pletcher, R. H., Munipalli, R., Shankar, V. Computational Fluid Mechanics and Heat Transfer. Boca Raton: CRC Press, 2020.
7. Patankar, S. V. Numerical Heat Transfer and Fluid Flow. New York: Hemisphere Publishing Corp., 1980.
8. Wilson, B. M., Smith, B. L. Uncertainty on PIV mean and fluctuating velocity due to bias and random errors. Measurement Science and Technology. 2013, vol. 24, no. 3, p. 035302.
9. Hart, D. P. PIV error correction. Experiments in Fluids. 2000, vol. 29, no. 1, pp. 13–22.
10. Mohammad Karim A. A review of physics of moving contact line dynamics models and its applications in interfacial science //Journal of Applied Physics. – 2022. – Т. 132. – №. 8.
11. Youngs D. L. Time-dependent multi-material flow with large fluid distortion //Numerical methods for fluid dynamics. – 1982.

ОБ АЛГОРИТМЕ ПОИСКА ГЛОБАЛЬНОГО МИНИМУМА ДЛЯ ОДНОГО КЛАССА МНОГОМЕРНЫХ МНОГОЭКСТРЕМАЛЬНЫХ ФУНКЦИЙ[1*]

А.С. Зайцев, К.А. Баркалов

Нижегородский государственный университет им. Н.И. Лобачевского
Научно-образовательный математический центр «Математика технологий будущего»

В работе рассматривается задача поиска глобального минимума многомерной многоэкстремальной функции, принимающей неотрицательные значения и достигающей в области определения нулевого значения. По сути, эта задача является задачей поиском нуля многомерной функции. Предложена адаптация алгоритма поиска нуля неотрицательной функции для решения многомерных задач, редуцированных с помощью развёртки (кривой) Пеано. Проведены вычислительные эксперименты, подтверждающего его эффективность.

Ключевые слова: глобальная оптимизация, поиск нуля функции, поиск корня функции, тестовые задачи.

1. Введение

Настоящая работа, являющаяся продолжением работы [4], заключается в адаптации алгоритма поиска нуля неотрицательной функции (АПН) для решения многомерных задач. И приводятся результаты численных экспериментов для демонстрации его эффективности.

2. Постановка задачи

Классическая задача поиска минимума многомерной многоэкстремальной функции $\varphi(y)$, где $y \in R^n$ $(n > 1)$ в N-мерном гиперкубе D записывается в следующем виде:

$$\varphi(y^*) = min\{\varphi(y), y \in D\}, D = \{y \in R^n, a_i \le y_i \le b, i = 1, \dots, n\} \quad (1)$$

Дополнительно потребуем, чтобы функция $\varphi(x)$ удовлетворяла следующим условиям:

1) $\varphi(y) \ge 0, y \in D$ (неотрицательность функции);
2) Существует вектор $y^* \in D$ такой, что $\varphi(y^*) = 0$ (существование нуля функции);
3) Функция удовлетворяет условию:

$$|\varphi(y') - \varphi(y'')| \le L|y' - y''|, y', y'' \in D;\ 0 < L < \infty,$$

где значение L – не задано (липшицевость функции).

Функцию $\varphi(y)$, удовлетворяющую описанным выше условиям, будем обозначать $\psi(y)$, где $y \in R^n$ $(n > 1)$. Очевидно, что для рассматриваемой функции $\psi(y)$ $(y \in D)$ задача поиска глобального минимума эквивалентна задачи поиска нуля функции (поиска корня функции).

Важно заметить, что с ростом размерности задачи N вычислительные затраты будут возрастать экспоненциально. Но этого эффекта можно избежать если редуцировать поставленную многомерную задачу до одномерной. В данной работе для редукции размерности будут использоваться развёртки (кривые) Пеано $y(x), x \in [0,1]$ [1, 2, 3]. Т. е. отображение отрезка $[0,1]$ вещественной оси на область D. Таким образом исходная многомерная задача (1) сводится к одномерной задаче следующего вида:

$$\psi(y(x^*)) = min\{\psi(y(x)), x \in [0,1]\} \quad (2)$$

Функция $\psi(y(x))$ удовлетворяет следующим условиям:

[1*] Работа выполнена при поддержке Минобрнауки РФ (проект № FSWR-2023-0034).

1) $\psi(y(x)) \geq 0, x \in [0,1]$ (неотрицательность функции);
2) Существует точка $x^* \in [0,1]$ такая, что $\psi(y(x^*)) = 0$ (существование нуля функции);
3) Функция удовлетворяет условию:

$$|\psi(y(x')) - \varphi(y'')| \leq H\sqrt[N]{|x' - x''|}, x', x'' \in [0,1];\ 0 < H < \infty,$$

где значение H – не задано. Из [2] известно, что константа Гёльдера H связана с константой Липшица L соотношением $H = 4L\sqrt{N}$ (гёльдеровость функции).

Для решения редуцированной задачи применимы численные методы липшицевой оптимизации, например, метод Стронгина [1, 2] и его модификации, например, [3]. Но данные методы не учитывают 1-е и 2-е свойство рассматриваемой функции.

Алгоритм Поиска нуля (АПН) [4], учитывающий неотрицательность функции, требует адаптации для решения рассматриваемой задачи. В данной работе предлагается такая адаптация.

3. Описание алгоритма

Алгоритм поиска нуля функции $\psi(y(x))$ порождает последовательность точек $\{x^k\}$. В этих точках вычисляется значение функции $z^k = \psi(y(x^k)), k = 0, 1, 2, \ldots$. Предполагается, что погрешность пренебрежимо мала. Первые два испытания осуществляются в граничных точках, то есть $x^0 = 0$ и $x^1 = 1$ соответственно. А выбор точек $x^{k+1}, k \geq 1$, очередного $(k+1)$-го испытания осуществляются по следующим правилам:

1) Перенумеруем (нижнем индексом) точки x^i, $0 \leq i \leq k$, в порядке возрастания значения координаты:

$$0 = x_0 < x_1 < \ldots < x_k = 1 \tag{3}$$

2) Для каждого интервала $(x_i - x_{i-1}), 1 \leq i \leq k$, вычислим величину:

$$K(i) = \frac{z_i + z_{i-1}}{\Delta_i}, \tag{4}$$

где $\Delta_i = \sqrt[N]{x_i - x_{i-1}}$, $z_i = \psi(y(x_i))$, называемую характеристикой этого интервала

3) Определим интервал $(x_t - x_{t-1})$, которому соответствует минимальная характеристика K:

$$K(t) = min\{K(i), 1 \leq i \leq k\} \tag{5}$$

Если минимальное значение соответствует нескольким интервалам, то в качестве значения t выбирается наименьшее число, удовлетворяющее условию (3).

4) Определим точку x^{k+1} очередного испытания следующим образом:

$$x^{k+1} = \frac{x_t + x_{t-1}}{2}, \tag{6}$$

5) Проведём очередное испытание в точке x^{k+1}:

$$z^{k+1} = \psi(y(x^{k+1})) \tag{7}$$

и примем $k = k + 1$

6) Проверим условие остановки алгоритма:
   a) Число итераций k достигло значения k_{max} (исчерпание вычислительного ресурса)
   b) Было достигнуто значение

$$\psi_k^* = min\{\psi(y(x_i)): 0 \leq i \leq k\} \leq \sigma \tag{8}$$

где $\sigma > 0$ (предустановленная допустимая погрешность в оценке нулевого значения функции $\psi(y(x))$).

4. Вычислительные эксперименты

Для того чтобы убедиться в корректности работы предложенного алгоритма изобразим графики функций, минимумы которых известны, и работу алгоритма, то есть точки, в которых метод проводил испытания функции. Значение допустимой погрешности σ установим равным 0.01, вычислительный ресурс k_{max} установим равным 1250.

4.1 Вычислительный эксперимент №1

Известно, что некоторая функция $\psi(y^*) = 0$ при $y^* = (0.341732, 0.197620)$

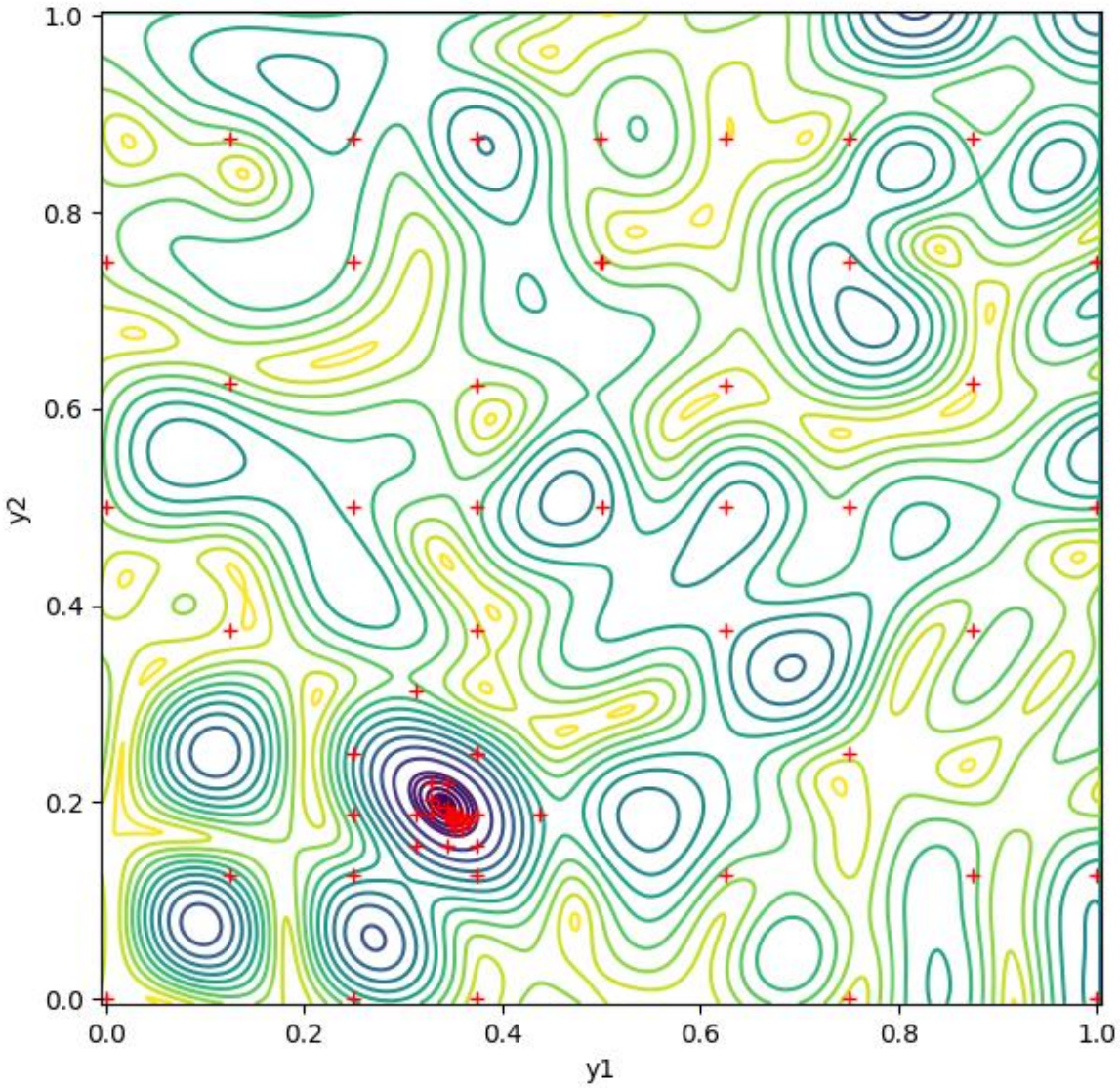


Рис. 1. График целевой функции и точек испытаний порождённых АПН

АПН нашёл точку $y^* = (0.343628, 0.195218)$ в которой значение функции $\psi(y^*) = 0.00650964$ за 76 итераций.

4.2 Вычислительный эксперимент №2

Известно, что некоторая функция $\psi(y^*) = 0$ при $y^* = (0.875874,0.653336)$

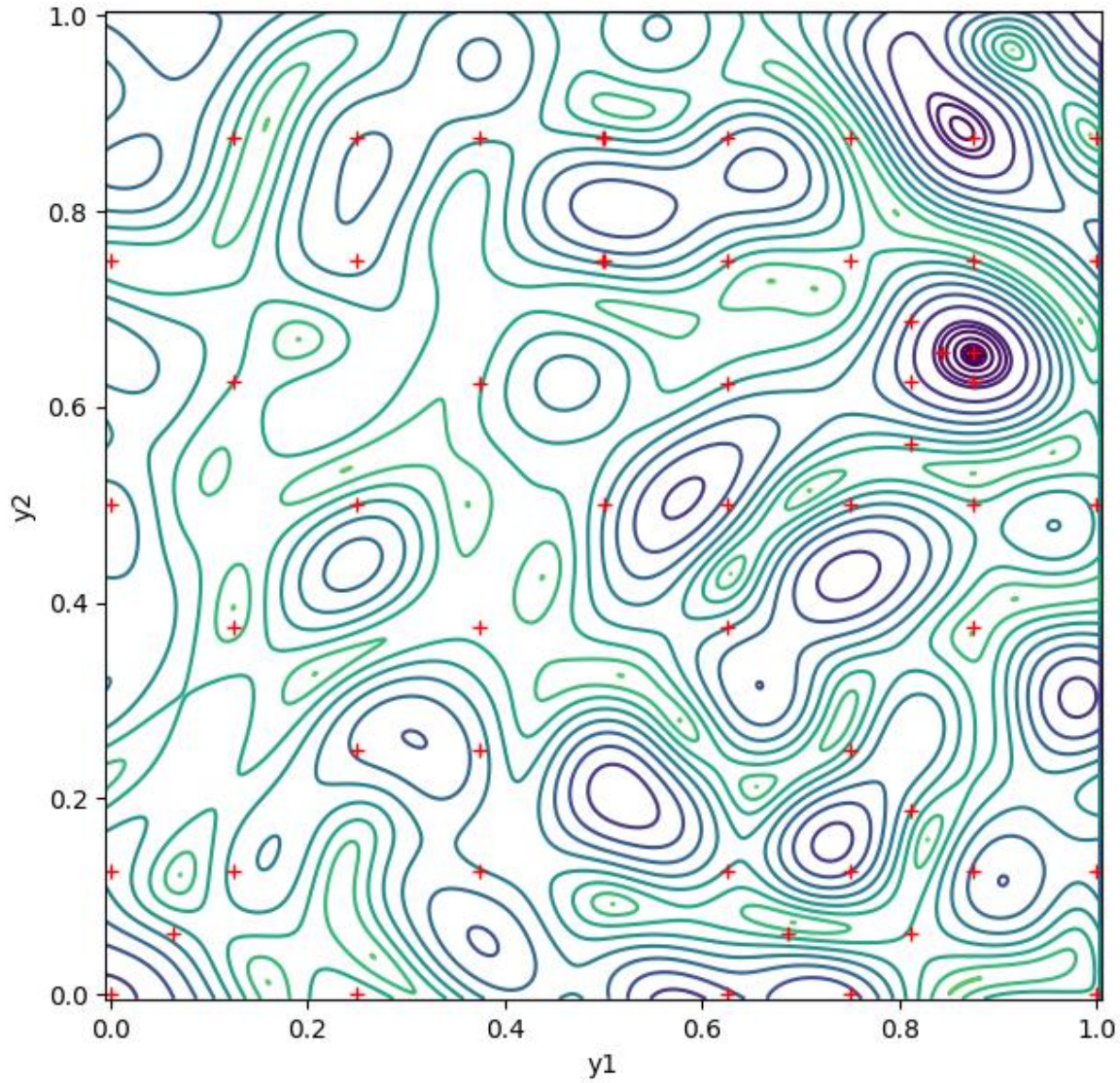


Рис. 2. График целевой функции и точек испытаний порождённых АПН

АПН нашёл точку $y^* = (0.874878, 0.656303)$ в которой значение функции $\psi(y^*) = 0.0081869$ за итераций 56.

4.3 Вычислительный эксперимент №3

Известно, что некоторая функция $\psi(y^*) = 0$ при $y^* = (0.336467, 0.909056)$

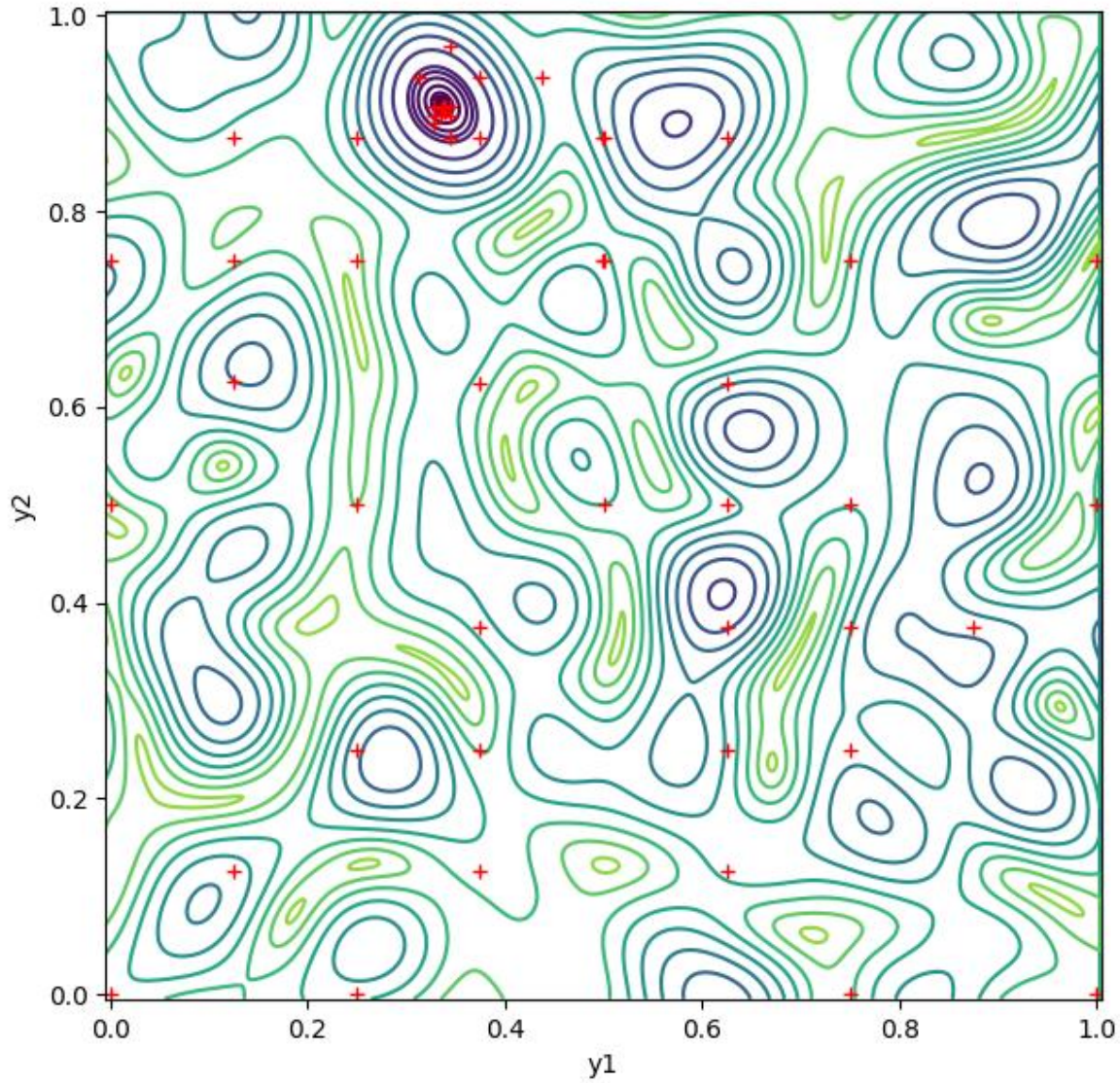


Рис. 3. График целевой функции и точек испытаний порождённых АПН

АПН нашёл точку $y^* = (0.337867, 0.906128)$ в которой значение функции $\psi(y^*) = 0.00646292$ за итераций 52.

5. Заключение

Предложенный алгоритм показал успешную работу при решении ряда тестовых задач поиска нуля неотрицательной двумерной функции. Проведенные эксперименты наглядно демонстрируют, что последовательность точек испытаний, порождаемая новым алгоритмом, является плотной в окрестности глобального минимума.

Литература

1. Стронгин Р.Г. Численные методы в многоэкстремальных задачах (информационно статистический подход). – М.: Наука, 1978, 240 стр.
2. Стронгин Р.Г. Поиск глобального оптимума. – М.: Знание, 1990.
3. Barkalov K.A., Strongin R.G., Bevzyuk S.A. Acceleration of Global Search by Implementing Dual Estimate for Lipschitz Constant // Lecture Notes in Computer Science. vol. 11974. 2020. P. 478−486.

4. Зайцев А.С., Баркалов К.А. Об алгоритме поиска глобального минимума для одного класса многоэкстремальных задач // Сборник трудов IV Молодёжной школы. Н.Новгород, Издательство ННГУ им. Н. И. Лобачевского, 2025. С. 39−42.

АЛГОРИТМИЧЕСКАЯ СЛОЖНОСТЬ ЗАДАЧ СОСТАВЛЕНИЯ РАСПИСАНИЙ В СИСТЕМАХ С УЧЕТОМ СТАТИЧЕСКОЙ И ДИНАМИЧЕСКОЙ СОСТАВЛЯЮЩИХ РАСХОДА ЭНЕРГИИ[1*]

А.О. Захаров, Ю.В. Захарова

Омский филиал Института математики им. С.Л. Соболева СО РАН

Рассматриваются задачи составления расписания с возможностью масштабирования скорости процессоров и энергопотребления. Энергопотребление состоит из двух компонентов: статического и динамического. Предложены приближенные алгоритмы с гарантированной константной оценкой точности для распараллеливаемых работ в двух случаях: для независимых работ и для работ, связанных отношением частичного порядка. Алгоритмы используют выпуклые модели для вычисления длительностей и методы списочного типа для построения допустимых решений.

Ключевые слова: расписание, энергия, алгоритм, сложность.

1. Постановка задачи

Множество распараллеливаемых работ $J=\{1,\dots,n\}$ необходимо выполнить на множестве процессоров $M=\{1,\dots,m\}$. Каждая работа $j\in J$ характеризуется числом требуемых процессоров $size_j \le m$ и объемом W_j (длительность работы на каждом процессоре при единичной скорости) [2]. В каждый момент времени каждый процессор может выполнять не более одной работы. Также рассматривается вариант задачи, когда задана верхняя граница на степень распараллеливания $\delta_j \le m$. В последнем случае необходимо при составлении расписания выбирать в том числе число используемых процессоров. Работы могут быть связаны отношением предшествования.

Процессоры могут работать с разной скоростью. Если процессор при выполнении работы $j\in J$ работает со скоростью s_j, то потребление энергии в единицу времени (мгновенная мощность) равно $(s_j)^{\alpha}$ ($\alpha>1$ – константа), что задает динамическую составляющую энергопотребления. За общее время работы системы $C_{\max}$ (длина расписания, makespan) формируется статическая часть энергопотребления с коэффициентом E_{stat}.

Требуется построить допустимое расписание выполнения работ, при котором общее потребление энергии $E_{\text{tot}} = \sum_{j\in J} size_j\,(s_j)^{\alpha-1}\,W_j + m\,C_{\max}\,E_{\text{stat}}$ минимизируется.

2. Вычислительная сложность и алгоритмы решения

Доказана NP-трудность задачи без частичного порядка и NP-трудность в сильном смысле задачи с частичным порядком в случае двух процессоров и однопроцессорных работ [6, 8]. Распараллеливаемость также приводит к NP-трудности в сильном смысле задачи [2].

Опишем метод построения ρ-приближенного решения для рассматриваемой задачи:

Этап 1. Построить и решить модель выпуклого программирования с целевой функцией

$$\sum_{j\in J} size_j\,(s_j)^{\alpha-1}\,W_j + m\,LB\,E_{\text{stat}},$$

где переменная LB отвечает за нижнюю границу на длину расписания и ограничения задают эту границу для каждого конкретного случая задачи индивидуальным способом, а переменные s_j определяют длительности работ $j\in J$.

Этап 2. Для длительностей работ, полученных на этапе 1 ($p_j = W_j/s_j$, $j\in J$), строится ρ -приближенное расписание для критерия минимизации длины расписания и это расписание будет ρ -приближенным для критерия минимизации расхода энергии.

Данный метод позволяет получить следующие результаты:

[1*] Исследование выполнено за счет гранта Российского научного фонда № 22-71-10015-П.

– независимые распараллеливаемые работы с фиксированным числом используемых процессоров ($size_j$ для работы $j \in J$)

$$\rho = 2 - 1/m,\ LB = \max\{1/m \sum_{j \in J} size_j\, p_j;\ \max_{j \in J} p_j\}$$ [4];

– распараллеливаемые работы с фиксированным числом используемых процессоров и заданным частичным порядком:

$$\rho = (2 - q)/(1-q),\ LB = \max\{\sum_{l \in L} P_l;\ 1/m \sum_{j \in J} size_j\, p_j\}$$ [5],

где $0<q<1$ такое, что $size_j \le qm$, $j \in J$, P_l – длительности цепочек $l \in L$ работ, связанных частичным порядком;

– независимые распараллеливаемые работы с верхней границей на степень распараллеливания (δ_j для работы $j \in J$):

$$\rho = 2 - 2/(m+1),\ LB = \max\{1/m \sum_{j \in J} \delta_j\, p_j;\ \max_{j \in J} p_j\}$$ [1];

– распараллеливаемые работы с верхней границей на степень распараллеливания и заданным частичным порядком:

$$\rho = 2 + \sqrt{5}/2,\ LB = \max\{\sum_{l \in L} P_l;\ 1/m \sum_{j \in J} \delta_j\, p_j\}$$ [7].

Литература

1. Belkhale K., Banerjee P. An approximate algorithm for the partitionable independent task scheduling problem // Proceedings of International Conference on Parallel Processing. 1990. Vol. 51. P. 72-75.
2. Drozdowski M. Scheduling for Parallel Processing. Springer-Verlag, London (2009).
3. Da Costa G. Hardware and application aware performance, power and energy models for modern HPC servers with DVFS // Sustainable Computing: Informatics and Systems. 2025. Vol. 46.
4. Feldmann A., Sgall J., Teng S.-H. Dynamic scheduling on parallel machines // Theoretical Computer Science. 1994. Vol. 130, iss. 1. P. 49-72.
5. Feldmann A., Kao M.-Y., Sgall J., Teng S.-H. Optimal online scheduling of parallel jobs with dependencies // Journal of Combinatorial Optimization. 1998. Vol. 1, iss. 4. P. 393-411.
6. Kononov A., Kovalenko Y. Makespan minimization for parallel jobs with energy constraint // MOTOR-2020, LNCS. 2020. Vol. 12095. P. 289-300.
7. Lepere R., Trystram D., Woeginger G. Approximation algorithms for scheduling malleable tasks under precedence constraints // International Journal of Foundations of Computer Science. 2002. Vol. 13, iss. 4. P. 613-627.
8. https://www2.informatik.uni-osnabrueck.de/knust/class/ (дата обращения: 29.09.2025).

КАМЕРЫ И СТЕНКИ В ПРОСТРАНСТВЕ ВЕЩЕСТВЕННЫХ АЛГЕБРАИЧЕСКИХ КРИВЫХ БИСТЕПЕНИ (4,3) НА ГИПЕРБОЛОИДЕ[1*]

В.И. Звонилов

Нижегородский государственный университет им. Н.И. Лобачевского

В работе завершено перечисление камер и стенок в пространстве вещественных алгебраических кривых бистепени (4,3) на гиперболоиде начатое ранее автором Основным техническим средством являются графы и скелеты вещественных тригональных кривых на поверхностях Хирцебруха, получаемых из кривых на гиперболоиде бирациональными преобразованиями.

Ключевые слова: пространства вещественных алгебраических кривых, кривые на гиперболоиде, тригональные кривые.

1. Определения

Пусть $[x_0 : x_1]$, $[y_0 : y_1]$ — однородные координаты на гиперболоиде $P^1 \times P^1$. *Вещественная алгебраическая кривая бистепени* (m_1, m_2) на гиперболоиде определяется уравнением $F(x_0, x_1; y_0, y_1) = 0$, где F — вещественный однородный многочлен степеней m_1 и m_2 по x_0, x_1 и y_0, y_1.

Обозначим через $C_{m,n}$ пространство вещественных алгебраических кривых бистепени (m, n) на гиперболоиде.

Вещественная схема неособой вещественной алгебраической кривой $A \subset \mathrm{P}^1 \times \mathrm{P}^1$ – это схема взаимного расположения компонент её вещественной части $\mathbb{R}A$. Стягиваемые в $\mathbb{R}\mathrm{P}^1 \times \mathbb{R}\mathrm{P}^1$ компоненты называются *овалами.* Все нестягиваемые компоненты N_1, ... , N_h реализуют один и тот же ненулевой класс (c_1,c_2) в $H_1(\mathbb{R}\mathrm{P}^1 \times \mathbb{R}\mathrm{P}^1)$, где c_1, c_2 взаимно просты. Вещественная схема кривой кодируется следующим образом: <(c_1,c_2), scheme$_1$, (c_1,c_2), scheme$_2$, ... , (c_1,c_2), scheme$_h$>, где scheme$_1$, scheme$_2$, ..., scheme$_h$ -- схемы расположения овалов, лежащих в связных компонентах поверхности $\mathbb{R}P^1 \times \mathbb{R}P^1 \backslash (N_1 \cup \ldots \cup N_h)$.

Вещественная алгебраическая кривая A принадлежит *типу I*, если множество $\mathbb{C}\widetilde{A} \backslash \mathbb{R}\widetilde{A}$ несвязно, где $\widetilde{A}$ – нормализация кривой A, и *типу II*, если оно связно. В данной работе под *комплексной схемой* кривой понимается её вещественная схема, наделённая типом кривой. Комплексные схемы с вещественной схемой <*B*> обозначаются через <*B*>$_{\mathrm{I}}$ и <*B*>$_{\mathrm{II}}$.

2. Поверхности Хирцебруха, преобразования Нагаты, тригональные кривые

Поверхность Хирцебруха $\Sigma_k, k \geq 0$, является пространством линейного расслоения $q: \Sigma_k \to P^1$, т.е. вещественной рациональной линейчатой поверхностью с исключительным вещественным сечением E_k, $E_k^2 = -k$. Поверхность $\Sigma_0 = P^1 \times P^1$ является гиперболоидом, в котором в качестве исключительного сечения можно взять любую кривую бистепени $(0,1)$. В пространстве $C_{m,n}$ при изучении кривых с единственной особенностью исключительным сечением считается кривая бистепени $(0,1)$, проходящая через особую точку кривой.

Положительным (отрицательным) преобразованием Нагаты называется послойное бирациональное преобразование $\Sigma_k \to \Sigma_{k+1}$ (соответственно, $\Sigma_k \to \Sigma_{k-1}$), состоящее из раздутия точки $p \in E_k$ (соответственно, $p \notin E_k$) и последующего стягивания в точку собственного прообраза слоя $q^{-1}(p)$.

[1*] Работа поддержана Минобрнауки РФ, проект FSWR-2023-0034.

Положительное (отрицательное) преобразование Нагаты переводит E_k в E_{k+1} (соответственно, в E_{k-1}).

Алгебраическая кривая $C \subset \Sigma$ называется тригональной, если сужение $q|_C$ является трёхлистным разветвлённым накрытием. Кривая называется собственной, если она не пересекается с исключительным сечением.

Положительное и отрицательное преобразования Нагаты устанавливают взаимосвязь между (тригональными) кривыми бистепени $(m, 3)$ на гиперболоиде и собственными тригональными кривыми: кривой $C \subset \Sigma_0$ отвечает её образ $N(C) \subset \Sigma_m$ при положительном преобразовании Нагаты, раздувающем точки пересечения кривой C с кривой $E_0 \subset \Sigma_0 = P^1 \times P^1$.

3. Граф тригональной кривой

В некоторой аффинной карте на Σ_k исключительное сечение E_k и собственная тригональная кривая C определяются уравнениями $y = \infty$ и $y^3 + b(x)y + w(x) = 0$, где b и w — сечения расслоения q, которые можно рассматривать как однородные многочлены степеней $2k$ и $3k$. Рациональная функция $j = j_C = \frac{4b^3}{\Delta} = 1 - \frac{27w^2}{\Delta}$ называется j-инвариантом кривой C.

Пусть круг D — фактор-пространство прямой $\mathbb{C}P^1$ по комплексному сопряжению, а $\mathrm{pr}: \mathbb{C}P^1 \to D$ — проекция. В образе j-инварианта $j_C: \mathbb{C}P^1 \to \mathbb{C}P^1 = \mathbb{C} \cup \{\infty\}$ вещественной тригональной кривой $C \subset \Sigma_k$ раскрасим $\mathbb{R}P^1$ следующим образом: 0, 1, и ∞ — это, соответственно, чёрная, белая и крест-вершина; $(\infty, 0)$, $(0,1)$ и $(1, \infty)$ — это, соответственно, сплошное, жирное и пунктирное ребро. Прообраз $\Gamma_C = \mathrm{pr}(j_C^{-1}(\mathbb{R}P^1))$, наделённый соответствующей раскраской, называется *графом тригональной кривой* C.

4. Скелет вещественной тригональной кривой

Объединение замыканий некоторых одинаково раскрашенных вещественных ребер графа Γ_C называется *сегментом*, если оно гомеоморфно отрезку. Пунктирный (жирный) сегмент называется максимальным если его концами являются две неособые (т.е. отвечающие простым корням дискриминанта d) крест-вершины (соответственно, две чёрные вершины). Максимальный пунктирный сегмент графа негиперболической кривой, содержащий четное/нечетное количество белых вершин, называется *овалом/зигзагом*; он является проекцией овала/зигзага соответствующей тригональной кривой. Максимальный пунктирный/жирный сегмент с четным/нечетным количеством белых вершин называется *волной/скачком*.

Как и в [1], можно доказать, что для тех особых тригональных кривых на Σ_3, Σ_4, которые изучаются в работе, можно считать, что все особые точки являются вещественными и уединёнными.

Скелет негиперболической тригональной кривой, допускающей лишь уединённые особые точки, получается из её графа склейкой каждого овала/зигзага в вещественную белую вершину, волны/скачка – в вещественную чёрную вершину скелета, сохранением всех крест-вершин и мнимых чёрных/белых вершин и наконец, удалением всех рёбер, кроме пунктирных. На рёбрах скелета, не являющихся частями разреза (подробнее см. [1, п. 4.6, п. 5]), указаны ориентации. Особые крест-вершины обозначаются двойным крестом.

5. Основной результат

Вещественные схемы неособых кривых бистепени (4,3) на гиперболоиде перечислены в работе [2]: <(3,±m)>, m = 0, 2, 4, <(1,±m), l>, $0 \le l \le 6$, m = 0, 2, и <(1,±4)>. Четыре схемы <(1,±m), l>, l = 0, 2 , m = 0, 2, могут иметь как тип I, так и тип II.

В работе автора [3, теоремы 1, 2] утверждалось, что *камера/стенка в пространстве* $C_{4,3}$ *однозначно определяется комплексной схемой кривой из этой камеры/стенки* (см. рис.1). Для доказательства этого использовался подход, предложенный в [4] для перечисления камер и стенок в пространстве плоских вещественных квинтик, и основанный на теореме 1 работы [3], в которой перечислялись все стенки в пространстве $C_{4,3}$. Доказательство теоремы 1 было начато

автором в [3]; в [5] был указан пробел в этом доказательстве, который был недавно устранён в работе [1] с использованием графов и скелетов вещественных тригональных кривых на Σ_3, полученных из особых кривых бистепени (4,3) на гиперболоиде.

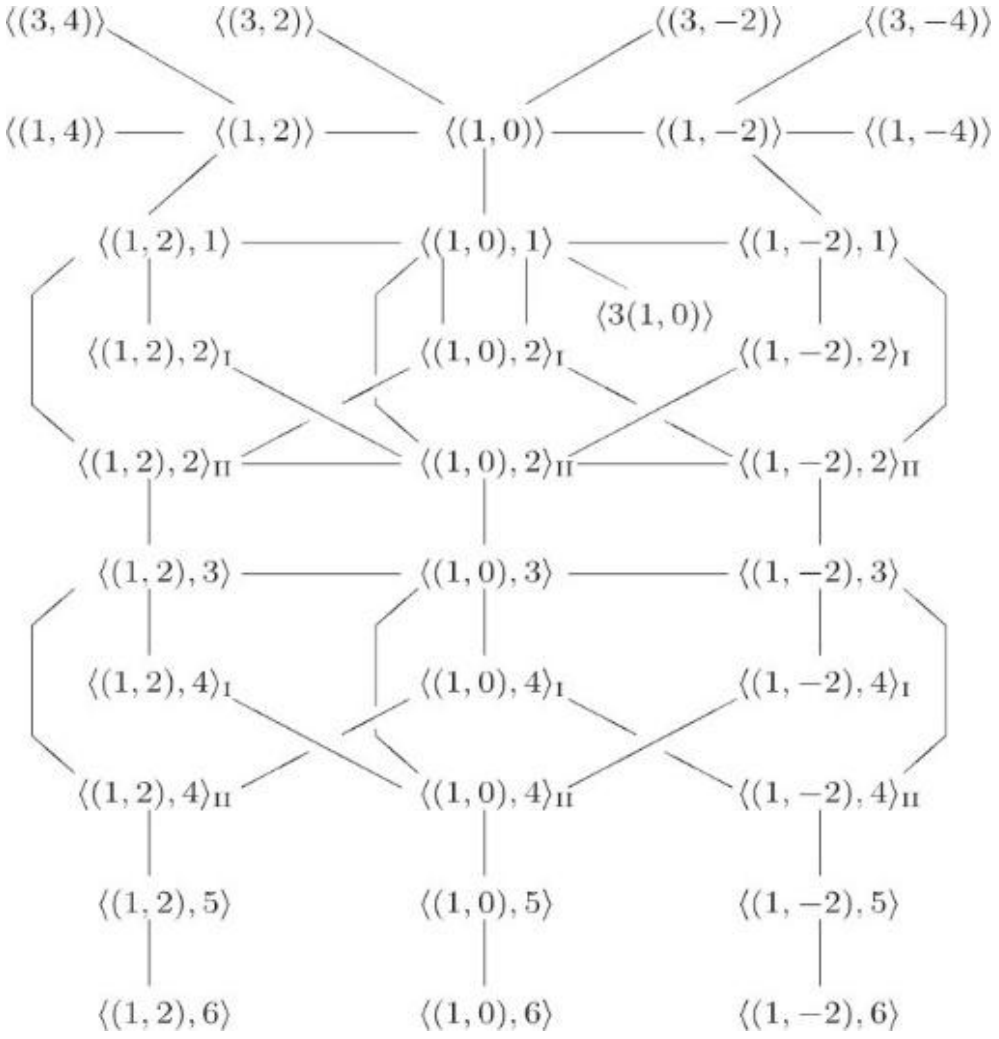


Рис. 1. Камеры и стенки пространства кривых бистепени (4,3) на гиперболоиде

Перечисление стенок/камер в пространстве $C_{4,3}$ даётся на языке графов и скелетов вещественных тригональных кривых на Σ_3, Σ_4, полученных из неособых кривых бистепени (4,3) на гиперболоиде, следующим образом. После трёхкратного/четырёхкратного положительного преобразования Нагаты такая кривая даёт собственную кривую на Σ_3, Σ_4, для которой доказывается, что она определяется каноническим графом или скелетом, задающим стенку/камеру.

На рис. 2–4 указаны вещественные схемы стенок (в обозначениях работы [3]) и соответствующие канонические графы и скелеты (точка a -- образ при преобразовании Нагаты вертикального слоя на гиперболоиде, проходящего через особую точку кривой из стенки).

Для камер канонические графы гиперболических кривых, отвечающих кривым на гиперболоиде со схемами <(3,±m)>, m = 0, 2, 4, указаны на рис. 5. Канонические скелеты негиперболических кривых, отвечающих кривым на гиперболоиде со схемами <(1,±m), l>, $0 \le l \le 6$, m = 0, 2, и <(1,±4)> указаны на рис. 6.

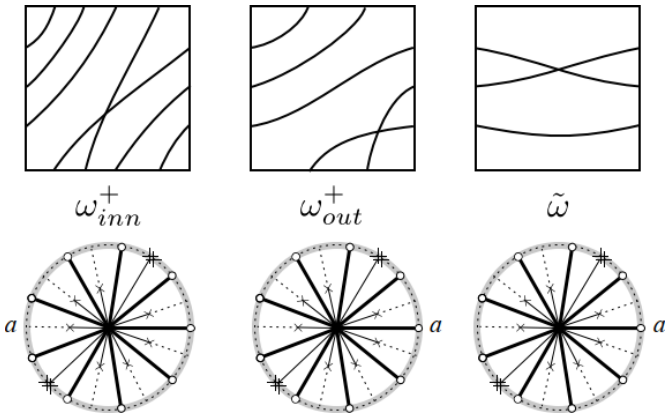


Рис. 2. Стенки: канонические графы гиперболических кривых

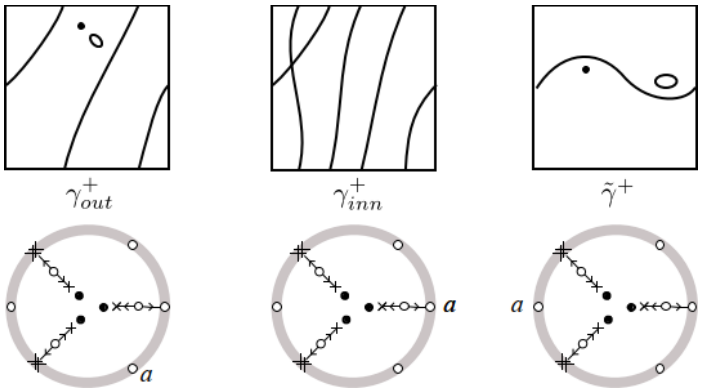


Рис. 3. Стенки: канонические графы негиперболических кривых

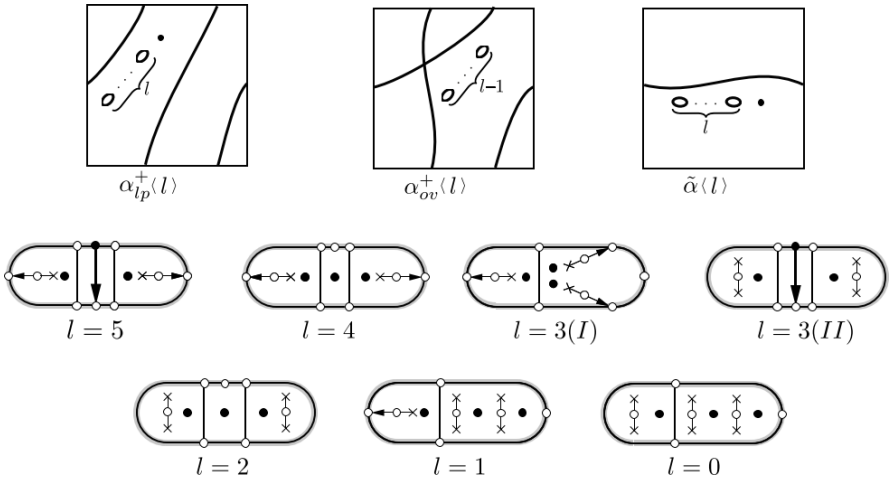


Рис. 4. Стенки: канонические графы негиперболических кривых

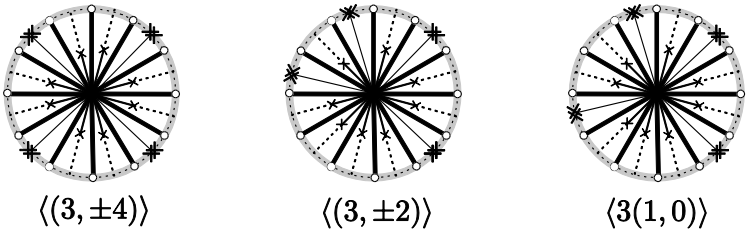


Рис. 5. Камеры: канонические графы гиперболических кривых

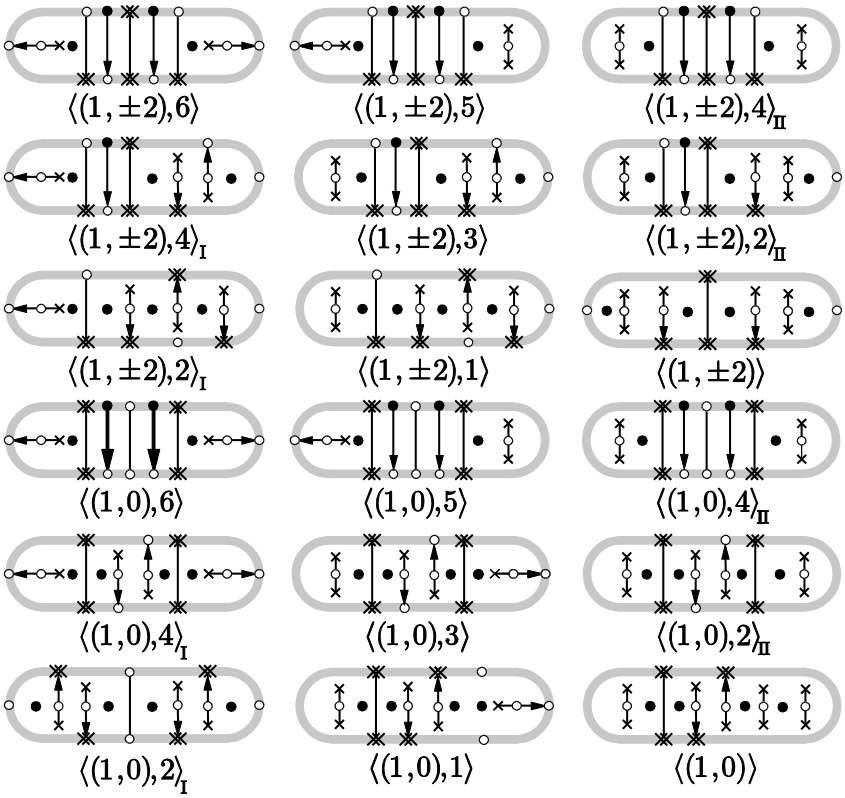


Рис. 6. Камеры: канонические скелеты негиперболических кривых

Работа выполнена в рамках государственного задания FSWR-2023-0034, а также по программе РНОМЦ «Математика технологий будущего».

Литература

1. Звонилов В.И. Графы тригональных кривых и жесткие изотопии особых веществен-ных алгебраических кривых бистепени (4,3) на гиперболоиде // Математика и теорети-ческие компьютерные науки. 2025. Т. 3. Вып. 1. С. 26–51.
2. Звонилов В.И. Комплексные топологические инварианты вещественных алгебраиче-ских кривых на гиперболоиде и эллипсоиде // Алгебра и анализ. 1991. Т. 3. Вып. 5. С. 88–108.
3. Звонилов В.И. Жесткая изотопическая классификация вещественных алгебраических кри-вых бистепени (4,3) на гиперболоиде // Вестник Сыктывкарского университета. 1999. Сер. 1. Вып. 3. С. 81–88.
4. Degtyarev A., Itenberg I., Kharlamov V. Real Enriques surfaces. Lecture Notes in Math., 2000, vol. 1746 , Springer-Verlag, 259 p.
5. Звонилов В. И. Жесткая изотопическая классификация вещественных алгебраических кри-вых бистепени (4,3) на гиперболоиде. Приложение // Вестник Сыктывкарского уни-верситета. 2003. Сер. 1. Вып. 5. С. 239–242.

РЕШЕНИЕ ЗАДАЧИ О НАЗНАЧЕНИЯХ ФИКСИРОВАННОЙ РАЗМЕРНОСТИ ПРИ ПОМОЩИ НЕЙРОННЫХ СЕТЕЙ

Д.К. Кармаев, Л.Г. Афраймович

Нижегородский государственный университет им. Н.И. Лобачевского

В работе исследуются возможности применения машинного обучения для решения двухиндексной задачи о назначениях фиксированной размерности. Дано описание реализованных в ходе работы нейросетевых моделей, вариантов алгоритмов обработки входных и выходных данных. Сделаны оценки эффективности нейросетевых моделей в сравнении с жадным алгоритмом. Сделаны выводы о наиболее перспективных подходах к созданию и обучению нейросети.

Ключевые слова: задачи комбинаторной оптимизации, линейная задача о назначениях, машинное обучение, свёрточные нейронные сети.

1. Введение

Многие задачи комбинаторной оптимизации характеризуются экспоненциальным или же факториальным ростом размера пространства возможных решений по отношению к размеру задачи. Это делает затруднительным применение алгоритмов поиска точного решения для случаев большого размера задачи, критичных к временной и пространственной вычислительной сложности.

Перспективным путём решения данной проблемы является разработка алгоритмов со сравнительно низкой вычислительной сложностью, позволяющих в большинстве случаев найти решение, близкое к оптимальному. Одним из подходов, рассмотренных в данной работе, является использование для поиска решения нейронных сетей. Данный подход исследуется на примере двухиндексной задачи о назначениях.

2. Описание задачи о назначениях

В ходе работы было рассмотрена задача распределения LSAP. Она может быть сформулирована как задача целочисленного линейного программирования:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \to max$$

с условиями

$$\sum_{j=1}^{n} x_{ij} = 1, \quad j = \overline{1,n}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad i = \overline{1,n}$$

$$x_{ij} \in \{0,1\}, \quad i,j = \overline{1,n}$$

Задача предполагает использование матрицы $C = \{c_{ij}\}$ в качестве входных и матрицы $X = \{x_{ij}\}$ в качестве выходных данных.

Для поиска оптимального решения двумерной задачи о назначениях может использоваться венгерский алгоритм полиномиальной сложности [3]. Несмотря на это, подход, описанный в данной работе, планируется обобщить для решения NP-трудной трёхиндексной задачи.

Существование алгоритмов полиномиального сведения NP-трудных задач к задаче о назначениях позволяет распространить выводы об эффективности решения трёхмерной задачи о назначения на прочие NP-трудные задачи.

3. Нейросетевая модель

В статье [1] приведены данные, демонстрирующие эффективность применения свёрточной архитектуры нейросети для решения задачи комбинаторной оптимизации. В ходе работы была реализована модель, основанная на применении такого типа нейросети.

Условие задачи задаётся как матрица размера $n * n$ со значениями из промежутка $[0, 1)$. В зависимости от размерности решаемой задачи меняется только размерность входного и выходного слоёв, без изменения внутренней архитектуры, что позволяет оценить возможность увеличения размерности задачи при схожем числе параметров.

Программная реализация алгоритмов осуществлялась на языке Python с использованием библиотеки машинного обучения ThensorFlow и библиотеки Keras.

3.1. Архитектура свёрточной нейронной сети

Архитектура нейросети представляет собой облегчённый вариант сети VGG с 11 весовыми слоями. Конфигурация промежуточных слоёв указана в Таблице 1. В качестве функции активации во всех слоях, кроме последнего, используется relu, в последнем слое используется функция softmax. Используется функция потерь categorical crossentropy и регуляризация SGD.

Таблица 1. Архитектура промежуточных слоёв

Тип слоя	Число каналов	Размер ядра	Количество
Convolution2D	8	3*3	1
Convolution2D	16	3*3	1
Convolution2D	32	3*3	2
Convolution2D	64	3*3	4
Dense	512	-	1
Dense	512	-	1

Было установлено, что нейросеть со сравнимым числом параметров, состоящая только из свёрточных слоёв, не даёт сколько-нибудь приемлемой точности решения. Обязательность добавления полносвязных слоёв означает невозможность применения единожды обученной нейросети для решения задач размерности большей, чем в наборе обучающих данных.

3.2. Алгоритм исключения коллизий

Решения, получаемые при помощи нейросети, могут не удовлетворять ограничению $\sum_{i=1}^{n} x_{ij} = 1, \quad i = \overline{1,n}$. Для преобразования решения к корректному используется алгоритм предотвращения столкновений. Выходные данные нейросетей обрабатываются таким образом, что каждому исполнителю назначается ровно одна работа.

Алгоритм последовательно выбирает максимальные элементы из матрицы $n \times n$, образованной поэлементным умножением матриц входных и выходных значений, и присваивает 1 элементу матрицы решения с тем же индексом. Строки и столбцы, уже содержащие значение 1, в дальнейшем не изменяются.

Чем больше точность получаемых решений, тем реже возникает необходимость в применении алгоритма исключения коллизий. В случае уровня эффективности нейросети, близкого к оптимальному, коллизий не возникает.

3.3. Обучение и тестирование

Исходные данные для одной задачи размерности n представляют собой матрицу $n * n$ с элементами из диапазона $[0, 1)$. При генерации данных обычным способом значения задаются случайно, на основе равномерного распределения. При генерации геометрическим способом на плоскости генерируются два набора по n точек. Значением элемента матрицы a_{ij} является геометрическое расстояние от точки i первого набора до точки j второго набора.

Для обучения использовалось 50 наборов по 1000 примеров, обучение производилось в 30 эпох.

4. Оценки эффективности модели

Для оценки точности решения после обучения набора нейросетей используются средняя эффективность $\frac{1}{n}\sum_{i=1}^{n}\frac{C(X_i)}{C(X_i^*)}$,

где: n – количество проверочных примеров, $C(X)=\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij}$ – значение целевой функции, X_i^* – оптимальное решение, X_i – найденное решение.

Точность решений, полученных с использованием нейросетей, сравниваются с точностью решений, полученных при помощи жадного алгоритма.

4.1. Полученные результаты

В Таблице 2 представлены полученные значения средней эффективности получаемых решений для различных размерностей задачи.

Таблица 2. Средняя эффективность

Размер задачи	Метод решения	Метод генерации данных	
		Обычный	Геометрический
10	Нейросетевая модель	0,94	0,97
	Жадный алгоритм	0,92	0,94
16	Нейросетевая модель	0,95	0,97
	Жадный алгоритм	0,93	0,95
32	Нейросетевая модель	0,96	0,98
	Жадный алгоритм	0,95	0,96

Продемонстрирована возможность повышения порядка задачи без существенного увеличения количества параметров, а следовательно, и числа операций в нейросети. Количество изменяемых параметров нейросети в случае размерности задачи n приведено в Таблице 3.

Таблица 3. Количество изменяемых параметров

Размерность	Число параметров
4	448560
10	491652

Модель сохраняет эффективность даже на больших размерностях, демонстрируя эффективность решения примерно на 1% лучшую, чем у жадного алгоритма. Полученные данные об эффективности нейросетевого подхода, алгоритмов преобразования входных и выходных данных могут быть использованы для моделей схожего принципа, например, для решения NP-трудной трёхиндексной задачи о назначениях.

Литература

1. Mengyuan Lee, Yuanhao Xiong, Guanding Yu, Geoffrey Ye Li. Deep Neural Networks for Linear Sum Assignment Problems // IEEE Wireless Communications Letters. December, 2018. Vol. 7, No. 6. P. 962–965. DOI: 10.1109/LWC.2018.2843359.
2. Shenghe Xu, Shivendra S. Panwar, Murali Kodialam, T.V. Lakshman. Deep Neural Network Approximated Dynamic Programming for Combinatorial Optimization. The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20). 2020. P. 1684–1691. DOI: 10.1609/aaai.v34i02.5531.
3. J. Munkres. Algorithms for the Assignment and Transportation Problems // Journal of the Society for Industrial and Applied Mathematics. March, 1957. Vol. 5, No. 1. P. 32–38. DOI: 10.1137/0105003.

ПРИНЯТИЕ РЕШЕНИЙ ПРИ ОПТИМИЗАЦИИ УПРАВЛЕНИЯ ОБРАЗОВАТЕЛЬНЫМ ПРОЦЕССОМ С ИСПОЛЬЗОВАНИЕМ МОДЕЛЕЙ ТЕОРИИ ИГР

Е.В. Кондрашова

Московский государственный строительный университет (НИУ МГСУ)

Объектом исследования является образовательный процесс, который рассматривается с позиции оптимизации управления при помощи теоретико-игровых моделей. В результате исследования были построены несколько моделей, как например, теоретико-игровая модель для выбора оптимальной стратегии ВУЗа, где администрация, основываясь на мотивации со стороны студенческого сообщества, принимает решения относительно выбора системы оценивания. После аналитического расчета был реализован программный код для автоматизации вычисления оптимальной стратегии, что позволило повысить точность и надежность расчета и обеспечить легкую адаптируемость под новые данные.

Ключевые слова: теоретико-игровые модели, математические модели, управление образовательным процессом, оптимальная стратегия, принятие решений.

1. Введение

В современном мире образование играет ключевую роль в развитии общества и экономики. От того, насколько грамотно организован учебный процесс, зависит не только качество подготовки будущих специалистов, но и конкурентоспособность образовательных учреждений. Современная образовательная система сталкивается с рядом проблем, таких как неэффективное распределение и использование ресурсов, дефицит коммуникации между участниками образовательного процесса и низкая адаптивность к непрерывно изменяющимся социальным и технологическим условиям [1].

В управлении образовательным процессом часто возникают ситуации, когда участники (студенты, преподаватели, администрация) имеют различные цели. При принятии управленческих решений необходимо учитывать, что действия одной стороны напрямую влияют на успех или неудачу других [2]. Именно такие ситуации принято называть «конфликтными ситуациями». Для выявления оптимального поведения участников в таких ситуациях при принятии решений возможно использование аппарата теории игр [3-5]. Теоретико-игровые подходы помогают анализировать взаимодействие участников, прогнозировать их поведение и приходить к оптимальным стратегиям. Однако моделирование реальных образовательных процессов сталкивается с трудностями, связанными со сложностью точного математического описания процессов и влиянием внешних факторов.

Далее рассматривается реализация одной из моделей оптимизации управления образовательным процессом.

2. Модель выбора оптимальной системы оценивания

Модель использует понятие игры с природой – парной игры с ограниченным числом стратегий, в которой решения принимает администрация ВУЗа для выбора оптимальной системы оценивания. В играх с природой выбор оптимальной стратегии основан на различных критериях принятия решений [6]. Участниками игры являются два игрока: администрация вуза разрабатывает и принимает различные системы оценивания, и студенческое общество, которое может оказаться в одном из своих состояний касательно формы посещения и отношения к занятиям. В рамках модели были определены возможные стратегии для каждой из сторон. Так, например, администрация выбирает одну из следующих систем оценивания: традиционная си-

стема оценивания, где на оценку влияет только итоговый экзамен; система накопленного балла, где весовой коэффициент за экзамен больше, чем весовой коэффициент за работу в семестре и т.д. Состояния студенческого сообщества могут быть определены как: игнорирование посещения учебных занятий; посещение только практических занятий при игнорировании лекционных занятий и т.д. Всего было сформировано пять стратегий (состояний) студенческого сообщества и пять стратегий администрации.

Для определения значений элементов матрицы выигрышей, был проведен опрос среди студентов различных курсов и специальностей подготовки, на основании которого были оценены выигрыши в зависимости от возможной ситуации. Матрица выигрышей составлялась с использованием данных опроса, учитывая, что для администрации ВУЗа выигрышем является максимальная заинтересованность и оценка мотивации со стороны студентов. Мотивация отражала, насколько система оценивания стимулирует учебный процесс и желание студентов учиться, эффективность – насколько каждая из систем оценивания способствует повышению качества образования и повышению результатов студентов. Оценка проводилась по десятибалльной шкале, где 1 – крайне неудовлетворительно, а 10 – максимально удовлетворительно. Было получено две матрицы (для мотивации и эффективности), при составлении которых использовались средние значения, так как близкие значения между средним, модой и медианой продемонстрировали, что распределение данных, симметрично, и среднее значение отражает центральную тенденцию данных.

В результате проведенных расчетов по различным критериям выбора оптимальной стратегии – по критериям Вальда, Сэвиджа, Гурвица, Байеса и Лапласа – было установлено, что как по показателям мотивации, так и по показателям эффективности обучения наилучшие результаты демонстрирует стратегия – выбор системы оценивания без экзамена, при которой оценка формируется только на основе работы в семестре.

3. Заключение

После аналитического расчета был реализован программный код для автоматизации вычисления оптимальной стратегии, что позволило повысить точность и надежность расчета и обеспечить легкую адаптируемость под новые данные. При этом оптимальность стратегии в контексте каждой задачи может зависеть от определенных критериев.

Теоретико-игровые модели предоставляют возможность формализовать взаимодействия между участниками образовательного процесса, что позволяет количественно анализировать стратегии, используемые различными участниками. Выбор (формулировку) стратегий и составление матрицы выигрышей следует делать, исходя из конкретной ситуации.

Литература

1. Ghic G. Applications of Games Theory in Analyzing Teaching Process / G. Ghic, C.J. Grigorescu // Procedia – Social and Behavioral Sciences. – 2014. – Т. 116. – С. 3588–3592.
2. Сигалова О.М. Теоретико-игровое моделирование экономического поведения в условиях отсутствия общего знания рациональности // Вестник экономического факультета МГУ. Научные исследования экономического факультета. Электронный журнал. – 2012. – Т. 4, № 2. – С. 49-71.
3. Ho E. Game Theory in Defence Applications: A Review / E. Ho, A. Rajagopalan, A. Skvortsov // Sensors. – 2022. – Т. 22, № 3.
4. Сачкова Н.А. Теория игр как объясняющая схема социальных процессов: на материале педагогики // Образование и наука. – 2010. – № 11 (79). – ISSN 2310-5828.
5. Чеботарёва Т.А., Рассказов Д.Е. Практическое применение теории игр в современной экономике и управлении // Вестник Института мировых цивилизаций. – 2018. – Т. 9, № 1 (18). – С. 65–69.
6. Aumann R.J. Lectures of game theory. – URL: https://perso.uclouvain.be/pierre.dehez/Aumann/Aumann-Lectures.pdf (дата обращения: 05.08.2025). – ISBN 0-8133-7578-9. – 128 с.

ОБ ОПЫТЕ ПРИМЕНЕНИЯ АЛГОРИТМА МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ ДЛЯ НАСТРОЙКИ ГИПЕРПАРАМЕТРОВ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

С.Ю. Коннов, Е.А. Козинов

Нижегородский государственный университет им. Н.И. Лобачевского

Задача настройки гиперпараметров моделей машинного обучения часто представляет собой сложный процесс перебора значений, влияющих как на точность модели, так и на скорость её работы. Для оптимизации данных критериев могут применяться различные методы многокритериальной оптимизации. В рамках статьи для решения поставленной задачи рассматривается авторский алгоритм поиска множества Парето на основе алгоритма Стронгина. Метод включает в себя использование различных моделей машинного обучения для повышения эффективности аппроксимации множества Парето. Эффективность работы алгоритма оценивается на задаче настройки гиперпараметров модели XGBoost. В статье приведено сравнение качества работы алгоритма с применением различных моделей.

Ключевые слова: настройка гиперпараметров, задачи многокритериальной оптимизации, алгоритмы глобального поиска, машинное обучение, множество Парето

1. Введение

Процесс проектирования и обучения моделей машинного обучения неразрывно связан с настройкой изначальных гиперпараметров, от которых зависит как точность, так и эффективность работы модели. Подбор оптимальных значений гиперпараметров представляет собой сложную задачу, требующую баланса между качеством модели и вычислительными затратами. Существуют различные традиционные подходы к решению данной задачи. Поиск по сетке предлагает зафиксировать наборы гиперпараметров фиксированным количеством значений. При проведении испытаний модель обучается и тестируется на каждой комбинации гиперпараметров. В методе случайного поиска вместо набора фиксированных значений для каждого гиперпараметра задаётся своё случайное распределение. Комбинация гиперпараметров составляется семплированием этих распределений. Более сложным методом является Байесовская оптимизация. Данный метод использует вероятностную модель зависимости между гиперпараметрами и целевой функцией. На каждой итерации алгоритм выбирает новые точки в пространстве гиперпараметров, максимизирующие ожидаемое улучшение целевой метрики [1,2].

Проблема традиционных подходов к решению задачи заключается в том, что они зачастую ориентированы на однокритериальную оптимизацию. Для решения задачи с несколькими целями, учитывающей как точность, так и скорость работы модели, могут применяться различные алгоритмы многокритериальной оптимизации.

Целью данной работы является применение алгоритма многокритериальной оптимизации, ранее реализованного нами в работе [3], для задачи настройки гиперпараметров модели машинного обучения.

2. Метод решения

В общем виде задача многокритериальной оптимизации может быть сформулирована следующим образом:

$$\Phi(y) = \left(f_1(y), f_2(y), \dots, f_s(y)\right) \to min, y \in D, \quad (1)$$

где $f_i(y)$, $i \in [1, s]$ – критерии эффективности, $y_i, i \in [1, N]$ – набор варьируемых параметров, N – размерность решаемой задачи, s – число оптимизируемых критериев. Каждый критерий

в такой постановке задачи является многоэкстремальной функцией, заданной в виде «черного ящика» и удовлетворяющей условию Липшица

$$|f_i(y') - f_i(y'')| \leq L_i\|y' - y''\|, y', y'' \in D, i \in [1, s], \quad (2)$$

константы Липшица $L_i, i \in [1, s]$ заранее неизвестны.

В работе [3] нами был разработан алгоритм для решения задачи многокритериальной оптимизации. Его можно разделить на две составляющих. Первая основана на алгоритме Стронгина [4]. Задача с несколькими критериями (1) скаляризуется в семейство однокритериальных задач отличающихся набором параметров свертки. Размерность задач редуцируется с помощью кривых Пеано. Во время работы алгоритма на очередной итерации выделяются интервалы между проведенными испытаниями по одномерным координатам. Для каждого интервала подчитывается характеристика, на основании которой принимается решение о добавлении следующей точки проведения испытания. Алгоритм работает до достижения требуемой точности или до ограничения по количеству испытаний.

Вторая составляющая основана на использовании методов машинного обучения для оценки множества Парето. На этапе выбора интервала нового испытания к характеристикам добавляется значение, величина которого зависит от степени уверенности принадлежности образующих интервал точек к области Парето. Во время работы алгоритма все полученные точки проведения испытаний разделяются на два класса. Первый класс включает в себя точки, принадлежащие множеству Парето. Все остальные точки относятся ко второму классу. Для оценки добавляемых значений использовались следующие подходы:

1. Подход с построением разделяющей гиперплоскости между выделенными классами и поиском расстояния от каждой точки проведения испытания до найденной гиперплоскости. Для нахождения гиперплоскости в вычислительных экспериментах использовалась модель SVC с линейным ядром.

2. Подход с поиском вероятности принадлежности каждой точки к области Парето. Для этого варианта в вычислительных экспериментах использовались модели SVC с разными типами ядер (линейное, полиномиальное, RBF) и нейросетевая полносвязная модель с двумя скрытыми слоями размерностью 128 и 64 узлов соответственно.

В данной работе был использован следующий подход по сведению задачи подбора гиперпараметров к задаче многокритериальной оптимизации. Набор гиперпараметров интерпретируется как варьируемые параметры задачи (1). Критериями задачи являются среднее время работы модели и точность вывода на валидационной выборке. На каждой итерации алгоритма глобального поиска происходит обучение, валидация и сбор значений критериев модели с заданными гиперпараметрами.

3. Результаты вычислительных экспериментов

Для демонстрации решения задачи настройки гиперпараметров была выбрана модель XGBoost, применяемая для многоклассовой классификации. В качестве входных данных был использован набор данных PASCAL VOC 2007 [5]. Он включает в себя набор изображений с 20 категориями объектов. Каждое изображение может включать в себя объект из одной и более категорий. В качестве варьируемых гиперпараметров были выбраны количество арбитров (от 20 до 300) и глубина каждого дерева (от 3 до 10).

Точность работы построенной модели на выбранном наборе данных определялась по формуле средней точности (mean average precision, mAP)

$$\mathrm{m}AP = \frac{1}{N}\sum_{i=1}^{N} AP_i, \quad (3)$$

где N – число классов, $AP_i, i \in [1, N)$ – площадь под кривой «точность-полнота» (Precision-Recall) для класса i.

Вычислительные эксперименты проводились на узлах суперкомпьютера «Лобачевский» Нижегородского государственного университета с двумя процессорами Intel Xeon Silver 4310T 2.3 GHz, 64 GB RAM. Использовались библиотеки машинного обучения SciKit-Learn 0.24.2, PyTorch 2.5.1 и XGBoost 2.1.1. Предложенный алгоритм реализован на языке Python с использованием фреймворка iOpt [6].

В испытаниях участвовали вариации алгоритма без использования методов машинного обучения (MGSA) и с использованием различных вероятностных моделей (MGSA Linear SVC, MGSA Poly SVC, MGSA RBF SVC, MGSA NN). Для каждого запуска было выставлено ограничение в 400 испытаний.

На рисунке 1 представлены кривые множества Парето, найденные при работе алгоритма с различными моделями машинного обучения. Кривая, расположенная левее и выше соответствует лучшей работе алгоритма. Рисунок 2 демонстрирует значения площадей под графиками из рисунка 1, ограниченные снизу значением точности в 0.74, справа значением времени в 0.075 сек. Чем больше значение площади под графиком, тем лучше.

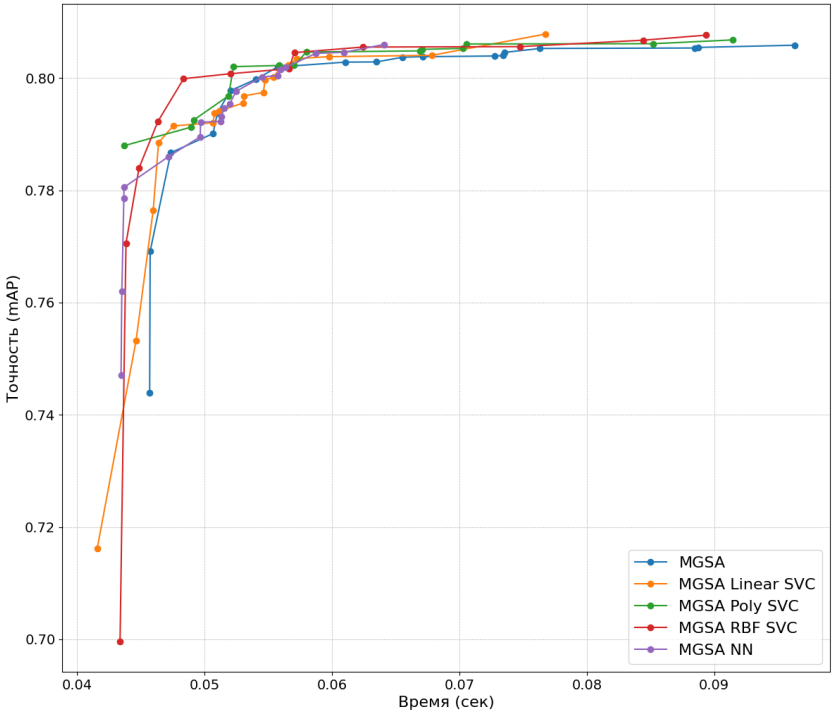


Рис.1. Значения множеств Парето для различных запусков алгоритма MGSA

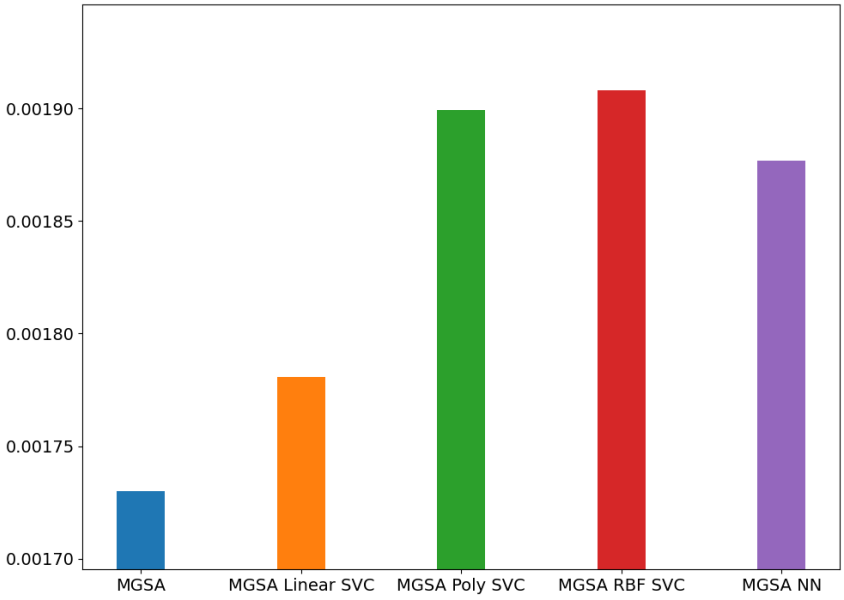


Рис.2. Сравнение значений площадей под графиками для различных запусков алгоритма MGSA

В среднем практически на всей области критериев вариации алгоритма MGSA с использованием моделей машинного обучения предлагают более оптимальные значения критериев. Со-

гласно графикам и значениям площадей под ними, наиболее оптимальные результаты были получены при работе алгоритма с моделью SVC, использующей ядро RBF. Близкие результаты также показала модель SVC с ядром Poly SVC.

Следует также отметить, что алгоритм MGSA позволяет для каждой точки из множества критериев предложить определённый набор гиперпараметров, при которых данных значений критериев можно будет достичь. Пример подобных параметров для алгоритма MGSA с моделью SVC (ядро RBF) представлен в таблице 1.

Таблица. 1. Значения гиперпараметров для модели XGBoost и соответствующие им значения критериев

Количество арбитров	Глубина деревьев	Время работы	Точность (mAP)
20	2	0.043	0.699
48	4	0.043	0.770
96	3	0.044	0.783
101	4	0.046	0.792
189	3	0.048	0.799
229	4	0.052	0.800
245	3	0.056	0.801
249	5	0.057	0.804
296	4	0.062	0.805
278	5	0.074	0.805
282	6	0.084	0.806
282	7	0.089	0.807

Таблица значений подтверждает утверждение о том, что при выбранных гиперпараметрах время работы и точность являются конфликтующими параметрами: при увеличении значений параметров (и усложнении модели) растут как время работы, так и точность. Можно дополнительно отметить порог точности в 0.804, при котором дальнейшее увеличение сложности модели практически не даёт эффекта по точности, однако существенно увеличивает время её работы. Аналогично, упрощение модели со временем работы 0.048 секунд приводит к значительному снижению точности при минимальном увеличении скорости.

Таким образом, можно сделать вывод, что задача настройки гиперпараметров модели машинного обучения может быть успешно сведена к задаче многокритериальной оптимизации. Алгоритм MGSA позволяет найти различные наборы гиперпараметров, в зависимости от которых будут меняться точность и скорость работы модели. Следует также отметить, что подходы с использованием методов машинного обучения позволяют получить более оптимальные результаты вычислений за одно и то же количество итераций. В дальнейшем планируется продолжить работы по исследованию применения алгоритма MGSA для различных моделей машинного обучения, а также сравнить полученные результаты с другими алгоритмами многокритериальной оптимизации (Optuna, NSGAII [7, 8]).

Литература

1. Анафиев А.С., Карюк А.С. Обзор подходов к решению задачи оптимизации гиперпараметров для алгоритмов машинного обучения // Таврический вестник информатики и математики. 2022. №. 2(55). С. 30-37.
2. Горшенин А.К., Кузьмин В.Ю. Оптимизация гиперпараметров нейронных сетей с использованием высокопроизводительных вычислений для предсказания осадков // Информатика и её применения. 2019. №. 1(13). С. 75-81.
3. Konnov S. et al. Comparative Efficiency of Machine Learning Models for Enhancing Algorithms in Solving Multiextremal Multicriteria Problems // International Conference on Optimization and Applications. Cham : Springer Nature Switzerland, 2024. P. 109-124.
4. Стронгин Р.Г. Численные методы в многоэкстремальных задачах (информационно-статистические алгоритмы). М.: Главная редакция физико-математической литературы издательства "Наука", 1978. 240 с.

5. Everingham M. et al. The Pascal Visual Object Classes (VOC) Challenge. Int. J. Comput. Vis. 2010. Vol. 88. P. 303-338. DOI: 10.1007/s11263-009-0275-4.
6. Исходный код алгоритма многокритериальной оптимизации iOpt. URL: https://github.com/aimclub/iOpt (Дата обращения: 12.10.25).
7. Akiba T. et al. Optuna: A next-generation hyperparameter optimization framework // Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019. P. 2623-2631.
8. Deb K., Pratap A., Agarwal, S. Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II // IEEE Trans. Evol. Comput. 2002. Vol. 6, No. 2. P. 182-197. DOI: 10.1109/4235.996017.

ИТЕРАТИВНОЕ УТОЧНЕНИЕ ПРИЗНАКОВ В СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЯХ С ПОМОЩЬЮ РЕКУРРЕНТНЫХ БЛОКОВ[1*]

Р.А. Кононов[1,2], О.В. Масленников[1,2], В.И. Некоркин[1,2]

[1]Институт прикладной физики им. А.В. Гапонова-Грехова РАН,
[2]Нижегородский государственный университет им. Н. И. Лобачевского

В работе исследуется архитектура сверточных нейронных сетей с адаптивной вычислительной сложностью для задач классификации изображений. Мы вводим рекуррентный блок, который позволяет итеративно уточнять внутреннее представление данных. Это дает возможность динамически управлять компромиссом между точностью классификации и вычислительными затратами, варьируя число итераций. Проанализированы два варианта архитектуры: рекуррентная свертка в начале сети, которая позволяет достичь существенного прироста точности ценой увеличения вычислений и рекуррентный классификатор, который действует как эффективный регуляризатор, улучшая точность базовой модели прямого распространения даже при выполнении всего одной итерации.

Ключевые слова: сверточные нейронные сети, адаптивные вычисления, рекуррентные блоки, компромисс точность-скорость, динамическая архитектура, классификация изображений, регуляризация

1. Введение

Стандартные сверточные нейронные сети (CNN) обладают фиксированной вычислительной сложностью: для любого входного изображения выполняется одно и то же количество операций. Такой подход неэффективен, поскольку простые и очевидные изображения требуют меньше обработки, чем сложные или зашумленные. В последние годы активно развивается направление сетей с адаптивными вычислениями, способных динамически изменять объем вычислений в зависимости от сложности входных данных или требуемой точности [1, 2]. Одним из ранних подходов в этой области стали сети с «ранним выходом» (early exiting), которые добавляют промежуточные классификаторы на разных уровнях глубины сети, позволяя получить быстрый ответ для простых случаев [3].

Более элегантным способом реализации адаптивных вычислений является введение рекуррентности в архитектуру прямого распространения. Вместо увеличения глубины сети за счет добавления новых слоев, можно многократно применять один и тот же блок, итеративно уточняя карту признаков. Такой подход не только позволяет эффективно переиспользовать параметры, но и имеет биологические аналоги: зрительная кора головного мозга также содержит многочисленные рекуррентные связи для обработки визуальной информации [4]. Идеи итеративного уточнения с помощью рекуррентных сверточных слоев были успешно применены для задач распознавания объектов ранее [5].

В данной работе мы исследуем простую, но эффективную реализацию этой идеи на основе рекуррентного блока с остаточной связью (residual connection). Мы рассматриваем две архитектуры: с рекуррентностью на уровне низкоуровневых признаков (сверточный слой) и на уровне высокоуровневых признаков (полносвязный слой классификатора). Наша цель – продемонстрировать, как число рекуррентных итераций напрямую влияет на компромисс между точностью и вычислительными затратами, а также исследовать регуляризующий эффект, который такой метод обучения оказывает на модель.

[1*] Работа выполнена при поддержке гранта Российского научного фонда № 23-72-10088.

2. Модель

Для решения задачи классификации изображений из набора данных CIFAR-10 мы использовали базовую сверточную архитектуру, в которую вносили рекуррентные модификации. Ниже описаны архитектура, устройство рекуррентного блока и процедура обучения.

2.1 Базовая архитектура

Основа представляет собой сеть прямого распространения, состоящую из следующих блоков (см. рис. 1):

Сверточный блок 1: Входная свертка (3 канала -> 96 каналов), ReLU.

Сверточный блок 2: Свертка (96 каналов -> 96 каналов), ReLU.

Сверточный блок 3 и 4: Два последовательных сверточных слоя с увеличенным шагом (stride) для уменьшения пространственной размерности карты признаков. После каждого слоя следует нормализация по пачке и функция активации ReLU. Эти слои также уменьшают число карт признаков с 96 до 32.

Классификатор: Трехслойный полносвязный перцептрон (MLP), принимающий на вход сжатое представление признаков и выдающий логиты для 10 классов.

В наших экспериментах мы поочередно преобразовывали в рекуррентный либо *сверточный блок 2*, либо первый полносвязный слой классификатора. На рисунке 2 эти блоки обозначены пунктирной стрелкой. Также для эти блоки снабжены обходной связью, это необходимо для более качественного обучения в рекуррентном режиме.

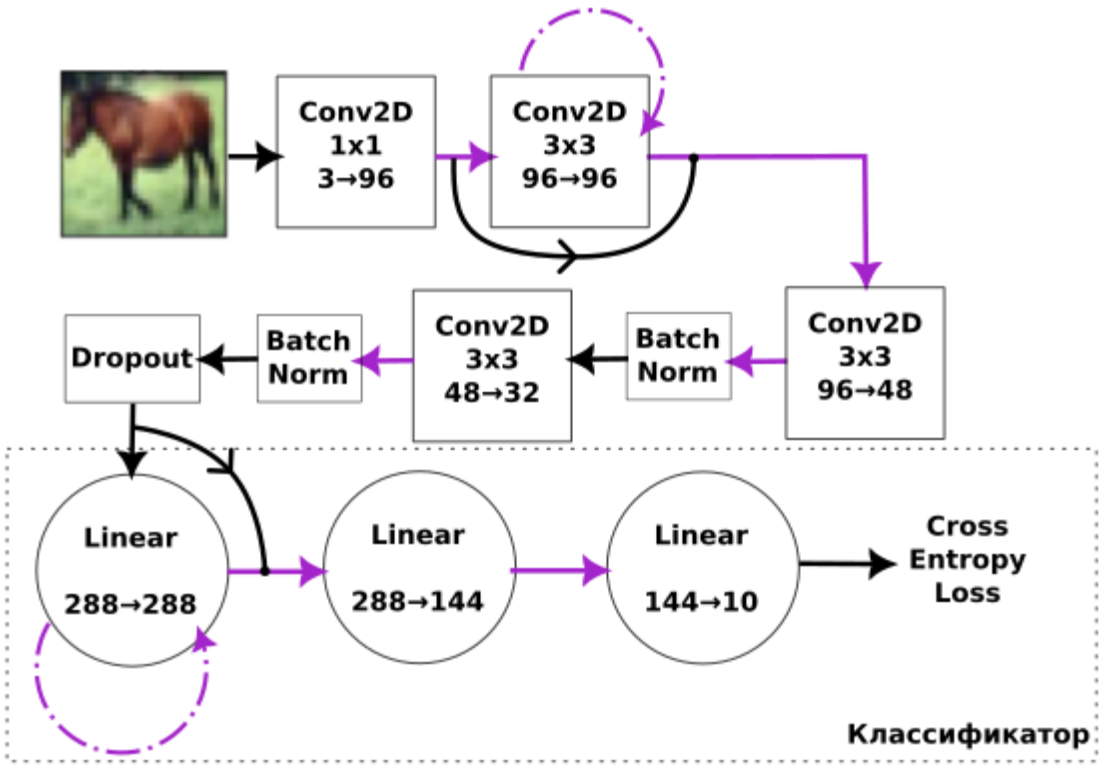


Рис. 1. Схема архитектуры, которая рассматривается в работе. Фиолетовыми стрелками обозначено применение функции активации ReLU

2.2. Рекуррентный блок

Преобразование стандартного блока в рекуррентный заключается в его итеративном применении с использованием остаточной связи. Пусть $F(x)$ – это операция, выполняемая блоком

(например, свертка + ReLU). Тогда состояние y на временном шаге t+1 вычисляется на основе состояния на шаге t следующим образом:

$$y_{t+1} = F(y_t) + y_t$$

где $y_0 = x$ – это вектор признаков, поданный на вход рекуррентного блока. Таким образом, блок представляет собой дискретную динамическую систему. На каждом шаге t итерации мы можем получить промежуточный результат y_t и передать его дальше по сети для получения ответа.

2.3. Процедура обучения

Обучение проводилось на пачках размером 200 изображений с использованием оптимизатора RMSProp. В процессе обучения рекуррентный блок разворачивался на T=13 итераций. Для каждой итерации $t \in [1, T]$ вычислялся выход сети и соответствующая ему функция потерь (кросс-энтропия). Итоговая функция потерь, которая минимизировалась, представляла собой сумму логарифмов от функций потерь по всем итерациям:

$$L_{total} = \sum_{t=1}^{T} ln(Loss(Output(y_t), Target))$$

Такой подход стимулирует сеть выдавать корректный результат как можно раньше, но при этом позволяет ей уточнять его на последующих итерациях. Обучение останавливалось при начале роста ошибки на валидационной выборке для предотвращения переобучения.

3. Результаты

На рисунке 2 показана зависимость ошибки классификации на валидационной выборке от числа операций с плавающей запятой (FLOPs), необходимых на проход по модели для: базовой сети прямого распространения, сети с рекуррентной сверткой и сети с рекуррентным классификатором. Каждая точка на траекториях рекуррентных моделей соответствует одной дополнительной итерации обработки. На рисунке 2 представлены значения функции ошибки для частных, но типичных представителей из обученных нами ансамблей.

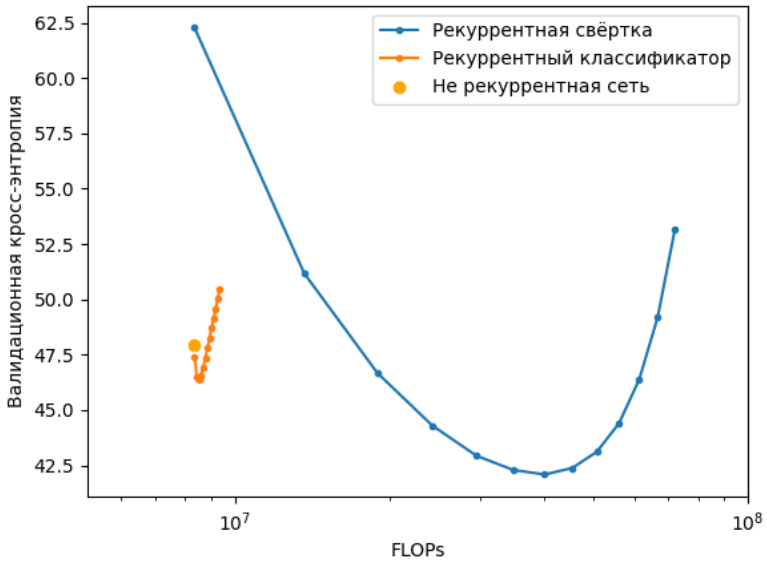


Рис. 2. Функция ошибки рассчитанная на валидационной выборке у рассматриваемых в работе сетей. На оси абсцисс отложено кол-во операций с плавающей запятой в логарифмическом масштабе

3.1. Сеть с рекуррентной сверткой

Эта модель демонстрирует наибольший потенциал для повышения точности.

Компромисс: На первой итерации (что эквивалентно одному проходу) ее точность уступает базовой модели. Однако с каждой последующей итерацией ошибка монотонно снижается, и на 3-5 итерациях модель превосходит базовую, в итоге достигая наименьшей ошибки среди всех рассмотренных архитектур.

Высокая цена: Этот прирост точности требует значительных вычислительных ресурсов. Чтобы достичь точности, сопоставимой с базовой моделью, требуется примерно втрое больше FLOPs. Это делает данный подход эффективным в сценариях, где максимальная точность важнее скорости.

3.2. Сеть с рекуррентным классификатором

Эта модель показывает иное поведение.

Эффект регуляризации: Самый примечательный результат заключается в том, что на первой же итерации (t=1) ошибка этой модели ниже, чем у базовой сети прямого распространения, хотя количество операций идентично. Это прямое следствие нашей процедуры обучения (суммирование потерь по всем итерациям), которая действует как сильный регуляризатор и заставляет сеть находить более качественное решение для одного прямого прохода.

Скромное улучшение: Последующие итерации приносят лишь незначительное снижение ошибки по сравнению с рекуррентной сверткой.

Таким образом, два подхода решают разные задачи. Рекуррентная свертка является инструментом для достижения максимальной точности при наличии вычислительных ресурсов. В то же время введение рекуррентного блока в классификатор с предложенной схемой обучения – это эффективный метод регуляризации, позволяющий улучшить качество базовой модели прямого распространения без каких-либо дополнительных затрат на этапе инференса (если используется только одна итерация).

4. Выводы

В данном исследовании была продемонстрирована эффективность использования рекуррентных блоков для создания сверточных нейронных сетей с адаптивной вычислительной сложностью. Мы показали, что такой подход позволяет гибко управлять компромиссом между точностью классификации и затраченными ресурсами.

Ключевые выводы работы:

- Рекуррентность на уровне сверток позволяет итеративно улучшать качество представления признаков и достигать значительно более высокой итоговой точности, чем у базовой сети. Однако этот прирост требует существенного увеличения количества вычислений.
- Рекуррентность на уровне классификатора, в сочетании с предложенной процедурой обучения, выступает в роли мощного регуляризатора. Главным результатом этого подхода является улучшение качества базовой модели даже при однократном прямом проходе (на первой итерации), что позволяет получить более точную и эффективную архитектуру без дополнительных затрат на этапе инференса.

Таким образом, введение рекуррентности является гибким инструментом: его можно использовать либо для достижения максимальной точности в задачах, не критичных ко времени ответа, либо для улучшения качества и обобщающей способности стандартных моделей прямого распространения.

Литература

1. Banino A., Balaguer J., Blundell C. PonderNet: Learning to Ponder // 8th ICML Workshop on Automated Machine Learning (AutoML). – 2021.
2. Dehghani, M., Gouws, S., Vinyals, O., Uszkoreit, J., & Kaiser, L. Universal Transformers. // International Conference on Learning Representations – 2019.
3. Teerapittayanon S., McDanel B., Kung H. T. Branchynet: Fast inference via early exiting from deep neural networks // 23rd international conference on pattern recognition (ICPR). – IEEE, 2016. – С. 2464-2469.
4. Soo W. et al. Recurrent neural network dynamical systems for biological vision // Advances in Neural Information Processing Systems. – 2024. – Т. 37. – С. 135966-135982.
5. Liang M., Hu X. Recurrent convolutional neural network for object recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – С. 3367-3375.

ФУНДАМЕНТАЛЬНЫЕ ПРЕДСТАВЛЕНИЯ ОРТОГОНАЛЬНОЙ АЛГЕБРЫ ЛИ И НОВЫЕ ПРОСТЫЕ ПОДАЛГЕБРЫ НЕАЛЬТЕРНИРУЮЩИХ ГАМИЛЬТОНОВЫХ АЛГЕБР ЛИ[1*]

М.И. Кузнецов, А.В. Кондратьева

Нижегородский государственный университет им. Н.И. Лобачевского

В докладе рассматривается действие ортогональной алгебры Ли *o(V)* на внешних степенях векторного пространства *V* размерности *n* с заданной невырожденной ортогональной формой. Найдены две исключительные простые подалгебры Ли $P_1(6)$, $P_2(6)$ в *P(6)*, размерности 2^5-1 и 2^6-1, соответственно, содержащие инвариантное подпространство пространства $\Lambda^3 V$, которые существуют только в случае 6 неизвестных. Основное поле – совершенное поле характеристика два.

Ключевые слова: совершенное поле характеристики два, неальтернирующие гамильтоновы алгебры Ли, фундаментальные представления.

Пусть *V* – *n*-мерное векторное пространство над совершенным полем *K* характеристики *2*, { , } – невырожденная симметрическая неальтернирующая билинейная форма на *V*, которую будем называть ортогональной. Ортогональная алгебра Ли *o(V)* состоит из всех линейных операторов *A* на *V* таких, что $\{Ax, y\} + \{x, Ay\} = 0$ для любых $x, y \in V$.

Для наших целей удобно рассматривать модель пары *(V, o(V))* как неположительную часть неальтернирующей гамильтоновой алгебры Ли $P(n) = L = L_{-1} + L_0 + L_1 + \ldots$, где *P(n)* – пространство многочленов *O(n)* в разделенных степенях от переменных $x_1, \ldots, x_n$ без свободного члена со скобкой Пуассона

$$\{f, g\} = \sum_{i=1}^{n} \partial_i f \partial_i g$$

Т.е. $L_{-1} = V = \langle x_1, \ldots, x_n \rangle$ и $L_0 = o(V) = \langle x_i x_j, x_i^{(2)}, i,j = 1,\ldots,n \rangle$, где $\{x_1, \ldots, x_n\}$ – ортонормированный базис относительно формы $\{x, y\}$, $x, y \in V$. Здесь $\{x, y\} \in K$. Обозначим $L_0' = o'(V) = \langle x_i x_j, i,j = 1,\ldots,n \rangle$.

Строение алгебры Ли *o(V)* над совершенным полем *K* описано в [1]. Определение и свойства неальтернирующих гамильтоновых алгебр Ли приведены в [1]-[4]. Некоторые свойства L_0-модулей L_k рассматривались в [5].

В докладе внешние степени стандартного представления алгебры Ли *o(V)* называются фундаментальными по аналогии с фундаментальными представлениями алгебры Ли *sl(V)*.

Теорема 1. *Пусть K – совершенное поле характеристики два, V – n-мерное векторное пространство над K, { , } – невырожденная симметрическая неальтернирующая билинейная форма на V, o(V) – соответствующая ортогональная алгебра Ли. Тогда*

1. $\Lambda^r V$ *– неприводимый o(V)-модуль для r= 1,…, n-1.*

2. Если n= 2t, то при $r \neq t$, $\Lambda^r V$ *– неприводимый so(V)-модуль. При r= t,* $\Lambda^t V$ *содержит единственный нетривиальный so(V)-подмодуль M,* $\dim M = \frac{1}{2}\binom{n}{t}$.

3. Если n= 2t+1, то $\Lambda^r V$ *– неприводимый so(V)-модуль для r= 1,…, n-1.*

4. С точностью до изоморфизма, фундаментальными модулями над so(V) являются модули $\Lambda^r V$, $r = 1, \ldots, \left[\frac{n}{2}\right]$. *На* $\Lambda^r V$ *существует инвариантная относительно so(V) ортогональная форма.*

Для исследования простых конечномерных 1-градуированных алгебр Ли представляет интерес описание однородных конечномерных простых подалгебр Ли $G = G_{-1} + G_0 + G_1 + \ldots$ неальтернирующей гамильтоновой алгебры Ли $P(n) = L_{-1} + L_0 + L_1 + \ldots$, таких, что $G_{-1} = L_{-1}$,

[1*] Работа выполнена при финансовой поддержке Министерства образования и науки РФ, проект FSWR-2023-0034, и научно-образовательного математического центра «Математика технологий будущего».

$L_0' \subseteq G_0 \subseteq L_0$, $G_1 \subset L_1$. Для этого нужна информация о структуре $so(L_{-1})$-подмодулей пространства L_1. Мы ограничиваемся здесь построением двух исключительных простых подалгебр, существование которых следует из теоремы 1. Согласно этой теореме только при $n=6$ пространство L_1 содержит подмодуль $M = \langle x_{i_1}x_{i_2}x_{i_3} + x_{i_4}x_{i_5}x_{i_6}\rangle$. Обозначим через $P_1(6)$ подалгебру алгебры Ли $P(6)$, порожденную локальной частью $G_{-1} + G_0 + G_1$, где

$$G_{-1} = \langle x_1, \dots x_6\rangle$$
$$G_0 = \langle x_i x_j \quad i,j = 1, \dots ,6\rangle$$
$$G_1 = M$$

Алгебра $P_1(6) = G_{-1} + G_0 + G_1$ является простой алгеброй Ли, $dim P_1(6) = 6+15+10 = 2^5-1$.

Пространство L_1 содержит L_0'-подмодуль $\langle zx_1, \dots zx_n\rangle$, где $z = x_1^{(2)} + \dots + x_n^{(2)}$. Строим подалгебру $P_2(6)$, порожденную локальной частью $G_{-1} + G_0 + G_1$, где

$$G_{-1} = \langle x_1, \dots x_6\rangle$$
$$G_0 = \langle x_i x_j \quad i,j = 1, \dots ,6\rangle + \langle z\rangle$$
$$G_1 = M + \langle zx_1, \dots zx_6\rangle$$

Алгебра $P_2(6) = G_{-1} + G_0 + G_1$ является простой алгеброй Ли, $dim P_2(6) = 6+16+16+15+10 = 2^6-1$.

Алгебры Ли $P_1(6)$ и $P_2(6)$ не имеют аналогов при $n \neq 6$.

Литература

1. Кондратьева А. В., Кузнецов М. И. Фильтрованные деформации градуированных неальтернирующих гамильтоновых алгебр Ли // Изв. вузов. Матем. 2024. Т.9. С.100-105.
2. Кондратьева А. В., Кузнецов М. И. Неальтернирующие гамильтоновы формы над алгеброй разделенных степеней в характеристике 2 // Изв. вузов. Матем. 2023. Т.6. С.95-100.
3. Kondrateva A. V. Non-alternating Hamiltonian Lie algebras in three variables // Lobachevskii Journal of Mathematics. 2021. V.42(12). P.2841-2853.
4. Lin Lei. Non-alternating hamiltonian algebra P(n,m) of characteristic two // Communications in Algebra. 1993. V.21(2). P.399-411.
5. Кондратьева А. В., Кузнецов М. И. К теореме вложения фильтрованных деформаций градуированных неальтернирующих гамильтоновых алгебр Ли // Журнал СВМО. 2024. Т.26(4). С. 392-403.

ВЫЧИСЛЕНИЕ ЛОКАЛЬНЫХ ДЕФОРМАЦИЙ РАСШИРЕНИЙ ПРОСТЫХ АЛГЕБР ЛИ ТИПА А С ПОМОЩЬЮ ВНЕШНЕГО ДИФФЕРЕНЦИРОВАНИЯ НАД ПОЛЕМ ХАРАКТЕРИСТИКИ 2[1*]

М.И. Кузнецов, М.М. Рабиа

Нижегородский государственный университет им. Н.И. Лобачевского

В докладе рассматриваются полупростые алгебры Ли $\boldsymbol{L}$ с простым ядром, получающиеся добавлением внешнего дифференцирования к простой алгебре Ли типа А. Приводятся результаты вычисления второй группы когомологий алгебр Ли $\boldsymbol{L}$ с коэффициентами в присоединенном модуле над совершенным полем характеристики два

Ключевые слова: совершенное поле характеристики два, полупростые алгебры Ли, когомологии, локальные деформации

В связи с классификацией простых алгебр Ли над полями характеристик 2 и 3 представляет интерес описание простых алгебр Ли малой размерности, которые обычно исключаются из рассмотрения при доказательстве общих утверждений (см, например, [1]). Результаты компьютерного поиска таких алгебр над полем $\mathbb{Z}_2$ приведены в работах [2] – [4]. В [5] 14-мерные алгебры Ли из полученного списка сопоставляются с деформациями известных простых алгебр Ли.

Деформации простых классических алгебр Ли исследуются в работах [6] – [10]. В докладе рассматриваются алгебры Ли $L = L' + < D >$, где L' – простая классическая алгебра Ли типа A, т.е. $L' = psl(n)$, $D -$ внешнее дифференцирование алгебры L'. Все дифференцирования классических алгебр Ли в характеристике 2 найдены в работе [11]. В частности, для рассматриваемого случая n четно и для $n > 4$ с точностью до изоморфизма существует только одна алгебра Ли $L = pgl(n)$. Можно доказать, что если $n = 4$, то с точностью до изоморфизма существуют две неизоморфные алгебры Ли L – $pgl(4)$ и алгебра, с нильпотентным дифференцированием D.

В докладе представлены следующие результаты:

1) если n четно, $n > 2, L = pgl(n)$, то $H^2(L, L) = 0$, т.е. алгебра Ли $pgl(n)$ является жесткой;

2) если $L = L' + < D >, L' = psl(4), D -$ нильпотентное дифференцирование алгебры L', то $H^2(L, L) \neq 0$. Найдены все весовые подпространства второй группы когомологий.

Литература

1. Kondrateva A.V., Kuznesov M.I. Filtered deformations of graded non-alternating Hamiltonian Lie algebras // Russian Math. 2024. V. 68(9). P.86-90.
2. Eick B. Some new simple Lie algebras in characteristic 2 / J. Symbolic Comput.2010. V. 45(9), P. 983-95.
3. Eick B., Moede T. Computing subalgebras and Z_2-gradings of simple Lie algebras over finite fields // Commun. Math. 2022. V. 30(2). P.37-50.
4. Cushing D., Stagg G.W., Stewart D.I. A Prolog assisted search for new simple Lie algebras found by computer and non-alternating Hamiltonian algebras // Math. Comp. 2024. V. 93. P.1473-1495.
5. Kondrateva A.V., Kuznetsov M.I. New simple Lie algebras found by computer and non-alternating Hamiltonian algebras // Commun. Comp. Inform. Sci. 2025. V. 2363. P.185-197.
6. Кузнецов М.И., Чебочко Н.Г. Деформации классических алгебр Ли // Матем. сб. 2000. Т.191(8). С.69-88.

[1*] Работа выполнена при частичной финансовой поддержке Министерства образования и науки РФ, проект FSWR-2023-0034, и научно-образовательного математического центра «Математика технологий будущего».

7. Чебочко Н.Г. Деформации классических алгебр Ли с однородной системой корней в характеристике два. I // Матем. сб. 2005. Т. 196(9). С.125-156.
8. Chebochko N.G., Kuznetsov M.I. Integrable cocycles and global deformations of Lie algebra of type G2 in characteristic 2 // Commun. Algebra. 2017. V. 45(7). P. 2969-2977.
9. Kuznetsov M.I., Kondrateva A.V., Chebochko N.G. Sumple 14-dimensional Lie algebras in characteristic two // J. Math. Sci. 2019. V. 240(4). P.474-480.
10. Kuznetsov M.I., Chebochko N.G. Global deformations of a Lie algebra of type A_5 in characteristic 2 // Lobachevskii J. Math. 2020. V. 41(2). P.94-98.
11. Пермяков Д.С. Дифференцирования классических алгебр Ли над полем характеристики 2 // Вестник ННГУ. Серия Математика. 2005. Вып. 1(3). С.123-134.

МОДЕЛИРОВАНИЕ ТОНКИХ КОМПОЗИТНЫХ ПЛАСТИН ПОД ВИБРАЦИОННОЙ НАГРУЗКОЙ С УТОЧНЕНИЕМ ПАРАМЕТРОВ МАТЕРИАЛА

С.А. Лавренков, А.В. Васюков

Московский физико-технический институт (Национальный исследовательский университет)

Данная работа посвящена проблеме определения амплитудно-частотной характеристики (АЧХ) колебаний пластины из композитного материала. Прямая задача поиска АЧХ по заданным параметрам материала сформулирована как серия краевых задач для системы уравнений движения тонкого образца в вибрационном стенде, решаемых методом конечных элементов. Упругие параметры материала определяются с помощью решения обратной задачи, сформулированной в терминах минимизации функционала ошибки между численной и экспериментально полученной АЧХ колебаний для тестового образца. В работе представлены результаты моделирования для нескольких пластин, с которыми были проведены эксперименты на вибрационном стенде.

Ключевые слова: вибрационные испытания, композитные материалы, обратная задача, автоматическое дифференцирование.

1. Введение

Композитные материалы часто применяются во многих отраслях машиностроения, в том числе, в аэрокосмической. Специфика этой области заключается в особом внимании к способности структурных элементов выдерживать вибрационные нагрузки, возникающие при выводе космического корабля на орбиту, а также в применении уникальных и дорогостоящих материалов для максимально возможного снижения массы конструкций без потери в жёсткости. Эти особенности приводят к тому, что проведение наземных экспериментов с конструкционными элементами становится экономически нецелесообразным, и возникает необходимость в численном моделировании на этапе проектирования. Однако, в силу мелкосерийности производства специализированных композитов, классические методы определения упругих параметров материала, необходимых для моделирования, также являются дорогостоящими.

В данной работе предлагается метод моделирования тонких композитных структур под вибрационными нагрузками, который позволяет определять частотный отклик для материалов с неизвестными упругими параметрами. Для получения недостающих данных используются результаты эксперимента на вибрационном стенде с пластиной, изготовленной из небольшого количества исследуемого материала. Такой подход позволяет уменьшить количество материала, необходимого для определения характеристик, и обеспечить возможность проведения экспериментов в различных конфигурациях, так как испытания не разрушают образец.

Существуют несколько работ, в которых используются похожие подходы к решению обратных задач, связанных с пластинами и неизвестными параметрами материала. Работа [1] фокусируется на нелинейности колебаний композитных пластин при схожей экспериментальной постановке; однако в ней используется более дорогое и сложное оборудование для получения частотного отклика. В работе [2] рассматривается математическая теория для решения обратной задачи для пластины с несколькими слоями с различными характеристиками. В [3] исследуется влияние атмосферы на затухание колебаний пластин.

Используемый в настоящей работе метод был впервые предложен в статье [4]. Данная работа расширяет область применимости метода, а также демонстрирует его применимость к реальным данным.

2. Методология исследования

Для построения математической модели рассматриваются прямая и обратная задачи колебаний пластины на вибрационном стенде. Экспериментальная установка, позволяющая получить амплитудно-частотную характеристику колебаний, состоит из вибростенда (генератора синусоидальных колебаний с изменяющейся частотой), пластины из композитного материала и двух синхронизированных механических акселерометров, один из которых закреплён на пластине. Тогда, прямая задача формулируется следующим образом: при заданных параметрах геометрии пластины, её плотности, упругих модулях, параметрах акселерометра (размеры и масса) получить на заданном диапазоне частот значения частотного отклика. Обратная задача: по заданным геометрии пластины, её плотности, параметрам акселерометра, а также частотному отклику колебаний (измеренному экспериментально) в некотором диапазоне найти такие упругие параметры материала, при которых результат моделирования частотного отклика совпадает с экспериментальным.

2.1. Прямая задача

Математическая модель динамики тонкой композитной пластины строится в рамках гипотез Кирхгофа-Лява и классической теории ламинатов. Трёхмерная система уравнений упругости сводится к системе трёх уравнений относительно компонент вектора перемещений двумерной срединной плоскости [5].

Для анализа установившихся колебаний осуществляется переход в частотную область, что позволяет исключить зависимость от времени и сформулировать задачу с параметром — частотой колебаний. Краевая задача дополняется граничными условиями, адекватно описывающими экспериментальное закрепление образца в вибростенде: на зажатом конце задаются условия Дирихле, на остальной границе — условия Неймана. Решение серии таких параметрических краевых задач позволяет построить частотный отклик системы, то есть решить прямую задачу.

Системы уравнений дискретизуются с помощью метода конечных элементов, затем полученные системы линейных уравнений решаются солвером для разреженных матриц UMFPACK. Также для ускорения вычислений используется распараллеливание по частоте колебаний с помощью OpenMP, в силу независимости краевых задач внутри одной серии. Алгоритм решения прямой задачи реализован с помощью библиотеки jax для языка Python, позволяющей использовать автоматическое дифференцирование солвера прямой задачи.

2.2. Обратная задача

Обратная задача формулируется как задача поиска минимума функционала ошибки между численным решением и экспериментальным частотным откликом:

$$\min_{\theta} L(\theta) = \min_{\theta} \left[\frac{1}{N} \sum_{n=1}^{N} \left(\log|F(\omega_n, \theta)| - \log\left|F_{\exp}(\omega_n)\right| \right) \right]$$

где θ — вектор неизвестных упругих модулей, F_{exp} — экспериментально полученный частотный отклик, F — численное решение, $\{\omega_n\}_{n=1}^{N}$ — дискретный набор частот.

Минимизация функционала производится численно, алгоритмом Simplicial Homology Global Optimisation (SHGO) [6] для глобальной оптимизации и Sequential Least Squares Programming (SLSQP) методом для локальной оптимизации, позволяющим накладывать физически обусловленные ограничения на значения упругих модулей. При этом для вычисления градиентов используется автоматическое дифференцирования солвера прямой задачи.

3. Результаты

Предложенный метод был применён к результатам серии экспериментов с прямоугольными пластинами из трёх углерод-эпоксидных материалов.

Для двух различных образцов, вырезанных под одним углом к углеродным нитям, из трёхслойного материала с ориентацией армирующих волокон (-45°, 45°, -45°) были получены АЧХ с использованием двух акселерометров разной массы. Для одной из пластин была решен обратная задача, и полученные параметры материала были применены для моделирования эксперимента с другой пластиной. При таком подходе было выявлено, что предложенный метод позволяет хорошо приближать реальный эксперимент при условии, что акселерометр, закреплённый на пластине, достаточно лёгкий. При использовании более тяжёлого датчика численная модель соответствует эксперименту лишь в диапазоне низких частот, и с увеличением частоты точность значительно снижается.

Аналогичные вычисления были проведены для двух других пластин с материалом с укладкой нитей под (0°, 30°, 60°).

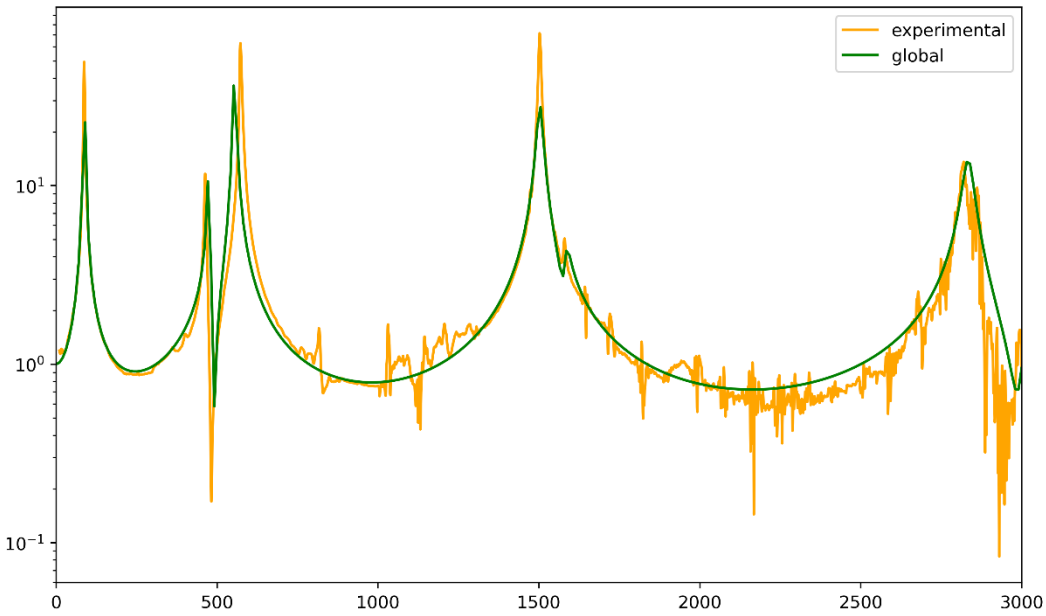


Рис. 1. Пример экспериментальной АЧХ для многослойной композитной пластины и численного решения с параметрами, полученными после решения обратной задачи

Также были проведены расчёты, соответствующие двум экспериментам с перфорированной круглыми отверстиями пластиной из материала с 26 перпендикулярно уложенными слоями. По результатам расчёта было показано, что численный метод может быть применён и к пластинам со сложной геометрией.

Литература

1. Markov V.T., Tuganbaev A.A. Rings essential over their centers // Comm. Algebra. − 2019. − Vol. 47, Issue 4. − P. 189−194. DOI: 10.1080/00927872.2018.1513012.
1. Nonlinear vibration study of fiber-reinforced composite thin plate with strain dependent property based on strain energy density function method / Li H., Xue P., Zhang T., and Wen B. // Mechanics of Advanced Materials and Structures. 2020. Vol. 27, No. 9. P. 761–773.
2. Romanov V.G., Weng C.I, Chen T.C. An inverse problem related to layered elastic plate // Journal of Inverse and Ill posed Problems. 2000. Vol. 8, No. 4. P. 421–447.
3. Exploring the resonant vibration of thin plates: Reconstruction of chladni patterns and determination of resonant wave numbers / Tuan P.H., Wen C.P., Chiang P.Y., Yu Y.T., Liang H.C., Huang K.F., and Chen Y.F. // The Journal of the Acoustical Society of America. 2015. Vol. 137, No. 4. P. 2113-2123.
4. Aksenov V., Vasyukov A., Beklemysheva K. Acquiring elastic properties of thin composite structure from vibrational testing data // Journal of Inverse and Ill-posed Problems. 2024. Vol. 32, No. 3. P. 467–484.

5. Reddy J. N. Mechanics of Laminated Composite Plates and Shells. — Boca Raton, FL: CRC Press, 2003.
6. Endres S., Sandrock C., Focke W. A simplicial homology algorithm for lipschitz optimization // Journal of Global Optimization. 2018. Vol. 72. P. 181-217.

РАСШИРЕНИЯ ЦЕНТРАЛЬНО СУЩЕСТВЕННЫХ КОЛЕЦ[1*]

О.В. Любимцев[1], А.А. Туганбаев[2]

[1]Нижегородский государственный университет им. Н.И. Лобачевского,
[2]Национальный исследовательский университет «МЭИ»

Кольцо R с ненулевой единицей называется центрально существенным, если для любого его ненулевого элемента a существуют такие ненулевые центральные элементы x и y, что ax = y. В работе описаны почти вполне первичные центрально существенные кольца в терминах идеальных расширений, центрально существенные расширения Доро и тривиальные расширения.

Ключевые слова: центрально существенное кольцо, почти вполне первичное кольцо, расширение Доро, тривиальное расширение.

1. Введение

Все кольца, рассматриваемые в работе, являются ассоциативными с ненулевой единицей. Кольцо R называется центрально существенным, если для любого его ненулевого элемента a существуют такие ненулевые центральные элементы x и y, что ax = y. Ясно, что кольцо R с центром Z(R) = Z является центрально существенным в точности тогда, когда модуль RZ − существенное расширение модуля ZZ. Класс центрально существенных колец является строгим расширением класса коммутативных колец. Центрально существенные кольца изучались в работах [1], [2], [3] и других. В работе [1, Proposition 2.8] доказано, что полупервичные центрально существенные кольца коммутативны.

Идеальное расширение кольца I через кольцо Q есть кольцо R, которое содержит идеал J, изоморфный I (как кольцо), и $R/J \cong Q$. Идеальные расширения изучались, например, в работе [4]. Кольцо называется *вполне первичным*, если всякий его идеал первичен и *почти вполне первичным*, если всякий его ненулевой собственный идеал первичен; см. [5], [6]. Приведем пример некоммутативного почти вполне первичного центрально существенного кольца S, которое является расширением своего центрального идеала I через поле F. Напомним, что *производной* из кольца R в (R, R)-бимодуль M называется отображение $d\colon R \to M$, удовлетворяющее условиям $d(x) + d(y) = d(x + y)$ и $d(xy) = xd(y) + d(x)y$ для всех $x, y \in R$.

Если M − (R, R)-бимодуль, то производные $d_1, d_2\colon R \to M$ называются *несравнимыми*, если их ядра несравнимы по включению, т.е., если существуют такие элементы $x_1, x_2 \in R$, что $d_1(x_1) \neq 0$, $d_1(x_2) = 0$, $d_2(x_1) = 0$, $d_2(x_2) \neq 0$; см. [4].

Пример 1. Пусть R − кольцо, F − подкольцо в R, являющееся полем. Рассмотрим две несравнимые производные $d_1, d_2\colon F \to R$, причем $d_1(a)d_2(b) \neq 0$, если $d_1(a), d_2(b) \neq 0$. Определим кольцо матриц

$$S = \left\{ \begin{pmatrix} f & d_1(f) & g \\ 0 & f & d_2(f) \\ 0 & 0 & f \end{pmatrix}, \text{ где } f \in F,\ g \in R \right\}.$$

Тогда кольцо S некоммутативно и является расширением центрального идеала

$$I = \left\{ \begin{pmatrix} 0 & 0 & g \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ где } f \in F,\ g \in R \right\}.$$

через поле F; см. [4, Proposition 7 and Corollary 8]. Кроме того, R − почти вполне первичное центрально существенное кольцо (см. Теорему 2 ниже).

[1*] Работа О.В. Любимцева поддержана Минобрнауки РФ, проект FSWR-2023-0034, и научно-образовательного математического центра «Математика технологий будущего».
Исследование А.А. Туганбаева выполнено за счет гранта Российского научного фонда № 22-11-00052, https://rscf.ru/project/22-11-00052.

Пусть R и S − два кольца. Предположим, что S – такой (R, R)-бимодуль, что действие R на S согласовано с умножением в S, т.е. $r(xy) = (rx)y$, $(xr)y = x(ry)$, $(xy)r = x(yr)$ для всех $r \in R$ и $x, y \in S$. Тогда $D(R, S) = R \oplus S$ − кольцо с умножением

$$(r_1, m_1)\cdot(r_2, m_2) = (r_1r_2, r_1m_2 + m_1r_2 + m_1m_2),$$

где $r_1, r_2 \in R$, $m_1, m_2 \in S$, называемое *расширением Доро* $D(R, S)$ R через S. Если $S = M$ только (R, R)-бимодуль, то полагая $m_1m_2 = 0$ для всех $m_1, m_2 \in M$, получим тривиальное расширение $T(R, M)$ R через M с умножением: $(r_1, m_1)\cdot(r_2, m_2) = (r_1r_2, r_1m_2 + m_1r_2)$. Тогда $T(R,M) \cong \left\{\begin{pmatrix} r & m \\ 0 & r \end{pmatrix} : r \in R, m \in M\right\}$, и $T(R, R) \cong R[x] / (x^2)$, где (x^2) − идеал, порожденный x^2. Обозначим

$$C_M(R) = \{r \in R \mid mr = rm \ \ \forall m \in M\}$$

и

$$CR(M) = \{m \in M \mid mr = rm \ \ \forall r \in R\}.$$

Нетрудно видеть, что $CM(R)$ и $CR(M)$ − подкольцо и подмодуль в R и M, соответственно. Определим также подмножества T и N в R и M:

$$T = \{r \in R \mid \ \exists\, c \in CM(R) \cap Z(R): 0 \neq rc \in CM(R) \cap Z(R)\ \},$$
$$N = \{m \in M \mid \ \exists\, c \in CM(R) \cap Z(R): 0 \neq mc \in CR(M)\ \}.$$

2. Основные результаты

Теорема 2.

Пусть R − некоммутативное почти вполне первичное кольцо. Следующие условия эквивалентны:

1) R − центрально существенное кольцо;

2) R является кольцом с единственным ненулевым собственным (левым, правым) идеалом P, который лежит в центре кольца, и факторкольцо R/P − поле;

3) R − расширение центрального идеала I через поле F;

4) R − расширение центрального идеала I через поле F, которое имеет две несравнимые производные из F в некоторое поле расширения F.

Замечание 3. Если в обозначениях теоремы 2(3, 4) характеристика поля F равна нулю, то в формулировку теоремы можно добавить эквивалентное условие (см. [4, Theorem 2]):

5) Степень трансцендентности поля F над своим простым подполем больше 1.

Предложение 4. Пусть R − почти вполне первичное центрально существенное кольцо. Тогда аддитивная группа $R+$ кольца R изоморфна $\oplus_\alpha Q$ или $\oplus_\alpha Z_p$ для некоторого простого p, где α − произвольный кардинал.

Теорема 5.

1) Пусть R − коммутативное кольцо, S − (R, R)-бимодуль, $RS = SR$ и $D(R, S)$ − расширение Доро R через S. Кольцо S является центрально существенным кольцом в точности тогда, когда $D(R, S)$ − центрально существенное кольцо.

2) Тривиальное расширение $T(R, M)$ центрально существенно в точности тогда, когда

$$N = M \text{ и } T = R.$$

Следствие 6. Кольцо R является центрально существенным в точности тогда, когда центрально существенно кольцо $D(\mathbb{Z}, R)$. Таким образом, любое центрально существенное кольцо вкладывается в центрально существенное кольцо с 1.

Следствие 7. Если $T(R, M)$ центрально существенное кольцо, то кольцо R центрально существенно.

Следствие 8. Пусть I − идеал в кольце R. Кольцо $T(R, I)$ центрально существенно в точности тогда, когда центрально существенно кольцо R.

3. Замечания и дополнительные примеры

Замечание 9. Существуют почти вполне первичные кольца, имеющие различные ненулевые собственные идеалы; см. [6, Example 2.4]. Более того, существуют вполне первичные некоммутативные кольца, имеющие бесконечно много идеалов; см. [5].

Замечание 10. Пусть D − множество всех делителей нуля кольца R. Из определения центрально существенного кольца следует, что если $\{0\} \neq D \subset Z(R)$, то кольцо R центрально существенно. Из [7, Example 1.1] следует, что существуют некоммутативные центрально существенные кольца, у которых все делители нуля лежат в центре кольца.

Пример 11. Пусть $\mathbb{Z}$ − кольцо целых чисел, $\mathbb{Q}$ − поле рациональных чисел, $\mathbb{Q}[i]$ − поле частных кольца целых гауссовых чисел $\mathbb{Z}[i]$. Обозначим через φ: $q_1 + q_2 i \to q_1 - q_2 i$ − автоморфизм поля $\mathbb{Q}[i]$. Пусть R_X и R_Y − локализации $\mathbb{Z}[i]$ по простым идеалам $X = (2 + i)$ и $Y = (2 - i)$, соответственно. Положим $R = R_X \cap R_Y$. Кольцо R совпадает с множеством всех несократимых рациональных гауссовых дробей, знаменатели которых не делятся на $2 + i$ и на $2 - i$. Обозначим через M правый R-модуль $\mathbb{Q}[i] / R_X$. Мы также превратим M в левый R-модуль по правилу $rm = m\varphi(r)$ для всех элементов $r \in R$ и $m \in M$. Непосредственно проверяется, что M является (R, R)-бимодулем. Рассмотрим тривиальное расширение $S = T(R, M)$ кольца R через модуль M. Тогда $C_M(R) = \mathbb{Q}$. Если $0 \neq z = q_1 + q_2 i$, где $q_2 \neq 0$, то z не лежит в T. По теореме 5(2) S не является центрально существенным кольцом.

Приведем пример центрально существенного тривиального расширения $S = T(R, M)$, где ${}_R M \neq M_R$.

Пример 12. Будем строить искомое расширение на базе примера 1.1 из работы [7]. Пусть $\mathbb{Z}[x, y]$ − кольцо многочленов от двух переменных x и y над $\mathbb{Z}$, $K = \mathbb{Z}[x, y] / (y^2)$ и $R = K[t, \sigma]$ − кольцо дифференциальных многочленов $\sum t^i \cdot a_i$, $a_i \in K$ с производной σ, определяемой как $\sigma(x) = y$ и $\sigma(y) = 0$. Так как $y^2 = 0$, то имеем $(f(x) + g(x)y) \cdot t = t \cdot (f(x) + g(x)y) + f'(x)y$. Рассмотрим кольцо

$$S = \left\{ \begin{pmatrix} f(x,y) & g(t) \\ 0 & f(x,y) \end{pmatrix}, \text{ где } f(x,y) \in K, g(t) \in R \right\}.$$

Так как

$$\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & xt \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & tx \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix},$$

то кольцо S некоммутативно. Кроме того, S является центрально существенным кольцом. Действительно, пусть ненулевая матрица $\begin{pmatrix} k & f(t) \\ 0 & k \end{pmatrix}$ не принадлежит центру $Z(S)$ кольца S, $k = f(x) + g(x)y$. Если $f(x) = 0$, то $g(x)y \in C_M(R)$. Если $f(x) \neq 0$, то $0 \neq ky \in C_M(R)$, $y \in C_M(R)$. Значит, $T = R$. Далее, если $f(t)$ не принадлежит $C_R(M)$, то $f(t)$ содержит моном с ненулевым коэффициентом $h(x)$ и $0 \neq f(t)y \in C_R(M)$. Следовательно, $N = M$ и S является центрально существенным кольцом по теореме 5(2).

Литература

1. Markov V.T., Tuganbaev A.A. Rings essential over their centers // Comm. Algebra. − 2019. − Vol. 47, Issue 4. − P. 189−194. DOI: 10.1080/00927872.2018.1513012.
2. Markov, V.T., Tuganbaev, A.A. Uniserial Noetherian Centrally Essential Rings // Comm. Algebra. − 2020. − Vol. 48, Issue 1. − P. 149−53. DOI: 10.1080/00927872.2019.1635607.
3. Lyubimtsev O.V., Tuganbaev A.A. Centrally essential factor rings and subdirect indecomposability // Comm. Algebra. − 2024. − Vol. 52, Issue 4. − P. 1614−1621. DOI: 10.1080/00927872.2023.2267688.
4. Jelisiejew J. On commutativity of ideal extensions // Comm. Algebra. − 2016. − Vol. 44, Issue 5. − P. 1931−1940. DOI: 10.1080/00927872.2015.1027394.
5. Blair W.D., Tsutsui H. Fully Prime Rings // Comm. Algebra. − 1994. − Vol. 22, Issue 13. − P. 5389−5400. DOI: 10.1080/00927879408825136.
6. Tsutsui H. Fully prime rings. II // Comm. Algebra. − 1996. − Vol. 22, Issue 9. − P. 2981−2989. DOI: 10.1080/00927879608825725.
7. Klein A.A., Bell H.E. On central and noncentral zero divisors // Comm. Algebra. − 1998. − Vol. 26, Issue 4. − P. 1277−1292. DOI: 10.1080/00927879808826198.

БИФУРКАЦИИ ПРЕДЕЛЬНЫХ ЦИКЛОВ И ТОРОВ В СИСТЕМЕ ВЗАИМОСВЯЗАННЫХ ОСЦИЛЛЯТОРОВ БАУТИНА[1*]

А.А. Маркелов[1,2], А.С. Дмитричев[2], В.И. Некоркин[1,2]

[1] Нижегородский государственный университет им. Н. И. Лобачевского,
[2] Институт прикладной физики им. А.В. Гапонова-Грехова РАН

Приводятся результаты исследования динамики ансамбля двух взаимосвязанных осцилляторов Баутина. Представлены двухпараметрические бифуркационные диаграммы, соответствуюшие однородным и неоднородным решениям, для которых амплитуды колебаний совпадают или различаются соответственно. Бифуркационная диаграмма для однородных решений включает кривые седло-узловых бифуркаций циклов и торов, бифуркаций Андронова-Хопфа (суб- и суперкритической), бифуркации Хопф-Хопф и слияния торов. Бифуркационные диаграммы для неоднородных решений включают кривые бифуркаций типа «вилка», седло-узловых бифуркаций циклов, а также бифуркаций Неймарка-Сакера и слияния торов, которые кроме того демонстрируют удвоение периода.

Ключевые слова: осциллятор Баутина, малый ансамбль, диффузионная связь, бифуркации.

1. Введение

Осциллятор Баутина является нормальной формой бифуркации Баутина, которая представляет собой обобщение бифуркации Андронова-Хопфа на случай, когда первая ляпуновская величина обращается в ноль, а вторая – не равна нулю [1]. Связанные осцилляторы Баутина служат эффективной моделью для исследования разнообразных нелинейных динамических эффектов, например, синхронизации и подавления колебаний, наблюдаемых в различных сетевых системах [2,3].

Здесь мы представляем результаты исследования поведения ансамбля из двух связанных осцилляторов Баутина, динамика которого описывается следующей системой:

$$\begin{cases} \dot{z}_1 = (\alpha + i\omega_1)z_1 + \beta z_1|z_1|^2 - z_1|z_1|^4 + c(z_2 - z_1), \\ \dot{z}_2 = (\alpha + i\omega_2)z_2 + \beta z_2|z_2|^2 - z_2|z_2|^4 + c(z_1 - z_2), \end{cases} \quad (1)$$

где $z_i(t) = x_i(t) + iy_i(t), i = 1,2$ характеризуют состояние i-го осциллятора в момент времени t; α и β – действительные параметры системы задающие его динамику, ω_i – собственная частота осциллятора, $c > 0$ – параметр, характеризующий силу связи между осцилляторами.

Система (1) при любых значениях параметров в $\mathbb{R}^4$ имеет единственное состояние равновесия в начале координат 0, которое в зависимости от значений параметров может быть устойчивым или неустойчивым фокусом, а также седло-фокусом.

2. Однородные решения

Решения вида $z_i(t) = r(t)\exp(i\varphi_i(t)), i = 1,2$ будем называть однородными. Зафиксируем параметры α и β, а параметры c и $\Delta = \omega_2 - \omega_1$ будем рассматривать как контрольные. На рис.1 представлено разбиение плоскости параметров (c, Δ) на области с различным числом предельных циклов и торов ($\beta = 0.9, \alpha = 0.1$), которое осуществляется следующими бифуркационными кривыми:

$SNIC_{1-3}$: $\{\Delta = 2c, 0 < c < \alpha\}$ – седло-узловой бифуркацией предельных циклов L_1 и L_3,

[1*] Работа выполнена при поддержке Министерством науки и высшего образования РФ в рамках государственного задания ИПФ РАН, проект FFUF-2024-0011.

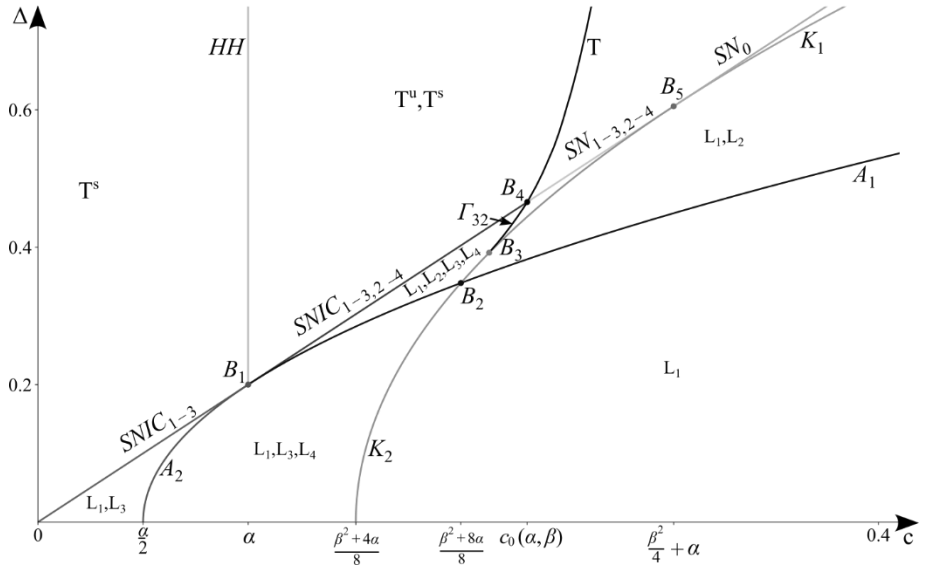


Рис.1. Бифуркационная диаграмма однородных решений системы (1). Здесь: L_1 – устойчивый предельный цикл, $L_{2,3}$ – седловые предельные циклы, L_4 – неустойчивый предельный цикл, T^s, T^u– устойчивый и неустойчивый двумерные инвариантные торы

A_2: $\left\{\Delta = 2\sqrt{2\alpha c - \alpha^2}, \frac{\alpha}{2} < c < \alpha\right\}$ – субкритической бифуркацией Андронова-Хопфа цикла L_4,

HH: $\{c = \alpha, \Delta > 2c\}$ – Хопф-Хопф бифуркацией инвариантного тора T^u,

A_1: $\left\{\Delta = 2\sqrt{2\alpha c - \alpha^2}, c > \alpha\right\}$ – субкритической бифуркацией Андронова-Хопфа цикла L_2,

$SNIC_{1-3,2-4} \cup SN_{1-3,2-4}$: $\{\Delta = 2c, \alpha < c < \frac{\beta^2}{4} + \alpha\}$ – седло-узловыми бифуркациями предельных циклов $L_1; L_3$ и $L_2; L_4$,

K_2: $\left\{\Delta = \frac{\sqrt{16c^2 - (4(\alpha - c) + \beta^2)^2}}{2}, \frac{1}{2}(\frac{\beta^2}{4} + \alpha) < c < \frac{\beta^2}{4} + \alpha\right\}$ – седло-узловой бифуркацией предельных циклов L_3 и L_4,

K_1: $\left\{\Delta = \frac{\sqrt{16c^2 - (4(\alpha - c) + \beta^2)^2}}{2}, c > \frac{\beta^2}{4} + \alpha\right\}$ – седло-узловой бифуркацией предельных циклов L_1 и L_2,

Γ_{32} – бифуркацией образованной многообразиями седловых предельных циклов L_3 и L_2,

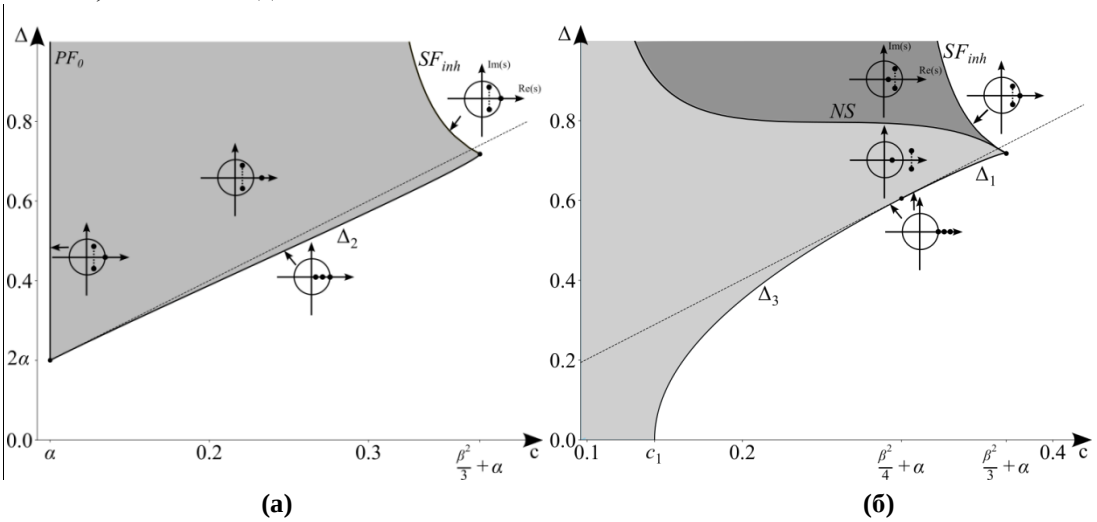
T – бифуркацией слияния торов T^s и T^u.

3. Неоднородные решения

В отличие от однородных, неоднородные решения характеризуются различными амплитудами колебаний $\mathrm{r}_1(\mathrm{t}) \neq \mathrm{r}_2(\mathrm{t})$.

Показано, что имеются два типа неоднородных периодических решений (обозначим их I_1, I_2), возникающих парами в результате бифуркаций типа «вилка» (pitchfork). Их возникновение связано с изменением устойчивости многообразий однородных седловых циклов L_3 и L_2. Данные решения появляются при пересечении кривых Δ_2(рис. 2а) и Δ_1, Δ_3 (рис. 2б). Выделена область существования каждой пары неоднородных циклов и проведён анализ их устойчивости. Область существования пары неоднородных циклов I_1 (рис. 2а), кроме кривой Δ_2, ограничена линией PF_0 и кривой $\mathrm{SF}_{\mathrm{inh}}$. При пересечении линии PF_0, происходит рождение неустойчивого однородного инвариантного тора T^{u} и двух неоднородных седловых циклов в результате изменения устойчивости состояния равновесия в начале координат. Кривая $\mathrm{SF}_{\mathrm{inh}}$ в свою очередь, отвечает седло-узловой бифуркации между обеими парами неоднородных циклов. Установлено, что пара I_1 не меняет устойчивость во всей области существования, оставаясь седловой.

Область существования пары неоднородных циклов I_2 (рис. 2) кроме кривых Δ_1, Δ_3 ограничена только кривой SF_{inh}. Установлено, что данная пара меняет свою устойчивость на кривой NS, на которой пара комплексно-сопряженных мультипликаторов попадает на единичную окружность, т. е происходит бифуркация Неймарка-Сакера. Выше этой кривой пара I_2 является устойчивой, а ниже – седловой.



(а) **(б)**

Рис. 2. Области существования пар неоднородных предельных циклов,рождающихся на кривых Δ_2(а) и Δ_1, Δ_3(б)

Помимо предельных циклов, в системе наблюдаются двумерные инвариантные торы, также рождающиеся парами. Области их существования приведены на рис. 3. Одна из пар инвариантных торов возникает в результате бифуркации Неймарка-Сакера, происходящей при пересечении кривой $NS = NS^+ \cup NS^-$ (рис. 2б и 3). При этом при пересечении кривой NS^+ имеет место суперкритическая бифуркация, и в сторону уменьшения параметра Δ, рождаются устойчивые торы, существующие в области 1. В свою очередь, при пересечении кривой NS^- имеет место субкритическая бифуркация и, в сторону увеличения параметра Δ, рождаются неустойчивые торы, существующие в области 2. Ещё одна пара устойчивых неоднородных инвариантных торов формируется при пересечении кривой T в сторону увеличения параметра c, соответствующей слиянию устойчивого и неустойчивого однородных торов. Данная пара решений существует в областях 1 и 2. На кривой G_1 происходит разрушение соответствующих торов в результате каскадов бифуркаций удвоения периода и последующей их хаотизации. Кривая G_2 соответствует слиянию торов образованных в результате субкритической бифуркации Неймарка-Сакера и устойчивых торов образованных при пересечении кривой T.

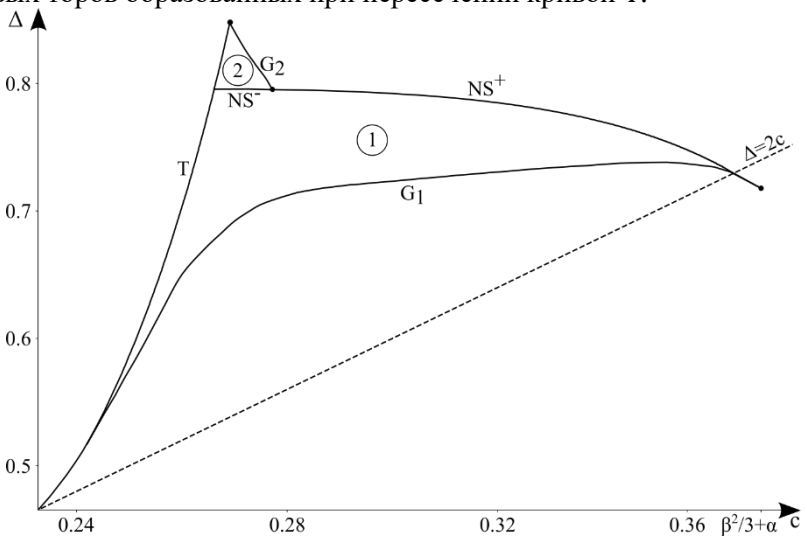


Рис. 3. Области существования неоднородных инвариантных торов

Заключение

Построено разбиение плоскости параметров ”сила связи - расстройка частот” на области с различным динамическим поведением осцилляторов. Исследованы бифуркации соответствующие рождению/исчезновению как однородных, так и неоднородных решений различного типа. Изучена область существования и устойчивость таких решений.

Литература

1. Kuznetsov Y.A. Elements of applied bifurcation theory. – New York: Springer, 1998, P. 307. DOI: 10.1007/978-0-387-22710-8_8.
2. Kato Y, Kori H. Weakly nonlinear analysis on synchronization and oscillation quenching of coupled mechanical oscillators // Scientific Reports. 2024. Vol. 14, No. 3. P. 1461 DOI: 10.1038/s41598-024-51843-9.
3. Emelianova A.A., Maslennikov O.V., Nekorkin V.I. Disordered quenching in arrays of coupled Bautin oscillators // Chaos: An Interdisciplinary Journal of Nonlinear Science. 2022. Vol. 32. No. 6 .P. 063126. DOI: 10.1063/5.0093947.

ОБ ОПЫТЕ РАЗРАБОТКИ ПЕРЕНОСИМЫХ ПРОГРАММ ДЛЯ НАУЧНЫХ ВЫЧИСЛЕНИЙ[1*]

И.Б. Мееров, С.А. Борисов, В.Д. Волокитин, А.В. Горшков, М.А. Козлов, Е.А. Панова, А.Ю. Пирова, А.В. Линев

Нижегородский государственный университет им. Н.И. Лобачевского

Проблема переносимости кода между различными аппаратными платформами возникает при разработке программных комплексов для научного и инженерного моделирования. Многие задачи, решаемые с использованием таких программных комплексов, являются вычислительно-трудоемкими и нередко требуют использования значительных ресурсов суперкомпьютерных систем. Современные суперкомпьютеры могут быть построены с использованием центральных и графических процессоров различной архитектуры. Существенные различия между архитектурами являются причиной проблем при разработке переносимых кодов, к которым предъявляются требования высокой производительности и эффективности использования вычислительных ресурсов. В докладе рассматриваются возможности разработки переносимых программ для гетерогенных вычислительных систем с использованием различных языков и библиотек параллельного программирования (SYCL/DPC++, Kokkos, Co-Array Fortran), обсуждаются особенности, достоинства и недостатки указанных подходов. Изложение проводится на примере двух существенно отличающихся друг от друга задач биоинформатики и вычислительной физики. Формулируются выводы и рекомендации по разработке переносимых программ с использованием указанных подходов.

Ключевые слова: высокопроизводительные вычисления, научное моделирование, переносимость программ, анализ и оптимизация производительности, переносимость производительности, гетерогенные вычислительные системы, SYCL, DPC++, Co-Array Fortran, Kokkos.

Архитектуры современных суперкомпьютерных систем постоянно развиваются и усложняются для повышения пиковой теоретической производительности, которая может быть вычислена аналитически. Вместе с тем, существенное значение для практического использования имеет реальная (достижимая) производительность, которая может быть оценена экспериментально с использованием широко известных тестов производительности, среди которых выделяются бенчмарки LINPACK [1], HPCG [2], Graph500 [3] и другие.

Многолетняя история замеров в перечне 500 наиболее производительных систем мира показывает, что несмотря на постоянное увеличение абсолютных значений достигнутой производительности, сохраняется огромная разница между результатами, полученными на разных задачах. Для многих суперкомпьютеров разница между результатами на тестах LINPACK и HPCG составляет два порядка [4]. Этот факт убедительно демонстрирует, что при всех несомненных технологических достижениях современные архитектуры плохо подходят для решения широкого спектра задач с низкой арифметической интенсивностью и нерегулярным доступом к памяти. Тому есть несколько причин. Во-первых, нельзя игнорировать тот факт, что на протяжении многих десятилетий тест LINPACK (прямое решение систем линейных уравнений с плотными матрицами) доминировал как основной инструмент сравнения и анализа, что несколько сместило фокус в развитии оборудования в сторону приложений, в которых вычисления преобладают над работой с памятью. По этой и другим причинам вычислительные возможности процессоров улучшались значительно быстрее, чем аналогичные характеристики подсистемы памяти, что усиливало и продолжает усиливать данный эффект. Во-вторых, на протяжении многих лет сообщество разработчиков программного обеспечения не уделяло достаточного внимания вопросам эффективности использования вычислительных ресурсов, пола-

[1*] Работа выполнена при поддержке Министерства образования и науки Российской Федерации, проект № FSWR-2023-0034.

гаясь исключительно на постоянный рост производительности суперкомпьютеров. В этой связи отметим важную дискуссию в работе [5], подчеркивающую тот факт, что возможности по дальнейшему взрывному росту производительности во многом исчерпаны, и, следовательно, настало время позаботиться об эффективности использования имеющихся ресурсов программными приложениями.

Данный вопрос особенно важен для прикладных кодов, решающих задачи из разных предметных областей методами численного моделирования на суперкомпьютерах. Такие коды, как правило, разрабатываются большими междисциплинарными коллективами, объединяющими специалистов из разных предметных областей. Из нашего опыта разработки соответствующих программных комплексов, а также из опыта обсуждения данной проблематики на конференциях, семинарах и в научной периодике известно, что на пути к созданию работающего ПО, которое может использоваться для решения задач, приходится преодолевать немало трудностей. Во-первых, существует проблема в коммуникациях между специалистами из разных предметных областей, нередко приводящая к непониманию и, соответственно, некорректной реализации, которую достаточно трудно протестировать и отладить. Во-вторых, сам по себе процесс диагностики таких кодов достаточно сложен: решение модельных задач с известным ответом далеко не всегда дает полную картину и уверенность в работоспособности ПО. В-третьих, проектирование таких программных комплексов также является нетривиальной задачей. При проектировании необходимо учитывать перспективы дальнейшего развития ПО, поскольку процесс научного поиска нередко требует реализации дополнительных вычислительных модулей, учитывающих различные эффекты и существенно расширяющих возможности использования. Важно отметить, что крайне желательно иметь возможность реализации таких расширений силами специалистов из прикладной области, не в полной мере владеющими навыками системного программирования и высокопроизводительных вычислений. Одним из примеров достаточно удачного проектирования является опыт разработки в ННГУ и ИПФ РАН кода PICADOR для численного моделирования лазерной плазмы [6], сочетающего оптимизированное вычислительное ядро с гибким интерфейсом, позволяющим дописывать дополнительные модули, почти не вникая в низкоуровневые детали организации параллелизма.

В ситуации, когда удается решить указанные выше проблемы, на первый план выходит следующий вопрос: как добиться возможности использования гетерогенных суперкомпьютеров, построенных с использованием устройств с существенно отличающейся архитектурой? На первый взгляд может показаться, что данная проблема сводится исключительно к переносимости кода (portability) и может быть легко решена с использованием одного из многочисленных языков (или библиотек) для гетерогенного программирования, среди которых выделим OpenCL [7], OpenACC [8], SYCL (DPC++) [9], Kokkos [10] и другие. Действительно, в целом проблема переносимости кода между разными программно-аппаратными платформами удовлетворительно решается, что потенциально может позволить реализовывать коды для научного моделирования с использованием соответствующих языков/библиотек, а далее собирать и запускать их на разных платформах. Однако при этом встает вопрос о переносимости производительности подобных реализаций (performance portability) и эффективности использования вычислительных ресурсов. Наш опыт показывает, что код, изначально написанный в логике систем, построенных на многоядерных центральных процессорах архитектур x86 или ARM, не очень хорошо переносится на GPU-архитектуры, и наоборот. Еще сложнее добиться приемлемой производительности при одновременном использовании всех устройств даже в случае реализации эффективной схемы балансировки вычислительной нагрузки. Проблема чаще всего связана с накладными расходами на передачу данных между разными устройствами. По нашему мнению, проблема переносимости производительности в общем случае на сегодняшний день не решена.

Доклад посвящен обсуждению этих вопросов в контексте накопленного в коллективе опыта по разработке переносимых вычислительных кодов с использованием разных средств для гетерогенного программирования [11-14]. В частности, мы изучаем особенности применения C++, DPC++, Kokkos при решении двух существенно отличающихся задач. Первая задача из области биоинформатики [14] заключается в поиске наиболее часто встречающихся последовательностей ДНК и сводится к использованию параллельных алгоритмов на строках. Вторая задача относится к области вычислительной физики и посвящена реализации метода конечных разностей во временной области (Finite-Difference Time-Domain) для численного интегрирова-

ния уравнений Максвелла [15]. Дополнительно в рассмотрение включен Co-Array Fortran, один из перспективных инструментов, являющихся абстракцией над низкоуровневыми способами реализации параллелизма в системах с общей и распределенной памятью. Co-Array Fortran пока не решает задачу переносимости кода на архитектуры GPU, однако является важным инструментом для части сообщества, программирующей на Fortran.

На рисунке 1 приведены результаты экспериментов в задаче об интегрировании уравнений Максвелла. Результаты показывают, что с использованием разных языков и библиотек для параллельного программирования возможно добиться приемлемой производительности, сравнимой с достаточно низкоуровневой платформенно-зависимой реализацией на C++. В докладе будет рассказано об особенностях использования DPC++, Kokkos и Co-Array Fortran, приведены методические рекомендации по разработке кодов для научного моделирования с использованием этих языков и библиотек, а также подробно изложены основные выводы по проделанной работе.

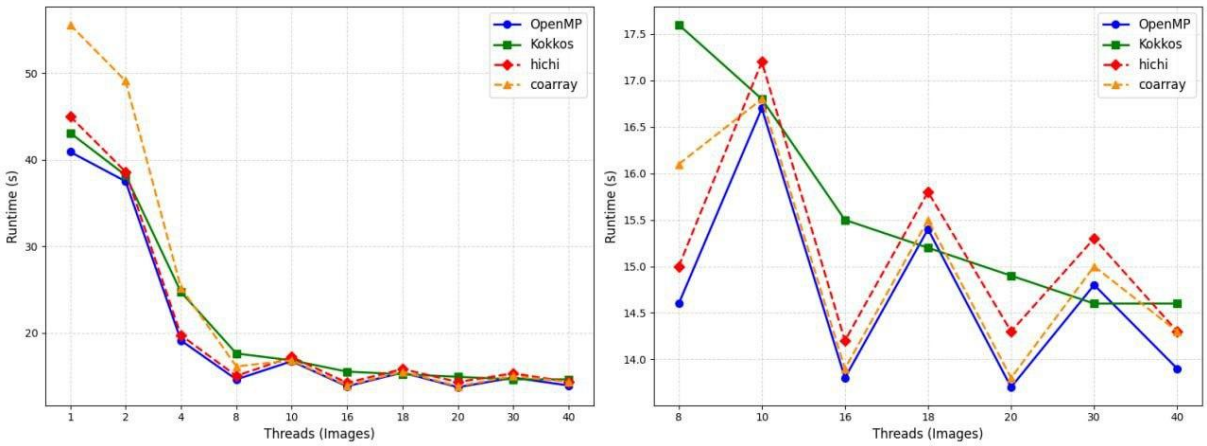


Рис. 1. Эксперименты по использованию разных языков и библиотек для параллельных вычислений в задаче интегрирования уравнений Максвелла методом FDTD. Разработаны версии кода на OpenMP, Kokkos и Co-Array Fortran. Результаты приведены в сравнении с реализацией из программного комплекса Hi-Chi (реализация на DPC++). На графиках приведено время работы с зависимости от числа используемых потоков. Эксперименты выполнены на двухпроцессорном узле (20 ядер, 40 потоков) суперкомпьютера Лобачевский

Литература

1. Dongarra J. J., Luszczek P., Petitet A. The LINPACK Benchmark: past, present and future // Concurrency and Computation: practice and experience. – 2003. – Т. 15. – №. 9. – С. 803-820.
2. Dongarra J., Heroux M. A., Luszczek P. High-performance conjugate-gradient benchmark: A new metric for ranking high-performance computing systems // The International Journal of High Performance Computing Applications. – 2016. – Т. 30. – №. 1. – С. 3-10.
3. Murphy R. C. et al. Introducing the graph 500 // Cray Users Group (CUG). – 2010. – Т. 19. – №. 45-74. – С. 22.
4. Dongarra J., Keyes D. The co-evolution of computational physics and high-performance computing // Nature Reviews Physics. – 2024. – Т. 6. – №. 10. – С. 621-627.
5. Leiserson C. E. et al. There’s plenty of room at the Top: What will drive computer performance after Moore’s law? // Science. – 2020. – Т. 368. – №. 6495. – eaam9744.
6. Surmin I. A. et al. Моделирование плазмы методом частиц в ячейках с использованием сопроцессоров Intel Xeon Phi // Вычислительные методы и программирование. – 2014. – Т. 15. – С. 530-536.
7. Munshi A. et al. OpenCL programming guide. – Pearson Education, 2011.
8. Farber R. Parallel programming with OpenACC. – Newnes, 2016.
9. Reinders J. et al. Data parallel C++: mastering DPC++ for programming of heterogeneous systems using C++ and SYCL. – Springer Nature, 2021.

10. Trott C. R. et al. Kokkos 3: Programming model extensions for the exascale era // IEEE Transactions on Parallel and Distributed Systems. – 2021. – Т. 33. – №. 4. – С. 805-817.
11. Volokitin V. et al. High performance implementation of Boris particle pusher on DPC++. A first look at oneAPI // International Conference on Parallel Computing Technologies. – Cham: Springer International Publishing, 2021. – С. 288-300.
12. Kurnikova A. et al. High-Performance Graph Coloring on Intel CPUs and GPUs Using SYCL and KOKKOS // International Conference on Mathematical Modeling and Supercomputer Technologies. – Cham : Springer Nature Switzerland, 2022. – С. 253-265.
13. Panova E. et al. Black-Scholes Option Pricing on Intel CPUs and GPUs: Implementation on SYCL and Optimization Techniques //Russian Supercomputing Days. – Cham : Springer International Publishing, 2022. – С. 48-62.
14. Козлов М. А., Панова Е. А., Мееров И. Б. Реализация поиска наиболее часто встречающихся последовательностей ДНК с использованием библиотеки Kokkos // Проблемы информатики. – 2024. – №. 2 (63). – С. 58-71.
15. Taflove A., Hagness S. C., Piket-May M. Computational electromagnetics: the finite-difference time-domain method // The Electrical Engineering Handbook. – 2004. – С. 629-670.

ГИБРИДНЫЙ ЭВОЛЮЦИОННЫЙ АЛГОРИТМ ДЛЯ ЗАДАЧИ 2-CORRELATION CLUSTERING[1*]

Моршинин А.В.

Институт математики им. С.Л. Соболева СО РАН

Рассматривается вариант задачи CORRELATION CLUSTERING, известной также как ЗАДАЧА КЛАСТЕРИЗАЦИИ ВЕРШИН ГРАФА, в которой число кластеров не превосходит 2. Эта задача является NP-трудной. Приведен обзор известных алгоритмов её приближённого решения, а также предложен новый алгоритм, опирающийся на эволюционные вычисления. Представлены результаты вычислительного эксперимента, направленного на сравнение эффективности изучаемых алгоритмов.

Ключевые слова: кластерный граф, NP-трудная задача, приближенный алгоритм, гибридный эволюционный алгоритм.

1. Постановка задачи

В задаче *кластеризации* требуется разбить заданное множество объектов на подмножества (кластеры) опираясь на сходство объектов друг с другом. Задача CORRELATION CLUSTERING [1, 2] является наглядной формализацией этой задачи: похожим объектам соответствуют ребра графа, а вершины взаимно однозначно соответствуют объектам.

Будем рассматривать только *обыкновенные графы*, то есть графы без петель и кратных ребер. Обыкновенный граф называется *кластерным*, если каждая его компонента связности является полным графом. Расстоянием $d(G_1, G_2)$ между двумя помеченными графами $G_1 = (V, E_1)$ и $G_2 = (V, E_2)$ является мощность симметрической разности $E_1 \Delta E_2$. Окрестностью $N_G(v)$ вершины v в графе G называется множество вершин этого графа, смежных с v. Обозначим через $\mathbf{CGS}_{\leq 2}(V)$ множество всех кластерных графов на множестве вершин V, имеющих не более двух кластеров. $C = C(X, Y)$ – кластерный граф с не более чем 2 кластерами (возможно $Y = \emptyset$).

$\mathbf{CC}_{\leq 2}$. Для произвольного графа $G = (V, E)$ найти ближайший к G кластерный граф $C^* \in \mathbf{CGS}_{\leq 2}(V)$, то есть граф, для которого величина $d(G, C^*)$ минимальна среди всех графов из $\mathbf{CGS}_{\leq 2}(V)$.

Известно, что эта задача является NP-трудной [1, 2]. В работе предложен новый гибридный эволюционный алгоритм, а также проведен вычислительный эксперимент, направленный на сравнение эффективности изучаемых алгоритмов.

2. Алгоритмы решения

2.1. Известные алгоритмы

В 2004 г. Бансал, Блюм и Чаула [1] предложили 3-приближенный алгоритм для задачи $\mathbf{CC}_{\leq 2}$.

Вход: граф $G = (V, E)$.
Выход: кластерный граф $C_N \in \mathbf{CGS}_{\leq 2}(V)$.
Шаг 0. Положить $F = \emptyset$.
Шаг 1. Для каждой вершины $v \in V$ определить кластерный граф $C_v = C(X, Y)$, где $X = \{v\} \cup N_G(v)$, $Y = V \setminus X$. Положить $F = F \cup \{C_v\}$.
Шаг 2. Среди всех графов из множества F выбрать ближайший к G кластерный граф C_N.

Рис. 1. Алгоритм **N** (Neighborhood)

[1*] Исследование выполнено за счет гранта Российского научного фонда № 22-71-10015-П.

Нетрудно заметить, что трудоемкость алгоритма **N** – $O(|V|^2)$.

В 2008 г. Колман, Саундерсон и Вирт [3] предложили следующую процедуру локального поиска.

Итерация *k*.
1. Для каждой вершины вычислить величину уменьшения целевой функции при её переносе в противоположный кластер.
2. Если для всех вершин вычисленная величина неположительна ($\Delta \le 0$), **завершить** алгоритм и вернуть текущее решение.
3. Иначе перенести вершину с максимальным положительным уменьшением в противоположный кластер и **перейти к итерации *k* + 1**.

Рис. 2. Процедура **LS** (Local Search)

Они также предложили 2-приближенный алгоритм для задачи $\mathbf{CC}_{\le 2}$.

Вход: граф $G = (V, E)$.
Выход: кластерный граф $C_{NmLS} \in \mathbf{CGS}_{\le 2}(V)$.
Шаг 1. Пусть F – множество всех кластерных графов, построенных алгоритмом **N** на шаге 1. Применить процедуру **LS** к каждому графу из F.
Шаг 2. Среди всех локальных оптимумов выбрать ближайший к G кластерный граф C_{NmLS}.

Рис. 3. Алгоритм **N+mLS** (Neighborhood + multiple Local Search)

Заметим, что трудоемкость процедуры **LS** – $O(|V|^3)$, а алгоритма **N+mLS** – $O(|V|^4)$.

Достаточно легко получить еще один 3-приближенный алгоритм, использующий процедуру **LS** однократно.

Вход: граф $G = (V, E)$.
Выход: кластерный граф $C_{N1LS} \in \mathbf{CGS}_{\le 2}(V)$.
Шаг 1. Пусть F – множество всех кластерных графов, построенных алгоритмом **N** на шаге 1. Применить процедуру **LS** к ближайшему к G кластерному графу из F. Обозначить этот граф через C_{N1LS}.

Рис. 4. Алгоритм **N+1LS** (Neighborhood + single Local Search)

Трудоемкость этого алгоритма – $O(|V|^3)$.

2.2. Эволюционный алгоритм

Алгоритм **N+mLS** обладает хорошей оценкой точности, но характеризуется высокой трудоёмкостью. Перспективной представляется идея сохранить преимущества процедуры **LS**, применяя её к ограниченному, не зависящему от размера графа множеству допустимых решений. Данная идея реализована в следующем *гибридном эволюционном алгоритме* [4], общая схема которого описана ниже.

Вход: граф $G = (V, E)$.
Выход: кластерный граф $C_E \in \mathbf{CGS}_{\le 2}(V)$.
Шаг 0. Положить начальную популяцию Π^0 равной множеству всех кластерных графов, построенных алгоритмом **N** на шаге 1. Положить $C_E = C_{N1LS}$, где C_{N1LS} – решение, построенное алгоритмом **N+1LS**.
Шаг 1. Пока не **выполнен критерий остановки**, повторять следующее:
Итерация *k*. Положить $\Pi^k = \emptyset$. Пока в популяции не M особей, повторять:
1. Выбрать p_1, p_2 из $\Pi^{(k-1)}$ оператором селекции.
2. Применить к p_1, p_2 оператор кроссинговера, получить c.
3. Применить оператор мутации к c, получить c'.
4. Добавить c' в популяцию Π^k.

Если в популяции есть особь C, лучшая, чем C_E, то положить $C_E = C$. Если **выполнен критерий встряски**, выполнить встряску популяции.

Рис. 5. Алгоритм **E** (Evolution)

Опишем основные параметры алгоритма **E**.

Кодирование решений. Разбиение множества вершин графа $G = (V, E)$ на два кластера кодируется бинарным вектором длины $|V|$. При этом значение i-й координаты вектора определяет принадлежность вершины v_i: если оно равно 0, вершина относится к кластеру X, иначе – к кластеру Y. Без потери общности можно считать, что вершина v_1 всегда принадлежит кластеру X.

Размер популяции. Размер популяции M равен 128, за исключением нулевой популяции, которая формируется с помощью алгоритма **N**.

Критерий остановки. Существуют различные критерии остановки вычислений: по истечении времени, по достижении заданного числа итераций и другие. В алгоритме **E** используется следующий критерий: если в течение последних 6 итераций не происходило обновление рекорда, вычисления останавливаются.

Оператор селекции. Используется турнирная селекция с размером турнира 10. В качестве функции приспособленности используется целевая функция задачи.

Оператор кроссинговера. Для поддержания разнообразия популяции между итерациями применяется следующая процедура скрещивания. Сначала формируются кластеры, включающие вершины, принадлежащие одному кластеру у обоих родителей. На этом этапе может образоваться до 4 кластеров. Затем, пока количество кластеров превышает 2, на каждом шаге объединяются два кластера, чье слияние минимально увеличивает целевую функцию. Когда остается 2 кластера, выполняется проверка: если их объединение уменьшает целевую функцию, кластеры сливаются. В результате получается допустимое решение, содержащее не более двух кластеров.

Оператор мутации совпадает с процедурой **LS**.

Встряска и критерий встряски. В процессе эволюции популяция стремится к заполнению локальными оптимумами. Для выхода из них применяется процедура встряски [5], которая активируется каждую пятую итерацию. Процедура выполняется 5 раз для каждой особи: на каждом шаге каждая вершина с вероятностью 0.1 переносится в противоположный кластер независимо от изменения целевой функции.

Легко видеть, что алгоритм **E** является 3-приближенным алгоритмом решения задачи $\mathbf{CC}_{\leq 2}$ за счет начального рекорда, заданного алгоритмом **N+1LS**, а его трудоемкость – $O(|V|^3)$.

3. Вычислительный эксперимент

Для оценки эффективности представленных алгоритмов был проведен вычислительный эксперимент. Заметим, что для любого графа значение целевой функции решений, полученных алгоритмами **N+mLS**, **N+1LS** и **E**, не превышает значения для алгоритма **N**, последний был выбран в качестве бейслайна. Было вычислено среднее процентное улучшение (уменьшение) целевой функции остальных алгоритмов относительно этого бейслайна.

Эксперименты проводились на наборе данных вопросов и ответов с сайта Stack Overflow (https://www.kaggle.com/datasets/stackoverflow/stacksample). Каждый объект в наборе данных помечен тегами (Java, JavaScript, ML и др.). Для адаптации данных к задаче кластеризации была введена мера схожести: для пары объектов вычисляется коэффициент Жаккара [6] между множествами их тегов. Объекты считаются схожими, если этот коэффициент превышает 0.5. Для упрощения эксперимента были отобраны только вопросы с тегами из топ-10 по популярности на Stack Overflow.

Исследовались графы размером 1000, 2000, 3000, 4000 и 5000 вершин. Для генерации графа заданного размера из набора данных равномерно случайно выбиралось соответствующее количество объектов, после чего ребра формировались согласно описанной мере схожести. Для каждого размера графа было решено по 20 тестовых задач.

Все алгоритмы были реализованы на языке C++20 в среде разработки JetBrains CLion 2025.2. Вычисления выполнялись на персональном компьютере с процессором Intel Core i7-9700 (тактовая частота 3.0 ГГц) и объёмом оперативной памяти 32 ГБ.

Таблица 1. Среднее снижение значения целевой функции относительно бейслайна, %

	N+1LS	**N+mLS**	**E**
1000	21.23	23.44	23.74
2000	20.31	23.15	23.32
3000	20.10	22.99	23.17
4000	20.56	23.70	23.84
5000	20.97	23.67	23.73

Согласно данным таблицы 1, наилучшие результаты относительно бейслайна показывает алгоритм **E**. На втором месте находится алгоритм **N+mLS**, тогда как алгоритм **N+1LS** демонстрирует наихудшие результаты. Отметим, что во всех тестовых примерах алгоритм **E** находил наилучшее решение среди все исследуемых алгоритмов.

Таблица 2. Среднее время работы алгоритмов, сек

	N	**N+1LS**	**N+mLS**	**E**
1000	0	0	1	3.25
2000	2.2	2.2	12	14.25
3000	9.4	9.45	46.5	41.1
4000	21.2	21.4	103.45	72.3
5000	39.75	40	191.95	113.75

Из таблицы 2 видно, что с ростом числа вершин алгоритм **E** становится быстрее алгоритма **N+mLS**. Время работы алгоритмов **N** и **N+1LS** растет медленнее.

Таким образом алгоритм **E** является оптимальным выбором при решении задачи $\mathbf{CC}_{\leq 2}$. Качество его решений значительно выше, чем у алгоритмов **N** и **N+1LS**, хотя и незначительно выше, чем у алгоритма **N+mLS**. Однако время его работы на больших графах меньше, чем у алгоритма **N+mLS** и разрыв растет с увеличением числа вершин.

Литература

1. Bansal N., Blum A., Chawla S. Correlation clustering. Machine Learning. 2004. Vol. 56. P. 89-113.
2. Giotis I., Guruswami V. Correlation clustering with a fixed number of clusters. Theory of Computing. 2006. Vol. 2. No. 1. P. 249-266.
3. Coleman T., Saunderson J., Wirth A. A local-search 2-approximation for 2-correlation clustering. LNCS. 2008. Vol. 5193. P. 308-319.
4. Evolutionary Computation 1: Basic Algorithms and Operators. Taylor & Francis, New York. 2000.
5. Борисовский П.А. Параллельный алгоритм «иди с победителями» для некоторых задач составления расписаний. Дискретн. анализ и исслед. опер. 2023. Т. 30. № 4. С. 5-23.
6. Jaccard P. Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines. Bull. Soc. Vaudoise sci. Natur. 1901. Vol. 37. No. 140. P. 241-272.

О ТРЕХ ОПОРНЫХ ЗАДАЧАХ СУПЕРКОМПЬЮТЕРНОГО КОДИЗАЙНА[1*]

Д.А. Никитенко

Московский государственный университет имени М.В. Ломоносова

Вместе с развитием архитектур и вычислительных методов, вопрос совместного выбора эффективных сочетаний аппаратуры, вычислительных методов и способов их реализации становился все острее. На данный момент доступен широчайший спектр архитектур и огромный багаж вычислительных методов, основанных на них алгоритмов, а также множество ранее апробированных собственных и сторонних программных реализаций. В данной работе будут описаны три наиболее часто встречающихся, опорных постановок задач суперкомпьютерного кодизайна и будут предложены схемы решения этих задач на опыте разработчиков открытой энциклопедии алгоритмов AlgoWiki и сопряженного с ней сервиса составления рейтингов эффективных реализаций алгоритмов Algo500.

Ключевые слова: кодизайн, высокопроизводительные вычисления, параллельные вычисления, параллельные программы, эффективность реализаций алгоритмов, эффективность вычислительных архитектур, совместный анализ свойств алгоритмов и архитектур.

1. Введение

Суперкомпьютерный кодизайн [1–7], относительно недавно выделенный как отдельное направление исследований в области высокопроизводительных вычислений, существует, конечно же, чуть ли не с самого появления вычислительных систем. Вместе с развитием архитектур и вычислительной математики, вопрос выбора эффективных сочетаний аппаратуры, вычислительных методов и способов их реализации естественным образом становился все острее. На данный момент исследователям и пользователям вычислительных систем доступен широчайший спектр архитектур и огромный накопленный багаж вычислительных методов, основанных на них алгоритмов, а также множество ранее апробированных собственных и сторонних программных реализаций.

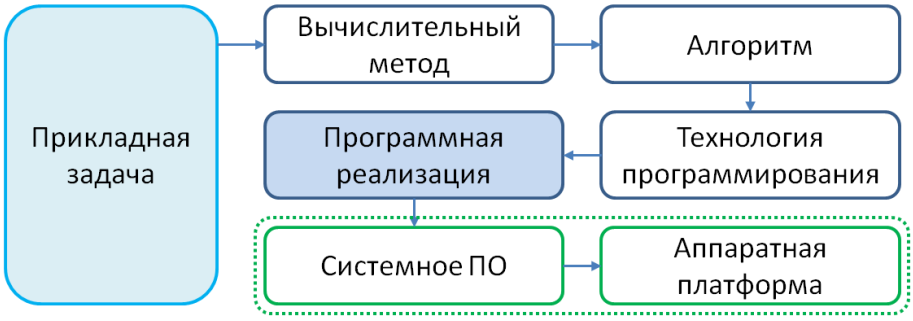


Рис. 1. Классическая цепочка решения прикладной задачи

Под суперкомпьютерным кодизайном понимают решение разного рода задач – от создания эффективного приложения до проектирования узкоспециализированных вычислительных систем, нацеленных на определенный класс приложений, или даже одно конкретное. Общее в этих задачах одно – для успешного их решения необходимо использовать совместный анализ

[1*] Исследование выполнено за счет гранта Российского научного фонда № 25–11–00181, https://rscf.ru/project/25-11-00181/. Работа выполнена с использованием оборудования Центра коллективного пользования сверхвысокопроизводительными вычислительными ресурсами МГУ имени М.В. Ломоносова [33].

свойств всего множества задействованных компонент - вычислительных методов, реализующих их алгоритмов, технологий программирования, программных реализаций, и, завершая, целевых аппаратных платформ. На рисунке 1 представлена классическая цепочка решения прикладных задач. технологии кодизайна привносят в эту цепочку дополнительные связи между этапами, в зависимости от того какие они вносят ограничения. Так, например, целевая SMP система почти однозначно определит выбор технологии межпроцессной коммуникации в пользу OpenMP, нежели MPI и т.д.

В данной работе будет приведен анализ трех наиболее часто встречающихся, опорных постановок задач суперкомпьютерного кодизайна и будут предложены схемы решения этих задач на опыте разработчиков открытой энциклопедии алгоритмов AlgoWiki [8] и сопряженного с ней сервиса составления рейтингов эффективных реализаций Algo500 [9, 10].

К указанным задачам можно отнести следующие:

– выбор вычислительного метода и алгоритма;
– выбор вычислительной платформы;
– создание/преобразование приложения с учетом используемого метода и елевой архитектуры.

Почему эти три задачи возникают чаще других, почему их можно назвать опорными – об этом пойдет речь в третьей главе.

Статья имеет следующую структуру. Во втором разделе описаны наиболее развитые средства описания алгоритмической основы и средств формирования рейтингов эффективности реализаций. В третьем разделе рассмотрены три опорные задачи суперкомпьютерного кодизайна с описанием возможных подходов к их решению, в заключении сформулированы основные выводы.

2. О формализации описаний свойств алгоритмов и эффективных реализаций – AlgoWiki и Algo500

2.1. Открытая энциклопедия алгоритмов AlgoWiki

Открытая энциклопедия алгоритмов AlgoWiki (http://algowiki-project.org) – наиболее развитый на данный момент проект, систематизирующий описания алгоритмов с целью эффективного их применения при написании параллельных программ [21–26]. На верхнем уровне структура описания алгоритмов и их реализаций опирается на иерархию – цепочку Задача(З)-Метод(М)-Алгоритм(А)-Реализация(Р) [27, 28], допустимые связи между ее элементами представлены на рисунке 2.

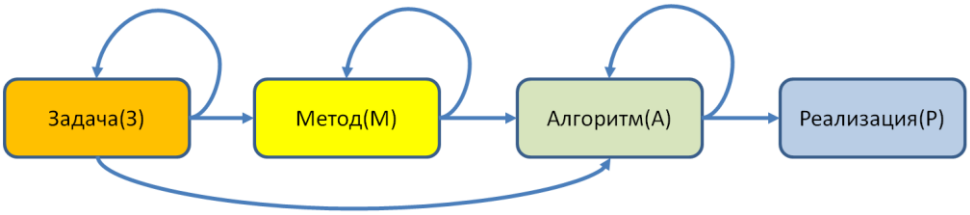


Рис. 2. Допустимые связи в иерархическом описании цепочки Задача(З) – Метод(М) – Алгоритм(А) – Реализация(Р) в энциклопедии AlgoWiki

Задача – любая прикладная задача, например «разложение матриц».

Метод – математический метод или группа методов, используемый для решения прикладных задач, например «метод Гаусса» для треугольного разложения матриц.

Алгоритм – непосредственно алгоритм или группа алгоритмов, реализующих указанный метод, например «LU-разложение методом Гаусса». Данный уровень описания является базовым для AlgoWiki.

Реализация – программная реализация конкретного алгоритма.

Максимальный уровень детализации описания приходится на уровень Алгоритмов, структура описания алгоритма содержит общее описание алгоритма, математическое описание алгоритма, описание вычислительного ядра алгоритма, макроструктуру алгоритма, схему реализа-

ции последовательной версии алгоритма, последовательную сложность алгоритма, информационный граф, оценку ресурса параллелизма, описание входных и выходных данных, свойства алгоритма.

На рисунке 3 представлен фрагмент реальной цепочки Задача-Метод-Алгоритм-Реализация с сайта проекта.



Рис. 3. Пример иерархического описания цепочки Задача(З) – Метод(М) – Алгоритм(А) – Реализация(Р) в энциклопедии AlgoWiki

Важно, что при всей богатейшей наполненности, проект ограничивается предоставлением данных для исследователей, но не предоставляет сервисов для сравнения тех или элементов цепочки.

2.2. Платформа для составления рейтингов производительности Algo500

Восполнение пробела, указанного в предыдущем параграфе, является целью проекта Algo500 [29–31]. Идея достаточно проста – связать имеющиеся формальные описания алгоритмов и методов в AlgoWiki или подобные с данными производительности программных реализаций на конкретных платформах, что влечет за собой и описание этих программно-аппаратных платформ и условий запусков, и на множестве полученных данных строить «срезы» – рейтинги производительности тех или иных методов или алгоритмов, либо производительности той или

иной платформы на множестве выбранных методов или алгоритмов. Это можно рассматривать как расширение набора известных суперкомпьютерных рейтингов (Top500, Graph500 и т.д. [29, 30]) на произвольное множество реализаций алгоритмов из энциклопедии AlgoWiki [32].

Как следствие, в вышеупомянутой цепочке появляется еще один компонент – описание компьютерной системы (рисунок 4), на которой был выполнен запуск конкретной программной реализации, для которого получены те или иные показатели производительности.

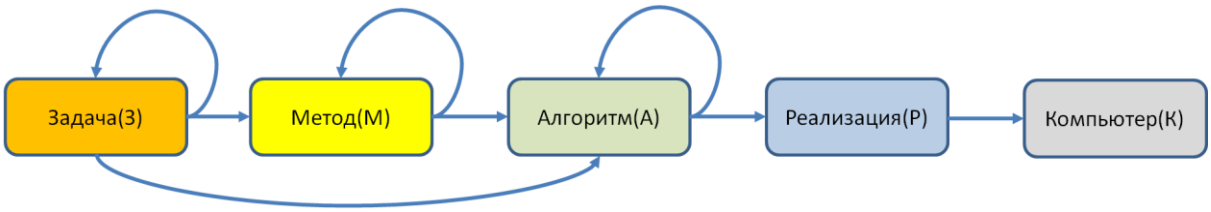


Рис. 4. Допустимые связи в иерархическом описании цепочки Задача(З)-Метод(М)-Алгоритм(А)-Реализация(Р)- Компьютер(К) в Algo500

Важно, что при этом возникает вопрос не только детализации описания вычислительных систем, но и описания конкретного подмножества системы, использованного для запуска конкретной реализации. На рисунке 5 представлена схема одной из версий описания систем в рамках проекта Algo500.

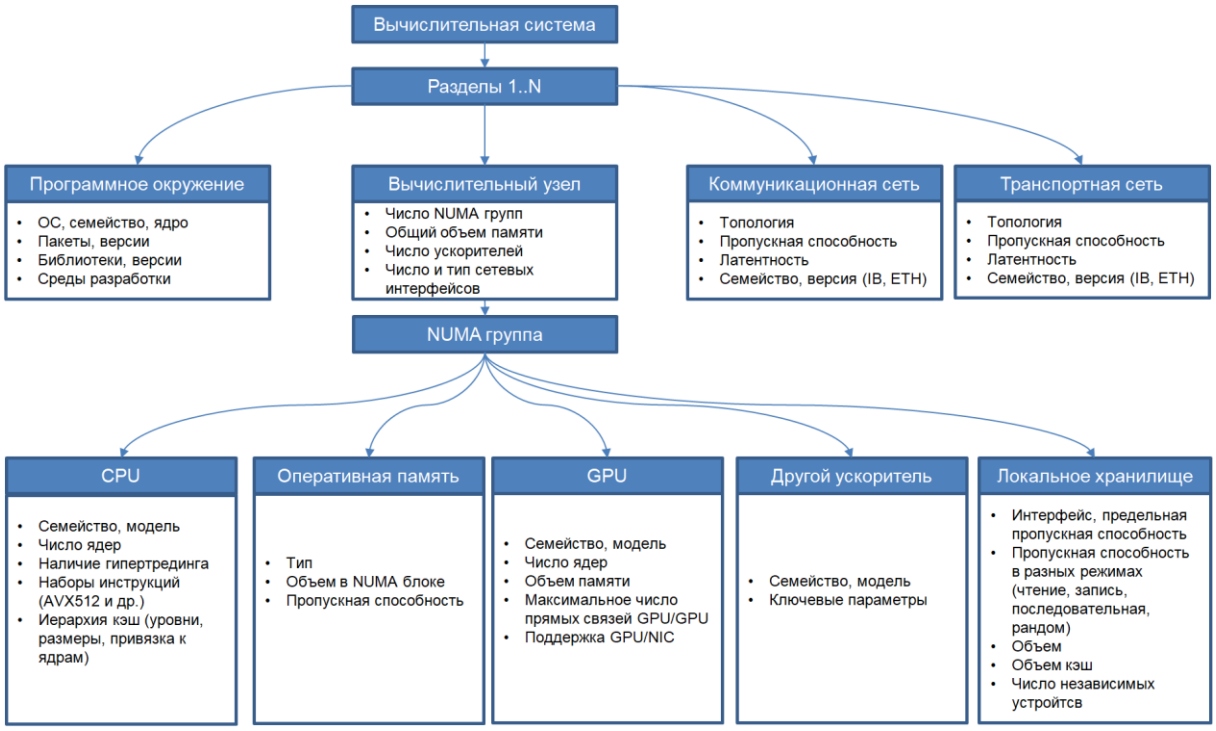


Рис. 5. Пример схемы описания значимых характеристик вычислительной системы в проекте Algo500

В рамках апробации в базы данных Algo500 заносились данные из известных рейтингов Top500, Graph500, Топ50. Кроме того, заносились данные множественных запусков на отдельных высокопроизводительных вычислительных системах МГУ, таких как Ломоносов-2, Ломоносов, Регата.

В целом проект стартовал в том виде, в котором его проектировали, но по результатам его разработке обнажилось множество проблем как технологического, так и идейного характера, которые привели к старту нового проекта, обобщающего заделы, полученные при создании, наполнении и поддержке как открытой энциклопедии AlgoWiki, так и проекта Algo500. Об основных целях и задачах, стоящих перед новым проектом, пойдет речь в следующей главе настоящей работы.

3. Три опорные задачи кодизайна и возможные способы их решения

У совместного анализа свойств вычислительных методов, алгоритмов, программно-аппаратных платформ может быть множество постановок задач. Практика показывает, что большинство этих задач сводится к следующим трем постановка, что объясняет частоту их обсуждения.

3.1. Выбор вычислительной платформы

Во-первых, это выбор вычислительной платформы при фиксированном методе и/или алгоритме (рисунок 6). Действительно, при наличии приоритетного метода или алгоритма (например, многолетняя собственная разработка, имеющаяся лицензия на программный пакет и т.п.) при многообразии доступных на сегодняшний день архитектур (SMP, кластерные системы различных топологий, GPU ускорители, векторные системы и т.д.) выбор целевой платформы возникает часто. У пользователей помимо возможных собственных систем различной архитектуры есть возможность арендовать вычислительные ресурсы практически любого вида.

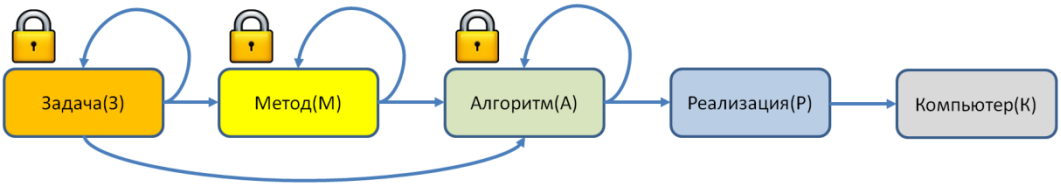


Рис. 6. Выбор вычислительно платформы для имеющейся задачи/метода/алгоритма

Некоторые из ограничений при таком выборе вытекают из свойств алгоритма, но все же это помогает определить лишь классы систем, на которых может быть эффективно решена данная задача. Для сравнения между собой близких по характеристикам систем уже необходимы данные практических экспериментов, а они, безусловно, зависят от качества программных реализаций. Тем не менее, если вести речь об энциклопедии алгоритмов, то особенно для наиболее популярных методов и алгоритмов, есть надежда иметь достаточное количество различных программных реализаций на разных платформах.

Таким образом, формально для подбора платформы необходимо 1) выбрать подходящие классы систем исходя из свойств алгоритма и 2) для выбранных классов построить сравнительный рейтинг известных референсных реализаций.

Выделение высокоуровневых свойств аппаратных платформ, являющихся ключевыми для успешной реализации алгоритмов, позволило бы частично автоматизировать данную последовательность действий.

Наиболее заметным примером решения такой задачи является создание специализированных суперкомпьютеров семейства Anton [31, 32].

3.2. Выбор вычислительного метода и алгоритма

Другая распространенная задача – оценка, какими приложениями, например, приложениями из какой области исследований, целесообразно загружать определенную архитектуру (рисунок 7).

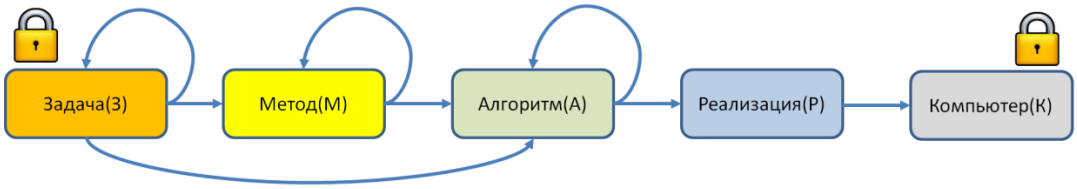


Рис. 7. Выбор вычислительного метода и алгоритма

В том числе, например, если у пользователей в организации имеется вычислительная система определенной архитектуры, то понимание ее особенностей может подсказать правильный выбор методов ил алгоритмов при решении конкретной прикладной задачи.

Для решения данной задачи можно построить рейтинг реализаций всех методов и алгоритмов, имеющих отношение либо к интересующей задаче, либо вообще всех, ограничив выборку характеристиками вычислительных систем на интересующем уровне абстракции. Это может быть пропускная способность памяти, число ядер в SMP узле, класс систем (кластерная, гибридная с GPU) и т.д.

Для реализации данного сценария в сервисе Algo500 не было описания подмножества системы, на которой был осуществлен запуск, а также не были реализованы классы и категории в описании систем.

3.4. Создать/преобразовать приложение с учетом метода и архитектуры

Третья постановка задачи касается практически каждого пользователя. Имея не только приоритетный метод или алгоритм из первой задачи, но и работающую реализацию для определенной вычислительной системы (рисунок 8), возможно, созданную с огромным трудом, иногда встает вопрос – а насколько эффективна эта реализация?

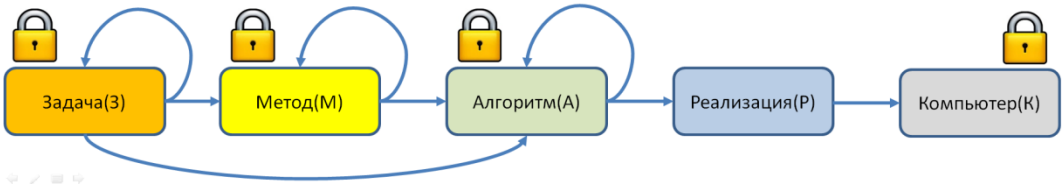


Рис. 8. Создание или модификация приложения

Для оценки производительности собственной реализации, а при нахождении более удачной – изучение ее особенностей, целесообразно было бы строить выборку одновременно по упомянутым в первой и второй задаче критериям.

При этом с одной стороны интересно прямое пересечение – на нахождение именно таких реализаций для таких платформ и методов/алгоритмов. С другой – интересно отдельно оценить удачность своей реализации с точки зрения используемого метода или выбранной целевой платформы. Как следствие, принять решение, быть может, стоит один из этих критериев сделать не столь жестким и рассмотреть развитие своего приложения в сторону использования новых методов или использования особенностей других архитектур.

Как бы там ни было, найдя некоторые реализации с более привлекательной производительностью/эффективностью, пользователю необходимо понять, что же он может изменить в своей программе, а для этого было бы крайне полезно посмотреть детали заинтересовавшей референсной реализации.

Такую возможность предоставил бы репозиторий референсных реализаций в дополнение к развитой системе построения выборок из первых двух задач.

4. Заключение

Суперкомпьютерный кодизайн – не просто абстрактное направление исследований в области высокопроизводительных вычислений, а каждодневная реальность большинства разработчиков параллельных приложений и суперкомпьютерных систем.

К выделенным трем задачи кодизайна (подбор метода/алгоритма при фиксированной целевой архитектуре, подбор архитектуры при фиксированном методе или алгоритме, поиск более эффективной реализации, использующей тот или иной алгоритм и ту или иную программно-аппаратную платформу) может быть сведено абсолютное большинство задач совместного анализа свойств алгоритмов, вычислительных методов и архитектур. От разработки приложений максимально учитывающих, например, структуру оперативной памяти, до разработки специализированных вычислительных систем для решения определенного класса задач.

Выделенные три опорные задачи с одной стороны, могут быть принципиально решены на базе разработанной открытой энциклопедии алгоритмов AlgoWiki и системы построения рейтингов производительности программных реализаций (а значит, и стоящих за ними алгоритмов и методов), однако уже первичный анализ показывает, что необходим ряд модификаций. В частности, это выделение экспертных оценок критически важных свойств целевой архитектуры для алгоритмов, формальное описание классов и разметка существующих описаний систем (например, SMP, системы с GPU, системы с значительным объемом оперативной памяти на вычислительное ядро, системы с параллельной файловой системой и т.п.), разработка методов описания подмножества систем, на которых запускались программные реализации и описание таких экспериментальных запусков в целом.

Решение поставленных задач позволит получить качественно новые инструменты анализа, доступные как высококвалифицированным специалистам в данной области, так и рядовым пользователям суперкомпьютерных систем, что должно вылиться и в повышение эффективности использования существующих высокопроизводительных вычислительных систем и программных комплексов, и повышение качества вновь создаваемых.

Литература

1. A. Geist and S. Dosanjh, “IESP Exascale Challenge: Co-Design of Architectures and Algorithms,”The International J. of High Performance Computing Applications 23 (4), 401–402 (2009). https://doi.org/10.1177/1094342009347766.
2. X. Hu, R. Murphy, S. Dosanjh, K. Olukotun, and S. Poole, “Hardware/software co-design for high performance computing: Challenges and opportunities,” IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis, CODES+ISSS, 63–64 (2010). https://doi.org/10.1145/1878961.1878975.
3. S. Dosanjh, R. Barrett, M. Heroux, and A. Rodrigues, “Achieving Exascale Computing through Hardware/Software Co-design,” EuroMPI, Lecture Notes in Computer Science 6960, 5–7 (2011). https://doi.org/10.1007/978-3-642-24449-0_2.
4. J. Ang, R. Brightwell, D. Donofrio, S. Dosanjh, K. S. Hemmert, A. Rodrigues, J. Shalf, and K. Wheeler, “Exascale computing and the role of co-design,” Advances in Parallel Computing, 20, 43–64 (2011). https://doi.org/10.3233/978-1-60750-803-8-43.
5. K. Chee, “Applied HW/SW Co-design: Using the Kendall Tau Algorithm for Adaptive Pacing,” Master’s Theses (2013). https://doi.org/10.15368/theses.2013.77 14. S. S. Dosanjh, R. F. Barrett, D. W. Doerfler, S. D. Hammond, K. S. Hemmert, M. Heroux, P. Lin, K. T. Pedretti, A. Rodrigues, T. G. Trucano, and J. Luitjens, “Exascale design space exploration and co-design,” Future Generation Computer Systems 30, 46–58 (2014). https://doi.org/10.1016/j.future.2013.04.018.
6. K. Mehta, I. Foster, S. Klasky, B. Allen, M. Wolf, J. Logan, E. Suchyta, J. Choi, K. Takahashi, I. Yakushin, and T. Munson, “A Codesign Framework for Online Data Analysis and Reduction,” EEE/ACM Workflows in Support of Large-Scale Science (WORKS), 11–20 (2019). https://doi.org/10.1109/WORKS49585.2019.00007.
7. Antonov, A.S., Maier, R.V., Nikitenko D.A., Voevodin V.V. An Approach to Solving the Problem of Supercomputer Co-design // Lobachevskii J Math. Vol. 45, N 7. Pp. 2965-2973 (2024). DOI: 10.1134/S1995080224603680.
8. Alexander Antonov, Jack Dongarra, Vladimir Voevodin. AlgoWiki Project as an Extension of the Top500 Methodology // Supercomputing Frontiers and Innovations, Vol.5, No.1 (2018). Pp.4-10. DOI: 10.14529/jsfi180101.
9. Antonov, A.S., Nikitenko, D.A. & Voevodin, V.V. Algo500 - A New Approach to the Joint Analysis of Algorithms and Computers // Lobachevskii J Math. Vol. 41, N 6. Pp. 1435-1443 (2020). DOI: 10.1134/S1995080220080041.
10. Antonov, A.S., Maier, R.V. Development and Implementation of the Algo500 Scalable Digital Platform Architecture // Lobachevskii J Math. Vol. 43, N 7. Pp. 837-847 (2022). DOI: 10.1134/S1995080222070058.
11. List of algorithms – Wikipedia. https://en.wikipedia.org/wiki/List_of_algorithms. (Дата обращения: 10 ноября 2025).

12. DSA Tutorial – Learn Data Structures and Algorithms – GeeksforGeeks. https://www.geeksforgeeks.org/dsa/dsa-tutorial-learn-data-structures-and-algorithms/. (Дата обращения: 10 ноября 2025).
13. Parallel Algorithms of the Standard Template Library. https://www.modernescpp.com/index.php/parallel-algorithm-of-the-standard-template-library/. (Дата обращения: 10 ноября 2025).
14. Алгоритмика. https://algorithmica.org/. (Дата обращения: 10 ноября 2025).
15. GitHub – algorithmica-org/algorithmica: A computer science textbook. https://github.com/algorithmica-org/algorithmica. (Дата обращения: 10 ноября 2025).
16. MAXimal :: algo. http://e-maxx.ru/algo/. (Дата обращения: 10 ноября 2025).
17. Main Page – Algorithms for Competitive Programming. https://cp-algorithms.com/. (Дата обращения: 10 ноября 2025).
18. Algorithm Repository. https://www.algorist.com/algorist.html. (Дата обращения: 10 ноября 2025).
19. Algocode wiki. https://wiki.algocode.ru. (Дата обращения: 10 ноября 2025).
20. Алгоритмы – IT wiki ru. https://www.it-wiki.com.ru/algorithms/. (Дата обращения: 10 ноября 2025).
21. M. Heroux and J. Dongarra, “Toward a New Metric for Ranking High Performance Computing Systems,” UTK EECS Tech Report and Sandia National Labs Report SAND2013-4744 (2013).
22. Home – | TOP500. https://www.top500.org. (Дата обращения: 10 ноября 2025).
23. Green500 Data. https://www.top500.org/lists/green500/2022/11/. (Дата обращения: 10 ноября 2025).
24. HPCG Benchmark. https://hpcg-benchmark.org/. (Дата обращения: 10 ноября 2025).
25. Graph 500 | large-scale benchmarks. https://graph500.org. (Дата обращения: 10 ноября 2025).
26. Green Graph500 – Graph 500. https://graph500.org/?page_id=446. (Дата обращения: 10 ноября 2025).
27. IO500 – SC23 – Research List. https://io500.org/. (Дата обращения: 10 ноября 2025).
28. Alexander Antonov, Jack Dongarra, Vladimir Voevodin. AlgoWiki Project as an Extension of the Top500 Methodology // Supercomputing Frontiers and Innovations, Vol.5, No.1 (2018). Pp.4-10. DOI: 10.14529/jsfi180101.
29. D. A. Nikitenko, “Hierarchical model of architecture of supercomputer systems for comparison and ranking,” Bulletin of the South Ural State University. Series "Computational Mathematics and Software Engineering" 11 (4), 5–18 (2022). https://doi.org/10.14529/cmse220401.
30. D. Nikitenko, A. Antonov, A. Zheltkov, and Vl. Voevodin, “Describing HPC System Architecture for Understanding Its Capabilities,” Communications in Computer and Information Science 1331, 425–435 (2020). https://doi.org/10.1007/978-3-030-64616-5_37.
31. D. E. Shaw et al., “ Anton2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer,” SC14 Int. Conf. High Perform. Comput. Networking, Storage Anal., 41–53 (2014). https://doi.org/10.1109/SC.2014.9.
32. D. E. Shaw et al., “ Anton 3: twenty microseconds of molecular dynamics simulation before lunch,” SC’21: Proceedings of the International Conference for High Performance OCComputing, Networking, Storage and Analysis (1), 1–11 (2021). https://doi.org/10.1145/3458817.3487397.
33. V. Voevodin, A. Antonov, D. Nikitenko, P. Shvets, S. Sobolev, I. Sidorov, K. Stefanov, V. Voevodin, and S. Zhumatiy, “ Supercomputer Lomonosov-2: Large Scale, Deep Monitoring and Fine Analytics for the User Community,” Supercomputing Frontiers and Innovations 6 2, 4–11 (2019). https://doi.org/10.14529/jsfi190201.

ОСОБЕННОСТИ СОСТЯЗАТЕЛЬНЫХ АТАК С ОГРАНИЧЕНИЯМИ ПРОТИВ СИСТЕМ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ, ОСНОВАННЫХ НА ПРИМЕНЕНИИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

А.Г. Никольская, Д.А. Рыболовлев

Институт системного программирования им. В.П. Иванникова РАН

В работе представлен анализ особенностей состязательных атак против моделей машинного обучения в доменах с ограничениями, в частности, в области обнаружения вторжений. Исследованы доменные ограничения, которым должны отвечать эффективные состязательные примеры против сетевых систем обнаружения вторжений. Рассмотрены пространства выполнения состязательной атаки с ограничениями. Определены направления дальнейших исследований.

Ключевые слова: система обнаружения вторжений; состязательная атака; машинное обучение; сетевой трафик.

1. Введение

Нарастающая популярность моделей машинного обучения в различных сферах применения, в том числе в системах обнаружения вторжений (СОВ, IDS), предназначенных для выявления компьютерных атак в сетевом трафике, приводит к необходимости защиты таких моделей от потенциальных угроз, одной из которых являются состязательные атаки (adversarial attacks). Целью такой разновидности состязательных атак, как атака «уклонения» (evasion attack), является нарушение нормального функционирования модели: путём добавления в классифицируемый объект малозаметных возмущений обеспечивается некорректный ответ ML-модели после обработки получившегося состязательного примера (adversarial example/sample), например, неправильная классификация такого примера. Подобные атаки можно использовать, в частности, для обхода IDS путём маскирования компьютерных атак в сетевом трафике. Уязвимость ML-моделей к атакам уклонения является критически важной, поскольку такие атаки могут быть осуществлены даже с минимальными знаниями об атакуемой модели.

Состязательные атаки возникли в контексте предметной области (домена) компьютерного зрения [1] и изначально были ориентированы на работу с пиксельными изображениями. Однако в некоторых предметных областях характер обрабатываемых данных (например, их неоднородность или наличие некоторой структуры) накладывает ограничения на вносимые возмущения. В таких доменах внесение произвольных и независимых возмущений в данные при помощи существующих методов состязательных атак может приводить к созданию нефункциональных, нереалистичных состязательных примеров, не соответствующих реальным объектам предметной области, которые можно встретить на практике. Это обуславливает необходимость использовать такие атаки, которые учитывают указанные ограничения, то есть состязательные атаки с ограничениями.

При использовании состязательных атак без учёта ограничений обрабатываемые IDS структурированные пакеты сетевого трафика могут потерять функциональность из-за нарушения лежащего в основе сетевого протокола. Например, даже если атака предложит состязательный пример, в котором значение признака «средняя длина пакета» равно -1 (то есть не попадает в область допустимых значений) или целочисленный признак «максимальная длина пакета» принимает вещественное значение, будет невозможно добиться таких значений на практике. Такой состязательный пример нереалистичен, поскольку соответствующий ему сетевой трафик не может быть воспроизведён в реальной сети. Важно отметить, что для обхода IDS на основе ML недостаточно создать пример, отвечающий доменным ограничениям (требованиям, условиям), то есть ограничениям, способствующим сохранению функциональности примера. Атаку-

ющий должен создавать состязательные примеры, сохраняющие функциональность как сетевого протокола, так и оригинальной компьютерной атаки, которую он пытается замаскировать в этом сетевом трафике при помощи состязательной атаки [2], то есть эффективные состязательные примеры. Таким образом, важными задачами при генерации состязательных примеров в области обнаружения вторжений в сетевом трафике являются проверки физической реализуемости найденного возмущения и сохранения действия атаки, причём последняя проверка в исследованиях обычно не проводится.

Оценка устойчивости ML-модели к состязательным атакам с использованием не отвечающих доменным требованиям состязательных примеров может не отражать реальную устойчивость модели [3]. В свою очередь, такая некорректная оценка может привести к завышенному выделению ресурсов на решение задачи защиты [4]. В случае предметной области обработки сетевого трафика также необходимо учесть траты ресурсов на проверку сгенерированных примеров: необходимо не только сгенерировать сетевой трафик, соответствующий найденным состязательным возмущениям, но и проверить его в реальной сети с сопутствующими временными затратами. Обработка нереалистичных состязательных примеров приведёт к лишним тратам времени и других ресурсов. Также открытым остаётся вопрос об эффективности применения реалистичных состязательных примеров в методах защиты ML-моделей от состязательных атак, например, в таком популярном методе защиты, как состязательное обучение (adversarial training), то есть обучение модели с добавлением корректно размеченных состязательных примеров.

Перечисленные проблемы обуславливают актуальность изучения состязательных атак с ограничениями против моделей машинного обучения для обнаружения компьютерных атак в сетевом трафике.

2. Состязательные атаки и доменные ограничения в задаче обнаружения сетевых вторжений

Большинство IDS проводят классификацию сетевого трафика на основе табличных данных, представляющих собой признаки, извлечённые из сетевых пакетов и/или потоков при помощи специальных инструментов [5]. Эти признаки имеют различные множества допустимых значений и типы, например, бинарные (флаги «SYN», «ACK» и др.), категориальные (протоколы «UDP», «TCP» и др.), вещественные (скорость потока данных «Flow Bytes/s» и др.) и целочисленные значения (максимальная длина пакета «Max Packet Length» и др.). Извлечённые признаки часто представляют собой некоторые статистические данные и могут зависеть друг от друга: например, при изменении признака «Flow Bytes/s» может меняться признак «Average Packet Size» (средняя длина поля данных пакета TCP/IP).

В настоящий момент отсутствует общепринятая классификация доменных ограничений. В некоторых работах [6, 7] предложено разделить доменные ограничения на синтаксические (диапазоны значений, тип признака и т.п.) и семантические, определяющие связи между признаками. Согласно [6, 8], семантические ограничения делят признаки на неизменяемые (IP-адрес «IP address», номер протокола «Protocol Number» и т.п.), изменяемые (длительность сессии «Flow Duration» и т.п.) и зависимые от них признаки («Average Packet Size» и т.п.). Недостаточно установить факт зависимости между признаками, так как зависимости могут быть прямыми, обратными, нелинейными [9]. Например, увеличение скорости передачи («Flow Bytes/s») может приводить к уменьшению длительности сессии («Flow Duration»), и если найденный состязательной атакой пример увеличивает оба признака, то такой пример не является осуществимым для большинства реальных случаев. При этом полное определение ограничений, особенно отношений между признаками, является нетривиальной задачей.

В целом, можно выделить два подхода к обеспечению соблюдения доменных ограничений: учёт ограничений в процессе генерации состязательных примеров или после генерации. Первый подход можно реализовать различными способами. Например, в статье [10] предлагают атаки со штрафами за нарушение доменных ограничений, используемыми в функции потерь в градиентном методе C-PGD или в целевых функциях в генетическом многоцелевом алгоритме MoEvA2. Другим примером является ограничение признакового пространства, когда состяза-

тельные возмущения вносятся только в определённые признаки; так, в фреймворке Tiki-Taka [11] используются только те признаки, которые не могут изменить семантику потока трафика. Реализовать ограничения на диапазоны значений можно при помощи «обрезки» значений (clipping) признаков на их верхней и нижней границах [9, 10, 12]. Второй подход подразумевает удаление не отвечающих доменным ограничениям сгенерированных примеров, например, при помощи фильтрации по правилам [9] или проверки функциональности пакета сетевого трафика, сгенерированного из полученного состязательного примера [12].

3. Особенности реализации состязательных атак с ограничениями

Для задачи поиска состязательных примеров с учётом доменных ограничений важным аспектом являются не только обрабатываемые данные, но и используемые модели. В силу того, что IDS работают преимущественно с неоднородными табличными данными, в них могут использоваться классические ML-алгоритмы, в частности, модели типа «случайный лес» (Random Forest, RF), демонстрирующие качество классификации не хуже нейронных сетей. Однако многие состязательные атаки являются градиентными методами, ориентированными на применение против нейронных сетей, например, метод IFGSM (BIM) [13] или его модификации CIFGSM [12] и C-IFGSM [9], учитывающие доменные ограничения. Такие атаки являются атаками «белого ящика», поскольку предполагают наличие знаний об атакуемой модели, достаточных для построения градиента функции потерь, и неприменимы напрямую к ансамблям решающих деревьев. Для атаки подобных моделей требуются специализированные методы или атаки «чёрного ящика», не предполагающие наличия знаний об атакуемой модели.

В состязательных атаках «чёрного ящика» может использоваться свойство переносимости состязательных примеров (adversarial sample transferability) – способность состязательных примеров сохранять свою эффективность для других моделей, помимо тех, для которых они были созданы [14]. При этом данное свойство не гарантирует того, что найденный пример приведёт к ошибочному решению целевой атакуемой модели. Например, в работе [15] показано, что успешность переноса примеров варьируется от 20% до 100% в некоторых случаях в зависимости от метода атаки, архитектур атакуемой модели и модели, для которой были сгенерированы примеры (т.н. модели замены или substitute model), достигая максимума при условии совпадения архитектур и использования специальной стратегии ранжирования состязательных примеров. При этом использование модели замены для создания отвечающих доменным ограничениям состязательных примеров может иметь дополнительные особенности. Так, в исследовании [4] многослойный перцептрон (MLP) в качестве модели замены позволил сгенерировать больше состязательных примеров, отвечающих ограничениям предметной области, по сравнению с моделями замены более сложной архитектуры, такими как свёрточная нейронная сеть (CNN) и CNN-BiLSTM. Более сложные границы принятия решений в сложных моделях в целом более устойчивы к состязательным атакам с учётом ограничений, однако некоторые методы атак отклоняются от этой тенденции [4].

В зависимости от пространства осуществления, атаки с ограничениями можно разделить на атаки в пространстве признаков (feature space), когда состязательные возмущения вносятся в извлечённые из обрабатываемых объектов признаки, и атаки в пространстве задачи (problem space) [2], когда состязательные возмущения вносятся в сами обрабатываемые объекты. Для задачи анализа сетевого трафика пространство задачи также называется пространством трафика (traffic space) [12], поскольку состязательные возмущения вносятся в пакеты сетевого трафика. Существование атак, в которых для генерации состязательных примеров используются как пространство признаков, так и пространство задачи, позволяет также выделить атаки, которые можно назвать гибридными. Например, в гибридных атаках CIFGSM [12] и FENCE [16] пакеты сетевого трафика изменяются в соответствии со значениями состязательных возмущений, полученных в пространстве признаков. В атаках же пространства трафика Deep PackGen [17] и TANTRA [18] изменяется сам сетевой трафик, при помощи глубокого обучения с подкреплением (DRL) в первом случае и нейронной сети, изменяющей межпакетные задержки трафика, во втором. Подобные атаки в пространстве трафика используют только допустимые изменения пакетов, с учётом или без учёта полезной нагрузки.

Важно отметить, что традиционные техники обратного отображения неэффективны в случае отображения извлечённых признаков в пакеты сетевого трафика, так как отображение не является ни обратимым, ни дифференцируемым [2]. Поэтому получение функционального объекта предметной области, которому соответствует состязательный пример, сгенерированный атакой в пространстве признаков, в общем случае является нетривиальной задачей.

Необходимо отметить, что в различных работах предприняты попытки формализовать задачу состязательной атаки с ограничениями, однако единая методология в настоящий момент не выработана, как и единые подходы к оценке успешности таких атак.

4. Заключение

Существующие методы генерации состязательных примеров в большинстве своём были предложены для предметной области обработки изображений и не учитывают ограничения предметной области обработки трафика. Это приводит к генерации неэффективных состязательных примеров, которые невозможно реализовать на практике. Таким образом, изучение состязательных атак с ограничениями против моделей машинного обучения для обнаружения компьютерных атак в сетевом трафике является актуальной задачей.

В качестве одного из основных направлений дальнейших исследований следует выделить развитие методов состязательных атак с ограничениями с целью повышения доли эффективных состязательных примеров среди всех сгенерированных при сохранении или уменьшении среднего времени генерации одного эффективного примера.

Эффективность применения состязательных примеров, отвечающих доменным требованиям, в различных методах защиты ML-моделей от состязательных атак также требует отдельного изучения. Для повышения устойчивости к состязательным атакам с ограничениями целесообразно сосредоточить усилия на разработке новых специфических методов защиты от таких атак или адаптации существующих механизмов защиты.

Литература

1. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I.J., & Fergus, R. (2013). Intriguing properties of neural networks. CoRR, abs/1312.6199.
2. Wang N., Chen Y., Xiao Y., Hu Y., Lou W. and Hou Y. T. MANDA: On Adversarial Example Detection for Network Intrusion Detection System. In IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 2, pp. 1139-1153, 1 March-April 2023. DOI: 10.1109/TDSC.2022.3148990.
3. Hamid Bostani, Zhengyu Zhao, Zhuoran Liu, and Veelasha Moonsamy. 2025. Level Up with ML Vulnerability Identification: Leveraging Domain Constraints in Feature Space for Robust Android Malware Detection. ACM Trans. Priv. Secur. 28, 2, Article 17 (May 2025), 32 pages. https://doi.org/10.1145/3711899.
4. Grini, Anass & Taheri, Oumaima & Khamlichi, Btissam & Fallah-Seghrouchni, Amal. (2025). Constrained Network Adversarial Attacks: Validity, Robustness, and Transferability. 10.48550/arXiv.2505.01328.
5. Гетьман А.И., Горюнов М.Н., Мацкевич А.Г., Рыболовлев Д.А. Методика сбора обучающего набора данных для модели обнаружения компьютерных атак. Труды Института системного программирования РАН. 2021;33(5):83-104 https://doi.org/10.15514/ISPRAS-2021-33(5)-5 / Getman A.I., Goryunov M.N., Matskevich A.G., Rybolovlev D.A. A Comparison of a Machine Learning-Based Intrusion Detection System and Signature-Based Systems. Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS). 2022;34(5):111-126. (In Russ.) https://doi.org/10.15514/ISPRAS-2022-34(5)-7.
6. Adeke J.M., Liu G., Zhao J., Wu N., Bashir H.M. Securing Network Traffic Classification Models against Adversarial Examples Using Derived Variables. Future Internet, vol. 15, no. 12: 405, 2023, p. 21. DOI: 10.3390/fi15120405.
7. Debicha I., Cochez B., Kenaza T., Debatty T., Dricot J.M., Mees W. Adv-Bot: Realistic Adversarial Botnet Attacks against Network Intrusion Detection Systems. Computers & Security, vol. 129, 103176, 2023. DOI: 10.1016/j.cose.2023.103176.

8. Teuffenbach M., Piatkowska E., Smith P. Subverting network intrusion detection: Crafting adversarial examples accounting for domain-specific constraints. In Proceedings of the Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, 25–28 August 2020; Proceedings 4; Springer: Berlin/Heidelberg, Germany, 2020; pp. 301–320.
9. Tian Y., Wang Y., Tong E., Niu W., Chang L., Chen Q.A., Li G., & Liu J. Exploring Data Correlation between Feature Pairs for Generating Constraint-based Adversarial Examples. 2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS), 2020, pp. 430-437. DOI: 10.1109/ICPADS51040.2020.00064.
10. Simonetto T., Dyrmishi S., Ghamizi S., Cordy M., & Traon Y. L. A Unified Framework for Adversarial Attack and Defense in Constrained Feature Space. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, {IJCAI-22}, 2022, pp. 1313-1319. DOI: 10.24963/ijcai.2022/183.
11. C. Zhang, X. Costa-Pérez and P. Patras, "Adversarial Attacks Against Deep Learning-Based Network Intrusion Detection Systems and Defense Mechanisms," in IEEE/ACM Transactions on Networking, vol. 30, no. 3, pp. 1294-1311, June 2022, doi: 10.1109/TNET.2021.3137084.
12. Maria Simidžioski. Adversarial Traffic Modifications for the Network Intrusion Detection Domain. A Practical Adversarial Network Traffic Crafting Approach. Master thesis. Delft University of Technology, 2021, 87 p.
13. Kurakin A., Goodfellow I. J., Bengio S. Adversarial examples in the physical world. In Artificial intelligence safety and security, 2018, pp. 99-112.
14. Papernot N., McDaniel P., & Goodfellow I. Transferability in machine learning: from phenomena to black-box attacks using adversarial samples. 2016, 13 p. ArXiv, abs/1605.07277.
15. Moshe Levy, Guy Amit, Yuval Elovici, and Yisroel Mirsky. 2024. Ranking the Transferability of Adversarial Examples. ACM Trans. Intell. Syst. Technol. 15, 5, Article 100 (October 2024), 21 pages. https://doi.org/10.1145/3670409.
16. Alesia Chernikova and Alina Oprea. 2022. FENCE: Feasible Evasion Attacks on Neural Networks in Constrained Environments. ACM Trans. Priv. Secur. 25, 4, Article 34 (November 2022), 34 pages. https://doi.org/10.1145/3544746.
17. Soumyadeep Hore, Jalal Ghadermazi, Diwas Paudel, Ankit Shah, Tapas Das, and Nathaniel Bastian. 2025. Deep PackGen: A Deep Reinforcement Learning Framework for Adversarial Network Packet Generation. ACM Trans. Priv. Secur. 28, 2, Article 15 (May 2025), 33 pages. https://doi.org/10.1145/3712307.
18. Y. Sharon, D. Berend, Y. Liu, A. Shabtai and Y. Elovici, "TANTRA: Timing-Based Adversarial Network Traffic Reshaping Attack," in IEEE Transactions on Information Forensics and Security, vol. 17, pp. 3225-3237, 2022, doi: 10.1109/TIFS.2022.3201377.

ИНТЕГРАЦИЯ ДОВЕРЕННОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ИНФРАСТРУКТУРУ ЗДРАВООХРАНЕНИЯ[1*]

В.П. Носов[1,2], А.А. Шестова[1], В.Е. Турлапов[1]

[1]Исследовательский центр в области искусственного интеллекта института ИТММ Нижегородского госуниверситета им.Н.И.Лобачевского
[2]Приволжский исследовательский медицинский университет (ПИМУ)

Предметом исследования является развитие применения искусственного интеллекта (ИИ) в реализации применения клинических рекомендаций (КР), несущих опыт доказательной медицины в клиническую практику медицины и являющихся основным инструментом повышения качества медицинской помощи. Показаны нерешенные проблемы применения ИИ: a) интеграция ИИ в инфраструктуру здравоохранения; b) обеспечение непрерывного обучения системы здравоохранения; c)сохранение низкой применяемости КР. Предложено новое определение доверенного ИИ и решение проблем на основе реализации КР не как традиционно справочного документа, а как главного управляющего дерева решений ИИ, обеспечивающего доверенность и 100-процентную применяемость КР.

Ключевые слова: доверенный искуственный интеллект, клинические рекомендации, здравоохранение, дерево решений, обновление дерева решений, сквозная цифровая технология, СЦТ, СЦТ искусственного интеллекта

1. Введение

Клинические рекомендации (КР) – вершина аналитической деятельности в медицине, ее продукт, основанный на концепции доказательной медицины и предназначенный для использования во врачебной практике и образовании. КР (англ. guideline, guidelines, clinical guidelines, clinical practice guidelines) охватывают вопросы профилактики, скрининга, диагностики, лечения, контроля состояния и реабилитации определенных заболеваний, являются основным инструментом повышения качества медицинской помощи [1]. Вне зависимости от нозологических форм, КР должны быть прозрачными (объяснимыми), надежными и периодически обновляться (пересматриваться) для развития и сохранения актуальности.

Практически с самого появления КР медицина поставила перед собой задачу унификации, стандартизации и обновления их структуры и содержания независимо от нозологии. Итогом стало создание коллаборации AGREE для решения этой проблемы и выпуск в 2001 году первой версии одноименных рекомендаций. В 2009 вышла вторая версия AGREE II, отражающая серьезную роль искусственного интеллекта (ИИ) в медицине, которая была обновлена в 2013 [2] из-за нарастания роли ИИ. Эта оценка методологического качества КР с использованием стандартного вопросника *«Appraisal of Guidelines for Research and Evaluation II» (AGREE II*), состоящего из 23 вопросов в 6 аспектах/доменах (процесс разработки КР; качество представления рекомендаций; полнота описания; клиническая валидность и т.д.) [2, 3] или его сокращенной версии AGREE Global Rating Scale (AGREE GRS) [4] стала правилом при всех типах разработки КР: *разработке КР de novo*; *принятии КР* (простейший); *адаптации КР*.

В 2002 году, с целью сократить дублирование усилий, также была создана Международная сеть КР (GIN), которая сегодня является реальным связующим звеном и насчитывает 119 организаций и 139 индивидуальных членов из 56 стран (по состоянию на 1 апреля 2025 года) [5]. Она предоставляет доступ к Международной библиотеке КР и другим ресурсам, а также шаблон для создания КР. А к 2010 году международной группой экспертов ADAPTE был разрабо-

[1*] Работа выполнена при поддержке Министерства экономического развития Российской Федерации (соглашение № 139-15-2025-004 от 17 апреля 2025 г., ИГК 000000Ц313925Р3Х0002).

тан Порядок адаптации КР. Его детальная процедура также свободно доступна на сайте GIN [6].

Благодаря ценности содержания КР, почти одновременно с их появлением делаются попытки сделать КР компьютеро-читаемыми путем повышения формализации КР и привлечения специальных алгоритмических языков. Первым на этом пути появляется подход GLIF (GuideLine Interchange Format) [7], который был вскоре сменен подходом (computer-interpretable guidelines) [8], который учитывал возможности ИИ для медицины и предлагал извлекать знания из КР. Этот подход несколько обогнал время и привлекал внимание со стороны ИИ и через почти десятилетие [9,10], хотя долгое время давал только частные решения задачи из-за недостаточности уровня развития как техники, так и ИИ, особенно в области больших лингвистических моделей (LLM).

В связи с тем, что в здравоохранении не оказалось на тот момент общепризнанных рекомендаций или стандартов для отчетности по решениям, принимаемым ИИ, в 2014 году для решения данной проблемы была создана международная рабочая группа RIGHT [11] и в 2017 году такое решение было опубликовано [12]. Оно предоставляет контрольный список из 22 пунктов собранным в 7 групп, дающий исчерпывающее объяснение и доказательное обоснование того, что должно быть, как в клиническом руководстве, так и в отчете по его исполнению, а также примеры качественного изложения информации по каждому пункту контрольного списка.

Таким образом в основном сложилось представление о содержании КР и способах, порядке и инструментах подготовки и обновления КР для мировой практики. Примером является российская публикация 2019 года, представившая детальное руководство, например, о трех способах подготовки КР с использованием инструментов ADAPTE, AGREE GRS и других [13]. Публикация стимулирована тем, что в 2012 году Минздрав РФ начал разработку клинических рекомендаций. В 2019 году, вступил в силу Федерального закона (ФЗ) № 489-ФЗ, закрепивший их обязательность с поэтапным переходом на них для всех медицинских организаций. Согласно Постановлению Правительства № 1968, с 1 января 2022 года начался этот поэтапный переход, завершившийся к 1 января 2025.

В итоге к 2022 году уровень развития ИИ и его продвижения в медицину значительно изменились. Выросло понимание важности и приоритет применения в медицине объяснимого ИИ (eXplainable Artificial Intelligence, XAI). Переход на очередной уровень развития обозначил обзор-прогноз национальной академии медицины США (NAM) [14] по состоянию, перспективам и опасностям применения ИИ в здравоохранении. В разделах 1-5 построен краткий обзор основ ИИ и его современных возможностей в медицине. Начиная с раздела 6, подробно рассмотрено внедрение ИИ в реальных клинических условиях, даны подборки публикаций по разделам. Основное внимание в перспективах уделено организационным проблемам и их решениям на основе ИИ, в том числе:

- Интеграция ИИ *в инфраструктуру* предоставления медицинских услуг, доступность получаемых данных, как для оценки эффективности самих приложений ИИ, так и для понимания как функционируют системы здравоохранения и каково текущее состояние пациентов.
- Обеспечение широкого лидерства, расширяя приверженность целям непрерывно обучающейся системы здравоохранения и содействуя широкому вовлечению ключевых заинтересованных сторон (врачей, ИТ-специалистов, руководства) и их ответственности в управлении системами ИИ в здравоохранении.
- Обеспечение совместимости между национальными и местными стандартами данных в их фиксированном и динамическом состоянии.
- Определение того что должно квалифицироваться как медицинское устройство, а что нет.

Другим значимым результатом явилась публикация [15], также 2022 года в Nature Medicine. Авторы указывали на то, что всё больше систем поддержки принятия клинических решений на основе ИИ демонстрируют многообещающие результаты в доклинических исследованиях, но лишь немногие из них демонстрируют затем реальную пользу для лечения пациентов. Проблема связана с необходимостью ранней клинической оценки фактической клинической эффективности системы ИИ (в том числе, оценки безопасности, человеческого фактора,

готовности крупномасштабных исследований). Предложено руководство по составлению отчётности о ранней клинической оценке систем ИИ (DECIDE-AI). Проведена экспертная оценка руководства экспертами из 20 заинтересованных сторон. В первом раунде оценки приняли участие 123 эксперта, во втором – 138, в совещании по достижению консенсуса – 16, в качественной оценке – 16. Руководство DECIDE-AI доработано по результатам оценки и включает 17 пунктов отчетности, посвященных ИИ (состоящих из 28 подпунктов), и десять общих пунктов отчетности, для каждого из которых предусмотрен раздел «Разработка и оценка».

В 2024 году появилась публикация [16], дополняющая руководство TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) от 2015 года, реальной включенностью ИИ (+AI) в принятие решения по прогнозу. Цель TRIPOD+AI – содействовать полной, точной и прозрачной отчетности об исследованиях, разрабатывающих модель прогнозирования или оценивающих ее эффективность.

Интересно выглядят две взаимно дополняющие отечественные публикации. Публикация [17] 2025 года посвящена интеграция КР в инструменты цифрового здравоохранения, используя метод *онтологического моделирования*, который решает задачу преобразования текстовых КР в машинопонимаемые клинические рекомендации (МКР). Ключевая идея состоит в двухуровневом представлении КР: *уровень врачей* – это иерархически структурированный текст, использующий «контейнеры данных», которые четко описывают наблюдения, вмешательства, условия их применения; *уровень интеллектуальных систем* – это формализованный граф знаний, с привязкой «контейнеров», цифровой медицинской карты пациента и правил обработки.

Публикация [18] 2024 года посвящена вопросам организации облачного сервиса дифференциальной диагностики и назначения персонифицированного лечения заболеваний сердца. Дифференциальная диагностика построена на использовании МКР и ограничена воспалительными заболеваниями. Сервис поддержки принятия диагностических решений со свойствами объяснимого искусственного интеллекта в кардиологии реализован на медицинском портале облачной платформы, которая позволяет масштабировать предложенное решение и обеспечивает доступ практикующих врачей со свободной регистрацией.

В 2025 публикован систематический обзор [19], который сравнил качество КР, руководств и стандартов, подготовленных разными коллективами. Из почти 5 тысяч публикаций было отобрано 11 наиболее представительных. Оценка выполнена по 6 параметрам (сфера применения и цель, участие заинтересованных сторон, тщательность разработки, ясность изложения, применимость, редакционная независимость) инструментом AGREE II. Обнаружено, что, хотя КР в целом соответствуют цели, но существуют значительные различия как в применимости и строгости разработки, так и с точки зрения вовлечения сторон медицины и ИИ. По качеству исполнения КР и прогностической модели ИИ хорошо зарекомендовали себя следующие руководства-требования к отчету КР и модели: TRIPOD+AI, DECIDE-AI, SPIRIT-AI (Протоколы клинических исследований, оценивающих вмешательство с участием ИИ) и CONSORT-AI (Отчёты о клинических исследованиях, оценивающих вмешательство с участием ИИ). Тем не менее, применяемость остается одной из основных проблем для КР.

В итоге обзора, отношения КР и ИИ, и их перспективы, в настоящее время представляются следующим образом:

1) Клинические рекомендации (КР) признаны во врачебных сообществах мира как основной источник, носитель и интегратор опыта доказательной медицины; они доступны врачу во многих медицинских информационных системах (МИС) для поддержки принятия решений.
2) Использование искусственного интеллекта (ИИ) почти повсеместно стало передовой клинической практикой и получило поддержку в специальных клинических рекомендациях многих групп экспертов медицины, поддержанных экспертами в области ИИ.
3) Объяснимость и доверенность действий ИИ принимается практикой медицины как обязательное свойство ИИ, применяемого медициной.
4) Применяемость остается одной из основных проблем для КР [14].
5) Будущее КР видится экспертам [14] как: a) интеграция ИИ в инфраструктуру предоставления медицинских услуг; b) обеспечение непрерывно обучения системы здравоохранения; c) доступность и обобщение получаемых результатов работы системы всем уровням здравоохранения, включая как оценку здоровья пациентов на уровне от клиники до государства, так и оценку эффективности самих приложений ИИ в здравоохранении.

2. Системное решение интеграции доверенного ИИ в инфраструктуру здравоохранения

В качестве системного решения интеграции доверенного ИИ в инфраструктуру здравоохранения предлагается использование КР как управляющего МИС дерева решений *доверенного ИИ* в организации клинической практики по каждой социально значимой нозологии.

Для полноценного понимания доверенности такого решения мы предлагаем новое определение Доверенного ИИ, обобщающее его существующие стандарты, например [20].

Определение: Доверенный ИИ, это искусственный интеллект, обладающий тремя свойствами, выраженными в категориях: 1) подконтрольность человеку; 2) безопасность; 3) объяснимость.

Специфика ИИ в медицине состоит в том, что для реальной доверенности отношений здесь необходима доверенность для всех трех категорий субъектов, участвующих в процессе. Это три субъекта: пациент, врач и ИИ. Каждый из них вступает сегодня в отношения с двумя другими, образуя треугольник «врач-пациент-ИИ». На наш взгляд наличие в тройке ИИ, как субъекта отношений, не должно встретить отторжения как факт реальной практики в современном обществе.

Пример раскрытия категорий в области медицины и здравоохранения:

Человек: персона, гражданин, пациент, врач; семья, коллектив; население поселения, города, государства; органы управления; издаваемые структурными единицами общества регламенты, приказы, постановления, законы государства; группы экспертов, привлекаемые для решения проблем диагностики, здоровья и управления; этика, культура.

Врач: персона, бригада скорой помощи (БСП), отделение, клиника, функциональная единица здравоохранения, Министерство здравоохранения; приказы, рекомендации и другие регламенты системы здравоохранения; эксперт или группа экспертов.

Пациент: персона; семья; коллектив; население территории: района, города, области; граждане государства.

ИИ (в медицине): система моделей искусственного интеллекта, моделирующая в медицине категорию Врач как доверенную.

Безопасность: защита от неправильной диагностики, неправильного лечения, неправильной реабилитации; ролевой доступ; надежность хранения данных; защита от внешнего вредоносного воздействия.

Из данного выше определения и раскрытия категорий очевидно, что обязательная для доверенного ИИ подконтрольность человеку, означает подконтрольность не одному человеку, а всем объектам, раскрывающим категорию Человек. При этом каждому объекту обеспечивается ролевой доступ, соответствующий его роли, и безопасность хранения данных. Например, Врач и ИИ действуют как автономные объекты, каждый из которых продуцирует свое документированное решение, доступное другому только в режиме чтения. Первым генерирует решение ИИ, на основе клинических рекомендаций, после чего врач знакомится с документом ИИ и может принять его или внести свои правки или замечания. А ИИ затем имеет возможность прочитать документ врача, зарегистрировать его для статистики оценки результата и, возможно, для дообучения.

Из определения доверенного ИИ также следует, что именно КР, в виде дерева объяснимых и надежных решений в клинической практике, наиболее эффективны в качестве управляющего дерева все решения которого уже одобрены профильными экспертами нозологии и приказом Минздрава. В таком случае сразу решается и обеспечение доверенности (подконтрольности ИИ человеку) и до сих пор нерешенная проблема 100%-ной применяемости клинических рекомендаций в клинической практике. На плечи ИИ при этом ложится и контроль за исполнением КР персоналом клиники, а также за полнотой сведения результатов всех клинических исследований по конкретному пациенту с учетом его пола, возраста, роста веса, сопутствующих заболеваний и вредных привычек. Именно поэтому в категории Безопасность из 6-и названных составляющих, первые 4 мы считаем обеспеченными доверенностью ИИ. Надежность хранения решаем на аппаратном уровне (логирование и мониторинг безопасности: резервное копирова-

ние, избыточность хранения, защита от сбоев). А защиту от внешнего вредоносного воздействия возлагаем на сервис безопасности инфраструктурного слоя сервера системы.

Алгоритмизация и программная реализация действующих Клинических рекомендаций для каждой социально значимой нозологии, как модель принятия и исполнения решений в системе здравоохранения, способны коренным образом изменить производительность и качество медицинских услуг населению прежде всего в области диагностики и лечения.

3. Алгоритмизация клинических рекомендаций для этапа диагностики на примере обобщенной нозологии Ишемическая болезнь сердца / Острый коронарный синдром

Алгоритм реализации клинических рекомендаций должен обладать свойствами доверенности и, таким образом, включать в себя доверенные действия и отношения каждого субъекта треугольника «врач-пациент-ИИ». Пример такого алгоритма показан на рис.1.

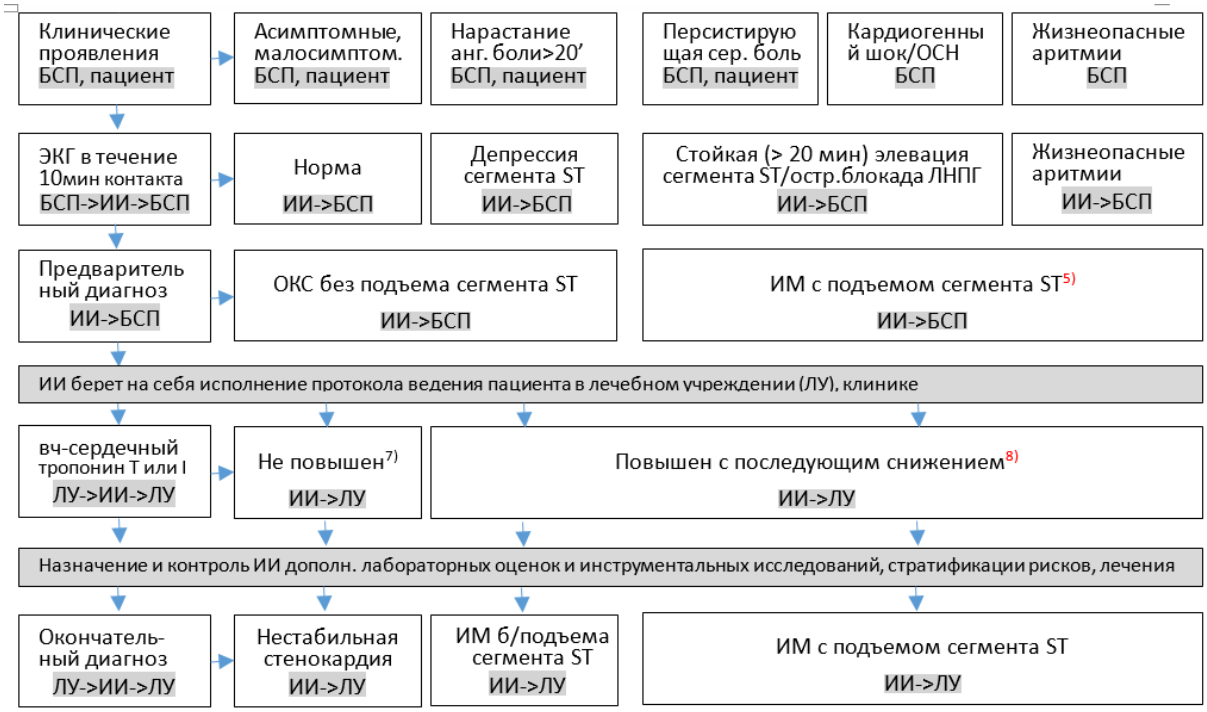


Рис. 1. Схема алгоритма реализации клинических рекомендаций нозологии Ишемическая болезнь сердца/острый коронарный синдром (ИБС-ОКС) для тройки «врач-пациент-ИИ» на примере этапа диагностики

В роли Врача здесь выступают бригада скорой помощи (БСП), приемный покой лечебного учреждения/клиники (ЛУ); персонал отделения ЛУ, принимающего пациента из приемного покоя ЛУ. Блоки схемы в левой колонке показывают развитие событий и участников событий (на сером фоне). Блоки справа в каждой горизонтали показывают возможные исходы событий. Пациент показан только на первой горизонтали, чтобы не перегружать схему: далее пациент также является качественно информируемой стороной взаимодействия и источником субъективной оценки своего состояния. ИИ задействован с момента регистрации ЭКГ в первые 10 мин контакта с БСП. С момента передачи пациента в клинику, ИИ берет на себя контроль за исполнением протокола ведения пациента в ЛУ (контроль высокочувствительного сердечного тропонина), а затем окончательного диагноза, лечения, дополнительных лабораторных оценок и инструментальных исследований. Каждый блок предусматривает свой комплекс мероприятий лечения, передаваемый на исполнение медперсоналу (показаны сносками как срочные-красные и регулярные-черные).

Алгоритм реализации этапа Лечение выглядит подобным образом как дерево решений и полностью поддерживается искусственным интеллектом, реализующим необходимый текущий мониторинг состояния пациента и публикующим ежедневный отчет о его состоянии.

4. Непрерывное обучение системы здравоохранения и сквозная цифровая технология для нее

Однако, системное решение интеграции доверенного ИИ в инфраструктуру здравоохранения окажется неполным без решения проблемы *непрерывного обучения системы здравоохранения*. Это одновременно означает и готовность к непрерывному обновлению КР по мере накопления статистики применения текущих КР и успешных/неуспешных прецедентов сложных клинических ситуаций. Это решение требует широкой межклинической интеграции результатов диагностики, лечения и реабилитации, а также поддержки процедуры обновления КР и замкнутости всего жизненного цикла КР со стороны доверенного ИИ. В этом случае наша МИС может получить статус сквозной цифровой технологии (СЦТ) в области здравоохранения, тиражируемой на произвольные нозологии.

Формы и содержание современной процедуры обновления КР достаточно полно описаны в публикации [13]. Порядок и сроки обновления регламентированы сегодня приказом МЗ РФ от 28.02.2019 года №104н с изменениями и дополнениями от 28.09.2023 года № 507н, введены в действие порядок и сроки одобрения и утверждения КР научно-практическим советом МЗ РФ и медицинскими профессиональными некоммерческими организациями.

Проект КР разрабатывается одной или несколькими уполномоченными медицинскими профессиональными некоммерческими ассоциациями и организациями, опираясь на доказательную медицину и учитывая критерии валидности, репрезентативности и применимости в клинике, с использованием федерального *реестра нормативно-справочной информации* из *единой государственной информационной системы здравоохранения* (ЕГИСЗ), и направляется в МЗ РФ.

Ответственным за рассмотрение проекта КР научно-практическим советом является профильный департамент Министерства здравоохранения Российской Федерации, который в случае положительного решения направляет КР в экспертную организацию (федеральное государственное бюджетное учреждение, находящееся в ведении МЗ РФ), для подготовки экспертного заключения по КР.

Экспертное заключение содержит научно обоснованную информацию об эпидемиологии, этиологии, патогенезе, современной классификации, клинической картине, лабораторно-инструментальной диагностике и терапии рассматриваемой нозологической формы. При этом медицинские услуги должны соответствовать номенклатуре медицинских услуг, лекарственные препараты для медицинского применения, медицинские изделия должны иметь государственную регистрацию, *названия фармакотерапевтических групп лекарственных препаратов - соответствовать анатомо-терапевтическо-химической классификации, рекомендованной ВОЗ,* а *показания к применению и противопоказания, способы применения и дозы лекарственных препаратов, должны соответствовать инструкции по применению*.

В случае реализации *непрерывного обучения и обновления системы здравоохранения* непрерывное (ежемесячное, еженедельное, ежесуточное, по критериям выработанным сообществом и утвержденным МЗ РФ) первичное обновление версии клинических рекомендаций должно быть переложено на ИИ. Версия КР готовится ИИ одновременно с проектом Экспертного заключения по ней. Поддержка актуальности всех справочных данных из ЕГИСЗ, необходимых для одобрения проекта КР и сопровождения КР после одобрения, может быть обеспечена подключением, к узлам и листам дерева решений КР, системы RAG (Retrieval Augmented Generation), обновляющей актуальные точные данные на каждый текущий момент времени. Текущее первичное обновление КР может быть доступно организациям для использования под ответственность организации. Утверждение проекта Экспертного заключения может осуществляться по мере необходимости также по утвержденным критериям.

Заключение

В статье дан анализ ключевых моментов развития Клинических рекомендации (КР) в практике медицины и здравоохранения, как основного документа – интегратора опыта доказательной медицины, предназначенного для использования во врачебной практике и образовании, охватывающего вопросы профилактики, скрининга, диагностики, лечения, контроля состояния и реабилитации заболеваний, основного инструмента повышения качества медицинской помощи. Дан анализ развития роли искусственного интеллекта как определяющей для развития применения клинических рекомендаций. Показаны проблемы развития ИИ, остающиеся нерешенными, а именно: a) интеграция ИИ *в инфраструктуру* предоставления медицинских услуг; b) обеспечение *непрерывного обучения системы здравоохранения*; c) сохранение низкой применяемости КР в здравоохранении. Предложено новое определение доверенного ИИ и решение указанных проблем на основе реализации КР не как традиционно справочного документа системы, а как главного управляющего дерева решений ИИ, обеспечивающего доверенность и 100-процентную применяемость КР как управляющей модели ИИ. В плане дальнейшей работы алгоритмизация непрерывного обучения системы здравоохранения.

Литература

1. Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ. 1999. https://doi.org/10.1136/bmj.318.7182.527.
2. The AGREE Collaboration. September 2001. – 26 pp [Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. Okt 2013. – 59 p. [https://www.agreetrust.org/wp-content/uploads/2013/10/AGREE-II-Users-Manual-and-23-item-instrument_2009_UPDATE_2013.pdf].
3. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in healthcare. CMAJ. 2010; 182(18):E839-842. https://doi.org/10.1503/cmaj.090449.
4. Brouwers MC, Kho ME, Browman GP, et al. The global rating scale complements the AGREE II in advancing the quality of practice guidelines. J Clin Epidemiol. 2012. 65(5):526-534. https://doi.org/10.1016/j.jclinepi.2011.10.008.
5. Guidelines International Network (GIN). Library & Resources. International Guideline Library (2002). https://g-i-n.net/international-guidelines-library [cited 2025 Aug 30].
6. Guidelines International Network (GIN). Library & Resources. Guideline adaptation (2010): A Resource Toolkit. Version 2.0. 95 pp. https://g-i-n.net/wp-content/uploads/2021/05/ADAPTE-Resource-toolkit-V2.1-March-2010-updated-disclaimer.pdf [cited 2025, Aug 30].
7. Ohno-Machado L, Gennari JH, Murphy SN, et al. The GuideLine Interchange Format: a model for representing guidelines. J Am Med Inform Assoc (JAMIA). 1998 Jul-Aug; 5(4):357-72. https://doi.org/10.1136/jamia.1998.0050357. PMID: 9670133; PMCID: PMC61313.
8. De Clercq P, Blom J, Korsten H, Hasman A (2004) Approaches for creating computer-interpretable guidelines that facilitate decision support. Artif Intell Med 31(1):1–27 https://doi.org/10.1016/j.artmed.2004.02.003.
9. Oliveira T., Novais P.J., Neves J. Guideline Formalization and Knowledge Representation for Clinical Decision Support. Advances in Distributed Computing and Artificial Intelligence Journal (ADCAIJ). July 2012. 1(2):1-12.
10. Oliveira, T., Novais, P. & Neves, J. Development and implementation of clinical guidelines: An artificial intelligence perspective. Artif Intell Rev 42, 999–1027 (2013). https://doi.org/10.1007/s10462-013-9402-2.
11. RIGHT Statement (the RIGHT Working Group, 2014). http://www.right-statement.org/ [cited 2025, Aug 30].
12. Chen Y., Yang K., Marušié A. et al.; for the RIGHT (Reporting Items for Practice Guidelines in Healthcare) Working Group. A Reporting Tool for Practice Guidelines in Health Care: The RIGHT Statement. Ann Intern Med. 2017. 166: 128–132. https://doi.org/10.7326/M16-1565.

13. Реброва О.Ю. О трех способах подготовки клинических рекомендаций. Инструменты ADAPTE и AGREE GRS. Проблемы Эндокринологии. 2019;65(3):197-203. https://doi.org/10.14341/probl10100.
14. National Academy of Medicine (NAM) 2022. Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, DC: The National Academies Press. -259pp. https://doi.org/10.17226/27111 6.DEPLOYING ARTIFICIAL INTELLIGENCE IN CLINICAL SETTINGS. Available from: https://www.ncbi.nlm.nih.gov/books/NBK605954/.
15. Vasey, B., Nagendran, M., Campbell, B. et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature Medicine 28, 924–933 (2022). https://doi.org/10.1038/s41591-022-01772-9.
16. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024 Apr 18; 385: q902. doi: 10.1136/bmj.q902. Erratum for: BMJ. 2024 Apr 16; 385: e078378. https://doi.org/10.1136/bmj-2023-078378.
17. Грибова В.В., Шалфеева Е.А., Петряева М.В., Окунь Д.Б. Интеграция клинических рекомендаций в инструменты цифрового здравоохранения: онтологическое моделирование. Сибирский журнал клинической и экспериментальной медицины. 2025. 40(3):36–49. https://doi.org/10.29001/2073-8552-2025-40-3-36-49.
18. Грибова В.В., Шалфеева Е.А., Петряева М.В., Окунь Д.Б., Федорищев Л.А., Ковалев Р.И. Облачный сервис дифференциальной диагностики и назначения персонифицированного лечения воспалительных заболеваний сердца. Программные системы: теория и приложения. 2023. 14(4):141–188. https://doi.org/10.25209/2079-3316-2023-14-4-141-188.
19. Shiferaw KB, Roloff M, Balaur I, Welter D, Waltemath D, Zeleke AA. Guidelines and standard frameworks for artificial intelligence in medicine: a systematic review. JAMIA Open. 2025 Jan 3;8(1):ooae155. https://doi.org/10.1093/jamiaopen/ooae155.
20. Ethics Guidelines for Trustworthy Artificial Intelligence (AI) / AI HLEG of European Comission, Apr 2019 – May 2021. Pp.1-39. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

ОБРАБОТКА И АНАЛИЗ ДАННЫХ ПОГЛОЩЕНИЯ АТМОСФЕРЫ НА ПОЛИГОНЕ «КАРА-ДАГ»[1*]

Е.А. Панкратова[1], И.Т. Бубукин[1], И.В. Ракуть[1,2], М.И. Агафонов[1,2], А.А. Яблоков[2,3], В.И. Бубукин[1], А.Л. Панкратов[2,3]

[1]Нижегородский государственный университет им. Н. И. Лобачевского,
[2]Нижегородский государственный технический университет им. Р. Е. Алексеева,
[3]Институт физики микроструктур РАН.

Проведены обработка и анализ данных поглощения атмосферы, полученных радиометрическими методами на полигоне у подножия горы Балалы-Кая в Юго-Восточном Крыму. Написана программа на языке Python, позволяющая построить гистограммы и провести анализ значений поглощения. Показано, что поглощение на полигоне «Кара-Даг» имеет низкий уровень, несмотря на малую высоту места.

Ключевые слова: полигон «Кара-Даг», астроклимат, поглощение атмосферы, радиометрические методы.

1. Введение

На данный момент на территории России нет крупных миллиметровых и субмиллиметровых телескопов, которые необходимы для исследования таких актуальных научных задач, как поиск новых химических соединений в космосе (в том числе органических молекул), изучение звездообразования во Вселенной, исследование эффекта Сюняева – Зельдовича в скоплениях галактик. Установка таких телескопов возможна на высокогорных полигонах с хорошим астроклиматом: большим количеством безоблачных дней и ночей в течение года и низким содержанием водяных паров в атмосфере, которое определяет уровень поглощения атмосферы [1].

В данной работе проведены обработка и анализ данных поглощения атмосферы за дневной и ночной период, полученных на полигоне «Кара-Даг» у подножия горы Балалы-Кая в Юго-Восточном Крыму за август, сентябрь и ноябрь 2019 года.

2. Обработка данных

Для получения данных использовался радиометрический комплекс, включающий радиометр на длине волны 3 мм. В рамках наблюдений применялись методы атмосферных разрезов «двух углов» и «десяти углов» [2,3].

Для обработки данных была написана программа на языке Python, позволяющая построить гистограммы, наглядно демонстрирующие процентное распределение поглощения атмосферы для длины волны 3 мм в течение выбранного месяца.

3. Анализ

На рис. 1 представлены гистограммы за август 2019 года для методов «двух углов» и «десяти углов». На них можно увидеть 2 ярко выраженных пика на значениях 0.19 и 0.3 Нп. Видно, что значения поглощения при методе «двух углов» немного меньше, чем при методе «десяти углов», это различие несущественно и зависит от выбора углов.

[1*] Работа выполнена при поддержке РНФ, грант № 25-79-20019.

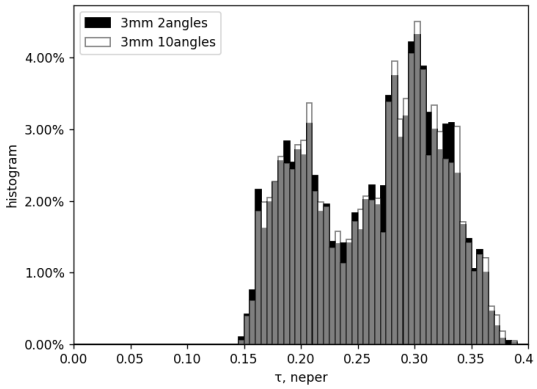


Рис. 1.

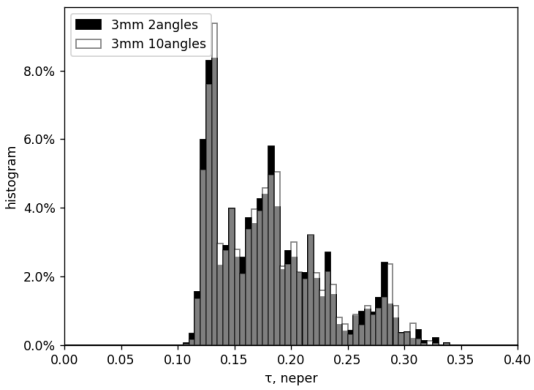


Рис. 2

На рис. 2 представлены гистограммы за ноябрь для методов «двух углов» и «десяти углов». На них можно увидеть 3 выраженных пика на значениях 0.13, 0.18 и 0.28 Нп. Значение самого левого пика соответствует относительной влажности порядка 10% и является крайне хорошим результатом.

3.1. Сравнение поглощения за периоды дня и ночи

В дневной и ночной период на территории полигона «Кара-Даг» преобладают различные господствующие ветра: днём преобладает ветер со стороны моря, который приносит влагу с воздушными массами, а ночью – сухой ветер с полуострова. За счет этого возникает различие поглощения атмосферы за разные периоды суток.

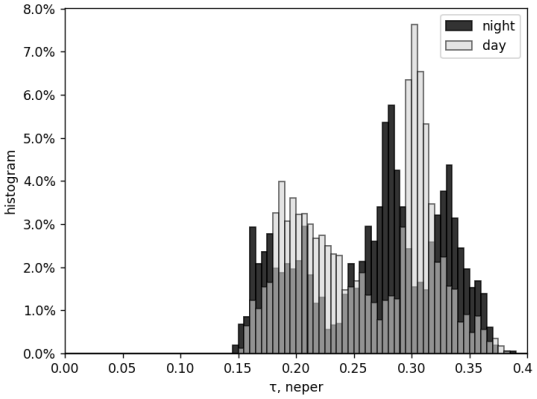


Рис. 3

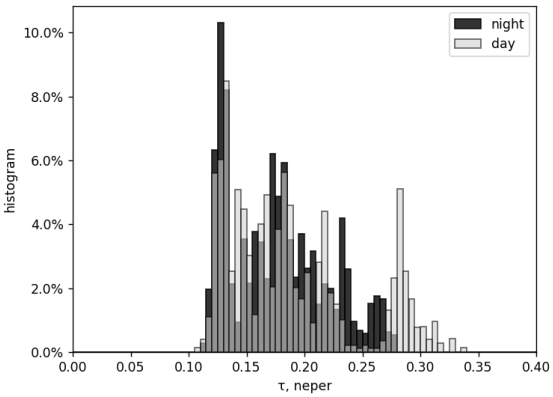


Рис. 4

Хорошо заметно, что в августе (рис. 3) значения поглощения имеют выраженную зависимость от периода дня и ночи. Ночью значения пиков ниже, что свидетельствует о меньшем поглощении атмосферы. Например, самый левый широкий пик находится в диапазоне от 0.15 до 0.22 Нп, что соответствует относительной влажности порядка 15-20%. Это означает, что периоды оптимального значения поглощения и работы радиотелескопов совпадают, что делает полигон «Кара-Даг» хорошей площадкой для реализации радиоастрономических задач.

На рис. 4, где представлены данные за ноябрь, левые пики для периодов дня и ночи совпадают и имеют малые значения, что означает небольшое количество влаги, приносимое за дневное время со стороны моря.

4. Выводы

Поглощение атмосферы на полигоне «Кара-Даг» у подножия горы Балалы-Кая в Юго-Восточном Крыму имеет низкий уровень. В силу различия направления господствующих ветров в дневной и ночной периоды значение поглощения в эти периоды отличается. В ночной период значения поглощения ниже, что важно для радиоастрономических измерений.

Данный полигон является перспективной площадкой для размещения субмиллиметровых телескопов, но требуется дополнительный набор статистики при долговременных измерениях в другие сезоны.

Литература

1. В.Б. Хайкин, А.Ю. Шиховцев, В.Е. Шмагин, М.К. Лебедев, Е.А. Копылов, В.П. Лукин, П.Г. Ковадло, О проекте евразийских субмиллиметровых телескопов (ESMT) и возможности применения адаптивной оптики для улучшения качества субмм изображений // Журнал радиоэлектроники. 2022. №7
2. И.Т. Бубукин, И.В. Ракуть, М.И. Агафонов, А.А. Яблоков, А.Л. Панкратов, Т.Ю. Горбунова, Р.В. Горбунов, Сравнительный анализ условий распространения миллиметровых радиоволн на радиоастрономических полигонах России и Узбекистана // Астрономический журнал. 2021. Т.98, №7. С. 581
3. И.Т. Бубукин, М.И. Агафонов, И.В. Ракуть, А.Л. Панкратов, А.А. Яблоков, А.В. Троцкий, А.С. Приймак, Р.В. Горбунов, Прототип двухволновой радиометрической системы миллиметрового диапазона длин волн для дистанционного зондирования атмосферы и особенности атмосферного поглощения на станции "Кара-Даг" по данным натурных измерений // Известия вузов. Радиофизика. 2019. Т.62, №7-8. С. 630.

ПОСТРОЕНИЕ ПРОГНОСТИЧЕСКОЙ МОДЕЛИ ВЫБОРА СПЕЦИАЛИЗАЦИИ СТУДЕНТАМИ БАКАЛАВРИАТА ИНСТИТУТА ИТММ НА ПРИМЕРЕ РЕАЛЬНЫХ ДАННЫХ

Е.В. Пройдакова, А.С. Бабанова

Нижегородский государственный университет им. Н.И. Лобачевского

В работе рассматривается задача исследования возможности построения прогностической модели выбора специализации студентами бакалавриата, которую можно в перспективе рассматривать как компонент системы поддержки принятия решений в управлении образовательным процессом. Тестировались различные прогностические модели на реальных данных, выгруженных из единой информационной системы ННГУ. Результаты исследования подтверждают возможность использования методов машинного обучения для прогнозирования выбора специализации студентов и оптимизации распределения по кафедрам. Предложенная схема продемонстрировала удовлетворительное качество классификации и устойчивость модели при ограниченном объеме данных. Обсуждаются варианты усовершенствования модели для повышения качества прогнозирования.

Ключевые слова: специализация студентов бакалавриата, методы машинного обучения, корреляционный анализ, регрессионная модель, поддержка принятия решений

1. Введение

Выбор специализации является значимым шагом при профессиональном самоопределении студентов и нетривиальной задачей для ВУЗа. Адекватный прогноз распределения студентов по кафедрам позволит заблаговременно планировать нагрузку преподавателей и перераспределять ресурсы. В настоящей работе предсказание выбора студентом кафедры при распределении на 3 курсе бакалавриата рассматривается не как изолированная задача классификации, а как потенциальный компонент системы поддержки принятия решений в управлении процессами внутри структурных подразделений ННГУ, в частности института ИТММ.

2. Сбор и подготовка данных

Выборка, на которой проводилось исследование, содержит данные, выгруженные из информационной системы «Галактика» и сведения о распределении студентов по кафедрам в 2022 году. Технология распределения студентов по кафедрам, действующая на базе института ИТММ следующая. После завершения трех семестров проводится анкетирование, в ходе которого студенты ранжируют кафедры по приоритетам. Высокий приоритет соответствует кафедре, поставленной студентом на первое место (наиболее предпочтительная), а самый низкий приоритет - на последнее (наименее предпочтительная).

Для защиты персональных данных применялось обезличивание: каждому студенту и преподавателю присвоен уникальный идентификатор. Первоначальный набор данных включал 413 записей и 183 признака. Каждая запись соответствует одному студенту. Исключались только те наблюдения, в которых отсутствовали значения целевой переменной - приоритета кафедры. Сохранение таких записей нецелесообразно, поскольку для них невозможно проверить корректность предсказания модели и рассчитать метрики качества классификации.

Признаки разделились на два типа:

– качественные (95 признаков): идентификаторы преподавателей лекции и (или) практики по дисциплинам первых трех семестров обучения, идентификатор студента, пол, направление подготовки, статус (в академическом отпуске/ учится), признак, указывающий гражданство (0 - иностранный студент, 1 - гражданин РФ), приоритеты выбора кафедр;

– количественные (88 признаков): средний балл студента за три сессии, итоговый балл ЕГЭ (Единый государственный экзамен) без учёта индивидуальных достижений, суммарный балл поступления (включая результаты ЕГЭ, внутренние испытания и индивидуальные достижения), сумма баллов за индивидуальные достижения, баллы ЕГЭ за конкретный предмет, оценки за экзамены и зачеты по дисциплинам за первые три семестра обучения.

Для обеспечения сопоставимости студентов с разными типами вступительных испытаний был создан новый показатель - относительное отклонение среднего балла поступления от среднего по группе абитуриентов (в процентах). Такой подход позволил нивелировать различия между студентами, поступившими по результатам ЕГЭ, и иностранными студентами, оценка которых производится по иной шкале.

Кроме того, вместо множества отдельных предметных оценок были рассчитаны агрегированные показатели по кафедрам. Для каждой кафедры вычислялись два признака - средний экзаменационный балл и средний зачётный балл по кафедральным дисциплинам. Такой подход позволил уменьшить размерность данных, сократить количество пропусков и повысить информативность признаков относительно академической успеваемости по направлениям подготовки. Из набора признаков были исключены данные о преподавателях, а также переменные с высокой взаимной корреляцией, включая отдельные баллы ЕГЭ, показатели индивидуальных достижений и схожие дублирующие параметры.

После предварительной очистки данных и удаления записей с пропущенными значениями в целевой переменной была сформирована итоговая выборка, включающая 300 наблюдений и 28 признаков.

Для дальнейшего исследования результатов распределения студентов, 8 кафедр были разбиты на три класса по содержательной направленности учебных планов:

Класс 1 – преобладают математические дисциплины (кафедра Алгебры, геометрии и дискретной математики, кафедра Дифференциальных уравнений, математического и численного анализа, кафедра Теории вероятностей и анализа данных);

Класс 2 – преобладают программистские дисциплины (кафедра математического обеспечения и суперкомпьютерных технологий);

Класс 3 – равные доли математических и программистских дисциплин (кафедра Информатики и автоматизации научных исследований, кафедра Прикладной математики, кафедра Теоретической, компьютерной и экспериментальной механики, кафедра Теории управления и динамики систем).

В работе фактически рассматривались две задачи: предсказание кафедры с высоким приоритетом и предсказание кафедры с низким приоритетом.

Отметим, что выборки для этих задач имеют разный объём: по высокому приоритету - 300 наблюдений (из них 210 в обучающей выборке и 90 в тестовой), по низкому приоритету - 249 наблюдений (из них 174 в обучающей выборке и 75 в тестовой).

Это связано с особенностями организации учебного процесса в институте ИТММ. Студентам, обучающимся по некоторым направлениям (например, 09.03.03 Прикладная информатика), доступна для распределения только кафедра Информатики и автоматизации научных исследований, поэтому их выбор фиксирован и всегда относится к высокому приоритету, а выбор низкого приоритета невозможен.

3. Отбор признаков

Корреляционный анализ позволил выявить признаки, дублирующие информацию, и исключить избыточные переменные. Из корреляционной матрицы (см. рис. 1) видно, что показатели «Общий балл» и «Балл без индивидуальных достижений» имеют почти функциональную линейную зависимость, также значения коэффициента корреляции, близкие к 1, наблюдаются между отдельными предметными баллами ЕГЭ и их суммарным значением.

Такие признаки не добавляют новой информации и лишь увеличивают размерность данных. Поэтому при дальнейшем анализе были сохранены только агрегированные показатели (например, суммарный балл ЕГЭ и средний балл за сессии), тогда как частные предметные и дублирующие значения были исключены.

Стоит отметить, что решение об исключении признаков принималось исходя из их информативности и взаимной зависимости, а не только строго по порогу значения коэффициента корреляции. Такой подход позволил сократить избыточность и снизить риск переобучения без потери значимой информации.

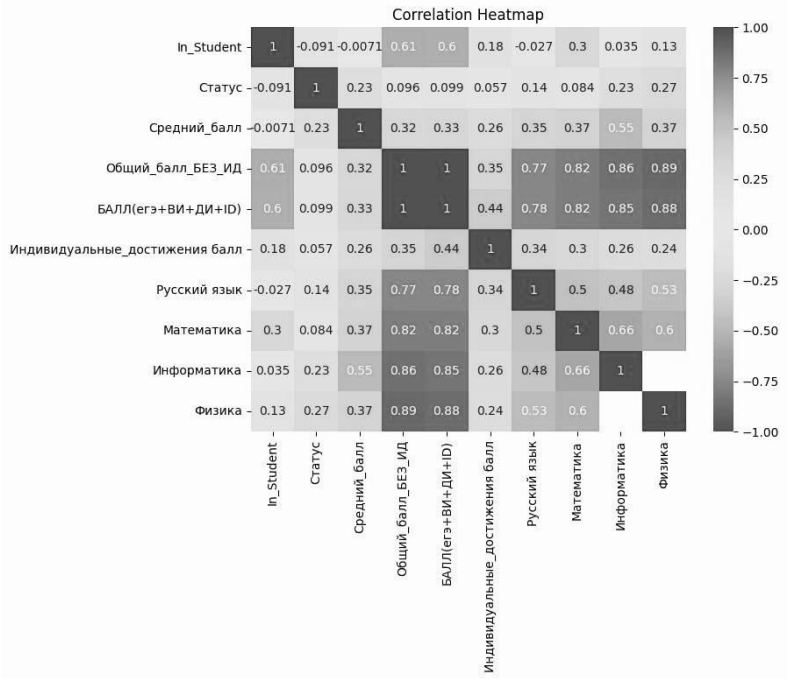


Рис. 1. Корреляционная матрица признаков: по осям - наименования признаков; интенсивность заливки ячеек отражает величину коэффициента корреляции

4. Методология

При дальнейшем анализе проверялись три модели: логистическая регрессия (One-vs-Rest, с балансировкой классов), Random Forest (представляет собой ансамблевый метод, основанный на построении множества деревьев решений[3]) и CatBoost (алгоритм машинного обучения, разработанный компанией Яндекс). Он предназначен для работы с табличными данными и особенно эффективен в задачах, где важную роль играют категориальные признаки [2]. Обучение проводилось с функцией потерь MultiClass. Для оценки качества моделей и сравнения различных алгоритмов используются метрики: Accuracy (доля правильных ответов), Precision (величина показывает долю объектов, которые были классифицированы как положительные и при этом действительно являющихся положительными), Recall (полнота), F1(гармоническое среднее Precision и Recall) и Macro F1 (усреднение F1 по всем классам)[1]. Ниже приведены используемые соотношения:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad Precision = \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN},$$
$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision \cdot Recall}, \quad Macro\ F_1 = \frac{1}{K}\sum_{k=1}^{K} 2 \cdot F_{1_k},$$

здесь TP-объекты, которые модель правильно отнесла к нужному классу, TN-объекты, правильно определенные как не принадлежащие данному классу, FP-объекты, которые модель ошибочно отнесла к этому классу, хотя на самом деле они относятся к другому, FN-объекты, которые модель не распознала как принадлежащие к данному классу, K- количество классов. Несмотря на относительно небольшой размер выборки (300 записей), данные были разделены на обучающую и тестовую в соотношении 80% к 20% соответственно. При случайном делении малые классы могут отсутствовать в одной из частей, что ведет к смещению оценки качества. Для устранения этой проблемы использовано стратифицированное разделение, сохраняющее пропорции классов. Такой подход обеспечивает корректное представление даже малых групп и

более надёжные метрики, особенно Macro F1, что позволяет корректно учитывать несбалансированные классы (см. табл. 1).

Таблица 1. Качество классификации без стратификации и со стратификацией

Задача	Стратификация	Accuracy	MacroF1
Высокий приоритет	Нет	0.71	0.65
Высокий приоритет	Есть	0.80	0.75
Низкий приоритет	Нет	0.71	0.44
Низкий приоритет	Есть	0.76	0.47

5. Сравнение моделей

Результаты, представленные в таблице 2 показывают, что CatBoost обеспечивает наилучшее качество классификации: Accuracy = 0.80, Macro F1 = 0.75 (для задачи высокого приоритета). Логистическая регрессия показала наименьшую точность (Macro F1 ≈ 0.52), Random Forest - промежуточные результаты.

Отметим, что высокий приоритет выбора кафедры предсказывается значительно лучше, чем низкий. Наиболее точно классифицируется программистский класс (F1 ≈ 0.87), хуже - математический (F1 ≈ 0.72) из-за малого числа наблюдений.

Таблица 2. Сравнение моделей(высокий приоритет)

Модель	Accuracy	MacroF1
Logistic Regression	0.61	0.52
Random Forest	0.73	0.65
CatBoost	0.80	0.75

Отдельно отметим, что исследование имеет ряд ограничений: малый размер выборки, особенно для математического класса, что снижает статистическую устойчивость выводов; сильный дисбаланс классов, усложняющий обучение моделей; ограниченный набор признаков - отсутствуют данные о посещаемости и пересдачах.

6. Заключение

В работе предложена методическая схема прогнозирования выбора специализации студентов при распределении по кафедрам на 3 курсе бакалавриата.

Выявлено, что метод CatBoost выделяется среди протестированных методов из-за способности эффективно работать с категориальными данными и предоставлять точные предсказания, минимизируя риск переобучения и обеспечивая высокую скорость обучения[2].

Направления развития исследования для повышения качества модели включают расширение выборки, введение новых признаков, применение методов балансировки (SMOTE, cost-sensitive learning).

Модель в перспективе может быть встроена в информационную систему института ИТММ и использоваться для прогнозирования распределения студентов по направлениям, планирования нагрузки преподавателей, а также для выявления направлений с переизбытком или дефицитом спроса. Это позволит повысить эффективность управления образовательным процессом.

Литература

1. Иванов, С. В. Метрики классификации задач машинного обучения / С. В. Иванов // Актуальные проблемы и перспективы развития экономики : труды XIX Всероссийской с международным участием научнопрактической конференции, Симферополь-Гурзуф, 15–17 октября 2020 года. – Симферополь: ИП Зуева Т. В., 2020. – С. 173-174. – EDN YLOAJV.
2. Яндекс.Практикум. Алгоритм градиентного бустинга CatBoost. URL: https://practicum.yandex.ru/blog/algoritm-gradientnogo-bustinga-catboost/ (дата обращения: 25.08.2025).

3. Кулинча, П. В. Использование случайного леса для классификации данных / П. В. Кулинча // Молодой ученый. – 2024. – № 24(523). – С. 88-91. – EDN QOOEIG.
4. Грас Дж. Data Science. Наука о данных с нуля. Пер. с англ. – Спб.:БХВ-Петербург, 2019.– 336 с.

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ БИОФАБРИКАЦИИ РАЗВЕТВЛЕННЫХ КРОВЕНОСНЫХ СОСУДОВ АКУСТИЧЕСКОЙ ЛЕВИТАЦИИ

Е.В. Пыряев, К.А. Беклемышева

Московский физико-технический институт

Рассматриваются вопросы моделирования явления акустической левитации применительно к созданию биопечатных конструкций. Имитационные эксперименты включают в себя решение системы акустических уравнений в жидкой среде на трёхмерной равномерной сетке. Проводятся два различных численных эксперимента для исследования подхода к созданию акустических полей для сложных трубчатых тканевых конструкций. Эксперименты с плоскими излучатели с отражателями параболической формы отражают дизайн инженерно масштабируемого и перспективного решения поставленной задачи.

Ключевые слова: биофабрикация, численное моделирование, акустическая левитация, потенциал Горькова, тканевая инженерия

1. Введение

На сегодняшний день технологии создания 2D- и 3D-конструкций из биоматериалов разделяются на две категории: биопечать и биофабрикация. Методы биопечати схожи по принципу с методом послойного наплавления (FDM): биоконструкт создаётся слой за слоем, каждый следующий накладывается поверх предыдущего. Биофабрикация же полагается на формирование всей конструкции из клеточного материала за один шаг. Обычно это предполагает использование определённых физических полей – акустических или магнитных, – для соединения гранул биоматериала в массив требуемой формы, сохраняемой после сращивания клеток между собой. Эта группа методов считается более перспективной, поскольку имеет шанс преодолеть ограничения биопечати и на пути к созданию более совершенных медицинских изделий.

Однако эти перспективы ограничены комплексом физических проблем, лежащих в основе механизмов точной манипуляции физическими полями. Контроль над их пространственным распределением представляет собой непростую задачу, поскольку аналитических решений для произвольных систем уравнений Максвелла или Навье-Стокса (даже с использованием акустического приближения в последнем) не существует. Однако существуют определенные подходы к преодолению этого теоретического пробела, а свойства задачи часто дают представление о поиске подходящего решения.

Наши исследования сосредоточены на использовании биофабричных конструкций в сфере сосудистой хирургии. Технологии шунтирования и стентирования, несмотря на свою надежность и широкое применение для лечения сосудистых заболеваний в настоящее время, попрежнему считаются неоптимальными из-за использования экзогенных деталей из металла. Эти устройства могут быть заменены биоконструктированными фрагментами кровеносных сосудов, поскольку их можно имплантировать для замещения проблемных тканей.

Традиционные методы биоконструкции включают так называемый скаффолдинг: биоконструкт создается с помощью каркаса из чужеродного материала, который служит скелетом для роста тканевых сфероидов. Хотя этот подход показал удовлетворительные результаты, конструкции, созданные с использованием скаффолдов, обладают нестабильными свойствами из-за недостатков процесса биофабрикации. Это ограничивает их запас прочности и снижает надёжность. Хотя технология скаффолдов используется и по сей день, поиск альтернативных методов без них продолжается.

Осуществление контроля над процессом биофабрикации без материальных скаффолдов – сложная задача, для которой пока не найдено универсального решения. Однако некоторые физические принципы позволяют манипулировать тканевым материалом. Например, совместное

использование магнитных и акустических полей позволяет собрать трубчатый биоконструкт из сфероидов ткани мочевого пузыря человека. Такой подход считается точным, технически осуществимым и адаптивным для каждого конкретного случая.

Хотя существуют исследования, посвященные использованию магнитных полей в биофабрикации, в нашей работе мы обсудим концепцию использования акустических полей именно для создания тканевых конструкций. Следует также отметить, что мы концентрируем наши усилия, прежде всего, на трубчатых биоконструкциях, поскольку они имеют очевидное клиническое применение и востребованы в медицинской промышленности. Для достижения необходимой точности управления акустическим полем требуется тщательное численное моделирование распространения звука. Мы используем сеточно-характеристический численный метод для решения акустических уравнений. Этот решатель был ранее проверен на базовой задаче акустической левитации [3].

2. Результаты работы

2.1. Цилиндрический пьезокристаллический излучатель

Дальнейшие исследования расширили применение акустической левитации в более релевантных случаях. В [1] исследователи использовали пьезокерамические цилиндрические и кольцевые ультразвуковые преобразователи. Этот тип преобразователей собирал тканевые сфероиды в трубчатые тканевые конструкции, генерируя стоячие акустические волны. Мы провели численное моделирование подобной акустической системы для дальнейшей проверки конвейера решений. В связи с автомодельностью задач акустической левитации по геометрическим размерам диаметр преобразователя был выбран таким образом, чтобы внутри могли поместиться три полуволны. Полученные в ходе расчётов результаты хорошо согласуются с аналитической оценкой решения на основе решения уравнения Бесселя.

2.2. Ансамбль из трёх цилиндрических излучателей

Кейс с одиночной трубчатой конструкцией легли в основу необычного эксперимента. Поскольку целью нашего исследования является возможность создания трубчатых конструкций с произвольным числом ветвей и управления их геометрией, была выдвинута простая гипотеза: три цилиндрических пьезокерамических ультразвуковых преобразователя необходимо сложить вместе в Y-образную форму так, чтобы их излучение в суперпозиции создавало потенциал, соответствующий желаемой формы ветвления.

Гипотеза была проверена моделированием. Внутри зоны излучения образовалась гладкая потенциальная яма, распределенная в пространстве по форме разветвляющегося сосуда. Яма имеет особенность в виде стенки волнообразной форма стенки: это связано с интерференцией пьезокристаллических излучателей.

2.3. Плоский излучатель и параболический отражатель

Дизайн предыдущего эксперимента Плоский излучатель имеет существенный недостаток: метод разветвления нескольких цилиндрических ультразвуковых преобразователей не может быть инженерно масштабирован для получения тканевых конструкций с более сложной структурой. Это потребовало бы изготовления произвольного количества пьезокристаллических излучателей, каждый из которых имеет свой размер и форму, что экономически невыгодно и громоздко. Был поставлен альтернативный численный эксперимент с целью моделирования процесса создания трубчатой конструкции без использования цилиндрических ультразвуковых преобразователей. В случае, когда наш арсенал излучателей ограничен лишь плоскими устройствами, подобное может быть возможно только при использовании отражателей специального типа, которые могли бы создавать акустический потенциал цилиндрической формы в комбинации с плоскими излучателями.

Было предложено провести численный эксперимент с плоским излучателем и противолежащим ему отражателем. Отличительной особенностью отражателя является параболический

вырез, выполненный таким образом, чтобы падающие волны после отражения собирались в фокусе параболы. Идея эксперимента заключалась в анализе интерференционной картины излучаемых и отраженных волн, уделяя особое внимание ситуации в фокусе параболы.

Как и следовало ожидать, пик амплитуды излучения формировался в фокусе параболы. Вокруг него образовалась потенциальная яма, принявшая форму шестиугольной призмы. Такую призму можно считать близкой к цилиндру из-за гладкости потенциального поля. Следует также отметить, что минимальная яма просачивается за пределы цилиндра на поверхность параболы. Это представляет собой проблему, поскольку может нарушить желаемое распределение тканевых сфероидов. Наши будущие исследования будут посвящены этому вопросу. Наша текущая гипотеза предполагает необходимость использования не одного акустического излучателя, а массива малогабаритных программируемых излучателей – антенной решетки. Таким образом, дальнейшее исследование с такими рефлекторами перспективно, поскольку обеспечивает решение, масштабируемое для широкого класса трубчатых тканевых конструкций с различным количеством ветвей.

Литература

1. V.A. Parfenov, E.V. Koudan, A.A. Krokhmal, E.A. Annenkova, S.V. Petrov, F.D.A.S. Pereira, P.A. Karalkin, E.K. Nezhurina, A.A. Gryadunova, E.A. Bulanova, O.A. Sapozhnikov, S.A. Tsysar, K. Liu, E. Oosterwijk, H. vanBeuningen, P. van derKraan, S. Granneman, H. Engelkamp, P. Christianen, V. Kasyanov, Y. D. Khesuani, V. A. Mironov, Biofabrication of a Functional Tubular Construct from Tissue Spheroids Using Magnetoacoustic Levitational Directed Assembly. Adv. Healthcare Mater. 2020, 9, 2000721. https://doi.org/10.1002/adhm.202000721.
2. K.A. Beklemysheva and I.B. Petrov, Damage Modeling in Hybrid Composites Subject to Low-Speed Impact, Mathematical Models and Computer Simulations 11, 469 478 (2019).
3. Pyriaev, E., Beklemysheva, K. (2025). Acoustic Levitation Applied to Tissue Engineering. In: Balandin, D., Barkalov, K., Meyerov, I. (eds) Mathematical Modeling and Supercomputer Technologies. MMST 2024. Communications in Computer and Information Science, vol 2363. Springer, Cham. https://doi.org/10.1007/978-3-031-80457-1_19.
4. Xie, W., Beiyu, W.: Dependence of acoustic levitation capabilities on geometric parameters. Phys. Rev. E, Stat. Nonlinear, Soft Matter Phys. 66, 026605 (2002). https://doi.org/10.1103/PhysRevE.66.026605.
5. L.P. Gor'kov, “Forces acting on a small particle in an acoustic field within an ideal fluid”, Dokl. Akad. Nauk SSSR, 140:1 (1961), 88–91.
6. Melde, K., Mark, A., Qiu, T. et al. Holograms for acoustics. Nature 537, 518–522 (2016). https://doi.org/10.1038/nature19755.

ИСПОЛЬЗОВАНИЕ ЦИКЛИЧЕСКОГО БУФЕРА ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ВОССТАНОВЛЕНИЯ ИСКАЖЕНИЙ ВОЛНОВОГО ФРОНТА В ОПТИЧЕСКИХ СИСТЕМАХ

Ю.А. Родимков

Нижегородский государственный университет им. Н.И. Лобачевского

Развитие технологий оптических систем позволило достичь экстремальных интенсивностей в экспериментах по лазерной физике. Однако автоматизация таких установок остаётся сложной задачей. В последние годы активно развивается ряд подходов, основанных на методах глубокого обучения, которые позволяют восстанавливать искажения волнового фронта. В данной работе предлагается подход к восстановлению искажений волнового фронта, основанный на использовании циклического буфера данных. В отличие от традиционных методов, где обучающая выборка фиксируется заранее, предложенный подход предусматривает динамическое обновление данных в процессе обучения.

Ключевые слова: нейронные сети, глубокое обучение, синтетические данные, переобучение, лазерная физика, оптические системы, искажения волнового фронта, калибровка лазерной установки.

1. Введение

достижения экстремальных интенсивностей в фокальной плоскости лазерных установок, включая современные петаваттные системы [1, 2], требуется высокая точность калибровки оптического тракта, в частности – устранение искажений волнового фронта. Подобные искажения возникают вследствие оптических аберраций и технологических несовершенств элементов системы, таких как линзы и зеркала. Для устранения таких искажений можно использовать адаптивную оптическую систему, для корректной работы которой обычно требуется её калибровка.

Существуют различные методы калибровки оборудования, использующие информацию об распределении энергии в фокусе, основанные на жадном поиске [3], генетических алгоритмах [4], стохастическом градиентном спуске [5, 6], а также алгоритме моделирования отжига [7, 8]. Однако к недостаткам этих методов относится относительно большое время работы и риск застревания в локальных минимумах. Поэтому в последние годы набирает популярность применение методов глубокого обучения [9, 10].

Одной из ключевых проблем методов глубокого обучения является переобучение. В данной работе предлагается метод обучения нейронной сети, в котором используется циклический буфер с генерацией синтетических данных в режиме реального времени с целью предотвращения переобучения и повышения точности восстановления искажений волнового фронта по сравнению с традиционным подходом, основанным на фиксированной выборке.

2. Постановка задачи

Принцип работы адаптивной оптической системы включает следующие этапы. Волновой фронт с искажениями p_1 поступает на деформируемое зеркало, после чего сигнал направляется на полупрозрачное зеркало. Часть излучения проходит далее для фокусировки, а другая часть поступает на датчик волнового фронта, который обычно состоит из оптических элементов. Датчик способен измерять искажения волнового фронта с высокой скоростью и точностью.

Однако у описанного подхода есть существенные недостатки. В частности, не учитывается влияние искажений, вносимых фокусирующим трактом p_2. Также искажения, возникающие в самом датчике волнового фронта, не влияют на область в фокусе, но учитываются датчиком.

Необходимо восстановить искажения волнового фронта $p = p_1 + p_2$ с помощью нейронной сети $NN: (ff_1, ff_2) \to p$, используя распределение энергии в фокальной плоскости ff_1 и её окрестности ff_2. Использование информации из окрестности фокуса ff_2 необходимо для устранения неопределённости. В противном случае одному и тому же распределению энергии в фокусе может соответствовать разные искажения волнового фронта.

3. Сбор синтетических данных

Одним из способов параметризации искажений волнового фронта p является использование полиномов Цернике. Лазерный импульс с искажённым волновым фронтом формировался как сумма ста полиномов Цернике с случайными коэффициентами a_{nm}. Для этого задавалась искажения лазерного импульса следующим образом:

$$W(p,\varphi) = \sum_{n=0}^{N} \sum_{m=-n}^{n} a_{nm} Z_n^m(p,\varphi)$$

$Z_n^m(p,\varphi) = R_n^m(p)\cos(m\varphi)$ и $Z_n^{-m}(p,\varphi) = R_n^m(p)\sin(m\varphi)$, R_n^m – радиальные многочлены, m и n – неотрицательные целые числа, удовлетворяющие условию $n \geq m$, φ – азимутальный угол, p – радиальное расстояние, $0 \leq p \leq 1$.

Амплитуда поля рассчитывается следующим образом:

$$E(p,\varphi) = e^{\frac{2\pi i W(p,\varphi)}{\lambda}},$$

где λ – длина волны, равная 10^{-6}. Далее с помощью преобразования Фурье поле распространяется в дальнюю зону, где рассчитывалась его интенсивность. Для распределения в фокусе $Z_0^0 = 0$. Распределение энергии в окрестности фокуса рассчитывалось аналогично, при этом $Z_0^0 = 0.5$. Полиномы Z_1^{-1} и Z_1^1 были равны нулю всегда.

С увеличением силы искажения распределение энергии в фокусе становится всё более хаотичным, среднеквадратичное отклонение волнового фронта соответственно увеличивается. При обучении изображения распределений энергии в фокусе и в окрестности нормировались на максимум отдельно для каждого изображения. Искажения волнового фронта нормировались на тройную длину волны.

4. Методология

Процесс обучения и обновления синтетических данных в циклическом буфере происходит асинхронно. Обучение представляет собой классическую процедуру, выполняемую на процессе мастере. В пакетном режиме обучающие примеры случайным образом, без повторений, выбираются из набора данных. Эта часть выполняется в критической секции. Далее происходит прямой и обратный проходы по нейронной сети.

Буфер в начальный момент хранит заранее собранную выборку. Такой подход направлен на борьбу с эффектом холодного старта, при котором модель может переобучаться на небольшой выборке на первых эпохах обучения, что замедляет сходимость.

После начала обучения запускается множество дополнительных процессов. В каждом из них производится генерация случайных коэффициентов полиномов Цернике, рассчитывается искажение волнового фронта, а затем выполняется численное моделирование.

Для уменьшения накладных расходов, связанных с синхронизацией, в каждом процессе сначала собирается 100 синтетических примеров. Как только пачка данных будет сгенерирована, процессы встают в очередь с процессом мастером. После этого данные последовательно добавляются в массивы для дальнейшей обработки. Часть, связанная с добавлением новых примеров, также находится в критической секции.

5. Результаты

Для решения задачи восстановления искажений волнового фронта использовалась архитектура U-Net [11]. Распределение энергии в фокусе и в его окрестности объединялось в двухканальное изображение. Для сравнительного анализа использовались три выборки, размеры кото-

рых составляли 100 000, 200 000 и 400 000 примеров. Обучение проводилось с использованием буфера, размер которого был равен 100 000 примеров. В качестве исходной выборки для инициализации использовалась выборка размером 100 000 примеров. Каждая конфигурация обучалась в течение 40 000 итераций для обеспечения единообразия. Для валидации и тестирования использовались отдельные выборки, содержащие 5 000 и 2 000 примеров соответственно.

На тренировочных данных наилучшая точность восстановления искажений волнового фронта достигается при использовании наименьшей по размеру выборки. Среди выборок фиксированного размера наилучший результат, как и ожидалось, был достигнут при использовании самой большой выборки, включающей 400 000 примеров. На тестовых данных лучшая точность демонстрирует нейронная сеть, обученная с использованием циклического буфера. Это подтверждает эффективность использования циклического буфера при работе с синтетическими данными, которые можно получить с помощью численного моделирования. По сравнению с выборкой аналогичного размера точность реконструкции искажений с помощью сверточной нейронной сети возросла в два раза. Кроме того, точность метода улучшилась примерно на 30% по сравнению с выборкой, размер которой в четыре раза превышал размер циклического буфера. Подробное описание полученных результатов будет приведено в докладе.

Литература

1. Danson C.N. et al. Petawatt and exawatt class lasers worldwide // High Power Laser Science and Engineering. – 2019. – Т. 7. – С. 54.
2. Соловьев А.А. и др. Исследования в области физики плазмы и ускорения частиц на петаваттном лазере PEARL // Успехи физических наук. – 2024. – Т. 194. – № 3. – С. 313–335.
3. Yang P. et al. 19-element sensorless adaptive optical system based on modified hill-climbing and genetic algorithms //3rd International Symposium on Advanced Optical Manufacturing and Testing Technologies: Optical Test and Measurement Technology and Equipment. – SPIE, 2007. – Т. 6723. – С. 36-42.
4. Poland S.P., Wright A.J., Girkin J.M. Evaluation of fitness parameters used in an iterative approach to aberration correction in optical sectioning microscopy //Applied optics. – 2008. – Т. 47. – №. 6. – С. 731-736.
5. Piatrou P., Roggemann M. Beaconless stochastic parallel gradient descent laser beam control: numerical experiments //Applied optics. – 2007. – Т. 46. – №. 27. – С. 6831-6842.
6. Котов А.В. и др. Адаптивная система коррекции оптических аберраций излучения мощных лазеров с динамическим определением эталонной формы волнового фронта //Квантовая электроника. – 2021. – Т. 51. – №. 7. – С. 593-596.
7. Zommer S. et al. Simulated annealing in ocular adaptive optics //Optics letters. – 2006. – Т. 31. – №. 7. – С. 939-941.
8. El-Agmy R. et al. Adaptive beam profile control using a simulated annealing algorithm //Optics Express. – 2005. – Т. 13. – №. 16. – С. 6085-6091.
9. Котов А.В. и др. Восстановление волнового фронта лазерного пучка на основе анализа распределения интенсивности в фокусе и вне фокуса с использованием свёрточной нейронной сети //ИЗВЕСТИЯ ВУЗОВ. РАДИОФИЗИКА. – 2024. – Т. 67. – №. 11-12. – С. 1013-1019.
10. He Y. et al. Deep learning wavefront sensing method for Shack-Hartmann sensors with sparse sub-apertures //Optics Express. – 2021. – Т. 29. – №. 11. – С. 17669-17682.
11. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation //International Conference on Medical image computing and computer-assisted intervention. – Cham : Springer international publishing, 2015. – С. 234-241.

ИМИТАЦИОННАЯ МОДЕЛЬ ГОРИЗОНТАЛЬНО РАСПОЛОЖЕННОГО ЦИЛИНДРИЧЕСКОГО РЕЗЕРВУАРА ДЛЯ ОЦЕНКИ ПОГРЕШНОСТИ ГРАДУИРОВОЧНОЙ ПРОЦЕДУРЫ ОБЪЁМНОГО МЕТОДА

С.В. Русаков[1], В.А. Старостин[2], И.И. Хейфец[1]

[1]Пермский государственный национальный исследовательский университет,
[2]Общество с ограниченной ответственностью «ВИСТЕРИЯ»

Настоящая работа посвящена оценке погрешности существующих методов градуировки горизонтальных цилиндрических топливных резервуаров. В ходе исследования была разработана цифровая модель такого резервуара, после чего с её помощью был проведён имитационный эксперимент по оценке погрешностей методов градуировки. Эксперимент выявил существенные погрешности, обусловленные методами.

Ключевые слова: градуировка ёмкостей, погрешность метода, имитационное моделирование.

1. Введение

В ходе эксплуатации цилиндрических ёмкостей на различных предприятиях топливно-энергетического комплекса, присутствует задача учёта хранимого топлива. Для выполнения этой задачи необходима периодическая градуировка ёмкостей. На текущий момент, общепринятым стандартом в области периодических градуировок цилиндрических резервуаров является ГОСТ 8.346-2000 [1]. Наиболее точные методы градуировки в рамках этого стандарта, объёмные методы, имеют существенные недостатки [2–5]: в частности – необходимость в выведении резервуара из эксплуатации и недостаточную точность. Также, следует отметить и негативное влияние человеческого фактора [2, 3] на качество градуировки. Цель настоящего исследования – оценить погрешность процедуры градуировки ёмкости по [1] для дальнейшей разработки статистического метода градуировки. Для достижения данной цели были поставлены задачи: разработка имитационной модели горизонтального цилиндрического резервуара, а также оценка влияния погрешности измерительных приборов на результаты градуировок. Актуальность данной работы состоит в получении нового градуировочного метода, лишённого вышеописанных недостатков, и, как следствие, в снижении издержек при хранении и распределении горюче-смазочных материалов на предприятиях топливно-энергетического комплекса.

2. Методология

В рамках настоящего исследования, анализируются объёмные методы стандарта [1]. Существует объёмный статический и объёмный динамический методы. Статический метод заключается в порционном наполнении резервуара дозами поверочной жидкости с одновременным измерением уровня и объёма при каждом изменении уровня на некую величину. Динамический метод подразумевает непрерывное наполнение с одновременной фиксацией уровня и объёма при изменении уровня.

Для обеспечения текущего контроля объема резервуаров, хранящих нефтепродукты используются два набора данных: объём отпущенного горючего и его уровень в резервуаре. При этом, эталоном являются результаты регулярно проводимой процедуры градуировки. В соответствии с [1] заполнение резервуара производится последовательно для достижения уровней $H_i = (1\ см) \cdot i,\ i = 0, 1, \dots, N$, где N - целая часть от выражения $2R/(1\ см)$, а R – радиус резервуара. При этом на каждом таком шаге фиксируется (измеряется) «накопленный объем»:

$$dV(H_i) = V(H_i) - V(H_{i-1}), \quad (1)$$

где $V(H_i)$ – объём, соответствующий уровню H_i, на основе которого вычисляется коэффициент вместимости – вместимость одного миллиметра высоты наполнения:

$$\theta_i = \frac{V(H_i) - V(H_{i-1})}{10}. \quad (2)$$

Выполним оценку погрешности данных градуировочной таблицы.

Обозначим $\widetilde{H}_i = H_i + \Delta H_i$- реальное значение величины H_i, которое определяется с точностью до 1 мм, следовательно $|\Delta H_i| \le |\Delta H| = 0.05$мм. В соответствии с линейной теорией погрешности [6] справедливо соотношение:

$$V(\widetilde{H}_i) = V(H_i) + \Delta H_i \frac{dV(H_i)}{dH}, \quad i = 1,\dots,n, \quad V(\widetilde{H}_0) = 0. \quad (3)$$

По мере заполнения резервуара на каждом сантиметровом шаге фиксируется значение «накопленного объема», содержащее допустимую относительную ошибку расходометра $|\delta V_i| \le \delta V = 0.15\%$. В этом случае

$$d\tilde{V}(\widetilde{H}_i) = \tilde{V}(\widetilde{H}_i) - \tilde{V}(\widetilde{H}_{i-1}) = dV(\widetilde{H}_i)(1 + \delta V_i), \quad i = 1,\dots,n, \quad \tilde{V}(\widetilde{H}_0) = 0. \quad (4)$$

Оценим погрешность вычисления коэффициента вместимости. Используя соотношение (3) получаем:

$$dV(\widetilde{H}_i) = V(\widetilde{H}_i) - V(\widetilde{H}_{i-1}) = \quad (5)$$
$$= V(H_i) - V(H_{i-1}) + \left(\Delta H_i \frac{dV(H_i)}{dH} - \Delta H_{i-1} \frac{dV(H_{i-1})}{dH}\right), \ i = 1,\dots,n.$$

Тогда, из выражений (4)-(5), получаем:

$$\frac{d\tilde{V}(\widetilde{H}_i)}{dV(H_i)} = \frac{\tilde{V}(\widetilde{H}_i) - \tilde{V}(\widetilde{H}_{i-1})}{V(H_i) - V(H_{i-1})} - 1 = \delta V_i + \frac{\Delta H_i \frac{dV(H_i)}{dH} - \Delta H_{i-1} \frac{dV(H_{i-1})}{dH}}{V(H_i) - V(H_{i-1})}, \ i = 1,\dots,n. \quad (6)$$

Справедливы соотношения

$$\frac{dV(H_i)}{dH} \approx \frac{V(H_i) - V(H_{i-1})}{H_i - H_{i-1}},$$
$$\frac{dV(H_{i-1})}{dH} \approx \frac{V(H_i) - V(H_{i-1})}{H_i - H_{i-1}},$$

поскольку их правые части являются конечными разностями «назад» и «вперед» соответственно. Тогда

$$\frac{\Delta H_i \frac{dV(H_i)}{dH}(H_i - H_{i-1})}{\big(V(H_i) - V(H_{i-1})\big)(H_i - H_{i-1})} \approx \frac{\Delta H_i}{H_i - H_{i-1}},$$
$$\frac{\Delta H_{i-1} \frac{dV(H_{i-1})}{dH}(H_i - H_{i-1})}{\big(V(H_i) - V(H_{i-1})\big)(H_i - H_{i-1})} \approx\approx \frac{\Delta H_{i-1}}{H_i - H_{i-1}}. \quad (7)$$

Окончательно из соотношений (6)-(7) получаем оценку относительной погрешности определения коэффициента вместимости:

$$\left|\frac{\tilde{V}(\widetilde{H}_i) - \tilde{V}(\widetilde{H}_{i-1})}{V(H_i) - V(H_{i-1})} - 1\right| \le |\delta V_i| + \frac{|\Delta H_i|}{H_i - H_{i-1}} + \frac{|\Delta H_{i-1}|}{H_i - H_{i-1}} \le \delta V + 2\frac{\Delta H}{H_i - H_{i-1}} = \quad (8)$$
$$= 0.15 \cdot 10^{-2} + 0.1.$$

Таким образом, с учётом допустимых погрешностей фиксации величин H_i и $dV(H_i)$, получаем, что относительная погрешность коэффициента вместимости может составлять до 10% и основной вклад в неё вносит именно погрешность определения уровня жидкости в резервуаре.

2.2. Имитационная модель

Для экспериментального подтверждения оценки (8) воспользуемся цифровой моделью горизонтально расположенного цилиндрического резервуара, которая представляет собой идеальный цилиндр радиусом R и длиной L, лежащий на боковой поверхности горизонтально. Уровень топлива H изменяется в пределах от 0 до $2R$. Общая схема модели представлена на рисунке 1.

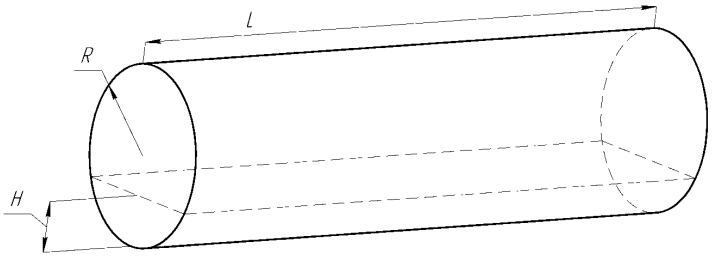


Рис.1. Общая схема цифровой модели

Нетрудно показать, что зависимость объёма резервуара от уровня – функция V(H), имеет вид:

$$V(H) = \begin{cases} \frac{1}{2}LR^2(\alpha - \sin\alpha), & \alpha = 2 \cdot arccos(1-h), \quad h \le 1, \\ \frac{1}{2}LR^2(2\pi - \beta + \sin\beta), & \beta = 2 \cdot arccos(h-1), \quad h > 1, \end{cases} \tag{9}$$

где $h = \frac{H}{R}$ - безразмерный относительный уровень. Кроме того,

$$\frac{dV(H)}{dH} = 2LR\sqrt{1-(1-h)^2}. \tag{10}$$

3. Результаты

С помощью описанной имитационной модели, проанализируем зависимость погрешность вычисления функций *V(H)* и *dV(H)*, получаемых при построении градуировочной таблицы объёмным методом.

При проведении вычислительного эксперимента будем вносить случайные возмущения в определении величин ΔH_i и δV_i по следующей схеме: $\Delta H_i = x\Delta H, \quad \delta V_i = x\delta V$, где *x* – случайная величина, равномерно распределенная на отрезке [-1,1]. На рисунке 2 представлены значения коэффициента вместимости, полученные в результате имитационного эксперимента для случая *R=200 см, L=500 см*. При этом используется так называемый рабочий диапазон уровней *50 см ≤Hi≤150 см* и рассмотрены 3 варианта:

$$\theta_i^{(0)} = \frac{dV(H_i)}{10}, \quad \theta_i^{(1)} = \frac{d\tilde{V}(\tilde{H}_i)}{10}, \quad \theta_i^{(2)} = \frac{d\tilde{V}(\tilde{H}_i)}{10(\tilde{H}_i - \tilde{H}_{i-1})}.$$

Вариант «0» соответствуют эталонному точному значению, в варианте «1» и уровень и объём взяты с учетом погрешности измерения, в варианте «2» учитывается тот факт, что в реальности разница в уровнях на каждом шаге не равна 1 см, а известна с точность до 1 мм.

Из рисунка 2 видно, что «0» и «2» варианты зрительно не различимы, в то же время вариант «1», имитирующий реальные измерения, даёт существенную погрешность. Так в случае приближение результатов регрессионным уравнением в виде полинома второго порядка получаем:

$$\begin{aligned} \theta_i^{(0)} &= 0{,}8654 + 5{,}362 \cdot 10^{-3}(H_i - 50) - 5{,}308 \cdot 10^{-5}(H_i - 50)^2, \\ \theta_i^{(1)} &= 0{,}8650 + 5{,}396 \cdot 10^{-3}(H_i - 50) - 5{,}351 \cdot 10^{-5}(H_i - 50)^2, \\ \theta_i^{(2)} &= 0{,}8654 + 5{,}360 \cdot 10^{-3}(H_i - 50) - 5{,}307 \cdot 10^{-5}(H_i - 50)^2. \end{aligned} \tag{11}$$

Откуда видно, что отличие варианта «2» от варианта «0» проявляется только в четвертой значащей цифре. Обработка серии 10-ти статистических испытаний показала, что среднее значение максимального по модулю отклонения варианта «1» от эталонного составляет 9.16% при дисперсии 0.569. Что подтверждает полученные ранее теоретические оценки. В тоже время для варианта «2» среднее значение максимального отклонения не превышает 0.3%.

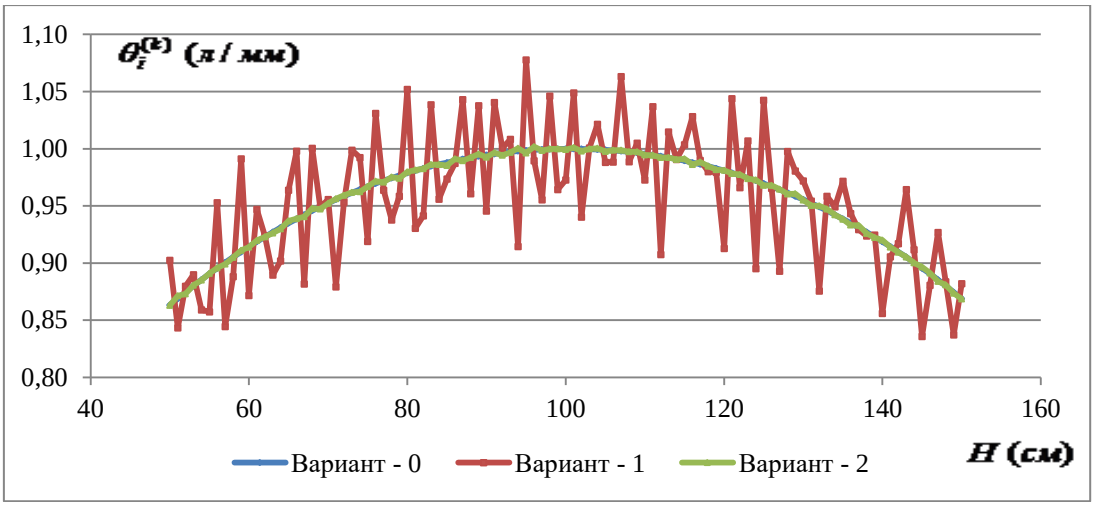


Рис.2. Значение коэффициента вместимости

Одним из основных способов уменьшения «случайной» погрешности является многократное повторение испытаний. Действительно, если провести измерения необходимые для построения градуировочной таблицы многократно (Nz>1 раз), то можно пересчитать значения соответствующих объёмов:

$$\tilde{V}_i^{(*)} = \frac{1}{N_s}\sum_{j=1}^{N_s} V\left(\tilde{H}_{i,j}\right) = V(H_i) + \Delta H_i^{(*)}\frac{\partial V(H_i)}{\partial H}, \quad i = i_0, \dots, i_n. \tag{12}$$

При этом ошибка, связанная с измерением уровней H частично нивелируется, так как её математическое ожидание равно 0:

$$\Delta H_i^{(*)} = \frac{1}{N_s}\left(\sum_{j=1}^{N_s}\Delta H_{i,j}\right), \quad \Delta H_i^{(*)} \xrightarrow[N_z\to\infty]{} 0, \quad i = i_0, \dots, i_n. \tag{13}$$

Помимо усреднения можно применять сглаживание, при этом воспользуемся процедурой, так называемого диффузионного сглаживания. Введем обозначение $V_i^{(0)} = \tilde{V}\left(\tilde{H}_i\right), \quad i = i_0, \dots, i_n$. Тогда процедура сглаживания выглядит следующим образом:

$$V_i^{(m+1)} = V_i^{(m)} + d\left(V_{i+1}^{(m)} - 2V_i^{(m)} + V_{i-1}^{(m)}\right), \quad i = i_0 - i_k, \dots, i_n + i_k, \quad m = 0,1,\dots \tag{14}$$

Итерационная процедура (14) является аналогом явной конечно-разностной схемы для уравнения диффузии, условие устойчивости которой d *<1/2*. Дополнительные узлы i_k позволяют частично нивелировать влияния граничных условий для разностного уравнения (13). Расчёты проводились со значением i_k*=10*. Итерационный процесс (14) продолжается при выполнении условия

$$\max_{i=i_0,\dots,i_n}\left|\frac{V_i^{(m)} - \tilde{V}\left(\tilde{H}_i\right)}{\tilde{V}\left(\tilde{H}_i\right)}\right| < 2\delta V,$$

после нарушении которого вычисления завершаются.

Результаты имитационного эксперимента, проведённого по описанной выше схеме, представлены в таблице 1, где используются следующие обозначения:

$$\Delta H^{(*)}(Nz) = \max_{i=50,\dots,150}\left|\Delta H_i^{(*)}\right|, \quad \delta V^{(*)}(Nz) = \max_{i=50,\dots,150}\left|\frac{V_i^{(*)}}{V(H_i)} - 1\right| \cdot 100\%,$$
$$M(Nz) = \max_{i=50,\dots,150}\left|\frac{\left(\tilde{V}_i^{(*)} - \tilde{V}_{i-1}^{(*)}\right) - \left(V(H_i) - V(H_{i-1})\right)}{V(H_i) - V(H_{i-1})}\right| \cdot 100\%. \tag{15}$$

Там же под обозначением 1* представлены результаты, полученные после сглаживания.

Таблица 1. Результаты имитационного эксперимента

Nz	ΔH(*)(Nz) см	δV(*)(Nz)%	M(Nz)%
1	0,0496 / 0,0004	0,1259 / 0,0115	9,16 / 0,569
1*	0,0496 / 0,0004	0,2349 / 0,0342	0,365 / 0,046
3	0,0407 / 0,0039	0,0751 / 0,0162	6,48 / 0,303
10	0,0252 / 0,0046	0,0444 / 0,0101	3,35 / 0,440
20	0,0170 / 0,0027	0,0305 / 0,0088	2,41 / 0,390
50	0,0113 / 0,0020	0,0153 / 0,0022	1,69 / 0,402
100	0,0073 / 0,0012	0,0130 / 0,0046	1,13 / 0,183

Отметим, что в таблице 1, помимо основного среднего значения через знак «/» указано стандартное отклонение.

4. Заключение

В ходе настоящего исследования была проведена оценка методов градуировки по ГОСТ 8.346-2000 [1] с помощью разработанной имитационной модели. Можно отметить, что многократное повторение процедуры градуировки уменьшает значение погрешности, но незначительно. Так при 100 испытаний, погрешность коэффициента вместимости уменьшается с 9% до 1%. В то же время процедура сглаживания экспериментальных данных при однократных измерениях, «ухудшая» погрешность вычисления функции *V(H)* в то же время приводит к снижению погрешности коэффициента вместимости до приемлемых уровней. Ниже приведено регрессионное уравнение для сглаженных данных

$$\theta_i^{(3)} = 0{,}8631 + 5{,}389 \cdot 10^{-3}(H_i - 50) - 5{,}324 \cdot 10^{-5}(H_i - 50)^2.$$

Таким образом, процедура построения градуировочной таблицы объёмным методом, даёт значительные погрешности при вычислении коэффициента вместимости, которые можно частично нивелировать использование сглаживания экспериментальных данных.

Литература

1. Danson C.N. et al. Petawatt and exawatt class lasers worldwide // High Power Laser Science and Engineering. – 2019. – Т. 7. – С. 54.
1. ГОСТ 8.346-2000. Резервуары стальные горизонтальные цилиндрические. Методика поверки. – Минск: ИПК Издательство стандартов, 2001. – 21 с.
2. Градуировка резервуаров: методы, ГОСТ, этапы, стоимость в 2024 году // СМС. URL: https://cmc3d.ru/articles/graduirovka-rezervuarov-s/ (дата обращения: 03.09.2025).
3. Градуировка резервуаров для нефтепродуктов на АЗС // АЗСК. URL: https://www.azsk74.ru/news/graduirovka-rezervuarov-dlya-nefteproduktov-na-azs/ (дата обращения: 03.09.2025).
4. Диагностика и вывод резервуара из эксплуатации // СНМаш. URL: https://snmash.ru/articles/192-diagnostika-i-vyvod-rezervuara-iz-ekspluatatsii.html (дата обращения: 03.09.2025).
5. Муленко И.Г., Пучка О.В., Рябко В.В., Пучка Е.О. К вопросу о точности методик измерения вместимости резервуаров // Компетентность / Competency (Russia). – 2022. – № 5. – С. 28–36. DOI: 10.24412/1993-8780-2022-5-28-36.
6. Демидович Б.П., Марон И.А. Основы вычислительной математики. – 7-е изд. – М.: Лань. 2009. – С. 17–52.

ЗОНЫ ИММУНИТЕТА МАЛОГО БИЗНЕСА В УСЛОВИЯХ ВНЕШНИХ ВОЗДЕЙСТВИЙ

В.П. Савельев, Н.И. Сутягина

Нижегородский государственный инженерно-экономический университет

В работе исследованы модели функционирования предприятий малого бизнеса по предоставлению товаров и услуг населению региона при наличии в регионе крупных предприятий, продукция и услуги которых реализуются, в основном, за пределами данного региона. Исследовано поведение соответствующих динамических систем в условиях внешних воздействий. В фазовом пространстве построены «зоны иммунитета», из которых траектории соответствующих динамических систем выйти не могут.

Ключевые слова: динамическая система, зона иммунитета, область достижимости, положительная полутраектория, предприятия малого бизнеса, состояние равновесия, фазовый портрет.

1. Введение

Ранее построены модели, показывающие, что появление в сельском регионе предприятий, продукция (услуги) которых востребованы за пределами региона, может привести к устойчивому функционированию предприятий малого бизнеса, оказывающих услуги жителям региона [1].

В настоящей работе исследуется вопрос о том, как могут повлиять на динамику малого бизнеса различные непредвиденные внешние воздействия. В фазовом пространстве построены такие области, «зоны иммунитета», которые динамическая система не покидает, несмотря на внешние воздействия. Понятие зоны иммунитета было использовано в работе при исследовании влияния приема лекарств на течение болезни [2].

2. Линейная модель

Будем считать, что трудоспособное население сельского региона состоит из двух частей. Соответственно, под состоянием системы будем понимать двумерную функцию времени $\big(x(t), y(t)\big)$, где $x(t)-$ объем предоставляемых услуг (в денежном выражении) предприятиями типа X, (малый бизнес) а $y(t)-$ объем произведенных товаров и предоставленных услуг предприятиями типа Y, продукция и услуги которых реализуются, в основном, за пределами данного региона.

Соответствующая система линейных дифференциальных уравнений, позволяющая анализировать сферу услуг жителям сельского региона, имеет вид

$$\begin{cases} \dot{x} = -ax + by, \\ \dot{y} = R + dx - cy, \end{cases} \tag{1}$$

где $ax-$ оплата труда работников внутреннего сервиса за вычетом их затрат на свои услуги, $by-$ доход системы внутреннего сервиса от работников предприятий типа Y, $dx-$ доход предприятий типа Y от работников внутреннего сервиса, $cy-$ оплата труда работников предприятий типа Y за вычетом их затрат на свой продукт, $R-$ доход предприятий типа Y.

При естественном предположении $ac - bd > 0$ состояние равновесия $M\left(\frac{bR}{ac-bd}, \frac{aR}{ac-bd}\right)$ системы (1) является устойчивым узлом.

Заметим, что состояние равновесия (x^*, y^*) расположено на прямой $y = \frac{a}{b}x$. Фазовый портрет системы (1) представлен на рисунке 1.

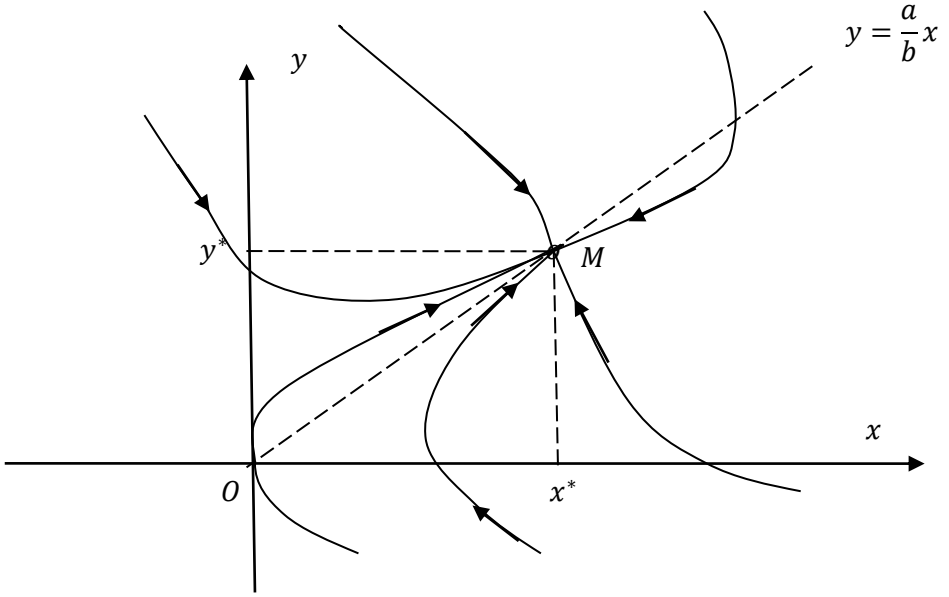


Рис. 1. Фазовый портрет линейной системы (1)

Поведение фазовых траекторий на прямой $y = \frac{a}{b}x$ определяется вектором $(\dot{x}, \dot{y}) = \left(0, R - x\frac{ac-bd}{b}\right)$. Это значит, что фазовые траектории системы пересекают прямую $y = \frac{a}{b}x$ сверху вниз при $x > x^*$, и снизу вверх при $x < x^*$.

Теперь предположим, что динамика малого бизнеса усложняется вследствие внешних непредсказуемых воздействий, и в результате этого описывается неавтономной системой уравнений

$$\begin{cases} \dot{x} = -ax + by, \\ \dot{y} = u(t) + dx - cy, \end{cases} \tag{2}$$

где $u(t) \in [0, R]$.

Заметим, что при $u(t) \equiv 0$ начало координат будет устойчивым узлом системы

$$\begin{cases} \dot{x} = -ax + by, \\ \dot{y} = dx - cy, \end{cases} \tag{3}$$

Фазовый портрет системы (3) таков же, как фазовый портрет системы (1), при этом фазовые траектории системы (3) пересекают прямую $y = \frac{a}{b}x$ сверху вниз при $x > 0$, и снизу вверх при $x < 0$.

Это означает, что из начала координат выходит положительная полутраектория $\Gamma_R^+(O)$ системы (1), пересекающая прямую $y = \frac{a}{b}x$ снизу вверх, и имеющая предельным множеством точку $M\left(\frac{bR}{ac-bd}, \frac{aR}{ac-bd}\right)$.

В то же время, из точки $M\left(\frac{bR}{ac-bd}, \frac{aR}{ac-bd}\right)$ выходит положительная полутраектория $\Gamma_o^+(M)$ системы (3), пересекающая прямую $y = \frac{a}{b}x$ сверху вниз, и имеющая предельным множеством начало координат. В результате получена область G, ограниченная положительными полутраекториями $\Gamma_o^+(M)$ и $\Gamma_R^+(O)$.

Если теперь рассматривать систему (2), как управляемую динамическую систему [2] с ограниченным управлением $0 \le u(t) \le R$, то область G есть не что иное, как область достижимости системы (2) из начала координат. Это означает, что любая допустимая траектория системы (2) может лишь войти в область G, но не может из нее выйти. Таким образом, никакие внешние воздействия (при условии $0 \le u(t) \le R$) не смогут вывести систему из области G, что позволяет трактовать эту область, как область иммунитета малого бизнеса к внешним воздействиям.

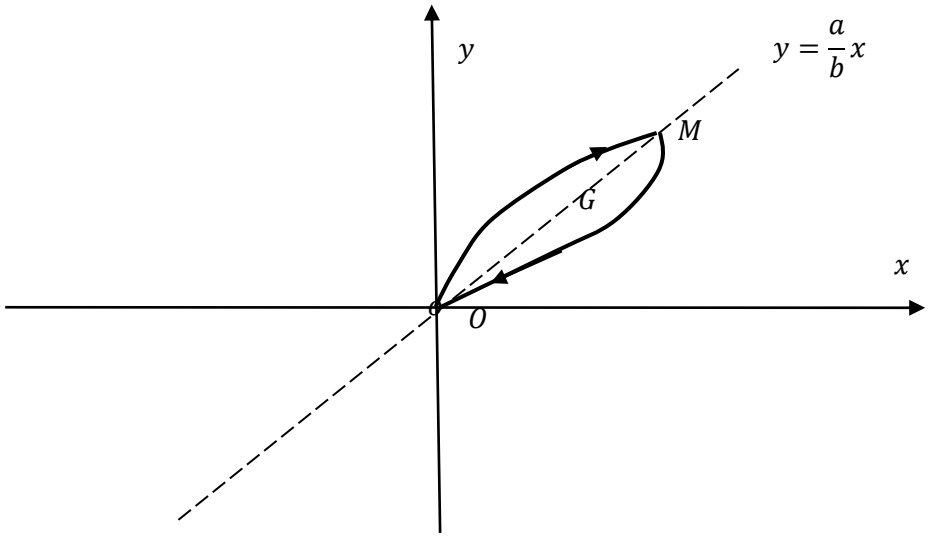


Рис. 2. Область достижимости системы (2) из начала координат

3. Нелинейная модель

Нелинейная модель регионального бизнеса [1] представляет собой систему нелинейных дифференциальных уравнений

$$\begin{cases} \dot{x} = -ax + bxy, \\ \dot{y} = R + dx - cy - bxy, \end{cases} \quad (4)$$

где $ax -$ оплата труда работников внутреннего сервиса за вычетом их затрат на свой продукт, $bxy -$ доход системы предприятий типа X от работников предприятий типа Y, $R -$ доход предприятий типа Y, $dx -$ доход предприятий типа Y от работников предприятий типа X, $cy -$ оплата труда работников предприятий типа Y за вычетом их затрат на свой продукт и затрат на внутренний сервис. Динамическая система (4) имеет два состояния равновесия: $P(0, \frac{R}{c})$ и $Q\left(\frac{Rb-ac}{b(a-d)}, \frac{a}{b}\right)$.

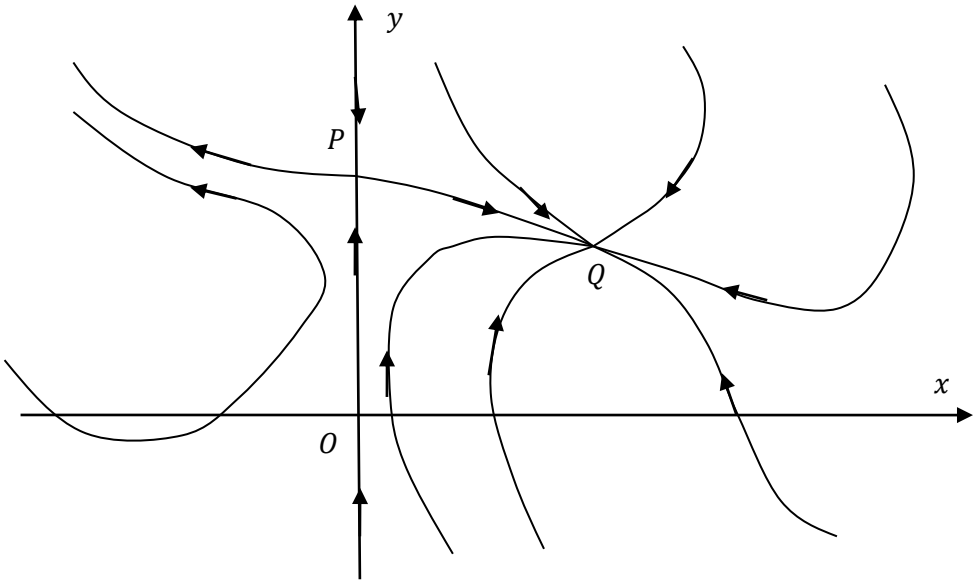


Рис. 3. Фазовый портрет системы (4) с устойчивым узлом

В предположении, что $Rb - ac > 0$, состояние равновесия $P(0, \frac{R}{c})$ является седлом, а состояние равновесия $Q\left(\frac{Rb-ac}{b(a-d)}, \frac{a}{b}\right)$ является устойчивым узлом или устойчивым фокусом. Соответствующий устойчивому узлу фазовый портрет системы (4) представлен на рисунке 3.

Теперь предположим, что динамика малого бизнеса усложняется вследствие внешних непредсказуемых воздействий, и в результате этого описывается неавтономной системой уравнений

$$\begin{cases} \dot{x} = -ax + bxy, \\ \dot{y} = u(t) + dx - cy - bxy, \end{cases} \tag{5}$$

где $u(t) \in [0, R]$.

Заметим, что при $u(t) \equiv 0$ начало координат будет устойчивым узлом системы

$$\begin{cases} \dot{x} = -ax + bxy, \\ \dot{y} = dx - cy - bxy, \end{cases} \tag{6}$$

а состояние равновесия $M(-\frac{ac}{b(a-d)}, \frac{a}{b})$ будет седлом.

Фазовый портрет системы (6) представлен на рисунке 4.

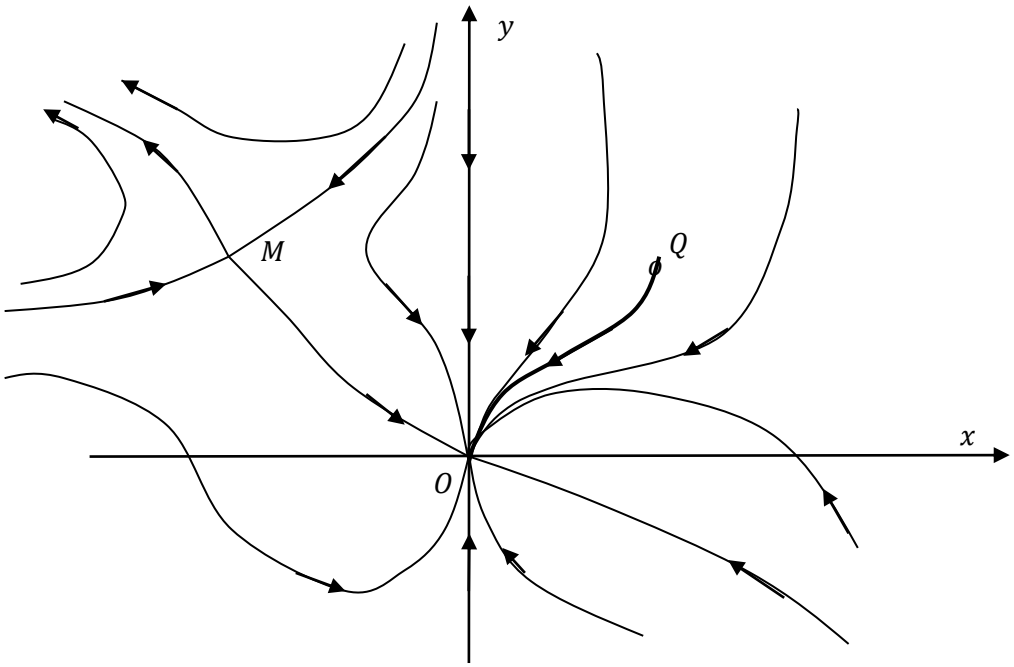


Рис. 4. Фазовый портрет системы (6)

Поведение траекторий на оси Ox определяется вектором $(\dot{x}, \dot{y}) = (-ax, dx)$. Это значит, что траектории пересекают ось Ox снизу вверх при $x > 0$ и сверху вниз при $x < 0$.

Заметим, что из точки $Q\left(\frac{Rb-ac}{b(a-d)}, \frac{a}{b}\right)$ выходит положительная траектория $\Gamma_o^+(Q)$ системы (6), имеющая предельным множеством начало координат. Если из начала координат выпустить положительную траекторию $\Gamma_R^+(O)$ системы (4), то ее предельным множеством будет седловая точка $P(0, \frac{R}{c})$, из которой, в свою очередь, выходит α-сепаратриса Γ^+, ведущая в устойчивый узел $Q\left(\frac{Rb-ac}{b(a-d)}, \frac{a}{b}\right)$ системы (4).

Если теперь рассматривать систему (5), как управляемую динамическую систему с ограниченным управлением $0 \le u(t) \le R$, то область D есть не что иное, как область достижимости системы (5) из начала координат.

Это означает, что любая допустимая траектория системы (5) может лишь войти в область G, но не может из нее выйти. Таким образом, никакие внешние воздействия (при условии $0 \le u(t) \le R$) не смогут вывести систему из области D, что позволяет трактовать эту область, как область иммунитета малого бизнеса к внешним воздействиям.

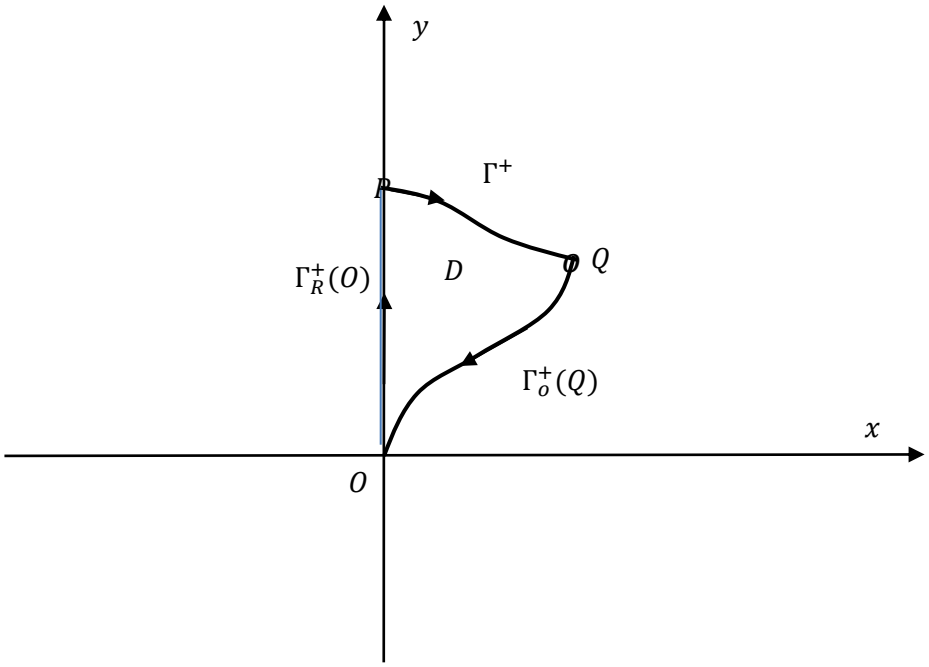


Рис. 5. Зона иммунитета системы (4)

Литература

1. Савельев В.П., Сутягина Н.И. «Простые модели динамики малого бизнеса», Труды XX Международной конференции «Математическое моделирование и суперкомпьютерные технологии», г. Н. Новгород, ноябрь 2020 г., С.329 – 332.
2. Бутенина Н.Н., Стародубровская Н.С. Области управляемости и зоны иммунитета в мате-матической модели сахарного диабета //VI Международная конференция «Мате-матика. Компьютер. Образование». Тезисы докладов. Пущино. 1999. С.48.

НЕКОТОРЫЕ ЭНДОМОРФИЗМЫ АБЕЛЕВЫХ ГРУПП

А. Сарвари

Московский педагогический государственный университет

Ранее нами описаны конечно порожденные абелевы группы, допускающие эндоморфизм, образ которого совпадает с его ядром [1]. Здесь решается аналогичная задача в классах коциклических, делимых (инъективных) и копериодических абелевых групп.

Ключевые слова: абелева группа, ядро эндоморфизма, образ эндоморфизма.

1. Введение

Под «группой» в работе всегда подразумевается абелева группа, записанная аддитивно. Все основные обозначения и определения стандартны и соответствуют книге Л. Фукса [2].

Изучение эндоморфизмов абелевых групп представляет для нас интерес, прежде всего, в связи с тем, что позволяет получить дополнительную информацию о самих абелевых группах. В данной работе изучаются абелевы группы, содержащие хотя бы один эндоморфизм, ядро которого совпадает с его образом. Главной задачей работы является описание абелевых групп A, содержащих эндоморфизм $\varphi\colon A \to A$, такой что $\ker\varphi = \operatorname{Im}\varphi$, в классах коциклических, делимых и копериодических групп.

2. Основные результаты

Теорема 1. Пусть A – абелева группа, тогда эндоморфизм $\varphi : A \to A$, такой что $\ker\varphi = \operatorname{Im}\varphi$ существует тогда и только тогда, когда в группе A найдется подгруппа B, такая что $A/B \cong B$.

Доказательство. Пусть существует эндоморфизм φ группы A, такой что $\ker\varphi = \operatorname{Im}\varphi$. Тогда по основной теореме о гомоморфизме для подгруппы $B = \ker\varphi$ группы A имеет место изоморфизм $A/B \cong B$.

Обратно, пусть $A/B \cong B$ для некоторой подгруппы группы A. Тогда построим цепочку гомоморфизмов:

$$A \xrightarrow{\alpha} A/B \xrightarrow{\beta} B \xrightarrow{\gamma} A,$$

где α – естественный эпиморфизм, β – изоморфизм и γ – вложение. Отсюда, если $\varphi = \alpha \circ \beta \circ \gamma$, то

$$\ker\varphi = B = \operatorname{Im}\varphi.$$

Следствие 1. Если $A \cong B \oplus B$, то существует эндоморфизм φ группы A, такой что $\ker\varphi = \operatorname{Im}\varphi$.

Следствие 2. Пусть A – конечная группа и $\varphi\colon A \to A$ эндоморфизм, такой что $\ker\varphi = \operatorname{Im}\varphi$, тогда порядок группы A является полным квадратом.

Напомним, что группа A называется *коциклической*, если существует элемент $a \in A$, такой что всякий гомоморфизм $\varphi : B \to A$, где $a \in \operatorname{Im}\varphi$, является эпиморфизмом. Хорошо известно (см, например, [2]), что группа A является коциклической тогда и только тогда, когда $A \cong \mathbb{Z}_{p^k}$ или $A \cong \mathbb{Z}_{p^\infty}$.

Теорема 2. Коциклическая группа A содержит эндноморфизм $\varphi : A \to A$, такой что $\ker\varphi = \operatorname{Im}\varphi$, тогда и только тогда, когда A – циклическая группа, порядок которой является четной степени некоторого простого числа, то есть $A \cong \mathbb{Z}_{p^{2n}}$.

Доказательство. Если $A = \langle a\rangle$ – циклическая группа, то по следствию 2 она имеет порядок p^{2n}, то есть $A \cong \mathbb{Z}_{p^{2n}}$. Обратно, если $|A| = p^{2n}$, то для подгруппы $B = \langle p^n a\rangle$ имеем:

$$A/B \cong \mathbb{Z}_{p^n} \cong B,$$

а значит, по теореме 1 найдется эндоморфизм $\varphi\colon A \to A$, такой что $\ker\varphi = \operatorname{Im}\varphi$.

Если же $A \cong \mathbb{Z}_p\infty$, то A не содержит подгруппы, удовлетворяющей условиям теоремы 1.

Так как для абелевых групп определено умножение элементов на натуральные числа, то логично было бы определить делимость элементов абелевых групп на натуральные числа. Группа D называется *делимой*, если любой ее элемент делится на любое натуральное число.

Теорема 3. Пусть A – делимая группа, тогда эндоморфизм $\varphi : A \to A$, такой что $\ker\varphi = \mathrm{Im}\,\varphi$ существует в том и только в том случае, когда $A \cong B \oplus B$ для некоторой подгруппы $B \subseteq A$.

Доказательство. В силу рассмотренного выше примера нам достаточно доказать только прямое утверждение. Пусть φ – эндоморфизм группы A, такой что $\ker\varphi = \mathrm{Im}\,\varphi$. Тогда по теореме 1 найдется подгруппа $B \subset A$, такая что $A/B \cong B$.

Поскольку факторгруппа делимой группы также является делимой группой, то и B – делимая группа. А тогда по основному свойству делимых групп B выделяется прямым слагаемым в группе A, то есть

$A \cong B \oplus A/B \cong B \oplus B$.

Хорошо известно строение делимых абелевых групп.

Теорема 4 [2]. Если D – делимая группа, то

$$A \cong \bigoplus_m \mathbb{Q} \oplus \bigoplus_{p \in P} \left[\bigoplus_{m_p} \mathbb{Z}_{p^\infty}\right],$$

где $\mathbb{Q}$ – группа рациональных чисел, $\mathbb{Z}_{p^\infty}$ – группа Прюфера; кардинальные числа m и m_p являются инвариантами группы D.

Следствие 3. Для делимой группы A существует эндоморфизм $\varphi : A \to A$, такой что $\ker\varphi = \mathrm{Im}\,\varphi$ тогда и только тогда, когда

$$A \cong \bigoplus_m \mathbb{Q} \oplus \bigoplus_{p \in P} \left[\bigoplus_{m_p} \mathbb{Z}_{p^\infty}\right],$$

где кардинальные числа m и m_p ($p \in P$) либо бесконечные, либо четные.

Абелева группа называется *копериодической*, если всякое ее расширение при помощи группы без кручения расщепляется. Это означает, что копериодическая группа выделяется прямым слагаемым из всякой группы, содержашей ее в качестве подгруппы, факторгруппа по которой является группой без кручения.

Хорошо известно, что произвольная копериодическая группа A однозначно представляется в виде

$$A = \prod_p A_p,$$

где A_p – редуцированная копериодическая группа, являющаяся p-адическим модулем, то есть модулем над кольцом целых p-адических чисел $\widehat{\mathbb{Z}}_p$.

Рассмотрим случай, когда A_p – циклический p-адический модуль, то есть $G_p = \mathbb{Z}_{p^{k_p}}$ или $G_p = \widehat{\mathbb{Z}}_p$. Введем следующее обозначение: *характеристикой* $\chi = (k_p)$ называется произвольная последовательность целых неотрицательных чисел и символов ∞, занумерованных простыми индексами. Построим группу $\mathbb{Z}_\chi = \prod_{p\in P} K_p$, где $K_p = \mathbb{Z}/p^{k_p}\mathbb{Z}$ при $k_p \neq \infty$ и $K_p = \widehat{\mathbb{Z}}_p$ при $k_p = \infty$. Отметим, что группа $\mathbb{Z}_\chi$ является циклическом модулем над кольцом целых полиадических чисел $\widehat{\mathbb{Z}} = \prod_{p\in P} \widehat{\mathbb{Z}}$.

Характеристика χ называется *кохарактеристикой* копериодической группы $\mathbb{Z}_\chi$.

Теорема 5. Группа $A = \mathbb{Z}_\chi$ содержит эндоморфизм $\varphi : A \to A$, такой что $\ker\varphi = \mathrm{Im}\,\varphi$ тогда и только тогда, когда ее кохарактеристика χ не содержит бесконечных и нечетных символов.

Доказательство. Пусть $A = \mathbb{Z}_\chi = \prod_{p\in P} K_p$ и $\varphi : A \to A$ – такой эндоморфизм, что $\ker\varphi = \mathrm{Im}\,\varphi$. Для каждого простого p рассмотрим ограничение $\varphi_p = \varphi|_{K_p}$ эндоморфизма φ на p-адическую компоненту K_p группы A. Так как группы K_p образуют жесткую систему (то есть $\mathrm{Hom}(K_p, K_q) = 0$ при $p \neq q$), то имеют место равенства

$$\mathrm{Im}\,\varphi = \prod_{p\in P} \mathrm{Im}\,\varphi_p \quad \text{и} \quad \ker\varphi = \prod_{p\in P} \ker\varphi_p,$$

что в соответствии с теоремой 1 и влечет справедливость данной теоремы.

Литература

1. Саврари А. Эндоморфизмы специального вида конечно порожденных абелевых групп // Чебышевский сборник: ФГБОУ ВО «ТГПУ им. Л. Н. Толстого», 2025. 240 - 249 с.
2. Фукс Л. Бесконечные абелевы группы. М.: Мир. Т. 1, 1974. 335 с.
3. Фукс Л. Бесконечные абелевы группы. М.: Мир. Т. 2, 1977. 416 с.
4. Крылов П.А., Михалев А.В., Туганбаев А.А. Абелевы группы и их кольца эндоморфизмов. М.: Факториал Пресс, 2006. 512 с.
5. Arnold D.M. Finite Rank Torsion Free Abelian Groups and Rings. LNM, 1982. 198 p.
6. Саврари А. Эндоморфизмы специального вида конечно порожденных абелевых групп // Алгебра, теория чисел, дискретная геометрия и многомасштабное моделирование: Современные проблемы, приложения и проблемы истории: Материалы XXIII Международной конференции, посвященной 80-летию профессора Александра Ивановича Галочкина и 75-летию профессора Владимира Григорьевича Чирского. Тула: Тул. гос. пед. ун-т им. Л. Н. Толстого, 2024. 67 с.
7. Саврари А. Примарные абелевы группы с нильпотентными эндоморфизмами индекса нильпотентности 2 // НАН Беларуси: Материалы XIV Беларуской Международной Математической Конференции, посвященной 65-летию Института математики. Минск: Нац. акад. нау. Беларуси. 2024. 56 - 57 с.
8. Крылов П.А., Туганбаев А.А., Царев А.В. sp-Группы и их кольца эндоморфизмов. Итоги науки и техн: Соврем. мат. и ее прил. Темат. обз. 2019. 68 с.

АЛГОРИТМ СПИСОЧНОГО ТИПА ДЛЯ РАЗМЕЩЕНИЯ ВИРТУАЛЬНЫХ МАШИН НА СЕРВЕРА С УЧЕТОМ NUMA-АРХИТЕКТУРЫ[1*]

М.Ю. Сахно

Институт математики им. С.Л. Соболева СО РАН, Омский филиал

В работе предлагается алгоритм списочного типа для размещения работ (виртуальных машин) на машинах (серверах) с NUMA-архитектурой. Алгоритм основан на быстрых онлайн-стратегиях типа «Fit». Рассмотрены различные стратегии упорядочения работ и выбора машины (First Fit, Random Fit и др.) и проанализирована их эффективность на основе реального набора данных, предоставленного компанией Хуавей. Результаты экспериментов показывают, что стратегия First Fit стабильно превосходит другие по качеству решений и времени выполнения. Также результаты показывают, что включение NUMA-ограничений непосредственно в процесс размещения существенно повышает эффективность распределения ресурсов.

Ключевые слова: построение расписаний, NUMA-архитектура, списочные алгоритмы, выделение ресурсов

1. Введение и постановка задачи

Рассматривается задача назначения работ на машины с NUMA-архитектурой (англ. Non-Uniform Memory Access [1]) при наличии нескольких ресурсов возобновимого типа. Такая задача соответствует задаче распределения виртуальных машин между серверами с NUMA-архитектурой. Ресурсы, которыми располагают сервера, представлены, как правило, ядрами процессора и объемом оперативной памяти [2, 3]. Эффективность алгоритмов размещения виртуальных машин напрямую определяет расходы на приобретение и эксплуатацию серверов, поэтому ключевой задачей является сокращение числа серверов, находящихся в активном состоянии (то есть размещающих хотя бы одну виртуальную машину).

Задача назначения работ на машины с NUMA-архитектурой при наличии нескольких ресурсов возобновимого типа представляет собой обобщение темпоральной векторной задачи об упаковке в контейнеры (англ. temporal vector bin packing) [4, 5]. Для случая с двумя типами ресурсов были предложены специализированные решения, такие как метод генерации столбцов [6] и генетический алгоритм [7], показавшие высокую эффективность на практических данных.

Рассмотрим постановку задачи. Имеется множество работ J, m машин с NUMA-архитектурой и R ресурсов возобновимого типа. Для каждой работы j известны моменты её начала s_j и окончания f_j, а также потребность a_{jr} в ресурсе типа r. Каждая машина состоит из N NUMA-узлов, при этом на каждом узле доступен объём c_r ресурса типа r. Работы размещаются на машинах, прерывания недопустимы. Запуск работы требует выделения соответствующего набора ресурсов машины, причём суммарный объём потребностей всех работ, выполняющихся одновременно на одной машине, не должен превышать её пропускной способности. Машина, на которой выполняется хотя бы одна работа, далее называется активной.

В зависимости от потребностей в ресурсах типы работ классифицируются как *маленькие* или *большие*. Работы маленьких типов размещаются на одном из NUMA-узлов машины, в то время как большие размещаются на нескольких NUMA-узлах, равномерно распределяя потребности между всеми NUMA-узлами. Таким образом, задача планирования выполнения работ на машинах с NUMA-архитектурой заключается в оптимизации назначения работ машинам таким образом, чтобы максимальное количество машин, которые одновременно задействованы в течение периода планирования, было сведено к минимуму.

[1*] Исследование выполнено за счет гранта Российского научного фонда № 22-71-10015-П, https://rscf.ru/project/22-71-10015-П/.

2. Алгоритм списочного типа LT_{fit} и вычислительный эксперимент

В работе предложен алгоритм списочного типа LT_{fit}, основанный на быстрых онлайн-алгоритмах семейства «Fit». В алгоритме LT_{fit} поочередно рассматриваются работы в выбранном порядке и каждая из них размещается на подходящую машину, обладающую достаточным объёмом ресурсов на всём интервале её активности. Если подходящая машина отсутствует, добавляется новая машина. Порядок работ и стратегия выбора машины определяют конкретную модификацию алгоритма. Работы могут упорядочиваться по времени начала (онлайн-режим) или по убыванию веса. Были исследованы различные стратегии выбора машины: First Fit, Next Fit, Random Fit и Best Fit. Адаптация алгоритма к NUMA-архитектуре может выполняться как во время его работы, так и отдельным этапом после основного распределения.

Вычислительный эксперимент проводился на открытом наборе данных, предоставленном компанией Хуавей[1]. Нижняя оценка, полученная на основе классической нижней границы Мартелло и Тота для задачи упаковки в контейнеры [8], для этого примера равна 885 серверам. Лучший результат (908 серверов) получен при использовании алгоритма LT_{fit} со стратегией First Fit, при этом вес виртуальной машины j определялся как $w_j = a_{j,cpu}\ a_{j,mem}$ и распределение по NUMA-узлам выполнялось в процессе размещения.

Литература

1. Manchanda N., Anand K. Non-uniform memory access (numa) // New York University. 2010.
2. Mann Z.A. Allocation of virtual machines in cloud data centers – a survey of problem models and optimization algorithm // Acm Computing Surveys. 2015. Vol. 48. No. 1. P. 1–34.
3. Talebian H., Gani A., Sookhak M., Abdelatif A.A., Yousafzai A., Vasilakos A.V., Yu F.R. Optimizing virtual machine placement in iaas data centers: taxonomy, review and open issues // Cluster Computing. 2020. Vol. 23. No. 2. P. 837–878.
4. Aydın N., Muter I., Ilker Birbil S. Multi-objective temporal bin packing problem: An application in cloud computing // Computers & Operations Research. 2020. Vol 121. P. 104959.
5. Cauwer M.D., Mehta D., O'Sullivan B. The temporal bin packing problem: An application to workload management in data centres // 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI). 2016. P. 157–164.
6. Ratushnyi A., Kochetov Y. A column generation based heuristic for a temporal bin packing problem // Mathematical Optimization Theory and Operations Research,. 2021. P. 96–110.
7. Sakhno M. A grouping genetic algorithm for the temporal vector bin packing problem // 19th International Asian School-Seminar on Optimization Problems of Complex Systems (OPCS). 2023. P. 94–99.
8. Martello S., Toth P. Lower bounds and reduction procedures for the bin packing problem // Discrete Applied Mathematic. 1990. Vol. 28. P. 59–70.

[1] https://vmagent.readthedocs.io/en/latest/simulator/dataset.html

О КОЛИЧЕСТВЕ БАЗИСОВ В КЛАССЕ БУЛЕВЫХ ФУНКЦИЙ

С.В. Сидоров, Т.Г. Смирнова

Нижегородский государственный университет им. Н. И. Лобачевского

Рассматривается вопрос о количестве базисов, состоящих из k логических функций ($k = 1,2,3,4$) от n переменных, в классе всех булевых функций P_2. Для $k = 1$ это известный результат о числе Шефферовых n-местных функций. Получены формулы для числа базисов мощностей $2, 3, 4$, описаны все таблицы Поста этих базисов и установлено, что для $k = 2, 3, 4$ существует соответственно $17, 22, 2$ типов таблиц Поста.

Ключевые слова: логическая функция, булева функция, полная система, базис, критерий полноты, Шефферова функция, замкнутый класс.

1. Введение

Приведём определения основных понятий и обозначений теории логических функций (см., например, [1–4]), которые будут использоваться в настоящей работе.

Функция $f(x_1, x_2, \dots, x_n)$, у которой каждая переменная принимает значения из множества $\{0, 1\}$ и сама функция принимает значения в этом множестве, называется *логической* функцией (или *булевой* функцией, или функцией *алгебры логики*). Областью определения такой функции является множество $\{0,1\}^n$ двоичных наборов длины n, а областью значений – множество $\{0, 1\}$, т. е. $f: \{0,1\}^n \to \{0,1\}$.

Множество всех логических функций $f(x_1, x_2, \dots, x_n)$, зависящих от n переменных, обозначим через $P_2(n)$. Ясно, что $|P_2(n)| = 2^{2^n}$. Заметим, что в это число включены и функции, зависящие от некоторых из переменных $x_1, x_2, \dots, x_n$ фиктивно[1], т.е. функции, фактически зависящие от меньшего числа переменных.

Система логических функций называется *полной*, если с помощью суперпозиций[2] функций этой системы может быть выражена любая логическая функция.

Пусть F – некоторое множество логических функций, тогда через $[F]$ обозначается замыкание системы F, т. е. множество всех суперпозиций функций из F. Если $F = [F]$, то F называется *замкнутым классом*.

Перечислим замкнутые классы, которые играют важную роль в теории логических функций.

T_0 – класс функций, сохраняющих константу 0, т.е. таких, что $f(0,0, \dots ,0) = 0$.

T_1 – класс функций, сохраняющих константу 1, т.е. таких, что $f(1,1, \dots ,1) = 1$.

S – класс самодвойственных функций, т.е. таких, что $f(x_1, x_2, \dots, x_n) = \overline{f}(\overline{x}_1, \overline{x}_2, \dots, \overline{x}_n)$. Самодвойственная функция на противоположных наборах принимает противоположные значения.

M – класс монотонных функций, т.е. таких, что $f(\alpha_1, \alpha_2, \dots, \alpha_n) \le f(\beta_1, \beta_2, \dots, \beta_n)$ для любых наборов $(\alpha_1, \alpha_2, \dots, \alpha_n)$ и $(\beta_1, \beta_2, \dots, \beta_n)$, для которых $\alpha_i \le \beta_i$ для всех $i = 1,2, \dots, n$.

L – класс линейных функций, т.е. таких, у которых полином Жегалкина не содержит произведений переменных.

Как известно [1–6], критерий полноты системы функций даёт следующая

---

[1] Функция зависит от переменной x_i *фиктивно*, если на всех парах наборов, различающихся лишь в i-й позиции, она принимает равные значения.

[2] *Суперпозицией* функций $f_1, \dots, f_m$ называется функция f, полученная с помощью подстановок этих функций друг в друга и переименования переменных.

Теорема 1. (*Критерий Поста*) *Для того чтобы система логических функций была полной, необходимо и достаточно, чтобы она целиком не содержалась ни в одном из пяти замкнутых классов* T_0, T_1, S, M *и* L.

Полная система называется *базисом*, если никакая её подсистема не является полной системой.

Логическая функция, не принадлежащая ни одному из пяти замкнутых классов T_0, T_1, S, M и L, называется *Шефферовой*. Любая Шефферова функция по Теореме 1 образует базис мощности 1.

Теорема 2. *Максимальное число функций в базисе равно четырём.*

В связи с понятием базиса возникают вопросы о числе различных базисов в P_2 данной мощности $k = 1,2,3,4$. Через $N_k(n)$ будем обозначать *число базисов мощности* k, содержащих функции от n переменных. Тогда $N_1(n)$ – число Шефферовых функций от n переменных. Известно [6], что $N_1(n) = 2^{2^n-2} - 2^{2^{n-1}-1}$.

2. Основные результаты

В зависимости от принадлежности функции замкнутым классам T_0, T_1, S, M и L, множество булевых функций разбивается на $2^5 = 32$ класса эквивалентности, причём 17 из них оказываются пустыми. Для вычисления количества базисов $N_k(n)$ мощности k найдём мощность каждого непустого класса эквивалентности, состоящего из функций от n переменных, и сведём полученные данные в таблицу 1.

Для классов $T_0T_1SM\overline{L}$ и $T_0T_1\overline{S}M\overline{L}$ определить точное значение числа логических функций от n переменных, им принадлежащим, не представляется возможным в силу того, что в настоящее время точное число монотонных функций известно только для $0 \le n \le 9$ [8], поэтому будем считать, что $|T_0T_1SM\overline{L}|_n = a$, $|T_0T_1\overline{S}M\overline{L}|_n = b$.

Таблица 1. Мощности непустых классов в $P_2(n)$ относительно замкнутых классов T_0, T_1, S, M *и* L

№	Класс эквивалентности	Мощность класса
1	$\overline{T}_0\overline{T}_1\overline{S}\,\overline{M}\,\overline{L}$	$2^{2^n-2} - 2^{2^{n-1}-1}$
2	$\overline{T}_0\overline{T}_1S\,\overline{M}\,\overline{L}$	$2^{2^{n-1}-1} - 2^{n-1}$
3	$\overline{T}_0\overline{T}_1S\,\overline{M}\,L$	2^{n-1}
4	$\overline{T}_0T_1\overline{S}\,\overline{M}\,\overline{L}$	$2^{2^n-2} - 2^{n-1}$
5	$\overline{T}_0T_1\overline{S}\,\overline{M}\,L$	$2^{n-1} - 1$
6	$\overline{T}_0T_1\overline{S}\,M\,L$	1
7	$T_0\overline{T}_1\overline{S}\,\overline{M}\,\overline{L}$	$2^{2^n-2} - 2^{n-1}$
8	$T_0\overline{T}_1\overline{S}\,\overline{M}\,L$	$2^{n-1} - 1$
9	$T_0\overline{T}_1\overline{S}\,M\,L$	1
10	$T_0T_1\overline{S}\,\overline{M}\,\overline{L}$	$2^{2^n-2} - 2^{2^{n-1}-1} - b$
11	$T_0T_1\overline{S}\,M\,\overline{L}$	b
12	$T_0T_1S\,\overline{M}\,\overline{L}$	$2^{2^{n-1}-1} - 2^{n-1} - a$
13	$T_0T_1S\,\overline{M}\,L$	$2^{n-1} - n$
14	$T_0T_1S\,M\,\overline{L}$	a
15	$T_0T_1S\,M\,L$	n

Заметим, что класс $\overline{T}_0\overline{T}_1\overline{S}\,\overline{M}\,\overline{L}$ (первая строка таблицы 1) – это, в точности, класс Шефферовых функций. Нетрудно понять, что отношение $\Delta_1(n) = \frac{N_1(n)}{|P_2(n)|}$ количества Шефферовых функций от n переменных к числу $|P_2(n)| = 2^{2^n}$ всех функций от n переменных стремится к $\frac{1}{4}$ при $n \to \infty$.

При исследовании полноты систем функций удобно пользоваться таблицей, которая называется *критериальной* или *таблицей Поста*. Каждый из пяти столбцов таблицы соответствует одному из замкнутых классов T_0, T_1, S, M и L, а строки таблицы соответствуют функциям исследуемой системы. На пересечении строки таблицы, соответствующей функции f, и столбца, соответствующего классу K, ставится знак плюс, если $f \in K$, и минус, если $f \notin K$. Система функций полна тогда и только тогда, когда в каждом столбце содержится хотя бы один знак минус.

Установлено, что для базисов мощностей $2, 3, 4$ существует соответственно $17, 22, 2$ типов таблиц Поста [5,7].

Теорема 3. *Число базисов, состоящих из двух функций от n переменных, равно*

$$N_2(n) = 2^{2^{n+1}-4} + 3 \cdot 2^{2^n+2^{n-1}-3} - 2^{2^n-2} - 3 \cdot 4^{n-1}.$$

При $n = 1,2,3,4$ получаем следующие значения $N_2(n)$:

$$N_2(1) = 0, \quad N_2(2) = 24, \quad N_2(3) = 5\,520, \quad N_2(4) = 274\,710\,336.$$

Следствие 1. *Для числа базисов в P_2, состоящих из двух функций от n переменных, верна следующая асимптотика:*

$$N_2(n) \sim 2^{2^{n+1}-4}.$$

Следствие 2. *Пусть*

$$\Delta_2(n) = \frac{N_2(n)}{\binom{|P_2(n)|}{2}}$$

– отношение числа базисов, состоящих из двух функций от n переменных, к числу всех неупорядоченных пар n-местных булевых функций. Тогда

$$\lim_{n\to\infty} \Delta_2(n) = \frac{1}{8}.$$

Теорема 4. *Число базисов, состоящих из трех функций от n переменных, равно*

$$N_3(n) = 2^{2^n+2n-4} + 2^{2^{n-1}+2n-2} - 3 \cdot 8^{n-1} + n + 2 - |M|_n,$$

где $|M|_n$ – число монотонных функций от n переменных.

При $n = 1,2,3,4$ получаем следующие значения $N_3(n)$:

$$N_3(1) = 0, \quad N_3(2) = 6, \quad N_3(3) = 1\,073, \quad N_3(4) = 1\,063\,262.$$

Следствие 3. *Для числа базисов в P_2, состоящих из трёх функций от n переменных, верна следующая асимптотика:*

$$N_3(n) \sim 2^{2^n+2n-4}.$$

Следствие 4. *Пусть*

$$\Delta_3(n) = \frac{N_3(n)}{\binom{|P_2(n)|}{3}}$$

– отношение числа базисов, состоящих из трёх функций от n переменных, к числу всех неупорядоченных троек n-местных булевых функций. Тогда

$$\lim_{n\to\infty} \Delta_3(n) = 0.$$

Теорема 5. *Число базисов, состоящих из четырех функций от n переменных, равно*

$$N_4(n) = (2^{n-1} - n) \cdot (|M|_n - n - 2),$$

где $|M|_n$ – число монотонных функций от n переменных.

При $n = 1,2,3,4$ получаем следующие значения $N_4(n)$:

$$N_4(1) = 0, \quad N_4(2) = 0, \quad N_4(3) = 15, \quad N_3(4) = 648.$$

Следствие 5. *Пусть*

$$\Delta_4(n) = \frac{N_4(n)}{\binom{|P_2(n)|}{4}}$$

– отношение числа базисов, состоящих из четырёх функций от n переменных, к числу всех неупорядоченных четвёрок n-местных булевых функций. Тогда

$$\lim_{n\to\infty} \Delta_4(n) = 0.$$

Литература

1. Яблонский С.В. Функциональные построения в k-значной логике // Труды МИАН СССР. 1958. Т. 51. С. 5–142.
2. Яблонский С.В. Введение в дискретную математику. – М.: Наука, 1986, 384 с.
3. Алексеев В.Е. Дискретная математика. Учебное пособие. – Нижний Новгород: Изд-во ННГУ им. Н.И. Лобачевского, 2017. – 139 с. [Электронный ресурс] Режим доступа: URL: http://www.unn.ru/books/met_files/Alekseev.pdf.
4. Гаврилов Г.П., Сапоженко А.А. Задачи и упражнения по дискретной математике. – М.: Физматлит, 2006, 416 с.
5. Яблонский С.В. О суперпозициях функций алгебры логики // Математический сборник. 1952. Т. 72. № 2. С. 329–348.
6. Гиндикин С.Г. Алгебра логики в задачах. – М.: Наука, 1972, 288 с.
7. Luka Krnič. Tipy bazisov algébry logiki (Types of bases in the algebra of logic). Glasnik matematičko-fizìčki i astronomski, ser. 2 vol. 20 (1965), pp. 23–32.
8. Jäkel C. A Computation of the Ninth Dedekind Number // Journal of Computational Algebra. 2023. Vol. 6-7. 100006. DOI: https://doi.org/10.1016/j.jaca.2023.100006.

КЛАССЫ ПОДОБИЯ ЦЕЛОЧИСЛЕННЫХ НИЛЬПОТЕНТНЫХ МАТРИЦ 4-ГО ПОРЯДКА С МИНИМАЛЬНЫМ МНОГОЧЛЕНОМ СТЕПЕНИ НЕ ВЫШЕ 3[1*]

С.В. Сидоров, Г.В. Уткин

Нижегородский государственный университет им. Н.И. Лобачевского

В работе получено описание структуры классов подобия над кольцом целых чисел для целочисленных нильпотентных матриц 4 порядка с минимальным многочленом степени не выше 3. Также получено описание канонических матриц классов подобия. Кроме того, для таких матриц получен критерий подобия жордановой форме.

Ключевые слова: подобие матриц, жорданова форма, кольцо целых чисел.

1. Введение

Определение 1. *Будем говорить, что матрица $A \in \mathbb{Z}^{n\times n}$ подобна матрице $B \in \mathbb{Z}^{n\times n}$ над кольцом целых чисел $\mathbb{Z}$, если существует такая унимодулярная матрица $X \in GL(n,\mathbb{Z})$, что $AX = XB$. При этом матрицу X называют трансформирующей. Факт подобия матриц A и B над кольцом целых чисел $\mathbb{Z}$ будем обозначать $A \sim B$.*

Очевидно, что если A подобна B над кольцом $\mathbb{Z}$, то они подобны и над полем $\mathbb{Q}$. Обратное неверно (см. примеры в [1, 2, 3] и ниже в настоящей работе). Существуют и более содержательные необходимые условия подобия над $\mathbb{Z}$. В частности, если $A \sim B$, то $\Delta_k(A) = \Delta_k(B)$, $k = 1, \dots, n$, где $\Delta_k(A)$ – НОД миноров k-го порядка матрицы A (см. [1]). К сожалению, они тоже не являются достаточными.

В данной работе изучаются семейство целочисленных нильпотентных матриц с минимальным многочленом степени не выше 3 и классы подобия над $\mathbb{Z}$.

2. Описание канонических матриц

Прямой суммой квадратных матриц $X_1, X_2, \dots, X_k$ будем называть блочно-диагональную матрицу:

$$X_1 \oplus \dots \oplus X_k = \begin{pmatrix} X_1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & X_k \end{pmatrix}.$$

Известно, что над полем $\mathbb{Q}$ любая целочисленная матрица, у которой все собственные числа целые, подобна жордановой форме, т.е. прямой сумме жордановых клеток. В частности, любая нильпотентная матрица порядка 4 подобна над $\mathbb{Q}$ одной из следующих жордановых матриц:

$$J_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ J_1 \oplus J_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ J_2 \oplus J_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$J_1 \oplus J_1 \oplus J_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ J_1 \oplus J_1 \oplus J_1 \oplus J_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

[1*] Работа выполнена при финансовой поддержке научно-образовательного математического центра «Математика технологий будущего».

В работе [2] были описаны классы подобия целочисленных матриц порядка n с минимальным многочленом $(\lambda_0 - \lambda)^2$. Частным случаем этого результата для нильпотентных матриц порядка 4 с минимальным многочленом λ^2 является следующая теорема.

Теорема 1. (*см.* [2]). *Пусть минимальный многочлен матрицы* $A \in \mathbb{Z}^{4\times4}$ *равен* λ^2. *Тогда*

1) *Если* $rank(A) = 1$, *то матрица A подобна над* $\mathbb{Z}$ *единственной матрице вида*

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

где $d \geq 1$, *причём* $d = \Delta_1(A)$;

2) *Если* $rank(A) = 2$, *то матрица A подобна над* $\mathbb{Z}$ *единственной матрице вида*

$$\begin{pmatrix} 0 & d_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_2 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

где $d_1, d_2 \geq 1$, *причём* $d_1 | \, d_2,\ d_1 = \Delta_1(A),\ d_2 = \frac{\Delta_2(A)}{\Delta_1(A)}$.

Следствие 1. *(см. [2]). Пусть минимальный многочлен матрицы* $A \in \mathbb{Z}^{4\times4}$ *равен* λ^2. *Тогда*

1) $A \sim J_1 \oplus J_1 \oplus J_2 \Leftrightarrow \text{rank(A)} = 1,\ \Delta_1(A) = 1$;
2) $A \sim J_2 \oplus J_2 \Leftrightarrow \text{rank(A)} = 2,\ \Delta_2(\text{A}) = 1$.

Основным результатом настоящей заметки является следующее описание классов подобия нильпотентных матриц порядка 4, у которых минимальный многочлен равен λ^3.

Теорема 2. *Пусть минимальный многочлен матрицы* $A \in \mathbb{Z}^{4\times4}$ *равен* λ^3. *Тогда* $rank(A) = 2$ *и матрица A подобна над* $\mathbb{Z}$ *единственной матрице вида*

$$R = \begin{pmatrix} 0 & 0 & 0 & s \\ 0 & 0 & a_1 & r \\ 0 & 0 & 0 & a_2 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

где $a_1, a_2 \geq 1, 0 \leq s \leq \frac{a_2}{2}, 0 \leq r < \text{НОД}(a_1, a_2, s)$.

Следствие 2. *Пусть минимальный многочлен матрицы* $A \in \mathbb{Z}^{4\times4}$ *равен* λ^3. *Тогда*

$$A \sim J_1 \oplus J_3 \Leftrightarrow \Delta_1(A^2) = 1.$$

Литература

1. Шевченко В.Н., Сидоров С.В. О подобии матриц второго порядка над кольцом целых чисел // Известия высших учебных заведений. Математика. 2006. № 4. С. 57– 64.
2. Сидоров С.В. О подобии некоторых целочисленных матриц с единственным собственным значением над кольцом целых чисел // Матем. заметки. 2019. Т. 105, № 5. С. 763–770.
3. Husert D. Similarity of integer matrices: PhD Thesis. Paderborn, 2017. 147 p.

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ ВОЗДЕЙСТВИЯ ДИНАМИЧЕСКОЙ НАГРУЗКИ НА КОНСТРУКЦИИ ИЗ КОМПОЗИЦИОННЫХ МАТЕРИАЛОВ

И.Е. Смирнов, А.В. Васюков

Московский физико-технический институт

В данной работе рассматривается метод математического моделирования распространения волновых процессов в композиционных конструкциях. Для аппроксимации пространственных производных применятся МКЭ, интегрирование по времени осуществляется методом Ньюмарка. Модель и методы использованы для моделирования ударного воздействия на элементы конструкции планера гражданского самолёта. Получена пространственно-временная картина распространения возмущений в образцах после удара.

Ключевые слова: динамическая задача, механика деформируемого твердого тела, метод конечных элементов.

1. Введение

Проблема ударных повреждений конструкция из композиционных материалов возникает в большом количестве областей, где такие материалы используются. Одним из ярких примеров является аэрокосмическая отрасль. Во время эксплуатации летательных аппаратов возможны удары различной энергии и скорости воздействующих инородных тел. При этом наибольший интерес представляют так называемые баллистические удары с энергией от 10 Дж до 20 кДж [3], так как такие повреждения чаще всего возникают непосредственно в процессе полёта. В рамках данной работы рассматриваются низкоскоростные низкоэнергетические удары, которые неизбежны при штатной эксплуатации (град, мелкий мусор, неаккуратное обращение при техническом обслуживании). Основными особенностями повреждений при таких ударах является отсутствие явных визуальных признаков и сложность обнаружения скрытых дефектов, а также существенно снижение прочности конструкции из-за риска повреждения волокон и расслоения композитов.

Важным фактором удара помимо энергии является скорость воздействующего тела. Тела большей массы, но меньшей скорости могут осуществлять схожее энергетическое воздействие с более легкими, но в то же время и более быстрыми телами. При этом характер распространения напряжений в композиционной конструкции и как следствие вид разрушения существенно различается [2].

Предлагаемый в данной работе численный метод, описывающий распространение волн в твёрдых изотропных средах, может использоваться для прогнозирования повреждений и проектирования устойчивых к ударам аэрокосмических композиционных конструкций.

2. Численный метод

Описание волновых процессов при малых упругих деформациях в приближении линейной упругости материала представляется системой уравнений равновесия и реологическим соотношением [1, 7, 9]:

$$\rho \cdot \ddot{u}_i = \nabla_j \cdot \sigma_{ij} + f_i,$$
$$\sigma_{ij} = C_{ij,kl} \cdot (\nabla_k u_l + \nabla_l u_k)$$

Здесь ρ – плотность сплошной среды, u_i – компоненты деформаций, σ_{ij} – компоненты тензора напряжений, $C_{ij,kl}$ – компоненты тензора малых деформаций, f_i – компоненты массовых сил на единицу объёма.

Представленная система уравнений с вектором искомых переменных $u = \{u_1, u_2, u_3\}$ рассматривается в некоторой пространственной области $\Omega \in R^3$ с необходимыми граничными условиями и во временной области $[0, T]$ с некоторыми начальными условиями. Для её решения можно использовать метод конечных элементов [10] для аппроксимации пространственных производных и метод Ньюмарка [6, 8] для интегрирования по времени.

Основная идея решения заключается в разбиении пространственной области на элементы и построении конечно-элементного пространства базисных пространственных функций известного вида. Каждая искомая переменная представляется линейной комбинацией базисных функций $(u_h)_i(t, x, y, z) = s(t)_{i,j} \cdot \varphi_j(x, y, z)$. Процесс отыскания решения исходной сводится к нахождению коэффициентов разложения $s(t)_{i,j}$ по базисным функциям. Таким образом, новый вектор искомых величин принимает вид $s = \{s_1 j, s_2 j, s_3 j\}$, и имеет размерность $3 \cdot J$, где J – количество базисных пространственных функций. Разбиение на конечные элементы – тетраэдры может производится при помощи специального программного модуля построения сеток [4].

Для решения методом конечных элементов, необходимо сформулировать слабую постановку задачи. Такая постановка для представленной исходной задачи имеет вид :

$$\rho \iiint\limits_{\Omega} \ddot{u}_h \cdot v_h + \iiint\limits_{\Omega} \varepsilon(v_h) : C : \varepsilon(u_h) - \iint\limits_{\partial\,\Omega} \sigma \cdot v_h = \iiint\limits_{\Omega} f_h \cdot v_h$$

Здесь $\varepsilon(u)_{lk} = (\nabla_k u_l + \nabla_l u_k)$ - компоненты тензора деформаций, v_h –тестовая функция из построенного конечно-элементного пространства. Приведённое уравнение должно выполняться для любой тестовой функции. f_h - разложение внешних сил по базисным функциям. Так как функции u_h, v_h состоят из линейной комбинации известных пространственных базисных функций, обозначенные пространственные интегралы можно найти [5] и перейти к системе ОДУ по времени относительно искомых коэффициентов s :

$$M \cdot \ddot{s} + K \cdot s = b$$

Полученную систему ОДУ по времени можно решать с помощью метода Ньюмарка [6, 8]:

$$s_{n+1} - s_n = \left(\frac{4}{\tau^2} M + K\right)^{-1} \cdot \left(b_{n+1} - b_n + \frac{4}{\tau} M \cdot \dot{s}_n + 2M \cdot \ddot{s}_n\right)$$

$$\dot{s}_{n+1} - \dot{s}_n = \frac{2}{\tau}(s_{n+1} - s_n) - 2\dot{s}_n$$

$$\ddot{s}_{n\,+1} = M^{-1} \cdot (b - K \cdot s_{n+1})$$

3. Результаты

Описанные модель и метод были применены для моделирования ударного воздействия на элементы конструкции планера гражданского самолёта. Были рассмотрены панель крыла и образец панели с частью стрингера. Была получена пространственно-временная картина распространения возмущений в образцах после удара.

Литература

1. Bedford, A. & Drumheller, D.S., 1994. Elastic Wave Propagation, John Wiley & Sons, Chichester.
2. FÄLLSTRÖM, K.E, LINDGREN, L.E., MOLIN, N.E. and WÅHLIN, A. Transient bending waves in anisotropic plates studied by hologram interferometry, J Exp Mech, 1989, 29, (4), pp 409-413.
3. G.A.O. Davies, R. Olsson. Impact on composite structures // The Aeronautical Journal, Vol. 108, Is. 1089, pp. 541-563.
4. Geuzaine C., Remacle J.-F. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities // International Journal for Numerical Methods in Engineering, 2009, vol. 79, issue 11, pp. 1309-1331.
5. HECHT, Frédéric. New development in FreeFem++. Journal of numerical mathematics, 2012, vol. 20, no 3-4, p. 251-266.

6. Hughes, T.J.R., 1987. The Finite Element Method—Linear Static and Dynamic Finite Element Analysis, Prentice Hall, Englewood Cliffs, NJ.
7. V. Novatskii, Elasticity Theory (Mir, Moscow, 1975) [in Russian].
8. Newmark N.M. A Method of Computation for Structural Dynamics // Journal of the Engineering Mechanics Division, 1959, vol. 85, issue 3, pp. 67-94.
9. L. I. Sedov, L. I. Sedov, A Course in Continuum Mechanics, Vol. 1 (Nauka, Moscow, 1970; WoltersNoordhoff, Groningen, 1971).
10. Zienkiewicz O.C. and Taylor R.L. Finite Element Method: Volume 1 - The Basis //5th Oxford: Butterworth-Heinemann, 2000.

РЕДУКЦИЯ ВЫСШЕГО ПОРЯДКА МОДЕЛИ КУРАМОТО С ЗАДЕРЖКОЙ ПО ВРЕМЕНИ[1*]

Л.А. Смирнов[1], В.О. Муняев[1], М.И. Болотов[1], И.В. Белых[2]

[1]Нижегородский государственный университет им. Н.И. Лобачевского,
[2]Университет штата Джорджия, США

В данной работе представлен новый метод редукции, преобразующий сети осцилляторов первого порядка с задержкой по времени в системы второго порядка без задержки, где параметр задержки проявляется как эффективная инерция. Полученная модель второго порядка демонстрирует отличное соответствие с исходной системой, точно воспроизводя статистические свойства решений, границы устойчивости и сложную динамику, такую как кластерные циклопные состояния. Данный подход предоставляет мощный и вычислительно эффективный инструмент для анализа сложной динамики сетей с запаздывающими взаимодействиями.

Ключевые слова: задержка по времени; фазовый осциллятор; модель Курамото; инерция; редукция.

1. Введение

Анализ сетей осцилляторов с задержками по времени крайне важен для понимания поведения реальных систем в биологии, нейронауках и технике. Однако бесконечномерная природа дифференциальных уравнений с запаздыванием делает их исследование аналитически сложным. Мы представляем новый метод редукции, который преобразует систему первого порядка с временной задержкой в модель второго порядка без задержки, где параметр задержки проявляется как эффективная инерция.

2. Описание модели

Мы рассматриваем ансамбль из N связанных фазовых осцилляторов с неоднородными взаимодействиями и единой задержкой по времени τ:

$$\dot{\boldsymbol{\theta}}_{\boldsymbol{j}} = \boldsymbol{\varpi} + \boldsymbol{\eta}_{\boldsymbol{j}}(\boldsymbol{t}) + \frac{\boldsymbol{\varkappa}}{\boldsymbol{N}}\sum_{\boldsymbol{k=1}}^{\boldsymbol{N}} \boldsymbol{F}_{\boldsymbol{jk}}\big[\boldsymbol{\theta}_{\boldsymbol{k}}(\boldsymbol{t}-\boldsymbol{\tau}) - \boldsymbol{\theta}_{\boldsymbol{j}}(\boldsymbol{t})\big], \tag{1}$$

где $\theta_j(t)$ – фаза осциллятора j, ϖ – средняя собственная частота, $\eta_j(t) = \varepsilon\omega_j + \varepsilon\zeta_j(t)$ включает малые отклонения частоты $\varepsilon\omega_j$ и внешнее воздействие $\varepsilon\zeta_j(t)$, ε – малый параметр, $\varkappa$ – глобальная сила связи, а $F_{jk}(\cdot)$ – 2π-периодические функции связи [1].

Применяя строгий асимптотический анализ в пределе слабой связи и слабой неоднородности, мы выводим редуцированную модель для динамики медленной фазы ϕ_j. Редукция показывает, что задержка τ непосредственно порождает инерционный член второго порядка:

$$\boldsymbol{\tau}\ddot{\boldsymbol{\phi}}_{\boldsymbol{j}} + \dot{\boldsymbol{\phi}}_{\boldsymbol{j}} = \left[\bar{\boldsymbol{\eta}}_{\boldsymbol{j}}(\boldsymbol{t}) + \frac{\boldsymbol{\varkappa}}{\boldsymbol{N}}\sum_{\boldsymbol{k=1}}^{\boldsymbol{N}} \boldsymbol{F}_{\boldsymbol{jk}}\big[\boldsymbol{\phi}_{\boldsymbol{k}} - \boldsymbol{\phi}_{\boldsymbol{j}} - \boldsymbol{\varpi}\boldsymbol{\tau}\big]\right] \times \left[\boldsymbol{1} - \frac{\boldsymbol{\tau}\boldsymbol{\varkappa}}{\boldsymbol{N}}\boldsymbol{F}'_{\boldsymbol{jk}}\big[\boldsymbol{\phi}_{\boldsymbol{k}} - \boldsymbol{\phi}_{\boldsymbol{j}} - \boldsymbol{\varpi}\boldsymbol{\tau}\big]\right] + \boldsymbol{\tau}\dot{\bar{\boldsymbol{\eta}}}_{\boldsymbol{j}}, \tag{2}$$

где $\bar{\eta}_j(t) = \varepsilon\omega_j + \varepsilon\bar{\zeta}_j(t)$, $\bar{\zeta}_j(t)$ – усреднение $\zeta_j(t)$ по периоду быстрых колебаний. Уравнение (2) представляет собой систему второго порядка без задержки. Член $\tau\ddot{\phi}_j$ добавляет влияние инерции, что коренным образом изменяет динамику редуцированной системы в отличие от передемпфированного случая редукции первого порядка.

[1*] Работа выполнена при поддержке Российского научного фонда (проект № 22-12-00348-П).

3. Результаты

Численное моделирование демонстрирует, что редукция второго порядка (2) точно воспроизводит сложную динамику исходной системы с задержкой (1), значительно превосходя предсказательную способность редукции первого порядка. Для сетей идентичных осцилляторов $\left(\eta_j(t) = 0\right)$ редуцированная модель верно воспроизводит статистические свойства решений. Она точно предсказывает вероятности реализации глобальной полной синхронизации и распределенных состояний (где распределение фаз равномерно) из случайных начальных условий. Редукция обеспечивает количественно точную аппроксимацию областей устойчивости как для синхронных, так и для распределенных состояний в пространстве параметров. Аналитически полученные границы устойчивости из редуцированной модели точно согласуются с таковыми для исходной системы (1). Эквивалентность поведения также распространяется на сложные динамические режимы, реализующиеся при наличии высших гармоник в функции связи. Редуцированная модель достоверно воспроизводит временную эволюцию параметров порядка для стационарных, бризерных и переключающихся циклопных состояний – сложных кластерных структур, представляющих собой два когерентных кластера и один уединенный осциллятор [2]. Применимость редукции не ограничивается глобальной топологией связи. Для случайной сетевой топологии «малый мир» [3] модель второго порядка успешно предсказывает распределение вероятностей реализации различных динамических режимов, что демонстрирует ее надежность и общую применимость.

Мы разработали строгий и мощный подход к редукции, который успешно отображает бесконечномерную динамику сетей осцилляторов с временной задержкой на конечномерную инерционную систему второго порядка. Отличное соответствие по статистическим метрикам, областям устойчивости и сложной переходной динамике подтверждает, что модель второго порядка является высокоточным и вычислительно эффективным инструментом для анализа и моделирования сетей с запаздывающими взаимодействиями.

Литература

1. Earl M.G., Strogatz S.H. Synchronization in oscillator networks with delayed coupling: A stability criterion // Physical Review E. 2003. Vol. 67, No. 3. P. 036204. DOI: 10.1103/PhysRevE.67.036204.
2. Bolotov M.I., Munyayev V.O., Smirnov L.A., Osipov G.V., Belykh I. Breathing and switching cyclops states in Kuramoto networks with higher-mode coupling // Physical Review E. 2024. Vol. 109, No. 5. P. 054202. DOI: 10.1103/PhysRevE.109.054202.
3. Ameli S., Karimian M., Shahbazi F. Time-delayed Kuramoto model in the Watts–Strogatz small-world networks // Chaos 2021. Vol. 31, No. 11. P. 113125. DOI: 10.1063/5.0064022.

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА ЛОКАЛЬНЫХ РЫНКОВ НЕДВИЖИМОСТИ

И.С. Созонов[1], Т.С. Бородина[1], А.А. Акобян[2], Л.А. Лейфер[2]

[1]Нижегородский государственный университет им. Н.И. Лобачевского,
[2]Приволжский центр методического и информационного обеспечения оценки ООО «Информ-оценка»

В данной работе рассматриваются подходы к автоматизации аналитических отчетов в сфере недвижимости с использованием методов математического моделирования, нейронных сетей и языковых моделей искусственного интеллекта. Реализовано веб-приложение, формирующее отчёты по объектам недвижимости в формате DOCX на основе структурированных JSON-данных. В качестве генератора аналитического текста применена языковая модель Gemini (Google Generative AI). Разработанная система объединяет методы моделирования структуры документа, обработку изображений, асинхронное взаимодействие с API и автоматизацию оформления итогового отчёта. Работа демонстрирует возможности интеграции нейросетевых технологий и суперкомпьютерных вычислений в процессах интеллектуального анализа данных и цифровизации экономики.

Ключевые слова: математическое моделирование, искусственный интеллект, генеративные модели, FastAPI, Python, автоматизация отчетности, рынок недвижимости, нейронные сети.

1. Введение

Современные тенденции цифровизации и рост объёмов данных формируют спрос на интеллектуальные системы [1, 5, 9–11], способные автоматически собирать, анализиро-вать и представлять информацию в удобной форме [2, 8]. В экономике и градостроитель-стве особое значение приобретают системы поддержки принятия решений, основанные на моделировании и обработке больших данных [2, 7].

Одной из наиболее трудоёмких задач является формирование аналитических отчётов по объектам недвижимости [6], включающих данные о характеристиках здания, местопо-ложении, рыночных аналогах, динамике цен и арендных ставок [2, 6].

Цель работы – разработка веб-приложения, автоматически генерирующего структурированные аналитические отчёты на основе данных и моделей искусственного интеллекта [1, 5, 8]. Основная идея – сформировать математическую модель [2, 7], описывающую процесс генерации отчёта, где входными параметрами являются данные об объекте, а вы-ходом – итоговый документ.

2. Постановка задачи и математическая модель

Задача автоматизации формирования отчета может быть описана как операторное преобразование:

$$R = F(D, S, M),$$

где D – структурированные данные об объекте недвижимости (в формате JSON), S – описание структуры отчета (иерархия разделов, шаблоны промптов, пути к данным), M – языковая модель (в данном случае – Gemini), R – результирующий документ в формате DOCX.

Каждый раздел i генерируется в виде функции:

$$T_i = M(P_i, D_i),$$

где P_i – шаблон промпта, D_i – соответствующие данные из входного файла.

Таким образом, итоговый документ является композицией нейросетевых функций от структурированных входов, а сам процесс – примером иерархического моделирования текста на основе данных [7, 9].

Для оптимизации взаимодействия с API введён асинхронный семафор [3–4, 8], обеспечивающий ограничение числа вызовов модели в единицу времени:

$$t_{i+1} - t_i \geq \frac{1}{L},$$

где L – лимит запросов в минуту (например, 60).

3. Архитектура и алгоритм работы системы

3.1. Общая архитектура

Разработанное веб-приложение реализовано по модели «клиент–сервер» [8] и включает два основных компонента:

- Серверная часть (Backend) – на Python с использованием `FastAPI`. Отвечает за обработку JSON-данных, взаимодействие с моделью Gemini, генерацию текста, форматирование и оформление отчета в формате DOCX/DOCM [3–4, 8].
- Клиентская часть (Frontend) – HTML/CSS/JavaScript-интерфейс, обеспечивающий запуск процесса генерации и автоматическое скачивание готового отчета [8].

Такое разделение обеспечивает масштабируемость, гибкость и возможность дальнейшего развития системы [3–4, 8].

3.2. Этапы работы алгоритма

1. **Загрузка данных:** пользователь загружает файл `input.json`, содержащий описание объекта недвижимости.
2. **Обработка входной структуры:** сервер извлекает данные из вложенных ключей и сопоставляет их с шаблонами разделов в файле `report_structure.json`.
3. **Генерация текста:** для каждого раздела формируется промпт, в который подставляются данные и динамические параметры (город, дата, год и т. д.). Промпт отправляется в модель Gemini, результат вставляется в документ.
4. **Вставка иллюстраций:** если во входных данных присутствуют изображения в формате base64, они декодируются и встраиваются в документ с сохранением пропорций.
5. **Форматирование и оформление:** документ дополняется титульным листом, оглавлением, нумерацией страниц и макросом VBA для обновления содержания при открытии.
6. **Автоматическая загрузка:** клиентская часть инициирует скачивание готового файла DOCM без дополнительных действий пользователя.

4. Реализация серверной части

4.1. Технологический стек

- **Python 3.12** – основной язык разработки.
- **FastAPI** – фреймворк для создания асинхронных веб-приложений.
- **python-docx** – библиотека для создания и форматирования документов Word.
- **Pillow** – для обработки и масштабирования изображений.
- **win32com.client** – автоматизация Word и внедрение VBA-макросов.
- **google-generativeai** – взаимодействие с моделью Gemini.
- **asyncio, base64, os, tempfile** – системные модули для асинхронности и работы с файлами.

4.2. Генерация и оформление документа

Библиотека `python-docx` используется для вставки заголовков (уровни 1 и 2), текста и изображений. Для каждого раздела применяется функция форматирования `format_text()`, обеспечивающая выравнивание по ширине, межстрочный интервал 1,5 и стиль Times New Roman 12 pt.

В конце документа добавляется VBA-макрос с функцией `AutoOpen()`, который при первом открытии автоматически обновляет оглавление, устанавливает обтекание текста вокруг изображений и корректирует поля.

5. Разработка клиентской части

Клиентская часть представляет собой минималистичный интерфейс из двух кнопок: «Сформировать отчёт» и «Скачать отчёт». После загрузки файла JSON данные сохраняются в `Session Storage`, а при нажатии на кнопку отправляются на сервер в теле POST-запроса. JavaScript обрабатывает ответ и инициирует скачивание готового файла, обеспечивая плавное взаимодействие без перезагрузки страницы.

Интерфейс ориентирован на пользователей без программных навыков, что делает систему удобной для аналитиков, оценщиков и риелторов [4, 8].

6. Оптимизация и устойчивость системы

На этапе оптимизации реализованы следующие механизмы:

- **Контроль скорости запросов:** семафор ограничивает частоту обращений к Gemini API, предотвращая ошибки 429 (`Too Many Requests`).
- **Обработка ошибок:** все операции вставки текста и изображений выполняются в блоках `try-except`, что повышает надежность работы.
- **Фильтрация пустых параграфов:** исключено дублирование пустых строк и повторная генерация в подразделах.
- **Масштабирование изображений:** сохранение пропорций и ограничение ширины до 80% страницы.

В результате документ формируется в строго установленном стиле (шрифт Times New Roman, межстрочный интервал 1,5, отступ первой строки 1,25 см) в соответствии с ГОСТ и ФСО [6], а качество и устойчивость генерации достигают уровня прикладных программных систем.

7. Тестирование и результаты экспериментов

Проведено многоэтапное тестирование:

- **Модульное тестирование:** проверены функции генерации текста, извлечения данных, обработки изображений.
- **Функциональное тестирование:** тестировались все этапы взаимодействия пользователя с интерфейсом.
- **Стресс-тестирование:** проверена работа при массовой генерации отчетов (до 30 разделов).

Среднее время полного формирования отчета – около 3 минут. Результаты показали сокращение трудозатрат аналитика на 70 - 80 % и повышение согласованности оформления на 90 % [2, 8].

8. Практическое применение и перспективы развития

Разработанная система уже применяется в оценочной деятельности для подготовки аналитических отчетов по локальным рынкам недвижимости на официальном сайте приволжского

центра методического и информационного обеспечения оценки ООО «Информ-оценка». Модель может быть адаптирована для других областей: финансового анализа, мониторинга рынков, бизнес-отчетности, образовательной аналитики [2, 5, 7].

Перспективы развития:

- Интеграция кластерных и параллельных вычислений для массовой генерации документов.
- Расширение архитектуры для поддержки нескольких языковых моделей (например, Gemini и Claude).
- Применение алгоритмов оценки качества текста с использованием метрик ROUGE и BERTScore.
- Разработка интерфейса обратной связи для экспертов по корректировке результатов генерации.

9. Заключение

В результате работы создана функциональная система, объединяющая методы математического моделирования и генеративного искусственного интеллекта для автоматизации аналитических отчетов. Сочетание модульной архитектуры, асинхронной обработки и VBA-автоматизации делает приложение гибким и масштабируемым инструментом для цифровизации аналитических процессов.

Представленный подход демонстрирует эффективность интеграции нейросетевых технологий в область информационного моделирования и автоматизации документооборота, а также открывает возможности для использования суперкомпьютерных ресурсов в масштабных сценариях обработки данных.

Литература

1. Вакуленко С.А. Практический курс по нейронным сетям / Вакуленко С.А., Жихарева А.А. СПб: Университет ИТМО, 2018. 71 с.
2. Золкин А.Л. Математическое моделирование и анализ данных. Санкт-Петербург: Лань, 2025. 128 с.
3. Макконнелл С. Совершенный код. Практическое руководство по разработке программного обеспечения. Санкт-Петербург: БХВ-Петербург, 2022. 896 с.
4. Мартин Р. Чистый код: создание, анализ и рефакторинг. Библиотека программиста. Санкт-Петербург: Питер, 2018. 464 с.
5. Ростовцев В.С. Искусственные нейронные сети: Учебник для вузов / В.С. Ростовцев. СПб.: Лань, 2024. 216 с.
6. Федеральный стандарт оценки "Оценка недвижимости (ФСО № 7)": собрание законодательства Российской Федерации, 1998, N 31, ст. 3813; 2006, N 31, ст. 3456; 2010, N 30, ст. 3998; 2011, N 1, ст. 43; N 29, ст. 4291; 2014, N 30, ст. 4226.
7. Федоткин И.М. Математическое моделирование технологических процессов. Москва: КД Либроком, 2018. 416 с.
8. Эспозито Д. Разработка современных веб-приложений: анализ предметных областей и технологий. Москва: Вильямс, 2017. 464 с.
9. Anirudh Koul, Siddha Ganju, Meher Kasam: Practical Deep Learning for Cloud, Mobile, and Edge: O'Reilly Media, Inc., 2019, 620 pp, ISBN: 9781492034865.
10. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618.
11. Simon O., Haykin: Neural Networks and Learning Machines: Pearson Education, 2011, 936 pp, ISBN: 0133002551.

КОЛЛЕКТИВНАЯ РАЗРАБОТКА БИБЛИОТЕКИ ДЛЯ ВЫВОДА РЕЗУЛЬТАТОВ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

А.А. Сорокин, С.М. Титов, А.Ю. Нестеров, А.А. Оболенский

Нижегородский государственный университет им. Н.И. Лобачевского

В работе рассматривается командная разработка библиотеки для инференса сверточных нейронных сетей (CNN) на примере архитектуры AlexNet. Акцент сделан на слаженную работу для создания гибкой и производительной системы, способной обрабатывать пользовательские данные с использованием предобученных моделей. Для обработки изображений интегрирована библиотека OpenCV, а параллельные вычисления оптимизированы с использованием TBB. Результаты проекта демонстрируют высокую эффективность: точность классификации на тестовой выборке MNIST достигла 98.01%, а время инференса на CPU составило 8 мс на одно изображение.

Ключевые слова: командная разработка, сверточные нейронные сети, AlexNet, инференс, направленный ациклический граф, параллельные вычисления, классификация изображений.

1. Введение

В настоящее время разработка многослойных нейронных сетей является очень популярным направлением промышленного программирования. Актуальность этого направления быстро растёт. Одним из самых распространённых, востребованных и эффективных типов многослойных нейронных сетей являются сверточные нейронные сети (CNN) [1], которые зачастую используются для решения задач распознавания и классификации. Это означает, что обученная сверточная нейронная сеть должна уметь принимать входное изображение и сопоставлять ему из заранее заданного набора классов тот класс, который лучше остальных характеризует это изображение. Подобное поведение сети обеспечивается последовательностью слоёв разных типов: одни из них выделяют признаки, необходимые для классификации, а другие обрабатывают их и определяют наиболее вероятный класс. Процесс разработки такого продукта, как сверточная нейронная сеть, сложен и поэтому требует создания чёткого плана действий команды разработчиков, основанного на ключевых целях и задачах реализуемого проекта.

В рамках нашего проекта, направленного на создание библиотеки для работы со сверточными нейронными сетями, мы стремимся организовать процесс разработки так, чтобы он соответствовал промышленным стандартам и был ориентирован на командную работу. Наша команда ставит перед собой задачу не только создать высококачественное программное обеспечение, но и обеспечить его удобство и доступность для конечных пользователей. Основной задачей работы является создание возможности для запуска и использования предобученных нейронных сетей с пользовательскими данными.

В процессе разработки архитектуры нашей библиотеки мы проанализировали реализацию инференса (вывода) в промышленных фреймворках, таких как TensorFlow Lite [2], ONNX Runtime [3], PyTorch [4] и OpenCV DNN [5]. Этот анализ позволил нам заимствовать ключевые архитектурные решения. Такой подход особенно важен для учебного проекта, разрабатываемого командой студентов, поскольку позволил найти баланс между производительностью, удобством разработки и простотой реализации в рамках поставленных задач. В нашем проекте мы также активно использовали сторонние библиотеки и фреймворки, например: OpenCV – для работы с изображениями, JSON [6] – для сериализации моделей, TBB [7] – для параллельных вычислений на CPU и GTest [8] – для модульного тестирования.

2. Процесс разработки проекта

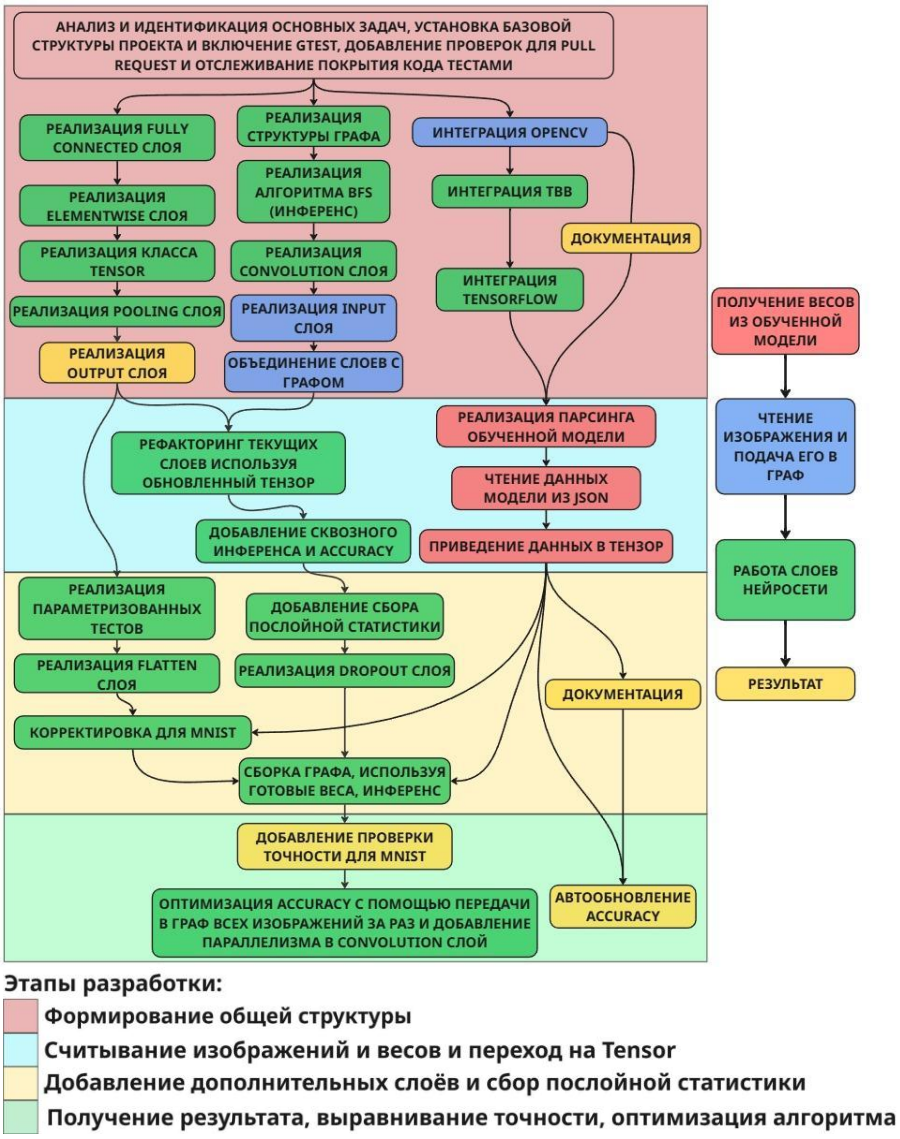


Рис. 1 Схема разработки библиотеки

На рисунке 1 отображены процессы разработки и взаимодействия команды на этапах создания библиотеки для инференса [19]. Далее более подробно описываются шаги, представленные на рисунке, а также проблематика командной работы.

В самом начале перед нашей командой стояла задача создания библиотеки, способной осуществлять инференс предобученных нейронных сетей. Первым делом необходимо было изучить этот вопрос в теории и разделить предстоящую работу на части для более эффективной работы в команде. В качестве ознакомления использовалась информация о работе с системой сборки CMake [9], с GitHub [10], операционной системой Linux [11], а также об использовании субмодулей в будущем проекте, предоставленная наставниками. Каждый изучал различные нейросетевые модели, используя такой инструмент, как Netron [12], чтобы понять структуру обученной CNN. В итоге было выделено четыре основных направления разработки.

1. Создание графа как структуры для обмена данными между узлами.
2. Подключение необходимых внешних библиотек и проверка их функционала.
3. Преобразование предобученной модели в подходящий нам формат хранения.
4. Разработка необходимых слоёв для работы с обученными данными.

Был создан общий репозиторий на GitHub, где по мере продвижения создавались и обсуждались новые задачи и их распределение между участниками. Также сразу была определена

структура проекта и организована его сборка через CMake. Код должен был быть в едином стиле – для этого были введены проверки стиля кода, а также множественные проверки возможных ошибок с помощью GitHub Actions. Каждый член команды сопровождал свой код тестами; для этого была подключена библиотека gtest и сервис Codecov. Эти детали командной разработки очень важны, так как они обеспечивают стабильность, качество и согласованность кода, а также упрощают совместную работу и дальнейшую поддержку проекта.

Важно, что на собраниях, которые проводились каждую неделю, мы делились полученным опытом и советовались при принятии большинства решений. Хотя работа и велась в нескольких относительно различных направлениях, в результате было необходимо всё собрать воедино.

На начальных этапах было важно, чтобы граф мог передавать данные от слоя к слою, а подключенные библиотеки обеспечивали полный необходимый функционал. Первыми результатами стали интеграция в проект библиотеки OpenCV для чтения изображений и написание основополагающих классов, таких как:

- Graph – основа библиотеки, структура хранения слоёв,
- Layer – шаблонный класс, основа всех будущих слоёв,
- Tensor – шаблонная структура хранения данных, которая передаётся от слоя к слою.

Далее основная часть команды перешла на разработку конкретных слоёв, используемых в сверточных нейронных сетях. Однако когда почти все слои были написаны, возникла проблема: необходимо было соединить базовые классы Layer и Graph, чтобы впервые запустить инференс с несколькими написанными слоями. Но Layer был шаблонным и требовал этого же от класса Graph. Из этого следовало, что в графе все слои могут быть только с одним типом данных, что вызывало противоречие, так как слои выполняют свои функции на разных уровнях. После анализа данной проблемы было решено переписать Tensor как класс без использования шаблонов. Соответственно, и слои, и граф благодаря этому лишались шаблонности.

После исправления всех недочётов впервые удалось запустить инференс, но только в виде end-to-end теста, или, другими словами, прохождения тензора по графу через все слои. Преодоление этого важного рубежа позволило брать тестовые данные и проверять верность работы частичного инференса.

Параллельно шла работа над интеграцией фреймворка TensorFlow для работы с моделью Alexnet_frozen.pb – обученной моделью классификации изображений . Однако из-за возникших сложностей настройки взаимодействия C++ и TensorFlow была произведена замена модели на Alexnet_model.h5, обученную распознавать цифры на датасете MNIST [13]. Для считывания структуры этой модели и её весов в наш формат тензоров был написан конвертер с помощью Python и формата-посредника JSON. Исходя из структуры новой модели, необходимо было дописать реализацию ещё нескольких слоёв; также были добавлены функции, собирающие статистику по каждому слою в процессе инференса.

Впоследствии нам удалось успешно запустить проход по графу для предсказания отдельных изображений, но ответы нейросети получались некорректными. После исправления ошибок, включая недостаточную точность конвертера модели в тензор, некорректную обработку изображений перед входным слоем, небольшие неточности в функциях softmax и topk для обработки ответов модели, а также некоторых других проблем, мы достигли точности чуть более 98% на 10 тысячах изображений, однако проверка такого объёма данных занимала около 20 минут, что значительно дольше по сравнению с другими реализациями аналогичных задач.

Следующей задачей стало ускорение времени выполнения простыми модификациями кода. Анализируя проблему, стало понятно, что 10 тыс. изображений считывались по одному во входной слой, и для каждого изображения заново строился весь граф. Некоторые слои не поддерживали формат хранения NCHW [14], из-за чего было невозможно подать в сеть сразу несколько изображений. После внесения всех необходимых исправлений производительность программы значительно возросла. Также был модифицирован сверточный слой – в нём появилось распараллеливание матричных операций, что дополнительно ускорило процесс инференса. В результате точность предсказаний модели на 10 тыс. изображений составила 98,01%, а время проверки – 79 секунд.

Важно отметить, что на протяжении всей разработки при достижении важных этапов обновлялась документация. Также в конце было добавлено автоматическое измерение производи-

тельности для каждого изменения в актуальной ветке библиотеки. Таким образом, в ReadMe всегда можно увидеть актуальную точность выполненной нами работы.

3. Архитектура библиотеки

Основополагающими компонентами библиотеки являются классы Graph, Shape и Tensor. Graph представляет собой реализацию направленного ациклического графа (Directed Acyclic Graph, DAG) [15]. В нашем проекте граф организован следующим образом:

- Узлы графа представлены объектами класса Layer, которые содержат операции над данными.
- Рёбра задаются методом makeConnection, который соединяет слои сети.
- Для выполнения прямого прохода по графу и выполнения инференса используется метод inference.
- Для обхода слоёв в правильной последовательности был реализован алгоритм, напоминающий поиск в ширину для генерации маршрута.

Класс Tensor является универсальной структурой, которая используется для хранения данных внутри слоя и для передачи данных между ними. В нашей библиотеке тензор – это N-мерная матрица, которая позволяет хранить элементы разных типов данных в одномерном динамическом массиве байт.

Класс Shape позволяет явно определить размерность тензора.

Также в библиотеке реализованы следующие слои:

- Input Layer: поддерживает форматы NCHW/NHWC через нормализацию и перестановку данных, обеспечивая гибкость для разных сетей.
- Convolutional Layer: использует padding, dilations и матричные операции с поддержкой NCHW, корректируя размер тензора.
- Fully Connected Layer: реализует умножение матрицы на вектор со смещением, обрабатывая тензоры и исключительные случаи.
- Elementwise Layer: применяет активационные функции (tanh, ReLU и др.) с параметрами alpha/beta для линейных преобразований [18].
- Pooling Layer: выполняет max/average pooling с расчётом выходного размера и проверкой границ тензора.
- Normalization Layer: стандартизирует данные в заданный диапазон для стабильности вычислений.
- Flatten Layer: преобразует многомерный тензор в одномерный вектор, что необходимо для обработки данных полносвязными слоями.
- Output Layer: частично сортирует результаты с помощью std::partial_sort для вывода итоговых предсказаний.

В проекте используется система сборки CMake для автоматизации процесса компиляции и сборки проекта. Это позволяет обеспечить переносимость проекта на различные операционные системы, такие как Windows, Linux и macOS, и архитектуры x64 и x86.

4. Результаты

В рамках проекта по созданию библиотеки для инференса нейронной сети AlexNet команда достигла поставленной цели благодаря слаженной работе и эффективной организации разработки. Проект выполнен на высоком уровне, что стало возможным благодаря чёткому распределению задач, регулярному взаимодействию внутри команды и использованию современных инструментов и технологий.

Ключевые результаты библиотеки:

1. Подтверждена точность 98,01% на тестовом наборе MNIST [13, 19]
2. Послойный замер скорости обработки набора из 10 000 изображений представлены в таблице 1.

Таблица 1. Скорость работы библиотеки

	Время работы слоя на машине, мс		
Слой	**Машина 1**	**Машина 2**	**Машина 3**
Input	18	30	18
Convolution 1	4365	6147	4795
Pooling 1	7136	7209	3597
Convolution 2	23801	30100	24834
Pooling 2	4008	4033	1950
Convolution 3	14587	18453	14389
Convolution 4	14504	18992	14360
Convolution 5	14440	19086	14375
Pooling 3	1001	1123	514
Flatten	171	203	104
Dense 1	309	295	333
Dense 2	194	207	198
Dropout	22	28	15
Dense 3	19	18	19
Общее время	84575	105924	79501

Данные о машинах из таблицы 1:

Машина 1 – Windows 10, процессор: Intel Core i7-11710 OEM, 32 ГБ RAM

Машина 2 – Windows 11, процессор: AMD Ryzen 7 5800H with Radeon Graphics, 16 ГБ RAM

Машина 3 – GitHub Ubuntu 24.04, процессор: AMD EPYC 7763 2 Cores, 15 ГБ RAM

Во время разработки был получен значительный прирост скорости относительно начальных этапов. Однако на данный момент существуют библиотеки, работающие намного быстрее [16, 17]. Например, библиотеки, использующие фреймворки TensorFlow/Keras с интерфейсом на языке Python, но ядром на C/C++, имеют время обработки того же количества изображений 0,6255с. с точностью 98,32%.

Важно отметить, что разработанная библиотека создавалась в первую очередь как учебный проект, направленный на глубокое понимание архитектуры нейронных сетей и процессов инференса. Код был реализован командой из трёх разработчиков под руководством экспертов, при этом основной задачей была корректная работа алгоритмов, а не максимальная оптимизация производительности. Каждый из нас получил огромный опыт в командной разработке, что значительно упростит адаптацию при дальнейшей работе.

В дальнейшем планируется развить и ускорить работу нашей библиотеки, используя более сложные алгоритмы оптимизации графа и различные виды параллелизма для конкретных слоёв. Также данная библиотека будет поддерживать более сложные сети, например семейства ONNX и YOLO, для чего будут расширены возможности графа и написаны новые слои.

Литература

1. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks // Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012). Lake Tahoe, Nevada, USA, December 3-6, 2012. 2012. Vol. 1. P. 1097–1105.
2. Abadi M., Barham P., Chen J. et al. TensorFlow: A System for Large-Scale Machine Learning // Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16). Savannah, GA, USA, November 2-4, 2016. 2016. P. 265–283.

3. ONNX Runtime: High-Performance Inference Engine [Электронный ресурс]. Microsoft, 2023. URL: https://onnxruntime.ai (дата обращения: 10.10.2023).
4. Paszke A., Gross S., Massa F. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library // Advances in Neural Information Processing Systems 32 (NeurIPS 2019). Vancouver, Canada, December 8-14, 2019. 2019. P. 8026–8037.
5. Bradski G. The OpenCV Library // Dr. Dobb's Journal of Software Tools. 2000. Vol. 25. No. 11. P. 120–125.
6. JSON for Modern C++ [Электронный ресурс] / N. Lohmann. 2023. URL: https://json.nlohmann.me (дата обращения: 17.05.2024).
7. Reinders J. Intel Threading Building Blocks: Outfitting C++ for Multi-core Processor Parallelism. Sebastopol: O'Reilly Media, 2007. 334 p.
8. Google Test: C++ Testing Framework [Электронный ресурс]. Google, 2023. URL: https://github.com/google/googletest (дата обращения: 19.09.2023).
9. CMake Documentation [Электронный ресурс]. Kitware, 2023. URL: https://cmake.org/documentation/ (дата обращения: 05.10.2023).
10. Chacon S., Straub B. Pro Git. 2nd ed. N. Y.: Apress, 2014. 456 p. URL: https://git-scm.com/book/en/v2 (дата обращения: 11.09.2023).
11. Kerrisk M. The Linux Programming Interface: A Linux and UNIX System Programming Handbook. San Francisco: No Starch Press, 2010. 1552 p.
12. Netron: Visualizer for Neural Networks [Электронный ресурс]. L. Roeder, 2023. URL: https://netron.app (дата обращения: 12.11.2023).
13. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-Based Learning Applied to Document Recognition // Proceedings of the IEEE. 1998. Vol. 86. No. 11. P. 2278–2324.
14. Chetlur S., Woolley C., Vandermersch P. et al. cuDNN: Efficient Primitives for Deep Learning // arXiv preprint. 2014. arXiv:1410.0759. URL: https://arxiv.org/abs/1410.0759 (дата обращения: 22.12.2023).
15. Cormen T.H., Leiserson C.E., Rivest R.L., Stein C. Introduction to Algorithms. 3rd ed. Cambridge: MIT Press, 2009. 1292 p.
16. Jia Y., Shelhamer E., Donahue J. et al. Caffe: Convolutional Architecture for Fast Feature Embedding // arXiv preprint. 2014. arXiv:1408.5093. URL: https://arxiv.org/abs/1408.5093 (дата обращения: 29.05.2024).
17. Vasilache N., Zinenko O., Theodoridis T. et al. Tensor Comprehensions: Framework-Agnostic High-Performance Machine Learning Abstractions // arXiv preprint. 2018. arXiv:1802.04730. URL: https://arxiv.org/abs/1802.04730 (дата обращения: 28.05.2024).
18. Nwankpa C., Ijomah W., Gachagan A., Marshall S. Activation Functions: Comparison of Trends in Practice and Research for Deep Learning // arXiv preprint. 2018. arXiv:1811.03378. URL: https://arxiv.org/abs/1811.03378 (дата обращения: 01.10.2025).
19. ITLabAI Research Group. Библиотека для вывода результатов сверточных нейронных сетей [Электронный ресурс] // GitHub. 2025. URL: https://github.com/embedded-dev-research/ITLabAI.

ПОСТРОЕНИЕ ОБОБЩЕННОГО H_∞ УПРАВЛЕНИЯ ДЛЯ СИСТЕМЫ МАТЬЕ НА КОНЕЧНОМ ИНТЕРВАЛЕ ПО АПРИОРНЫМ И ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ[1*]

А.В. Степанов

Нижегородский государственный архитектурно-строительный университет, Научно-технологический университет «Сириус»

В статье рассмотрен подход построения робастного управления на конечном интервале для линейных нестационарных систем с использованием априорной информации и экспериментальных данных. В отличие от традиционных подходов, метод не требует выполнения условия постоянного возбуждения или идентифицируемости системы. Предложенная схема объединяет преимущества моделей, построенных на основе данных, и робастного управления, основанного на априорных знаниях, обеспечивая улучшенные гарантии устойчивости и производительности. Эффективность подхода продемонстрирована на численном примере, основанном на дискретной модели уравнения Матье, описывающего параметрические колебания линейного осциллятора.

Ключевые слова: обобщенная $\boldsymbol{H}_\infty$-норма, неопределенность, робастное управление, экспериментальные данные, линейные матричные неравенства.

1. Введение

Современные методы синтеза робастных законов управления активно развиваются в направлении использования экспериментальных данных, полученных в процессе исследований [3,4,5]. Традиционные подходы в основном опираются на априорную информацию о системе. Однако, применение экспериментальных данных совместно с априорной информацией в итоге улучшает качество управления [9,10]. Особенно это актуально для систем с параметрической неопределённостью. Такие методы предполагают использование модели линейной дробно-частичной трансформации (LFT), где система разбивается на две подсистемы: известную и содержащую неопределённые параметры.

При синтезе робастного управления на основе экспериментальных данных предполагается, что возмущения и ошибки измерений имеют ограниченные значения, что позволяет корректно определить область неопределённых параметров через квадратичные неравенства. Однако прямое использование экспериментальных данных для синтеза робастных законов управления вызывает сложности, связанные с необходимостью выполнения ранговых условий, обеспечивающих идентифицируемость параметров.

В работах [3,4,5,7] были предложены новые подходы, позволяющие синтезировать законы управления по априорным и экспериментальным данным без необходимости выполнения условий постоянного возбуждения. В частности, использование обобщённой H_∞ -нормы, выраженной через квадратичные функции Ляпунова для двойственных систем, позволило преодолеть эти ограничения [9,10]. В данной статье описан подход к построению робастного управления для полностью неопределённых нестационарных систем на конечном интервале времени, используя как экспериментальные данные, так и априорные знания. Показано, что при этом не требуется выполнение условий идентифицируемости, что существенно снижает объем требуемых данных.

[1*] Работа выполнена при финансовой поддержке РНФ (проект № 24-11-20023).

2. Постановка задачи

2.1. Экспериментальные данные

Рассмотрим линейную нестационарную систему с неопределёнными параметрами:

$$\begin{aligned} x(t+1) &= A_t x(t) + B_t u(t) + w(t), \qquad x(0) = x_0 \\ z(t) &= C_t x(t) + D_t u(t), \qquad t = 0, \ldots. N\text{-}1 \end{aligned} \tag{2.1}$$

где $x(t) \in R^{n_x}$ - вектор состояния системы, $u(t) \in R^{n_u}$ - управляющее воздействие, $w(t) \in R^{n_w}$ – внешние возмущения, а $z(t) \in R^{n_z}$ - целевой выход. Матрицы A_t, B_t, C_t и D_t неизвестны и принадлежат множеству неопределённостей Δ_t, задаваемому на основе как априорной информации, так и данных экспериментов.

В общем плане требуется на основе данных, полученных из экспериментов, и априорной информации синтезировать линейные обратные связи по состоянию, которые оптимизируют уровень гашения начальных и/или внешних возмущений на конечном горизонте.

Информация о неизвестных матрицах системы (2.1) извлекается из конечного набора измерений ее траектории длиной N в результате проведения $L \geq 1$ экспериментов. Предположим, что в l-м эксперименте получены измерения состояния системы $x_{0,l}, x_{1,l}, \ldots, x_{N,l}$ и целевого выхода $z_{0,l}, z_{1,l}, \ldots, z_{N-1,l}$ при выбранных управлениях $u_{0,l}, u_{1,l}, \ldots, u_{N-1,l}$ и некоторых неизвестных возмущениях $w_{0,l}, w_{1,l}, \ldots, w_{N-1,l}$. Для каждого *t = 0, …. , N-1* составим матрицы

$$\Phi_t = \left(x_{t,1} \cdots x_{t,L}\right),\ U_t = \left(u_{t,1} \cdots u_{t,L}\right),\ W_t = \left(w_{t,1} \cdots w_{t,L}\right),\ Z_t = \left(z_{t,1} \cdots z_{t,L}\right) \tag{2.2}$$

для которых в силу уравнения объекта (2.1) выполняется

$$\begin{aligned} \Phi_{t+1} &= A_t^{(real)} \Phi_t + B_t^{(real)} U_t + W_t \\ Z_t &= C_t^{(real)} \Phi_t + D_t^{(real)} U_t, \end{aligned} \tag{2.3}$$

где $A_t^{(real)}, B_t^{(real)}, C_t^{(real)}, D_t^{(real)}$ – реальные неизвестные матрицы уравнений системы.

Обозначим

$$\Delta_t^{(real)} = \begin{pmatrix} A_t^{(real)} & B_t^{(real)} \\ C_t^{(real)} & D_t^{(real)} \end{pmatrix}, \quad \widehat{\Phi}_t = \begin{pmatrix} \Phi_t \\ U_t \end{pmatrix}, \widehat{\Phi}_{t+1} = \begin{pmatrix} \Phi_{t+1} \\ Z_t \end{pmatrix}, \widehat{\mathrm{W}}_t = \begin{pmatrix} \mathrm{W}_t \\ 0 \end{pmatrix}$$

и запишем уравнения (2.3) в виде линейной матричной регрессии

$$\begin{aligned} \widetilde{\Phi}_{t+1} &= \Delta_{real}^t \widehat{\Phi}_t + \widehat{W_t} \\ t &= 0, \ldots., N\text{-}1 \end{aligned} \tag{2.4}$$

Допустим, что возмущение в эксперименте удовлетворяет условию $\sum_{l=1}^{L} w_{t,l} w_{t,l}^T = W_t W_t^{\mathrm{T}} \leq \Omega_t$, т.е.

$$\widehat{W_t} \widehat{W_t}^T \leq \begin{pmatrix} \Omega_t & \star \\ 0 & 0 \end{pmatrix} = \widehat{\Omega_t} \tag{2.5}$$

В частности, если $\sum_{l=1}^{L} \left|w_{t,l}\right|^2 \leq \alpha_t^2$, то $\Omega_t = \alpha_t^2 I$. Если для всех t возмущение удовлетворяет ограничению $\left|\left|w(t)\right|\right|_\infty \leq d_v$ для некоторого заданного d_w, которое будем называть уровнем возмущения, то $\Omega_t = d_v^2 n_v L B_{v,t} B_{v,t}^T$.

Далее, из (2.5) следует, что матрицы Δ_t порядка $(n_x + n_z)$ x $(n_x + n_u)$, которые могли бы генерировать полученные в экспериментах матрицы Φ_t и Z_t при выбранных управлениях U_t и некоторых матрицах $\widehat{W_t}$, удовлетворяющих ограничению (2.5), характеризуются неравенствами

$$\left(\widetilde{\Phi}_{t+1} - \Delta_t\widehat{\Phi}_t\right)\left(\widetilde{\Phi}_{t+1} - \Delta_t\widehat{\Phi}_t\right)^{\mathrm{T}} \le \widehat{\Omega},\ \mathrm{t} = 0, \ldots, \mathrm{N}\text{-}1 \tag{2.6}$$

Представим эти неравенства в виде

$$\left(\Delta_t\, I_{n_x+n_z}\right)\Psi^{(1)}(t)\left(\Delta_t\, I_{n_x+n_z}\right)^{\mathrm{T}} \le 0,\ t = 0, \ldots, N-1 \tag{2.7}$$

Где симметрические матрицы $\Psi^{(1)}(t)$ порядка $(2n_x + n_u + n_z) \times (2n_x + n_u + n_z)$ разбиваются на блоки $\Psi_{ij}^{(1)}(t)$, i,j = 1,….4 соответствующего порядка и вычисляются как

$$\Psi^{(1)}(t) = \begin{pmatrix} \Phi_t\Phi_t^{\mathrm{T}} & * & | & * & * \\ U_t\Phi_t^{\mathrm{T}} & U_tU_t^{\mathrm{T}} & | & * & * \\ - - - & - - - & - - - & - - - & - - - \\ -\Phi_{t+1}\Phi_t^{\mathrm{T}} & -\Phi_{t+1}U_t^{\mathrm{T}} & | & \Phi_{t+1}\Phi_{t+1}^{\mathrm{T}} - \Omega_t & * \\ -Z_t\Phi_t^{\mathrm{T}} & -Z_tU_t^{\mathrm{T}} & | & Z_t\Phi_{t+1}^{\mathrm{T}} & Z_tZ_t^{\mathrm{T}} \end{pmatrix} \tag{2.8}$$

Обозначим $\boldsymbol{\Delta}_t^{(\rho)}$ множество матриц Δ_t, удовлетворяющих для данного t неравенству (2.7), для данного t.

2.2. Априорная информация

Пусть имеется дополнительная информация о том, что неизвестная матрица $\Delta_t^{(real)}$, t = 0,…, N-1 удовлетворяет ограничениям

$$(\Delta_t - \Delta_t^*)(\Delta_t - \Delta_t^*)^{\mathrm{T}} \le \rho_t^2 I,\ \Delta_t^* = \begin{pmatrix} A_t^* & B_t^* \\ C_t^* & D_t^* \end{pmatrix} = \begin{pmatrix} \Delta_{t,1}^* \\ \Delta_{t,2}^* \end{pmatrix} \tag{2.9}$$

в которых Δ_t^* и ρ_t - заданные матрицы и скаляры, характеризующие центры и размеры областей неопределенности. Запишем это неравенство в виде

$$(\Delta_t\, I)\,\Psi^{(2)}(t)(\Delta_t\, I)^T \le 0, \tag{2.10}$$

где матрица $\Psi^{(2)}$(t) состоит из блоков $\Psi_{i,j}^{(2)}(t)$, i,j = 1,..., 4 и имеет вид

$$\Psi^{(2)}(t) = \begin{pmatrix} I_{n_x} & \star & | & \star & \star \\ 0_{n_u \times n_x} & I_{n_u} & | & \star & \star \\ - - - & - - - & - - - & - - - & - - - \\ -A_t^* & -B_t^* & | & \Delta_{t,1}^*\Delta_{t,1}^{*\mathrm{T}} - \rho_t^2 I_{n_x} & \star \\ -C_t^* & -D_t^* & | & \Delta_{t,2}^*\Delta_{t,1}^{*\mathrm{T}} & \Delta_{t,2}^*\Delta_{t,2}^{*\mathrm{T}} - \rho_t^2 I_{n_z} \end{pmatrix} \tag{2.11}$$

Обозначим $\boldsymbol{\Delta}_t^{(a)}$ множество матриц Δ_t, удовлетворяющих для данного t неравенству (2.10), которое представляет собой “матричную сферу”. Кроме того, обозначим $\Delta_t = \boldsymbol{\Delta}_t^{(p)} = \boldsymbol{\Delta}_t^{(p)} \cap \boldsymbol{\Delta}_t^{(a)}$ множество матриц, удовлетворяющих неравенствам (2.7) и (2.10). Таким образом, $\boldsymbol{\Delta_t}$ - это множество всех матриц Δ_t, согласованных при данном t с экспериментальными и априорными данными. Очевидно, что $\boldsymbol{\Delta}_t^{(real)} \in \boldsymbol{\Delta_t}$. На рисунке 1 схематично изображены множества $\boldsymbol{\Delta}_t^{(p)}$ $\boldsymbol{\Delta}_t^{(a)}$ и их пересечение $\boldsymbol{\Delta_t}$.

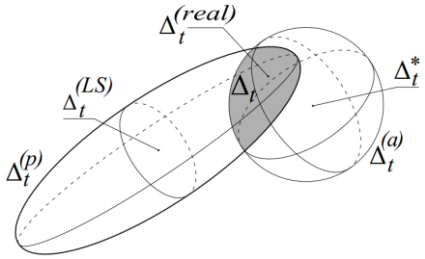


Рис.1. Множества $\boldsymbol{\Delta_t^{(p)}}$, $\boldsymbol{\Delta_t^{(a)}}$, $\boldsymbol{\Delta_t}$ матриц, согласованных с экспериментальными, априорными данными и их пересечением

Обозначим набор матриц $\Delta_{[0,N-1]} = (\Delta_0, \ldots, \Delta_{N-1})$, согласованных только с экспериментальными, только с априорными, а также с теми и другими данными для всех t = 0, …. , N-1 соответственно через $\boldsymbol{\Delta}_{[0,N-1]}^{(p)}$ = ($\boldsymbol{\Delta}_0^{(p)}$,…., $\boldsymbol{\Delta}_{N-1}^{(p)}$), $\boldsymbol{\Delta}_{[0,N-1]}^{(a)}$ = ($\boldsymbol{\Delta}_0^{(a)}$,…., $\boldsymbol{\Delta}_{N-1}^{(a)}$) и $\boldsymbol{\Delta_{[0.N-1]}}$ = $\boldsymbol{(\Delta_0, \ldots, \Delta_{N-1} = \Delta_{[0,N-1]}^{(p)} \cap \Delta_{[0,N-1]}^{(a)}}$.

2.3. Задача управления

Качество неопределенной системы (2.1), замкнутой управлением вида линейной нестационарной обратной связи по состоянию u(t) = Θ_tx(t), будем оценивать максимально возможным уровнем гашения начального и внешнего возмущений, т.е. верхней границей обобщенной H_∞-нормы замкнутой системы при всех матрицах системы, согласованных с экспериментальными и априорными данными:

$$\gamma_{g\infty}(\Theta_{[0,N-1]}; R, S) = \sup_{\Delta_{[0,N-1]} \in \boldsymbol{\Delta}_{[0,N-1]}} \sup_{x_0,\, w} \left(\frac{\|z\|^2_{[0,\,N-1]} + x^{\mathrm{T}}(N) S x(N)}{x_0^{\mathrm{T}} R^{-1} x_0 + \|w\|^2_{[0,\,N-1]}} \right)^{1/2}, \tag{2.12}$$

где R = R^T> 0 и S = S^T> 0 - весовые матрицы начального и терминального состояний соответственно $\|\xi\|^2_{[0,t]} = \sum_{i=0}^{t} |\xi(i)|^2$. Смысл этого критерия можно пояснить следующим образом. Состояние системы в текущий момент времени зависит линейно от начальных условий и возмущений и их увеличение приводит к увеличению переменных состояния. Для того чтобы характеризовать поведение системы при неопределенных начальных условиях и возмущениях, нормализуем соответствующий функционал величиной, равной указанной в знаменателе сумме, или, что то же самое для линейной системы, ограничим эту сумму единицей.

Когда начальное состояние нулевое, обобщенная H_∞- норма превращается в стандартную H_∞-норму

$$\gamma_{\infty}(\Theta_{[0,N-1]}; S) = \sup_{\Delta_{[0,N-1]} \in \boldsymbol{\Delta}_{[0,N-1]}} \sup_{w \neq 0} \frac{\left(\|z\|^2_{[0,\,N-1]} + x^{\mathrm{T}}(N) S x(N) \right)^{1/2}}{\|w\|_{[0,\,N-1]}}, \tag{2.13}$$

что соответствует R → 0 в обобщенной H_∞-норме (2.11). При отсутствии внешнего возмущения, т.е. при *w(t)* ≡ 0, критерий (2.11) принимает вид так называемой γ_0- нормы

$$\gamma_0(\Theta_{[0,N-1]}; R, S) = \sup_{\Delta_{[0,N-1]} \in \boldsymbol{\Delta}_{[0,N-1]}} \sup_{x_0 \neq 0} \left(\frac{\|z\|^2_{[0,\,N-1)} + x^{\mathrm{T}}(N) S x(N)}{x_0^{\mathrm{T}} R^{-1} x_0} \right)^{1/2}, \tag{2.14}$$

характеризующей максимальное значение квадратичного функционала на траекториях системы при начальных состояниях, принадлежащих эллипсоиду $x^T R^{-1} x$ <=1. Если в уравнении (2.1) имеем $C_t \equiv 0$ и $D_t \equiv 0$, то получим верхнюю границу так называемого максимального уклонения терминального состояния замкнутой неопределенной системы

$$\gamma_N(\Theta_{[0,N-1]}; R, S) = \sup_{\Delta_{[0,N-1]} \in \mathbf{\Delta}_{[0,N-1]}} \sup_{x_0,\, w} \left(\frac{x^{\mathrm{T}}(N) S x(N)}{x_0^{\mathrm{T}} R^{-1} x_0 + \|w\|^2_{[0,\,N-1]}} \right)^{1/2} \tag{2.15}$$

Задача заключается в том, чтобы, не имея и не строя математическую модель системы, синтезировать управление, при котором один из перечисленных выше критериев ограничен заданной константой, т.е. в общем случае $\gamma_{g\infty}(\Theta_{[0,N-1]}) \leq \gamma$.

Цель состоит в синтезе последовательности матриц управления Θ_t, обеспечивающих минимизацию верхней границы H_∞-нормы замкнутой системы с учётом переходных процессов для всех $\mathbf{\Delta_t}$.

3. Обобщенное H_∞- управление на конечном интервале

Приведем основные шаги получения по экспериментальным и априорным данным верхней границы обобщенной H_∞-нормы и соответствующих матриц параметров Θ законов управления для неопределенной замкнутой системы

$$\begin{gathered} x(t+1) = (A_t + B_t\Theta_t)x(t) + w(t), \\ z(t) = (C_t + D_t\Theta_t)x(t) \end{gathered} \tag{3.1}$$

Предполагая, что параметры Θ выбраны так, что замкнутая система устойчива, и учитывая введенные обозначения, представим эти уравнения в виде

$$\begin{gathered} x(t+1) = (I_{n_x}\ 0_{n_x \times n_z})\, \Delta_t \begin{pmatrix} I_{n_x} \\ \Theta_t \end{pmatrix} x(t) + w(t), \\ z(t) = (0_{n_z \times n_x}\ I_{n_z})\, \Delta_t \begin{pmatrix} I_{n_x} \\ \Theta_t \end{pmatrix} x(t), \end{gathered} \tag{3.2}$$

где Δ_t – неизвестная $(n_x + n_z) \times (n_x + n_u)$-матрица, Θ_t - $(n_u \times n_x)$ - матрица параметров регулятора. Уравнения двойственной системы имеют вид

$$\begin{gathered} x_d(t+1) = \begin{pmatrix} I \\ \Theta_{N-1-t} \end{pmatrix}^{\mathrm{T}} \Delta^{\mathrm{T}}_{N-1-t} \begin{pmatrix} I \\ 0 \end{pmatrix} x_d(t) + \begin{pmatrix} I \\ \Theta_{N-1-t} \end{pmatrix}^{\mathrm{T}} \Delta^{\mathrm{T}}_{N-1-t} \begin{pmatrix} 0 \\ I \end{pmatrix} w_d(t), \\ z_d(t) = x_d(t). \end{gathered} \tag{3.3}$$

Рассмотрим расширенную систему с дополнительными искусственными входом $w_\Delta(t) \in R^{n_x+n_u}$ и выходом $z_\Delta(t) \in R^{n_x+n_u}$, определяемую уравнениями

$$\begin{gathered} x_a(t+1) = \begin{pmatrix} I \\ \Theta_{N-1-t} \end{pmatrix}^{\mathrm{T}} w_\Delta(t), \\ z_a(t) = x_a(t), \quad z_\Delta(t) = \begin{pmatrix} I \\ 0 \end{pmatrix} x_a(t) + \begin{pmatrix} 0 \\ I \end{pmatrix} w_a(t), \end{gathered} \tag{3.4}$$

в которых $x_a(t) \in R^{n_x}$ - состояние, $w_a(t) \in R^{n_z}$ - возмущение, $z_a(t) \in R^{n_x}$- целевой выход. Допустим, что дополнительные входной и выходной сигналы в системе (3.4) при всех $t \geq 0$ удовлетворяют неравенству

$$\begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix}^{\mathrm{T}} \Psi^{(k)}(N-1-t) \begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix} \leqslant 0, \quad k = 1, 2, \tag{3.5}$$

где матрицы $\Psi_1^k(t)$ заданы в (2.8) и (2.11). Множество сигналов $w_\Delta(t)$ обозначим как $\boldsymbol{W_\Delta}$.

Определим уровень гашения возмущений с весовыми матрицами S > 0, R > 0 системы (3.4), (3.5) как:

$$\widehat{\gamma}_{g\infty}(S, R) = \sup_{w_\Delta \in \mathbf{W_\Delta}} \sup_{x_a(0),\, w_a} \left(\frac{\|z_a\|^2_{[0,\,N-1]} + x_a^{\mathrm{T}}(N) R x_a(N)}{x_a^{\mathrm{T}}(0) S^{-1} x_a(0) + \|w_a\|^2_{[0,\,N-1]}} \right)^{1/2}. \tag{3.6}$$

Заметим, что $w_\Delta(t) = \Delta^{\mathrm{T}} z_\Delta(t) \in \boldsymbol{W_\Delta}$ и, следовательно, система (3.3) при всех $\boldsymbol{\Delta}_{[0,N-1]} \in \boldsymbol{\Delta}$ “погружена” в расширенную систему (3.4), (3.5), а обобщённая H_∞-норма с весовыми матрицами S > 0 и R > 0 двойственной системы (3.3) не превышает уровень гашения возмущений. С

учетом леммы 3.1.[10] это позволяет получить верхнюю границу обобщенной H_∞-нормы неопределенной системы, используя соответствующее свойство расширенной системы.

Как показано в Теореме 4.1. [10] верхняя граница обобщенной H_∞-нормы неопределенной системы (2.1) при законе управления $u(t) = \Theta_t x(t)$, где $\Theta_{N-1-t} = Q_{t+1}P_{t+1}^{-1}$, t = 0,…N-1, удовлетворяет неравенству $\gamma_{g\infty}\left(\Theta_{[0,N-1]}; R, S\right) \leq \gamma$, если следующие линейные матричные неравенства разрешимы относительно $P_t = P_t^{\mathrm{T}} > 0$, Q_t, $\mu_t^{(k)} \geq 0$ k =1,2

$$\begin{pmatrix} -P_{t+1} & \star & \star & \star & \star \\ P_{t+1} & -\sum_{k=1}^{2}\mu_t^{(k)}\Psi_{11}^{(k)} & \star & \star & \star \\ Q_{t+1} & -\sum_{k=1}^{2}\mu_t^{(k)}\Psi_{21}^{(k)} & -\sum_{k=1}^{2}\mu_t^{(k)}\Psi_{22}^{(k)} & \star & \star \\ 0 & -\sum_{k=1}^{2}\mu_t^{(k)}\Psi_{31}^{(k)} & -\sum_{k=1}^{2}\mu_t^{(k)}\Psi_{32}^{(k)} & -P_t + I - \sum_{k=1}^{2}\mu_t^{(k)}\Psi_{33}^{(k)} & \star \\ 0 & -\sum_{k=1}^{2}\mu_t^{(k)}\Psi_{41}^{(k)} & -\sum_{k=1}^{2}\mu_t^{(k)}\Psi_{42}^{(k)} & -\sum_{k=1}^{2}\mu_t^{(k)}\Psi_{43}^{(k)} & -\gamma^2 I - \sum_{k=1}^{2}\mu_t^{(k)}\Psi_{44}^{(k)} \end{pmatrix} \leqslant 0,$$

$$t = 0, \ldots, N-1, \quad P_N = R, \quad P_0 \leqslant \gamma^2 S^{-1}, \tag{3.7}$$

В которых $\Psi_{ij}^{(k)} = \Psi_{ij}^{(k)}$(N-1-t), i, j = 1,…,4 обозначают соответствующие блоки матриц $\Psi^{(k)}$ (N-1-t), а матрицы $\Psi^{(\mathrm{k})}(t)$ заданы в (2.8) и (2.11).

Обозначим через $\gamma_{g\infty}^*$, $\gamma_{g\infty}^{(a)}$, $\gamma_{g\infty}^{(p)}$ мнимальные значения γ, для которых неравенства (3.7) разрешимы при $\mu_t^{(k)} \geq 0$, k = 1,2, при $\mu_t^{(1)_I} \equiv 0$, $\mu_t^{(2)} \geq 0$ и при $\mu_t^{(1)} \geq 0$ соответственно. Другими словами, это гарантированные значения обобщенной H_∞ −нормы замкнутой системы, которые могут быть достигнуты при законах управления, синтезированных по экспериментальным и априорным данным, только по априорным и только по экспериментальным данным соответственно. Из теоремы 4.1.[10] непосредственно следует неравенство

$$\gamma_{g\infty}^* \leq \min\{\gamma_{g\infty}^{(a)}, \gamma_{g\infty}^{(p)}\},$$

которое объясняет преимущество робастных законов управления, построенных по априорным и экспериментальным данным, над робастными законами управления, построенными только по априорным или только по экспериментальным данным. А именно, с одной стороны, когда априорная информация достаточно грубая, т.е. радиусы p_t матричных сфер в (2.9) достаточно велики и соответственно $\gamma_{g\infty}^{(a)}$ принимает большое значение, то $\gamma_{g\infty}^*$ может оказаться малой, если помехи в измерениях не очень большие, т. е. небольшими являются матричные эллипсоиды $\Delta_t^{(p)}$. С другой стороны, если помехи в измерениях оказываются достаточно большими и соответственно большой является $\gamma_{g\infty}^{(p)}$ или, более того, если не выполняется ранговое условие, и информационная матрица является вырожденной, а матричные эллипсоиды $\Delta_t^{(\mathrm{a})}$ оказываются неограниченными, то $\gamma_{g\infty}^{(p)}$ может, тем не менее, стать малой за счет малости радиусов матричных сфер при использовании априорной информации.

4. Иллюстративный пример

Для демонстрации работоспособности метода рассмотрим дискретную модель уравнения Матье, описывающего параметрические колебания линейного осциллятора:

$$\frac{\mathrm{d}^2\varphi}{d\,\tau^2} + \omega_0^2(1 + \varepsilon \sin\omega\,\tau)\varphi = \mathrm{u} + \mathrm{v},$$

Данные генерируются следующими уравнениями:

$$x(t+1) = \begin{pmatrix} 1 & h \\ -\omega_0^2[1+\varepsilon\sin(\omega th)]h & 1 \end{pmatrix} x(t) + \binom{0}{h} u(t) + \binom{0}{h} v$$

$$z(t) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} x(t) + \binom{0}{1} u(t), \qquad t = 0, \ldots, N-1$$

где $x(t) = col(\varphi(\mathrm{t}\cdot\mathrm{h}), \dot{\varphi}(\mathrm{t}\cdot\mathrm{h}))$, $|\nu(\mathrm{t}\cdot\mathrm{h})| \leq \mathrm{d}$, $\omega_0 = \pi$, $\omega = 2\pi$, $\varepsilon = 0.01$, h = 0.2, R = 0.1I, S=0.05I

В каждый момент времени существует 12 неизвестных параметров, что в сумме дает 120 параметров в рассматриваемом конечном интервал $N = 10$. В качестве центра матричной сферы $\Delta_t^{(a)}$ в априорной информации определим матрицу Δ_t^*, связанную с осциллятором при ε = 0. Возмущения, начальные условия и управляющие воздействия выбираются случайным образом из интервалов: [-d, d], [-1, 1] и [-1, 1] соответственно.

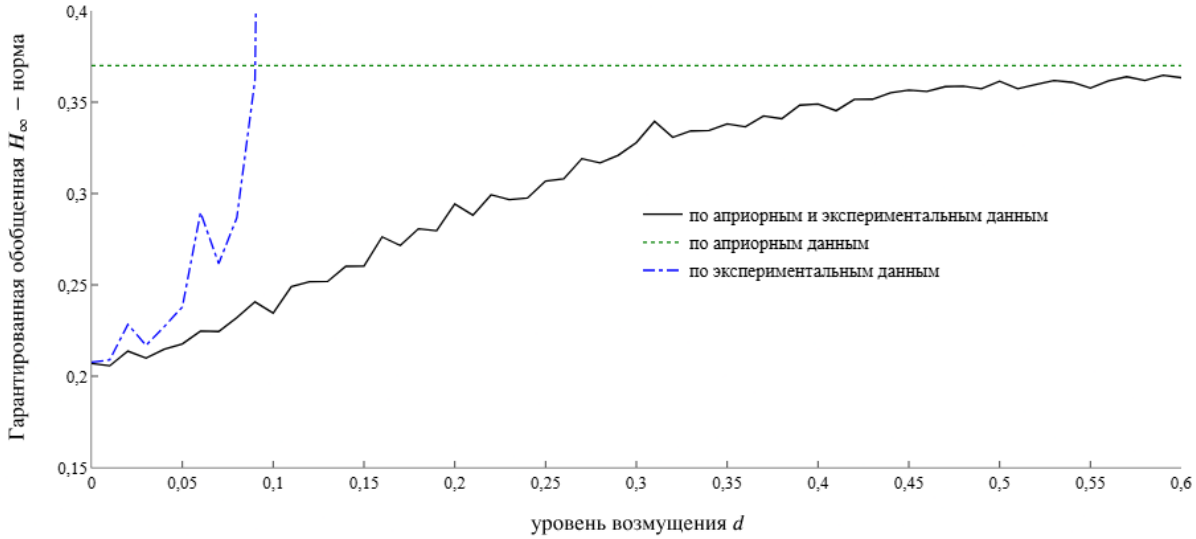


Рис.2. Гарантированные обобщенные H_∞-нормы с учетом переходных процессов при использовании различной информации в зависимости от уровня возмущения d

Рис. 2 изображает квадраты гарантированных H_∞-норм с учетом переходных процессов замкнутой системы при трех сценариях: использование только априорной информации (пунктирная линия $\gamma_{g\infty}^{(a)2}$=0.37 при p_t = 0.02, использование только данных из $L = 3$ экспериментов (штрих-пунктирная кривая ($\gamma_{g\infty}^{(p)}$) и использование как априорной информации, так и данных (сплошная кривая ($\gamma_{g\infty}^{*}$).

Значения $\gamma_{g\infty}^{(a)2}$ и $\gamma_{g\infty}^{(a)2}$для каждого уровня возмущения вычислялись с при использовании априорной информации и данных для усредненных значений 20 экспериментов в зависимости от уровня возмущения d. Когда уровень возмущения d приближается к нулю, данные дают точное описание неопределенности, и $\gamma_{g\infty}^{(*)}$ совпадает с H_∞- нормой с учетом переходных процессов для задачи, основанной на модели, равной γ^2=0.202.

С увеличением уровня возмущения d матричный эллипсоид $\mathbf{\Delta}_t^{(p)}$ и, следовательно, $\gamma_{g\infty}^{(p)2}$существенно растут. Тем не менее, за исключением очень малого уровня возмущения, $\gamma_{g\infty}^{(*)}$ значительно ниже, чем $\gamma_{g\infty}^{(p)2}$, и постоянно остается ниже $\gamma_{g\infty}^{(a)}$.

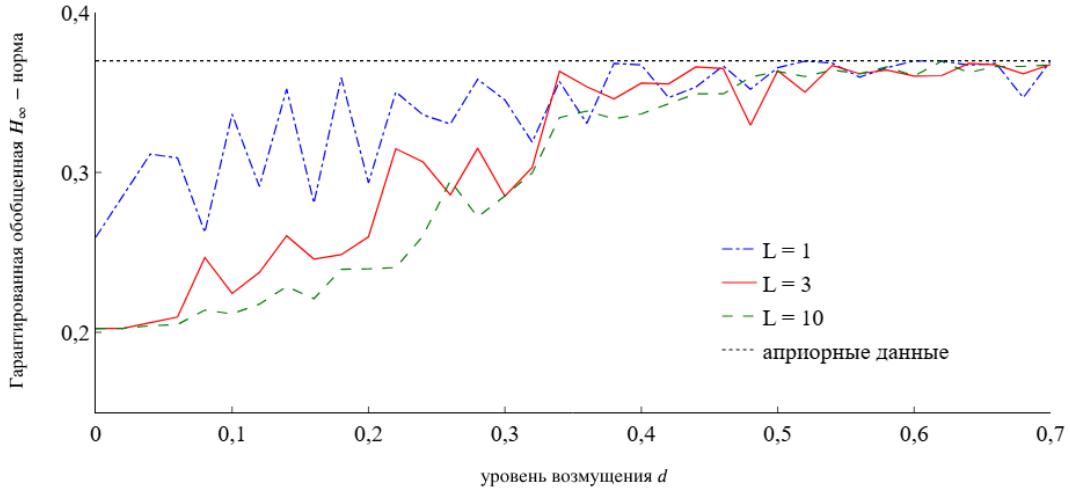


Рис. 3. Гарантированные обобщенные H_∞-нормы как функции уровня возмущения для различного числа экспериментов

На рисунке 3 показана динамика $\gamma_{g\infty}^*$ при использовании априорной информации и данных для усредненных значений 20 экспериментов в зависимости от уровня возмущения d: штрихпунктирная кривая для $L = 1$, сплошная кривая для $L = 3$ и пунктирная кривая для $L = 10$. Пунктирная линия представляет $\gamma_{g\infty}^{(a)2}$=0.37. Из Рис. 3 можно наблюдать, что, во-первых, даже когда $L = 1$ и, следовательно, условие ранга определенно нарушено (т.е. система неидентифицируема), $\gamma_{g\infty}^*$ меньше, чем $\gamma_{g\infty}^{(a)2}$. Во-вторых, для относительно малых уровней возмущения, с увеличением числа экспериментов, $\gamma_{g\infty}^*$ уменьшается. В-третьих, если уровень возмущения d довольно велик, то $\gamma_{g\infty}^*$ слабо зависит от числа экспериментов и приблизительно равен $\gamma_{g\infty}^{(a)2}$. На рисунке 4 показано $\gamma_{g\infty}^*$ при L=3 и различном радиусе матричной сферы априорных данных; показано влияние априорных знаний на эффективность алгоритма.

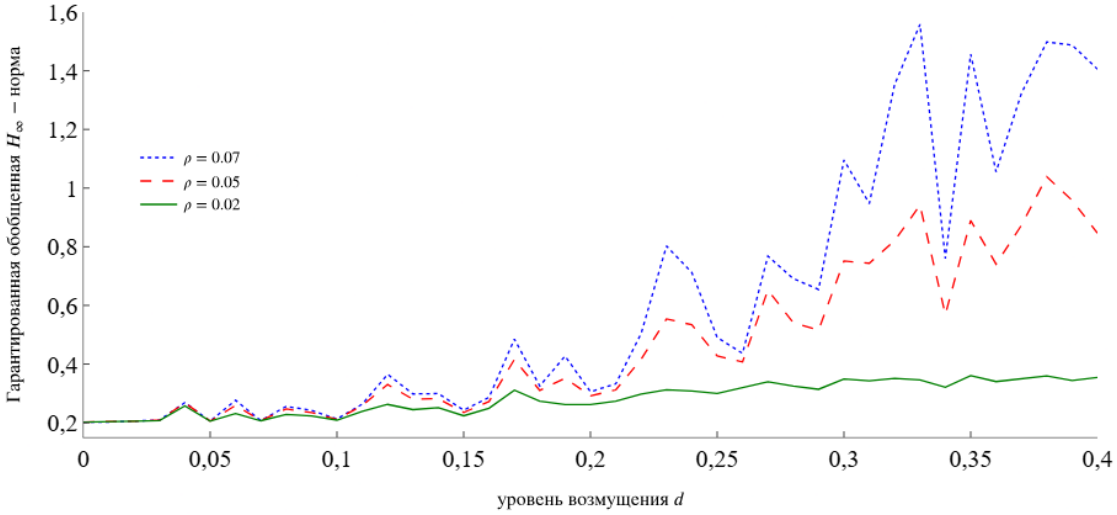


Рис.4. Гарантированные обобщенные H_∞-нормы как функции уровня возмущения при различных радиусах матричных сфер в априорной информации

5. Заключение

В работе предложен метод синтеза оптимальных законов управления для линейных нестационарных систем на конечном интервале времени, основанный на совместном использовании априорной информации и экспериментальных данных. Метод обеспечивает минимизацию обобщённой H_∞-нормы замкнутой системы и характеризует эффективность подавления внеш-

них и начальных возмущений. Показано, что предложенный подход сохраняет устойчивость и качество управления даже при ограниченном объёме данных и отсутствии условий идентифицируемости системы.

Литература

1. Willems J.C., Rapisarda P., Markovsky I., De Moor B. A note on persistency of excitation // Syst. Control Lett. 2005. V. 54. P. 325–329.
2. Petersen I. R. A stabilization algorithm for a class of uncertain linear systems // Syst. Control Lett. 1987. V. 8. P. 351–357.
3. De Persis C., Tesi P. Formulas for Data-Driven Control: Stabilization, Optimality and Robustness // IEEE Trans. Automat. Control. 2020. V. 65. No. 3. P. 909–924.
4. Waarde H. J., Eising J., Trentelman H.L., Camlibel M.K. Data Informativity: a New Perspective on Data-Driven Analysis and Control // IEEE Trans. Automat. Control. 2020. V. 65. No. 11. P. 4753–4768.
5. Waarde H.J., Camlibel M.K., Mesbahi M. From Noisy Data to Feedback Controllers: Nonconservative Design via a Matrix S-Lemma // IEEE Trans. Automat. Control. 2022. V. 67. No. 1. P. 162–175.
6. Якубович В.А. S-процедура в нелинейной теории управления // Вестн. Ленинград. ун.-та. Математика. 1977. Т. 4. С. 73-93.
7. Bisoffi A., De Persis C., Tesi P. Data-driven Control via Petersen's Lemma // Automatica. 2022. V. 145. Article 110537.
8. Berberich J., C.W. Scherer C.W., Allgower F. Combining Prior Knowledge and Data for Robust Controller Design // IEEE Trans. Automat. Control. 2023. V. 68. No. 8. P. 4618 - 4633.
9. Коган М.М., Степанов А.В. Синтез субоптимальных робастных регуляторов на основе априорных и экспериментальных данных // АиТ. 2023. № 8. С. 24 - 42.
10. Коган М.М., Степанов А.В. Как улучшить робастное управление линейной нестационарной системой с помощью экспериментальных данных // АиТ. 2024. № 6. С. 115 – 139.

ГРАФИЧЕСКИЙ УСКОРИТЕЛЬ GEFORCE 3090 И МЕТОД МОНТЕ-КАРЛО КАК УНИВЕРСАЛЬНОЕ РЕШЕНИЕ ОДНОЙ КЛАССИЧЕСКОЙ ОБРАТНОЙ ЗАДАЧИ

А.А. Тимофеев, М.А. Кривов

Московский государственный университет им. М.В. Ломоносова

Вычислительная мощность всего одного современного графического ускорителя оказывается вполне достаточной, чтобы методом ненаправленного поиска решить классическую задачу астрономии, которая ещё 30 лет назад считалась практически нерешаемой и требовала проведения серии упрощений и перехода к автомодельному решению. При движении в атмосфере Земли метеориты светятся менее 10 секунд, в то время как достоверная информация о финальной части их траекторий обычно отсутствует, что затрудняет поиск упавших фрагментов. К счастью, место их падения можно оценить, задействовав популярную систему ОДУ, которая весьма точно описывает движение произвольного объекта в атмосфере Земли. К несчастью, внеземная масса, форма и состав метеоритов неизвестны, поэтому не имеется возможности корректно сформулировать задачу Коши для этой системы ОДУ. В предлагаемом подходе мы с помощью метода Монте-Карло перебираем все неизвестные параметры, после чего траектории подобных виртуальных метеоритов сопоставляем с небольшой частью реально наблюдаемой. Отобрав и усреднив около 1000 наиболее точных совпадений, нам удаётся не только предсказать место падения, но и оценить плотность, массу и форму метеорита. А благодаря графическому ускорителю GeForce 3090, весь этот процесс занимает порядка нескольких минут.

Ключевые слова: метеориты, ОДУ, метод Адамса, метод Монте-Карло, GPGPU, CUDA, высокопроизводительные вычисления

1. Введение

В 1960-ые годы стали появляться сети астрономических станций по типу European Fireball Network, которые с некоторой периодичностью фотографируют ночное небо. Каждый запечатлённый ими метеор или метеорит обычно представлен в виде набора фотографий с разных ракурсов с 5-10 светящимися точками. Так как местоположение станций известно, по таким фотографиям можно достоверно оценить высоту и скорость движения небесных тел в соответствующие этим точкам моменты времени. После анализа эта информация систематизируется по годам, преобразуется в таблицы и регулярно публикуется в тематических журналах [1], хотя для получивших известность метеоритов также появляются отдельные расширенные отчёты [2].

При организации экспедиций для поиска фрагментов метеорита интерес представляют всего два вопроса: «куда примерно он упал?» и «какая у него остаточная масса?». После прекращения свечения метеорит продолжает движение в атмосфере Земли ещё десятки секунд, поэтому для экстраполяции данных от астрономических станций были предложены разные математические модели. Например, в одной из первых моделей оценивалась яркость свечения, что косвенно характеризует интенсивность горения и, как следствие, плотность метеорита и изменение его массы. Другая модель рассматривала только торможение, так как, зная скорость и ускорение, можно вычислить массу тела.

Более современные подходы [3, 4] основаны на решении универсальной системы ОДУ, описывающей движение произвольного объекта в атмосфере некоторой планеты, сопровождающиеся его горением. Также встречаются работы с модификациями этой системы ОДУ [5]. Переменными величинами являются высота над уровнем моря $h(t)$, масса $m(t)$, скорость $V(t)$ и угол её вектора относительно горизонта $\gamma(t)$. О планете необходимо знать радиус R, плотность атмосферы $\rho_{\text{atm}}(h)$ и ускорение свободного падения g. В качестве характеристик самого объек-

та выступают аэродинамические коэффициенты c_d и c_L, эффективная энтальпия разрушения H и коэффициент теплообмена с окружающей средой c_h. Информация о форме объекта и его плотности ρ также необходима, но только для определения площади миделева сечения $S(t)$. В итоге сама система ОДУ принимает следующий вид:

$$m\frac{dV}{dt} = -\frac{1}{2}c_d\rho_a V^2 S + m \cdot g \cdot sin\gamma$$
$$mV\frac{d\gamma}{dt} = m \cdot g \cdot cos\gamma - \frac{mV^2}{R}cos\gamma - \frac{1}{2}c_L\rho_a V^2 S$$
$$\frac{dh}{dt} = -V \cdot sin\gamma$$
$$H\frac{dm}{dt} = -\frac{1}{2}c_h\rho_{atm}V^3 S$$

Как уже отмечалось, в общем случае сформулировать для неё задачу Коши невозможно, так как отсутствует информация о $m(0)$, а для $\gamma(0)$ имеется только грубая оценка. Более того, пока не найдены фрагменты метеорита, неизвестными остаются величины H и ρ, а постоянные и также неизвестные коэффициенты c_d, c_L и c_h являются особенностями модели и не имеют физически обоснованных значений. В [3, 5] эти неизвестные частично заменяли на «разумные» с точки зрения авторов значения. В [3, 4] проводили математические преобразования и сводили число неизвестных до двух коэффициентов α и β, которые затем подбирались на основе табличных данных высоты и скорости.

2. Предложенный подход

В нашей предыдущей работе [6] был описан стохастических метод для решения рассмотренной системы ОДУ. Во-первых, по результатам анализа литературы была предложена физически обоснованная нижняя и верхняя оценка $\overrightarrow{x_{min}} \le \vec{x} \le \overrightarrow{x_{max}}$ для вектора неизвестных параметров $\vec{x} = (m_0, \rho, \gamma_0, c_d, c_L, c_h, H)$. Выбирая случайным образом $\vec{x}$ из этого диапазона, мы корректно формулируем задачу Коши, решение которой соответствует полёту некоторого виртуального метеорита. В отличие от модельных коэффициентов α и β из работ [3, 4], определённые таким образом виртуальные метеориты действительно могли бы существовать.

Во-вторых, мы используем явный многошаговый метод Адамса для численного решения системы ОДУ, переписанной в векторном виде $\frac{d\vec{y}}{dt} = f(t, \vec{y})$ относительно переменной $\vec{y}(t) = \big(m(t), V(t), \gamma(t), h(t), l(t)\big)$ с начальным условием $\vec{y}(0) = \big(x_m, V_0, x_\gamma, h_0, 0\big)$. Дополнительная вспомогательная величина $l(t)$ соответствует текущему расстоянию от начала свечения метеорита и необходима для определения предполагаемого места его падения с учётом кривизны Земли. Например, формула интегрирования для двухшагового варианта метода принимает следующий вид:

$$\vec{y}_i = \vec{y}_{i-1} + \left(\frac{3}{2}f(t_{i-1}, \vec{y}_{i-1}) - \frac{1}{2}f(t_{i-2}, \vec{y}_{i-2})\right) \cdot \Delta t$$

В-третьих, для оценки точности соответствия виртуального метеорита реальному мы сравниваем имеющиеся табличные данные скорости V_{tab} и высоты h_{tab} с найденным численным решением. Функционал соответствия может быть самым разным, в данной работе была использована его следующая форма, где индексы j соответствуют моментам времени, для которых в таблице V_{tab}/h_{tab} имеется нужная запись:

$$F(\vec{y}) = \frac{\sqrt{\frac{\sum\left(V_{tab}(\Delta t \cdot j) - V_j\right)^2}{V_{tab}(0)}} + \sqrt{\frac{\sum\left(h_{tab}(\Delta t \cdot j) - h_j\right)^2}{h_{tab}(0)}}}{\sqrt{N+1}}$$

В итоге рассматриваемый подход сводится к перебору очень большого количества виртуальных метеоритов, среди которых отбираются случаи с наименьшим значением функционала $F(\vec{y})$. Далее они ещё раз пересчитываются, но с использованием более точного трёхшагового метода Адамса, после чего усредняются по каждому оцениваемому параметру из вектора $\vec{x}$.

3. Программная реализация

С точки зрения структуры вычислений последовательная реализация рассмотренного алгоритма оказывается достаточно простой (рис. 1). На вход ему подаются подготовленные табличные данные из статей, подобных [1, 2], после чего от него требуется перебрать миллионы виртуальных метеоритов (строка 4) и сопоставить найденные траектории с фрагментом реальной (строка 17), оставив при это только около тысячи наилучших кандидатов (строка 18). Как и для всех методов Монте-Карло, наиболее важным аспектом является выбор качественного генератора случайных чисел (строки 1 и 5), в роли которого у нас выступает алгоритм вихря Мерсенна std::mt19937 из стандартной библиотеки C++17. Моделирование отдельно взятого виртуального метеорита может завершиться как при его полном сгорании или соударении с Землёй (строка 13), так и в случаях, когда время превышает некоторый разумный порог (строка 16).

```
1  int seed ← GenerateRngSeed();
2  tuple_int_real[] best_seeds;
3  real[] t_tab, V_tab, h_tab ← ReadTrajectory("Innisfree.txt");
4  for (int i : [0, N])
5      real[7] x ← GenerateVirtualMeteorite(seed);
6      real[5] y ← { x[0], V_tab[0], x[2], h_tab[0], 0 };
7      real t ← 0;
8      real[] V_virt, h_virt;
9      do
10         if (Contains(t_tab, t))
11             V_virt += y[1];
12             h_virt += y[4];
13         if (y[0] < 0.01 or y[3] <= 0.0) break;
14         y ← OneStepAdams(y, x, dt);
15         t ← t + dt;
16     while (t < t_end);
17     real f ← ComputeL2(V_virt, V_tab, h_virt, h_tab);
18     InsertSortAndTruncate(best_seeds, seed, f);
```

Рис. 1. Схема предложенного алгоритма на примере одношагового метода Адамса. Все исходные коды доступны в репозитории проекта (https://github.com/m-krivov/MeteoriteSimulator)

Для адаптации этого алгоритма под GPU достаточно переписать основной цикл (строка 4) в виде CUDA-ядра с аналогичным количеством нитей, а генератор случайных чисел заменить на вызовы библиотеки cuRAND, однако при этом возникает несколько существенных проблем. Например, для отбора наилучших виртуальных метеоритов требуется поддерживать в отсортированном состоянии массив с их зёрнами (строка 18), для чего необходимы достаточно медленные атомарные операции. Чтобы этого избежать, мы перешли к двухуровневой структуре алгоритма, заменив основной цикл (строка 4) на два. В первом из них, итерации которого соответствуют CUDA-нитям, создаются свои локальные массивы с четырьмя наилучшими кандидатами, которые в дальнейшем выгружаются в глобальную память для последующего слияния. Во втором (вложенном) цикле каждая нить перебирает несколько тысяч доставшихся ей виртуальных метеоритов, отбирая эти четыре кандидата.

Другой проблемой оказалось расхождение нитей внутри варпа, так как некоторые виртуальные метеориты могут досрочно сгореть или столкнуться с Землёй. Из-за особенностей архитектуры SIMT такие нити не смогут перейти к очередному метеориту и будут простаивать, ожидая своих соседей по варпу. В качестве возможного решения мы сделали альтернативный вариант вложенного цикла, в котором обход осуществляется не по отдельным метеоритам, а сразу по всем шагам метода Адамса. Таким образом, первая нить может только завершать моделирование своего первого метеорита, в то время как вторая уже успеет обработать три или четыре случая. Однако в подобной схеме возникает дополнительное их расхождение при генерации векторов $\vec{x}$ и $\vec{y}(0)$ (строки 5 и 6), поэтому данная оптимизация не является универсальным решением.

4. Производительность

Масштаб расчётов, необходимых для анализа полёта одного метеорита с помощью предложенного подхода, можно оценить по замерам из Табл. 1. Доработки и улучшения математической модели обычно осуществляются на базе последовательной версии, которая за три целевые минуты способна перебрать всего 1 млн. виртуальная метеоритов. Для отладки CUDA-ядер существует отдельная версия, которая использует тот же генератор случайных чисел на основе вихря Мерсенна и то же зерно, что и последовательная реализация. Это приводит к необходимости частых пересылок данных между памятью CPU и GPU, поэтому итоговое ускорение оказывается весьма скромным – порядка 3 раз.

Таблица 1. Сравнение времени расчётов при моделирования полёта метеорита Иннисфри [2] с использованием разных реализаций алгоритма. Так как проект изначально ориентирован на GPU, первые два столбца приведены в качестве иллюстрации. Все времена указаны в секундах

Перебираемые виртуальные метеориты	**Реализация для CPU**, AMD Ryzen 9 5950X, одно ядро	**Реализация для GPU**, генератор случайных чисел std::mt19937 (C++17)	**Реализация для GPU**, генератор случайных чисел cuRAND
100 тыс.	12.8	6.7	-
1 млн.	172.2	63.5	-
10 млн.	1270.6	631.8	-
344 млн.	-	-	53.5
688 млн.	-	-	131.6
1 032 млн.	-	-	199.9

С другой стороны, реализация, изначально разработанная под архитектуру GPU с использованием библиотеки cuRAND, за три минуты способна перебрать чуть менее 1 млрд. случаев. Ускорение относительно одного ядра CPU составляет около 900 раз, и подобную производительность было бы сложно достичь, даже проведя идеальное распараллеливание по всем ядрам (x16) и их векторным расширениям AVX2 (x8). Результаты моделирования не являются идентичными, так как на GPU каждый блок нитей независимо обрабатывает по $512 \cdot 1024$ случаев, отбирая из них ровно 256 кандидатов. Теоретически, это может привести к потере некоторых виртуальных метеоритов, идеально совпадающих с изучаемым. Однако подобная структура алгоритма позволяет осуществлять расчёты полностью внутри потокового мультипроцессора, так как каждый блок нитей всего один раз читает из глобальной памяти зёрна генератора случайных чисел и табличные данные высоты/скорости (1.1 килобайта), и также один раз сохраняет «хорошие» виртуальные метеориты (256 значений типа curandState, в сумме около 10 килобайт). Как следствие, все расчёты проходят без каких-либо барьерных синхронизаций, что для GPU является оптимальным сценарием.

Отдельный интерес представляет эффект от изменения структуры цикла внутри нити. В табл. 1 была использована версия алгоритма, когда каждая нить осуществляет обход всех доставшихся ей метеоритов, а потом выгружает результаты в глобальную память (далее «цикл по метеоритам»). Альтернативой ей является перемешивание всех метеоритов и переход к единому «циклу по шагам интегрирования». Как отмечалось ранее, это одновременно и уменьшает, и увеличивает расхождение нитей внутри варпа, поэтому выбор предпочтительной реализации можно сделать только по результатам тестов.

На рис. 2 приведены подобные замеры для случаев, когда моделирование проводилось с разным пределом времени полёта метеорита. Процесс их свечения в среднем длится лишь несколько секунд [1], а у крупных метеороидов, обычно попадающих в заметки новостных изданий, он редко превышает 10 секунд. Таким образом, для нашего метода имеются эталонные табличные данные всего для 3-10 секунд, и в этом диапазоне оптимизация с перемешиванием шагов интегрирования оказывается малополезной (рис. 2а). За столь небольшой период лишь немногие виртуальные метеориты сгорают полностью, поэтому расхождение нитей внутри варпа как таковое не наблюдается, в то время как переключения между метеоритами в случае «цикл по шагам интегрирования» лишь вносят дополнительные издержки.

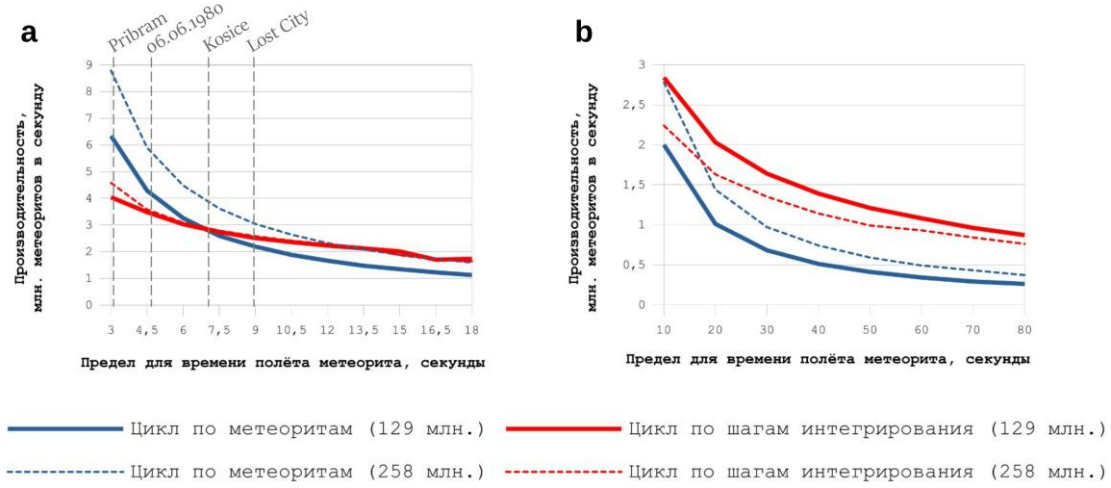


Рис. 2. Сравнение двух реализаций алгоритма при варьировании продолжительности времени полёта виртуальных метеоритов и их общего числа. В качестве иллюстрации на первом графике приведены моменты, когда некоторые реальные метеориты прекращали светиться, из-за чего достоверная информация об их дальнейшей траектории отсутствует

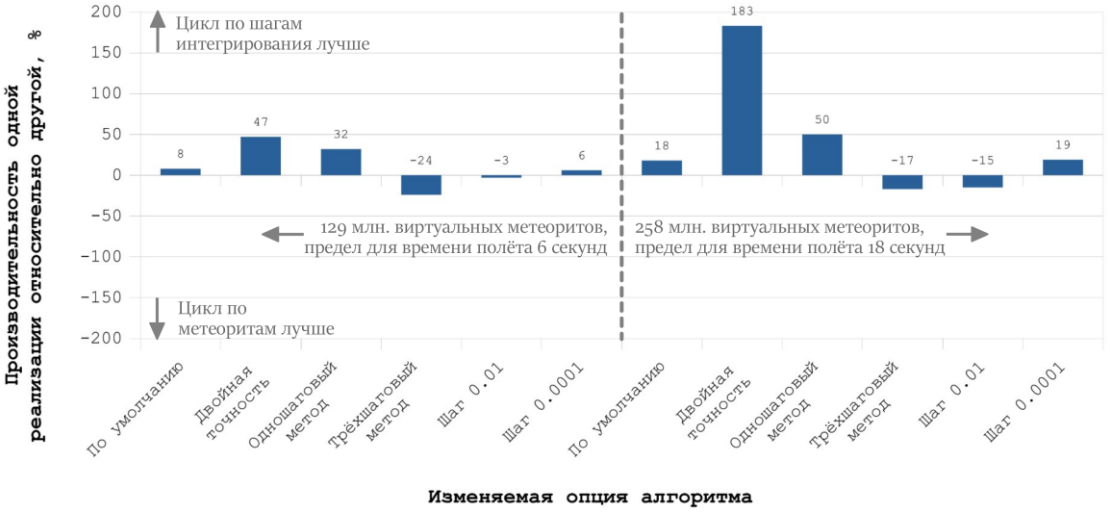


Рис. 3. Соотношение производительности двух реализаций алгоритма при использовании разных настраиваемых опций. Вариант «По умолчанию» соответствует случаю «Одинарная точность, Двухшаговый метод Адамса, Шаг по времени 10^{-3}»

Если же расширить предел времени с 10 до 80 секунд (Рис. 2b), то ситуация ожидаемо меняется. Некоторые виртуальные метеориты за это время сгорают, другие соударяются с поверхностью Земли, а каждый варп вынужден выполняться до тех пор, пока последний из его метеоритов не закончит свой полёт (или не будет превышено пороговое время). Как следствие, с некоторого момента польза от перемешивания шагов интегрирования действительно начинает превышать вносимые накладные расходы, что делает вторую реализацию более производительной. С практической точки зрения подобное искусственное увеличение порога времени полета имеет мало смысла. Однако, теоретически, оно может потребоваться для отсечения немногочисленных случаев, когда метеориты, подобно самолётам, начинают набирать высоту и улетают обратно в космос, так как в исходной системе ОДУ учитывается подъёмная сила.

Предел для времени, при котором имеет смысл переходить от «цикла по метеоритам» к «циклу по шагам интегрирования» зависит как от общего числа виртуальных метеоритов, так и опций алгоритма, поэтому на Рис. 3 дополнительно приведено соотношение производительности этих двух реализаций. Можно заметить, что наибольший эффект наблюдается при выборе двойной точности и, независимо, одношагового метода Адамса.

5. Заключение

Представленные результаты являются промежуточным, но достаточно важным шагом для развития разрабатываемого авторами открытого пакета для моделирования метеоритов. Стоит признать, что схожие оценки можно получить, заменив метод Монте-Карло на любой подходящий метод многомерной оптимизации [3], в результате чего потребность в рассмотренных высокопроизводительных вычислениях полностью пропадёт. Однако в таком случае ответом становится единственное значение, например, что угол начала траектории строго равен 64.01º [3], которое из-за ограничений математической модели не обязательно соответствует реальным 67º. Наш же стохастический подход позволяет оценить разброс значений и посчитать средне-квадратичное отклонение, что косвенно укажет на возможные неточности в результатах моделирования, или, наоборот, подтвердит достоверность расчётов.

Другой потенциально дискуссионной особенностью данной работы является перебор порядка 10^9 виртуальных метеоритов, хотя для оптимизации семи искомых параметров достаточно рассмотреть всего 10^6 – 10^7 случаев. Данный порог обусловлен необходимостью дальнейшего развития модели за счёт добавления вращения (ось и угловая скорость, четыре параметра), формы (эллипсоид, два параметра) и фрагментации метеорита [7] (пропорция, один параметр на каждый случай дробления). Последний аспект представляет наибольший интерес, так как в современных работах [4] фиксируют табличные данные высоты и скорости как самого метеорита, так и отделяющихся от него более мелких фрагментов. Таким образом, после обновления нашей математической модели количество варьируемых параметров возрастёт минимум до тринадцати, поэтому изначально заложенный порог в миллиарды виртуальных метеоритов является вполне обоснованным.

Литература

1. Halliday I., Griffin A.A., Blackwell A.T. Detailed data for 259 fireballs from the Canadian camera network and inferences concerning the influx of large meteoroids // Meteoritics & Planetary Science. 1996. V. 31. No. 2. P. 185–217.
2. Halliday I. Et al. The Innisfree meteorite fall - A photographic analysis of fragmentation, dynamics and luminosity // Meteoritics. 1981. Vol. 16, No. 2. P. 153–170.
3. Gavrilov A., Lebedev M. Determination of meteorite trajectories: observations and modeling // Computational Mathematics and Modeling. 2018. Vol. 29, No. 3. P. 265–274.
4. Gritsevich M. et al. A comprehensive study of distribution laws for the fragments of Košice meteorite // Meteoritics & planetary science. 2014. Vol. 49, No. 3, P. 328–345. DOI: 10.1111/maps.12252.
5. Иванков А.А., Финченко В.С., Численное исследование теплового разрушения метеорита «Челябинск» при входе в атмосферу Земли // Компьютерные Исследования и Моделирование. 2013. Т. 5. № 6. С. 941–956.
6. Krivov M.A. A Numerical Approach to Deriving Statistical Estimates of the Residual Mass, the Impact Point, and Other Meteorite Parameters from the Bright Section of the Trajectory // Computational Mathematics and Modeling. 2021. Vol. 32. P. 428–437. DOI: 10.1007/s10598-022-09544-3.
7. Брыкина И. Г. и др. О моделях фрагментации метеороидов в атмосфере // Физико-химическая кинетика в газовой динамике. 2019. Т. 20. №. 2. С. 1–10.

СТРУКТУРА ЛЕВОГО НУЛЬ-ПРОСТРАНСТВА МАТРИЦЫ ОГРАНИЧЕНИЙ ЧЕТЫРЁХИНДЕКСНОЙ ПЛАНАРНОЙ ТРАНСПОРТНОЙ ЗАДАЧИ

Е.Б. Титова, И.Н. Алпутов, О.А. Петров, Р.Д. Махнёв

Нижегородский государственный университет им. Н.И. Лобачевского

Для матрицы ограничений четырёхиндексной планарной транспортной задачи рассмотрена структура левого нуль-пространства, построен базис L данного пространства, получен спектр собственных чисел для матрицы $L \cdot L^{\mathrm{T}}$.

Ключевые слова: четырёхиндексная планарная транспортная задача, левое нуль-пространство, собственные числа.

1. Введение

Пусть $\mathbb{Q}-$ поле рациональных чисел, $\mathbb{Q}^n-$ линейное пространство столбцов с компонентами из $\mathbb{Q}$, а $\mathbb{Z}^n-$ столбцы из $\mathbb{Q}^n$ с целочисленными компонентами. Для $(m \times n)$-матрицы A с элементами из $\mathbb{Q}$ обозначим через $L_A = \{u \in \mathbb{Q}^m | uA = 0\}$ левое нуль-пространство матрицы A, а через $L-$ какой-нибудь его базис. Можно считать, что $L-$ целочисленная $(m - r \times m)$-матрица, где $r-$ ранг матрицы A. Множество $L_A^Z = L_A \cap \mathbb{Z}^m$ назовем левым модулем матрицы A. Множество целочисленных линейных комбинаций строк матрицы L содержится в L_A^Z, однако зачастую с ним не совпадает. Они совпадут тогда и только тогда, когда матрица L будет унимодулярная (т.е. НОД базисных миноров равен 1). В этом случае L будем называть базисом левого модуля матрицы A.

Обозначим через $M(A, b)$ множество рациональных решений системы $Ax = b$. Базис L левого нуль-пространства L_A позволяет описать множество всех векторов $b \in \mathbb{Q}^m$, при которых $M(A, b) \neq \emptyset$. По теореме Фредгольма [1] оно совпадает с множеством $\{b \in \mathbb{Q}^m | Lb = 0\}$. Аналогичный вопрос над кольцом целых чисел $\mathbb{Z}$ решается сложнее, однако имеет место следующий результат: если $b \in \mathbb{Z}^m$, $A-$ унимодулярна и $L-$ полная система векторов в L_A, то $M(A, b) \cap \mathbb{Z}^m \neq \emptyset$ тогда и только тогда, когда $Lb = 0$.

Известно, что если матрица A содержит базисную систему столбцов $A^{\mathcal{B}}$ $(|\mathcal{B}| = r)$, содержащую базисный минор, равный 1, то (см., например, [1]) $|\det A^{\mathcal{B}}(\mathcal{J})| = |\det L_A^Z(\bar{\mathcal{J}})|$, где $\bar{\mathcal{J}}-$ множество, дополнительное к $\mathcal{J}$. Таким образом, количественный анализ базисных миноров матриц $A^{\mathcal{B}}$ и L_A^Z, равносилен.

В данной работе для левого нуль-пространства L_A четырёхиндексной планарной транспортной задачи получен спектр собственных чисел для матрицы $L \cdot L^T$ и построен базис $L_{\bar{\mathcal{B}}}$.

2. Постановка задачи

Рассмотрим ограничения-равенства планарной четырёхиндексной транспортной задачи (см., например, [2]) с неизвестными $x_{j_1 j_2 j_3 j_4} \geq 0$ ($j_i = 1, \dots, n_i$ при $i = 1, 2, 3, 4$):

$$\sum_{j_1=0}^{n_1} x_{j_1 j_2 j_3 j_4} = b_{0 j_2 j_3 j_4},$$

$$\sum_{j_2=0}^{n_2} x_{j_1 j_2 j_3 j_4} = b_{j_1 0 j_3 j_4},$$

$$\sum_{j_3=0}^{n_3} x_{j_1 j_2 j_3 j_4} = b_{j_1 j_2 0 j_4},$$

$$\sum_{j_4=0}^{n_4} x_{j_1 j_2 j_3 j_4} = b_{j_1 j_2 j_3 0}$$

и обозначим через $T(n_1, n_2, n_3, n_4)$ ее матрицу.

Из [3] следует, что матрицу $T(n_1, n_2, n_3, n_4)$ можно представить в виде строчечных блоков:

$$T(n_1, n_2, n_3, n_4) = \begin{pmatrix} I_{n_1} \times E_{n_2} \times E_{n_3} \times E_{n_4} \\ E_{n_1} \times I_{n_2} \times E_{n_3} \times E_{n_4} \\ E_{n_1} \times E_{n_2} \times I_{n_3} \times E_{n_4} \\ E_{n_1} \times E_{n_2} \times E_{n_3} \times I_{n_4} \end{pmatrix},$$

где $A \times B -$ кронекерово произведение матриц A и B (определение и свойства см., например, в [4]), $E_n -$ единичная матрица n-го порядка, $I_n = (1, \dots, 1) - n$-мерная строка, все компоненты которой равны 1.

Матрица $T(n_1, n_2, n_3, n_4)$ имеет число столбцов $N_T = n_1 n_2 n_3 n_4$, число строк $M_T = n_1 n_2 n_3 + + n_1 n_2 n_4 + n_1 n_3 n_4 + n_2 n_3 n_4$ и ранг $r_T = n_1 n_2 n_3 + n_1 n_2 n_4 + n_1 n_3 n_4 + n_2 n_3 n_4 - n_1 n_2 - - n_1 n_3 - n_1 n_4 - n_2 n_3 - n_2 n_4 - n_3 n_4 + n_1 + n_2 + n_3 + n_4 - 1$. Матрица $T(n_1, n_2, n_3, n_4)$ содержит базисную систему столбцов $T^{\mathcal{B}}$ ($|\mathcal{B}| = r$), имеющую минор порядка r, равный 1.

3. Основные результаты

Рассмотрим левое нуль-пространства матрицы $T(n_1, n_2, n_3, n_4)$.

В [1] дано описание построения левого нуль-пространства. Следуя ему, матрицу $L(n_1, n_2, n_3, n_4)$ можно представить в блочном виде:

$$L(n_1, n_2, n_3, n_4) =$$
$$= \begin{pmatrix} -I_1 \times I_{n_2} \times E_{n_3} \times E_{n_4} & I_{n_1} \times I_1 \times E_{n_3} \times E_{n_4} & 0_{n_3 n_4 \times n_1 n_2 n_4} & 0_{n_3 n_4 \times n_1 n_2 n_3} \\ -I_1 \times E_{n_2} \times I_{n_3} \times E_{n_4} & 0_{n_2 n_4 \times n_1 n_3 n_4} & I_{n_1} \times E_{n_2} \times I_1 \times E_{n_4} & 0_{n_2 n_4 \times n_1 n_2 n_3} \\ 0_{n_1 n_4 \times n_2 n_3 n_4} & -E_{n_1} \times I_1 \times I_{n_3} \times E_{n_4} & E_{n_1} \times I_{n_2} \times I_1 \times E_{n_4} & 0_{n_1 n_4 \times n_1 n_2 n_3} \\ -I_1 \times E_{n_2} \times E_{n_3} \times I_{n_4} & 0_{n_2 n_3 \times n_1 n_3 n_4} & 0_{n_2 n_3 \times n_1 n_2 n_4} & I_{n_1} \times E_{n_2} \times E_{n_3} \times I_1 \\ 0_{n_1 n_3 \times n_2 n_3 n_4} & -E_{n_1} \times I_1 \times E_{n_3} \times I_{n_4} & 0_{n_1 n_3 \times n_1 n_2 n_4} & E_{n_1} \times I_{n_2} \times E_{n_3} \times I_1 \\ 0_{n_1 n_2 \times n_2 n_3 n_4} & 0_{n_1 n_2 \times n_1 n_3 n_4} & -E_{n_1} \times E_{n_2} \times I_1 \times I_{n_4} & E_{n_1} \times E_{n_2} \times I_{n_3} \times I_1 \end{pmatrix},$$

где $0_{m \times n} -$ нулевая $(m \times n)$-матрица, $I_1 -$ число 1.

Непосредственно перемножая матрицы, получаем, что

$$L(n_1, n_2, n_3, n_4) \cdot T(n_1, n_2, n_3, n_4) = 0.$$

Матрица $L(n_1, n_2, n_3, n_4)$ имеет число строк $M_L = n_1 n_2 + n_1 n_3 + n_1 n_4 + n_2 n_3 + n_2 n_4 + + n_3 n_4$, число столбцов $N_L = n_1 n_2 n_3 + n_1 n_2 n_4 + n_1 n_3 n_4 + n_2 n_3 n_4$ и имеет размерность $d_L(n_1, n_2, n_3) = M_T - r_T = n_1 n_2 + n_1 n_3 + n_1 n_4 + n_2 n_3 + n_2 n_4 + n_3 n_4 - n_1 - n_2 - n_3 - n_4 + 1$.

Утверждение 1. *Собственные числа матрицы* $L(n_1, n_2, n_3, n_4) \cdot L^T(n_1, n_2, n_3, n_4)$ *имеют следующий вид*:

$\lambda_1 = n_1 + n_2 + n_3 + n_4$ *кратности 3,*
$\lambda_2 = n_1 + n_2 + n_3$ *кратности* $2(n_4 - 1)$, $\lambda_3 = n_1 + n_2 + n_4$ *кратности* $2(n_3 - 1)$,
$\lambda_4 = n_1 + n_3 + n_4$ *кратности* $2(n_2 - 1)$, $\lambda_5 = n_2 + n_3 + n_4$ *кратности* $2(n_1 - 1)$,
$\lambda_6 = n_1 + n_2$ *кратности* $(n_3 - 1)(n_4 - 1)$, $\lambda_7 = n_1 + n_3$ *кратности* $(n_2 - 1)(n_4 - 1)$,
$\lambda_8 = n_1 + n_4$ *кратности* $(n_2 - 1)(n_3 - 1)$, $\lambda_9 = n_2 + n_3$ *кратности* $(n_1 - 1)(n_4 - 1)$,
$\lambda_{10} = n_2 + n_4$ *кратности* $(n_1 - 1)(n_3 - 1)$, $\lambda_{11} = n_3 + n_4$ *кратности* $(n_1 - 1)(n_2 - 1)$
и $\lambda_{12} = 0$ *кратности* $n_1 + n_2 + n_3 + n_4 - 1$.

Выделим в матрице $T(n_1, n_2, n_3, n_4)$ базисную систему столбцов $T^{\mathcal{B}}$ ($|\mathcal{B}| = r$), а в матрице $L(n_1, n_2, n_3, n_4)$ базисную систему строк $L_{\overline{\mathcal{B}}}$.

Для построения базиса $L_{\overline{\mathcal{B}}}$ достаточно удалить из матрицы $L(n_1, n_2, n_3, n_4)$ в блоке III последние n_4 строки, в блоке V - последние n_3 строки, каждую n_1 строку из блока VI и послед-

ние n_2 строки из блока VI (т. о., самая последняя строка удаляется дважды). После удаления линейно зависимых строк получим матрицу $L_{\overline{B}}$.

Утверждение 2. *Определитель Грама матрицы $L_{\overline{B}}$ равен*

$$\det(L_{\overline{B}} \cdot (L_{\overline{B}})^{\mathrm{T}}) = (n_1+n_2)^{(n_3-1)(n_4-1)} \cdot (n_1+n_3)^{(n_2-1)(n_4-1)} \cdot$$
$$\cdot (n_1+n_4)^{(n_2-1)(n_3-1)} \cdot (n_2+n_3)^{(n_1-1)(n_4-1)} \cdot (n_2+n_4)^{(n_1-1)(n_3-1)} \cdot$$
$$\cdot (n_3+n_4)^{(n_1-1)(n_2-1)} \cdot (n_1+n_2+n_3)^{2(n_4-1)} \cdot (n_1+n_2+n_4)^{2(n_3-1)} \cdot$$
$$\cdot (n_1+n_3+n_4)^{2(n_2-1)} \cdot (n_2+n_3+n_4)^{2(n_1-1)} \cdot (n_1+n_2+n_3+n_4)^3.$$

Ранее [5] Титовой Е.Б. рассмотрено левое нуль-пространство трёхиндексной планарной транспортной задачи, получен базис модуля этого пространства, для которого вычислен определитель матрицы Грама. На основании этих результатов проведено сравнение роста среднего значения суммы квадратов миноров рангового порядка базисной системы строк и базисной системы столбцов матрицы ограничений трёхиндексной планарной транспортной задачи.

Литература

1. Шевченко В.Н. Многогранники многоиндексных транспортных задач: алгебраический подход. // Материалы конференции «Дискретный анализ и исследование операций». Новосибирск: Изд-во ин-та математики, 2004. – С. 64–70.
2. Шевченко В.Н. Характеристические многочлены многоиндексных транспортных задач. // Дискретная математика Т.15. Вып. 2. 2003. С. 83–88.
3. Воеводин В.В., Кузнецов Ю.А. СМБ. Матрицы и вычисления. М.: Наука, 1984.
4. Мишина А.П., Проскуряков И.В. Высшая алгебра. Линейная алгебра, многочлены, общая алгебра. М.: Наука, 1965.
5. Титова Е.Б. Рост миноров рангового порядка матрицы ограничений трёхиндексной планарной транспортной задачи. // Труды XXIV Международной конференции «Математическое моделирование и суперкомпьютерные технологии», Нижний Новгород, 18 – 21 ноября 2024 года. – Нижний Новгород: Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, 2024. – С. 164–167.

ПРЯМОЙ ЭКСТРЕМАЛЬНЫЙ ПОДХОД ДЛЯ ИДЕНТИФИКАЦИИ ПАРАМЕТРОВ В ВИДЕ ВЕКТОРА ФУНКЦИЙ В ДИФФУЗИОННОЙ МОДЕЛИ

М.А. Толстых

Донецкий государственный университет

В работе рассматривается задача параметрической идентификации вектора функций в диффузионной модели распространения информации в социальных сетях. Предложена модификация метода с регулируемым направлением спуска (МРНС) для одновременной идентификации скалярного параметра и функциональных зависимостей. Модификация включает введение параметров регулирования направления спуска для каждой компоненты вектора управления. Численные эксперименты показали, что предложенный метод обеспечивает одновременную равномерную сходимость всех компонент управления к оптимальным значениям за 177 итераций. Достигнуто существенное улучшение сходимости по сравнению с классическими градиентными методами, которые не обеспечивают равномерной сходимости при восстановлении вектор-функций из-за разномасштабности компонент градиента.

Ключевые слова: идентификация параметров, метод с регулируемым направлением спуска, диффузионная модель, социальные сети, бесконечномерная оптимизация.

1. Введение

Современные задачи математического моделирования социальных сетей требуют решения обратных задач параметрической идентификации не только для скалярных величин, но и для функциональных зависимостей. При прямой идентификации классические градиентные алгоритмы демонстрируют недостаточную эффективность в бесконечномерных пространствах, проявляя неравномерность сходимости [1]. В настоящем исследовании предлагается усовершенствованная версия метода с регулируемым направлением спуска (МРНС) [2], устраняющая указанный недостаток за счёт введения специального параметра регулирования.

2. Постановка задачи идентификации

Рассматривается задача идентификации параметра $u = \{p, h(x), r(t)\}$ в рамках нелинейной диффузионной модели, описывающей распространение информационных потоков в социальных сетях [3]:

$$\frac{\partial v}{\partial t} - p\frac{\partial^2 v}{\partial x^2} - rh\theta(v-\varepsilon) = 0, \qquad \text{на области } \Omega = (x_a, x_b) \times (t_0, t_1).$$

В приведённом уравнении: t – время; x – метрика расстояния в графе социальной сети, характеризующая глубину проникновения информационного контента (например, новости); $v(x,t)$ – концентрация информации (число репостов); p – коэффициент диффузии, определяющий скорость проникновения новости в удалённые узлы сети; $r(t)$ – функция, описывающая динамику активности пользователей; $h(x)$ – пропускная способность кластера сети; θ – функция Хевисайда, обеспечивающая пороговый эффект ($\varepsilon > 0$) для исключения самопроизвольного возникновения информации.

Краевые и начальные условия формулируются следующим образом: $v = 1$ при $x = x_a, \frac{\partial v}{\partial x} = 0$ при $x = x_b, v = 0$ в начальный момент t_0.

Оптимальные значения компонент вектора $u = \{p, h(x), r(t)\}$ определяются путём решения вариационной задачи минимизации целевого функционала:

$$J(u) = \int_{\Omega} (v - v_e)^2 \delta(x - x_i) dx dt \to \min,$$

где v_e – эмпирические данные о состоянии информации в узлах графа, δ – дельта-функция Дирака.

Для данной задачи метод с регулируемым направлением спуска будет выглядеть следующим образом:

$$\begin{cases} p^{k+1} = p^k - b^k \beta_p \alpha_p \nabla_p J^k, \\ h^{(k+1)}(x) = h^k(x) - b^k \beta_h \alpha_h(x) \nabla_h J(h^k; x), \\ r^{(k+1)}(t) = r^k(t) - b^k \beta_r \alpha_r(t) \nabla_r J(r^k; t). \end{cases}$$

Здесь $\{\beta_p, \beta_h, \beta_r\}$ – набор весовых коэффициентов; $\{\alpha_p = 1, \alpha_h(x), \alpha_r(t)\}$ – параметры регулирования направления спуска.

Параметры адаптации $\alpha_h(x)$ и $\alpha_r(t)$ определяются соотношениями:

$$\alpha_h(x) = 0.2 \frac{h^0(x)}{|\nabla_h J(h^0; x)|},$$

$$\alpha_r(t) = 0.2 \frac{r^0(t)}{|\nabla_r J(r^0; t)|}.$$

Весовые коэффициенты $\beta_i,\ i = \{p, h, r\}$ рассчитываются на начальной итерации:

$$\beta_i = 0.1 \frac{\|u_i^0\|}{\|\alpha_i \cdot \nabla_i J^0\|}.$$

3. Численное решение

Задача решалась по неявной конечно-разностной схеме Кранка-Николсона с четырехточечным шаблоном методом прогонки. Пространственно-временная сетка задавалась значениями $n = 50, m = 500$. Это соответствовало расстоянию в пять рёбер графа сети на которое распространилась новость от источника, и времени $t_1 - t_0 = 72$ часа, $t_\Delta = 5$ часов.

Решение находилось формально обобщенным на бесконечномерные пространства методом наискорейшего спуска (МНС) и МРНС. МНС показал лишь интегральную сходимость за 225 итераций, в то время как МРНС показал равномерную сходимость за 177 итераций.

4. Выводы

Предложенная модификация МРНС обеспечила одновременную и равномерную сходимость всех компонент вектора управления к оптимальным значениям. Традиционные градиентные методы не позволяют достичь сопоставимой равномерности при восстановлении вектор-функций вследствие разномасштабности компонент градиента

Литература

1. Толстых М.А. Идентификация функции пропускной способности социальной сети // Вестник Московского университета. Сер. 15. Вычислительная математика и кибернетика. 2024. № 1. С. 61–66.
2. Толстых В.К. Прямой экстремальный подход для оптимизации систем с распределенными параметрами. Донецк: Юго-Восток, 1997. 150 с.
3. Толстых М. А. Математическая модель диффузии информации в социальной сети // Современные методы теории функций и смежные проблемы: Материалы международной конференции Воронежская зимняя математическая школа. 2025. С. 334–336.

РОТОБРИЗЕРНЫЕ ЦИКЛОПНЫЕ СОСТОЯНИЯ В АНСАМБЛЕ ФАЗОВЫХ ОСЦИЛЛЯТОРОВ СО СВЯЗЬЮ ПО ВЫСШИМ МОДАМ[1*]

М.М. Хамков, Л.А. Смирнов, М.И. Болотов

Нижегородский государственный университет им. Н.И. Лобачевского

Коллективная динамика ансамблей фазовых осцилляторов с инерцией и бигармонической связью демонстрирует сложные варианты синхронного поведения. В работе исследуется существование и устойчивость ротобризерных циклопных состояний. На основе модели Курамото-Сакагучи с высшими модами связи получены аналитические условия существования и устойчивости данного режима. Методами численного моделирования построены диаграммы устойчивости в пространстве параметров связи. Результаты имеют значение для понимания сложной синхронной динамики в сетях различной природы.

Ключевые слова: циклопный режим, ротобризерный режим, осциллятор, Курамото-Сакагучи, кластер, уединённый элемент.

1. Введение

Ансамбли фазовых осцилляторов являются фундаментальной моделью для описания коллективного поведения в системах, состоящих из большого числа элементов, – от нейронных до энергетических сетей. Модель Курамото–Сакагучи с инерцией может демонстрировать сложные варианты синхронного поведения, включая циклопные состояния, характеризующиеся сосуществованием двух когерентных кластеров и уединённого осциллятора. В данной работе исследуется возникновение и устойчивость ротобризерных циклопных режимов, при которых два кластера совершают вращение относительно уединённого элемента под воздействием связи, содержащей старшие гармоники [1].

2. Модель

Рассматривается ансамбль из N глобально связанных фазовых осцилляторов с инерцией μ и бигармонической функцией связи:

$$\boldsymbol{\mu\ddot{\theta}_k + \dot{\theta}_k = \omega + \frac{1}{N}\sum_{n=1}^{N}\sum_{q=1}^{2}\varepsilon_q \sin\left[q(\theta_n - \theta_k) - \alpha_q\right]}, \quad (1)$$

где θ_k – фаза k-го осциллятора ($k = 1, 2, \ldots, N$), μ – инерция (масса), ω – собственная частота, α_q – фазовый сдвиг q-й гармоники связи, ε_q – амплитуда q-й гармоники. Степень синхронизации количественно характеризуется параметрами порядка R_q, определяемыми как:

$$\boldsymbol{R_q = \frac{1}{N}\sum_{m=1}^{N} e^{iq\theta_m} = r_q e^{i\psi_q}}. \quad (2)$$

Для отталкивающей первой гармоники и нечётного N в системе формируется устойчивый ротобризерный циклопный режим [2]. Для ротобризерного режима межкластерные фазовые отстройки $x(t)$ и $y(t)$ демонстрируют неограниченный рост (убывание), то есть фазы двух кластеров вращаются относительно одиночного осциллятора (рис. 1).

[1*] Работа выполнена при поддержке Российского научного фонда (проект № 24-72-00105).

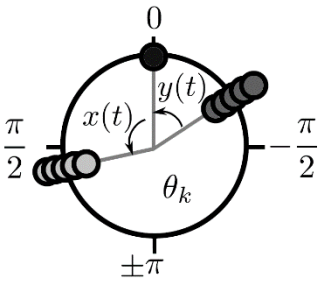


Рис. 1. Ротобризерный циклопный режим в ансамбле из $N = 11$ осцилляторов. Временная эволюция фазовых различий $x(t)$ и y(t) демонстрирует устойчивую вращательную динамику между двумя когерентными кластерами и одиночным осциллятором (фиксированным при $\theta_6 = 0$ для наглядности)

Динамика разностей фаз $x(t)$ и $y(t)$ описывается системой:

$$\mu\ddot{x} + \dot{x} = \sum_{q=1}^{2} \frac{\varepsilon_q}{N}\Big[\sin\alpha_q - \sin\big(qx + \alpha_q\big) - \frac{N-1}{2}\big[\sin\big(qx - \alpha_q\big) + \sin\big(qy - \alpha_q\big) + \sin\big(\alpha_q\big) + \sin\big(q(x-y) + \alpha_q\big)\big]\Big],$$

$$\mu\ddot{y} + \dot{y} = \sum_{q=1}^{2} \frac{\varepsilon_q}{N}\Big[\sin\alpha_q - \sin\big(qy + \alpha_q\big) - \frac{N-1}{2}\big[\sin\big(qx - \alpha_q\big) + \sin\big(qy - \alpha_q\big) + \sin\big(\alpha_q\big) + \sin\big(q(y-x) + \alpha_q\big)\big]\Big].$$

2. Результаты

Основным результатом является аналитическое и численное определение областей параметров, при которых существуют ротобризерные циклопные состояния (рис. 2). Этот режим проявляет устойчивость в широком диапазоне значений параметров второй гармоники, который существенно превосходит область устойчивости стационарного варианта циклопного режима.

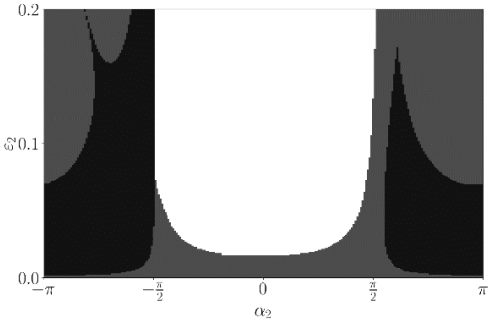


Рис. 2. Диаграмма устойчивости ротобризерного циклопного состояния. Чёрная и серая области соответствуют устойчивым и неустойчивым режимам. Параметры: $N = 11$, $\mu = 1.0$, $\varepsilon_1 = 1.0$, $\alpha_1 = 1.7$

Литература

1. Munyayev V.O., Bolotov M.I., Smirnov L.A., Osipov G.V., and Belykh I. Cyclops States in Repulsive Kuramoto Networks: The Role of Higher-Order Coupling. 2023. DOI: 10.1103/PhysRevLett.130.107201.
2. Bolotov M.I., Munyayev V.O., Smirnov L.A., Osipov G.V., and Belykh Igor. Breathing and switching cyclops states in Kuramoto networks with higher-mode coupling. 2024. DOI: 10.1103/PhysRevE.109.054202.

СЕГМЕНТАЦИЯ СЕРДЦА МЕТОДАМИ ГЛУБОКОГО ОБУЧЕНИЯ: ОБЗОР ПУБЛИКАЦИЙ[1*]

А.И. Хоменко, В.Е. Турлапов

Исследовательский центр в области искусственного интеллекта
Нижегородского государственного университета им. Н.И. Лобачевского

Сегментация структур сердца на медицинских изображениях является ключевой задачей количественного анализа и клинической диагностики. В работе представлен обзор методов сегментации сердца, отражающий эволюцию от классических подходов – активных контуров, моделей уровня поверхности и мультиатласных методов – к современным архитектурам на основе сверточных и трансформерных нейронных сетей, а также к foundation-моделям. Систематизированы открытые датасеты, основные метрики качества и клинические показатели. Обсуждаются перспективные направления развития: использование больших предобученных моделей, сопряжение сегментации с клиническими метриками и автоматизация рабочих процессов в медицине. Материал обзора формирует целостную картину области, соотнося методологические решения с требованиями клинической практики и намечая траектории для дальнейших исследований и внедрения.

Ключевые слова: сегментация сердца, медицинские изображения, глубокое обучение; U-Net/nnU-Net, трансформеры, SAM/MedSAM, Dice, HD95

1. Введение

Сегментация структур сердца (рис.1) на медицинских изображениях является ключевым инструментом количественной оценки функции миокарда и обоснования клинических решений. От её точности зависят объём полостей, масса миокарда, толщина стенок, ударный объём и фракция выброса [1]. Фракция выброса левого желудочка служит интегральным показателем сократимости и широко используется для диагностики сердечной недостаточности и стратификации риска. Автоматическое выделение границ камер на данных магнито-резонансной томографии и эхокардиографии обеспечивает воспроизводимый расчёт конечно-диастолического и конечно-систолического объёма, а также фракции выброса без ручной разметки [2].

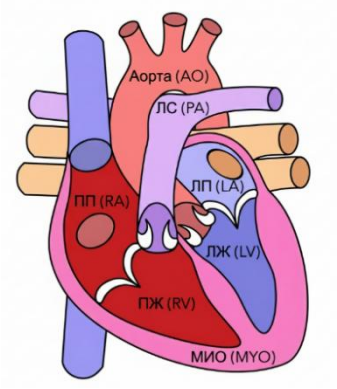


Рис. 1. Ключевые регионы интереса при сегментации сердца. Аорта (AO); ЛС (PA) – легочный ствол; ПП (RA) – правое предсердие; ЛП (LA) - левое предсердие; ПЖ (RV) – правый желудочек; ЛЖ (LV) – левый желудочек; МИО (MYO) – миокард
(источник: Wapcaplet, Wikimedia Commons, CC BY-SA 3.0)

[1*] Работа выполнена при поддержке Министерства экономического развития РФ (соглашение о предоставлении гранта № 139-15-2025-004 от 17 апреля 2025 г., ИГК 000000Ц313925Р3Х0002).

Результаты сегментации востребованы не только в диагностике и мониторинге, но и при планировании лечения. Оценка объёма и формы камер сердца и патологических образований поддерживает выбор стратегии вмешательства, подбор устройств и предоперационное моделирование, включая трёхмерную печать анатомических моделей [3]. Сегментация лежит в основе концепции цифрового двойника – персонализованной компьютерной модели сердца, имитирующей работу органа и прогнозирующей исходы терапии. Для построения такой модели требуется точная многокомпонентная разметка и регулярное обновление по мере поступления новых данных визуализации [4].

Ручная разметка изображений трудоёмка и подвержена субъективным колебаниям. Методы глубокого обучения позволяют автоматизировать сегментацию, повышая объективность и повторяемость измерений, ускоряя получение диагностических показателей и обеспечивая анализ крупных мультицентровых коллекций. Это имеет принципиальное значение на фоне высокой распространённости сердечно-сосудистых заболеваний.

2. Датасеты

Открытые наборы данных с верифицированной разметкой являются основой для развития и объективного сопоставления алгоритмов сегментации. В настоящем разделе дано краткое описание ключевых датасетов. Развёрнутые характеристики приведены в таблице 1.

ACDC. Кинематографическая МРТ короткой оси. Коллекция охватывает несколько клинических групп и содержит разметку эндокарда и эпикарда левого желудочка, а также эндокарда правого желудочка в конце диастолы и в конце систолы. Набор стал стандартом для валидации алгоритмов и вычисления объёмов камер и фракции выброса.

M&Ms. Мультицентровая и мультивендорная коллекция кинематографической МРТ. Набор предназначен для оценки переносимости моделей между центрами и типами оборудования. На нём систематически исследуется влияние смещения домена и эффективность методов адаптации.

MM-WHS 2017. Совмещённая коллекция КТ и трёхмерной МРТ для сегментации всего сердца. В разметку включены полости и миокард желудочков и предсердий, а также крупные сосуды. Набор стимулировал разработку трёхмерных алгоритмов и позволяет сопоставлять качество на разных модальностях.

LASC 2013/2018. Контрастные МРТ с отсроченным усилением для сегментации левого предсердия. По итогам соревнования 2018 года лучший алгоритм обеспечил показатель Dice около 0,93 и среднюю поверхностную ошибку около 0,7 мм [5]. Набор является реперной точкой для задач предсердий и вмешательств при фибрилляции.

CAMUS и **EchoNet-Dynamic**. Крупные коллекции эхокардиографии с видеопетлями стандартных проекций. Датасеты используются для обучения и проверки сегментации на ультразвуке и для автоматической оценки объёмов и фракции выброса на уровне реальной клинической практики.

HVSMR 2.0. Публичная коллекция трёхмерной МРТ при врождённых пороках сердца. Включены многокомпонентные разметки камер и крупных сосудов. Набор предназначен для проверки алгоритмов на нетипичной анатомии и сложных клинических сценариях [6].

Наряду с перечисленными наборами активно применяются коллекции для рубца и отёка миокарда, для коронарных артерий на КТ-ангиографии и для задач целого сердца. Широкая доступность данных по разным модальностям и клиническим сценариям ускорила прогресс в области и позволила системно изучать обобщаемость моделей на мультицентровых испытаниях.

Таблица 1. Ключевые открытые датасеты для задачи сегментации сердца

Датасет (год)	Модальность	Размер	Целевые структуры
Sunnybrook (2009) [7]	МРТ (SA cine)	45 пациентов	Левый желудочек (эндокард, эпикард)
LVSC (2011) [8]	МРТ (SA + LA SSFP)	200 исследований (100 train + 100 test)	Миокард левого желудочка (эндокард + эпикард)
RVSC (2012) [9]	МРТ (SA)	48 пациентов	Правый желудочек

			(эндокард ± эпикард)
LASC (2013) [10]	МРТ / КТ (3D)	60 исследований (30 МРТ + 30 КТ)	Левое предсердие
ACDC (2017) [11]	МРТ (SA cine 2D + 3D)	150 пациентов, 5 классов патологий	ЛЖ (эндо, эпи) и ПЖ (эндо)
MM-WHS (2017) [12]	МРТ + КТ (3D)	120 томов (60 КТ + 60 МРТ)	Четыре камеры сердца, миокард, сосуды
AtriaSeg / LA 2018 [13]	МРТ LGE (3D)	154 пациента (100 train + 54 test)	Левое предсердие (± рубцы)
CAMUS (2019) [14]	Эхо 2D видео (A2C/A4C)	500 пациентов	ЛЖ (эндо, эпи), ЛП (ED/ES)
MS-CMRSeg (2019) [15]	МРТ (LGE, T2, bSSFP SA)	45 пациентов	ЛЖ, ПЖ, миокард
CHD CT / CHD68 (2019) [16]	КТ (3D)	68 пациентов	ЛЖ, ПЖ, ЛП, ПП, миокард, аорта, лёгочная артерия
M&Ms (2020) [17]	МРТ (SA cine)	≈ 375 пациентов (6 центров)	ЛЖ, ПЖ, миокард
EchoNet-Dynamic (2020) [18]	Эхо 2D видео (A4C)	10 030 видео / пациентов	LVEF, EDV/ESV, контуры ЛЖ (ED/ES)
CMRxMotion (2022) [19]	МРТ cine с артефактами дыхания	40 добровольцев (~320 серий)	ЛЖ, ПЖ, миокард; оценка качества и сегментация

3. Сравнение моделей и результатов сегментации

Эволюция методов. До появления глубокого обучения с 2000 по 2015 годы применялись активные контуры [20], статистические модели формы и внешнего вида [21–23], уровневые методы для выделения границ [21–23], графовые оптимизации разметки [24] и распространение меток [25]. Существенный вклад внесли мультиатласные схемы с объединением результатов и голосованием меток [26–30]. Эти подходы сформировали основу автоматической сегментации сердца, однако их точность и переносимость ограничивались вариативностью данных и зависимостью от ручной настройки.

Прорыв обеспечили полносвёрточные сети U-образного типа. Архитектура U-Net радикально улучшила качество сегментации биомедицинских изображений [31], а 3D U-Net и родственные CNN закрепили доминирование глубокого обучения [32]. Конкурс ACDC задал эталонные данные и протокол оценки [11]. Ансамбль от Isensee et al. [33] предвосхитил самоконфигурируемые решения, реализованные в nnU-Net [34], что вывело качество к уровню экспертной разметки на типовых задачах. Далее появились гибриды с трансформерами, включая TransUNet и UNETR, которые усилили учёт глобального контекста [35, 36]. Публикация Segment Anything стимулировала медицинские адаптации, в том числе MedSAM и прямое применение SAM к кардиосегментации [37–39]. Параллельно развиваются модели состояний с продольной памятью, такие как Mamba-UNet и VM-UNet [40, 41], а также эффективные свёрточные энкодеры семейства ConvNeXt, представленные в SCANeXt [42].

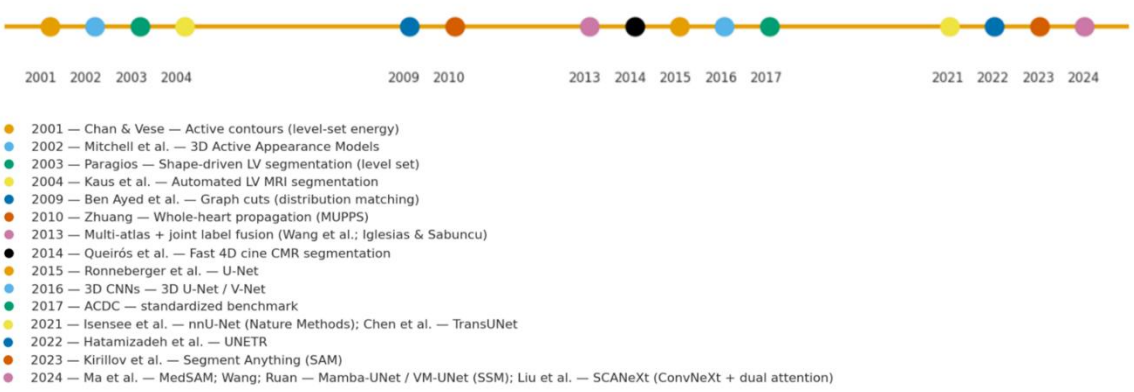


Рис. 2. Хронология ключевых методов и этапов в сегментации сердца (2001–2024)

Результаты сегментации на датасете ACDC. По итогам соревнования ACDC - лучшая планка средних значений Dice по трём структурам составила около 0.927 с использованием решений класса nnU-Net [11]. В последующие годы сообщались результаты в диапазоне от 0.91 до 0.92, что указывает на достижение устойчивого уровня качества. Отдельные работы заявляли более высокие значения при альтернативных протоколах обучения и оценки, однако на закрытом тесте рост оставался ограниченным. Новые модели уделяют больше внимания точности границ. Примером служит SAMba-UNet с комбинированным применением глобального сегментатора и модулей состояния, обеспечивший средний Dice около 0.91 и рекордно малое значение HD95 порядка 1.1 мм на ACDC [43]. Дальнейший прогресс связывается не только с повышением Dice, но и с уменьшением пограничных ошибок.

Обобщаемость и перенос домена. Обобщаемость и перенос домена. Высокие показатели на отдельных наборах не гарантируют устойчивости на внешних данных. При смене центра, производителя оборудования или протокола возникает доменный сдвиг, который снижает метрики перекрытия и повышает пограничные ошибки. Оценка алгоритмов должна включать проверку на независимых центрах и анализ не только Dice, но и метрик границы.

На рис. 3 показана динамика Dice при переносе между вендорами и наборами. На M&Ms без адаптации потери составляют 1-3 процентных пункта, что отражает чувствительность к оборудованию и параметрам съёмки. Совместное обучение на ACDC и MS-CMR с последующим тестированием на M&Ms обеспечивает отставание менее 0,5 процентного пункта по миокарду по сравнению с моделью, обученной непосредственно на целевом наборе [44].

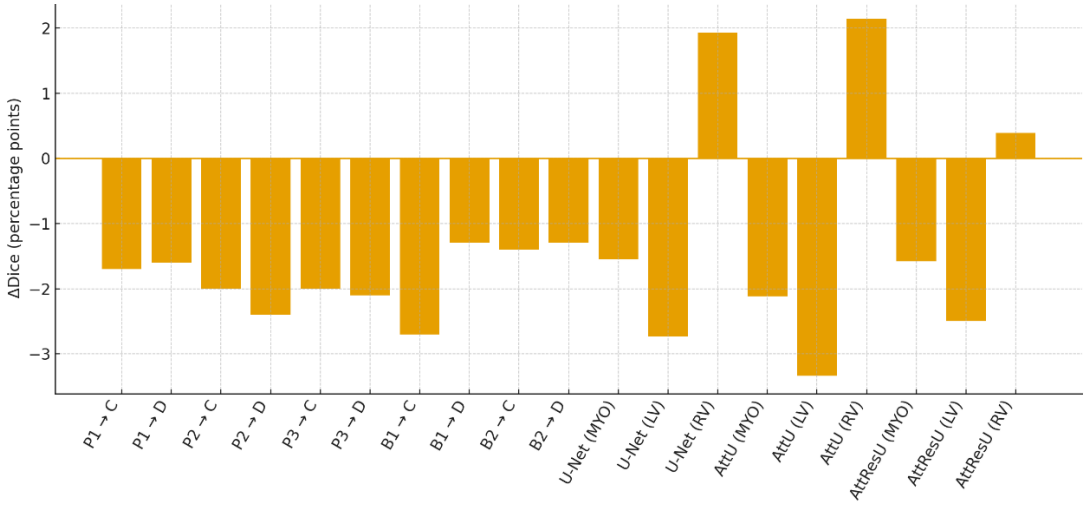


Рис. 3. Изменение метрики Dice при переносе моделей между доменами

Многоцентровое обучение, гармонизация интенсивностей и доменная адаптация уменьшают просадку и выводят качество к уровню целевого домена. При корректной стратегии переноса снижение не превышает одного процентного пункта [45]. Без адаптации падение достигает

двух-трёх процентов и более [46]. Эти результаты подтверждают необходимость явного учёта доменных различий при подготовке к клиническому внедрению.

Клиническая значимость сегментации. Высокая точность автоматической сегментации имеет прямое клиническое значение, поскольку по маскам вычисляют объёмы полостей, толщину стенок и фракцию выброса левого желудочка. Автоматическая обработка стандартизирует расчёт и повышает воспроизводимость показателей.

Независимые наборы данных подтверждают сопоставимость алгоритмов с экспертной разметкой. На выборке UK Biobank из шестисот МР-исследований средняя абсолютная ошибка по фракции выброса составила около 3,2 процента, что близко к межнаблюдательной вариабельности порядка трёх процентов [47]. В исследованиях кинопоследовательностей реального времени без задержки дыхания предварительно обученная нейросеть класса nnU-Net достигала ошибки по фракции выброса 2,6 процента в покое и 3,6 процента при физической нагрузке, при этом ошибки по объёму левого желудочка составляли около трёх миллилитров в покое и около одиннадцати миллилитров при нагрузке [48]. На эхокардиографии точность несколько ниже, однако остаётся клинически приемлемой: в конкурсе CAMUS средняя ошибка определения фракции выброса находилась на уровне около 4,7 процента [33]. Совокупность результатов иллюстрирует готовность автоматической сегментации к применению в задачах оценки систолической функции и объёмов полостей, что подтверждается материалами на рис. 4.

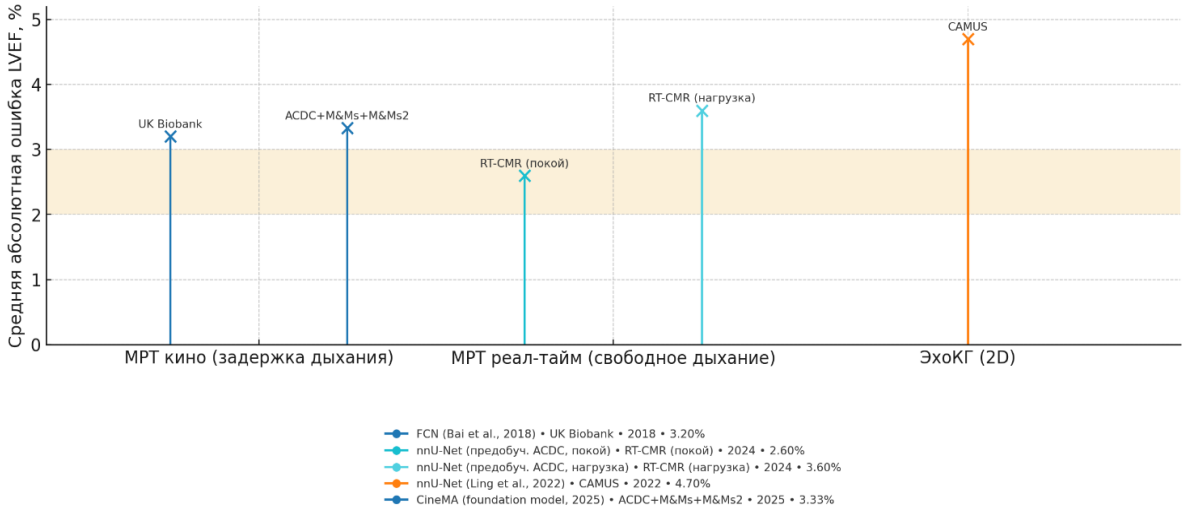


Рис. 4. Ошибка расчёта LVEF при автоматической сегментации для разных типов медицинских изображений сердца

Набирают силу крупные предобученные модели, объединяющие сегментацию и вычисление функциональных показателей. Модель CineMA после дообучения на наборах ACDC, M&Ms и M&Ms2 обеспечила среднюю абсолютную ошибку по фракции выброса 3,33 процента с диапазоном от 2,88 до 3,82 процента, а при применении без дообучения на внешних данных ошибка составила 4,83 процента с диапазоном от 4,58 до 5,21 процента. Ошибки по объёму левого желудочка находились в пределах от пяти до восьми миллилитров в зависимости от набора [49]. Эти результаты показывают, что по качеству сегментации и по точности производных клинических метрик современные методы приблизились к уровню экспертного согласия и способны поддерживать воспроизводимую количественную оценку функции сердца в клинической практике.

Современные архитектуры. Базовым шаблоном остаются U-образные свёрточные сети с каскадным объединением признаков. Дальнейшее развитие идёт по трём направлениям. Первое направление усиливает учёт глобального контекста за счёт механизмов самовнимания, что показано в TransUNet и UNETR [35, 36]. Второе направление повышает эффективность и глубину представлений с помощью современных свёрточных энкодеров, примером служит SCANeXt на основе ConvNeXt, демонстрирующий рост точности и устойчивости [42]. Третье направление использует модели состояний с продольной памятью, которые обеспечивают дальние зависимости при умеренной вычислительной стоимости, что иллюстрируют Mamba-UNet и VM-UNet

[50, 40]. Различия архитектур отражаются на согласованности границ и топологии масок, что подтверждается сравнительным анализом на наборе ACDC [51].

Ещё один вектор связан с крупномасштабным предобучением и появлением универсальных энкодеров. Идеи Segment Anything стимулировали специализированные медицинские адаптации, включая MedSAM и кардиологические варианты [38, 39, 40]. Доменно ориентированные foundation-подходы объединяют сегментацию и расчёт функциональных показателей, как показано на примере CineMA с устойчивостью на внешних данных после донастройки [49]. Конвергенция этих линий приводит к гибридным решениям, которые сочетают точность на границах, переносимость между доменами и практическую применимость в клинических рабочих процессах.

4. Выводы и заключение

Сегментация сердца на медицинских изображениях прошла этап стремительного развития и превратилась в зрелое направление с устойчивыми технологическими и методическими основами. Эволюция от классических подходов к глубоким нейронным сетям привела к тому, что автоматические методы обеспечивают качество, сопоставимое с экспертной разметкой для основных структур и типовых протоколов магнито-резонансной томографии, компьютерной томографии и эхокардиографии. Наличие открытых датасетов и соревнований сформировало общий язык сравнения и ускорило отбор эффективных архитектур. На практике де-факто стандартом стали U-образные свёрточные сети и их самонастраиваемые варианты, на основе которых строится значительная часть современных систем.

Одновременно выявились ограничения, определяющие повестку дальнейших исследований. К ключевым вызовам относятся переносимость между центрами и производителями оборудования, устойчивость на редких патологиях и тонких структурах, согласованность во времени на кинопоследовательностях и интерпретируемость результатов. Эти факторы порождают требования к оценке не только по метрикам перекрытия, но и по пограничным метрикам, а также к сопоставлению с клиническими показателями, которые используются в принятии решений. Возникает потребность во внешней валидации, в анализе ошибок на независимых данных и в контроле неопределённости предсказаний.

Развитие архитектур идёт по нескольким согласованным траекториям. Трансформерные компоненты расширяют охват глобального контекста и повышают согласованность контуров. Модели состояний обеспечивают дальние зависимости при умеренной вычислительной сложности и позволяют работать с трёхмерными и четырёхмерными объёмами. Современные свёрточные энкодеры укрепляют базовую линию за счёт глубокой и устойчивой иерархии признаков. Параллельно формируется класс больших предобученных моделей, которые дают универсальные энкодеры признаков и ускоряют адаптацию к новым данным. Набирают силу гибридные решения, где универсальный энкодер сочетается со специализированным декодером и постобработкой, а также интерактивные схемы с подсказками.

Сегментация всё теснее связывается с клиническими задачами. Маски сегментации используются для расчёта функциональных показателей, для планирования вмешательств и для построения персонализованных цифровых моделей сердца. Эта связка требует воспроизводимых протоколов, мониторинга качества и встраивания в рутинные рабочие контуры. Интеграция в информационные системы класса PACS, обеспечение аудита вычислений и управление версиями моделей становятся необходимыми условиями внедрения. Для повышения переносимости востребованы многоцентровое обучение, гармонизация данных, методы адаптации домена и подходы к совместному обучению без обмена первичными изображениями. Важна объяснимость данных, соответствие требованиям, включая защиту персональных данных, контролируемость решений человеком и мультиагентные доверенные схемы принятия решений.

Намечаются направления, способные задать следующий виток развития. К ним относится дальнейшая интеграция данных с объединением ЭКГ, анализа крови, коронарографии, ультразвука, томографии, гистологии, обработка динамических последовательностей, как упорядоченных сигналов с учётом физиологии. В отношении искусственного интеллекта (ИИ) – привлечение методов самообучения без учителя, методов синтеза разметки в синтетических датасетах, приемов компенсации дефицита разметки, корреляции сегментации с прогностическими

моделями риска и цифровыми двойниками пациента. Ожидается также усиление роли больших предобученных моделей с доказанной полнотой датасетов.

Литература

1. Chen C., Qin C., Qiu H., et al. Deep Learning for Cardiac Image Segmentation: A Review // Frontiers in Cardiovascular Medicine. 2020. Vol. 7. Art. 25. DOI: 10.3389/fcvm.2020.00025.
2. Campello V.M., Mansi T., Duchateau N., et al. Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation (M&Ms) Challenge // STACOM 2020. LNCS 12544. P. 204–213. DOI: 10.1007/978-3-030-68107-4_20.
3. Bernard O., Popescu D., Mansi T. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? // IEEE Trans. Med. Imaging. 2018. Vol. 37, No. 11. P. 2514–2525. DOI: 10.1109/TMI.2018.2837502.
4. Rudnicka Z., Tavares J., Ayllón M., et al. Cardiac healthcare digital twins supported by AI and XR: systematic review // Electronics. 2024. Vol. 13, No. 5. Art. 866. DOI: 10.3390/electronics13050866.
5. Xiong Z., Xia Q., Hu Z., et al. A global benchmark of algorithms for segmenting the left atrium from LGE-MRI // Medical Image Analysis. 2021. Vol. 67. Art. 101832. DOI: 10.1016/j.media.2020.101832.
6. Pace D.F., Contreras H.T.M., Romanowicz J., et al. HVSMR-2.0: a 3D cardiovascular MR dataset for whole-heart segmentation in congenital heart disease // Scientific Data. 2024. Vol. 11, No. 1. Art. 721. DOI: 10.1038/s41597-024-03469-9.
7. Radau P.E., Lu Y., Connelly K.A., Paul G., Dick A.J., Wright G.A. Evaluation framework for algorithms segmenting short-axis cardiac MRI // The MIDAS Journal – Cardiac MR Left Ventricle Segmentation Challenge. 2009. URL: http://hdl.handle.net/10380/3070 (дата обращения: 07.10.2025). DOI (Insight Journal): 10.54294/g80ruo.
8. Suinesiaputra A., Bluemke D.A., Cowan B.R., et al. LV Segmentation Challenge: a database and evaluation framework for cardiac MRI // IEEE Trans. Med. Imaging. 2014. Vol. 33, No. 2. P. 325–333.
9. Petitjean C., Zuluaga M.A., Bai W., et al. Right Ventricle Segmentation Challenge 2012 // Medical Image Analysis. 2015. Vol. 19, No. 1. P. 187–202.
10. Tobon-Gomez C., Geers A.J., Peters J., et al. Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets // IEEE Trans. Med. Imaging. 2015. Vol. 34, No. 7. P. 1460–1473.
11. Bernard O., Lalande A., Zotti C., et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: the ACDC challenge // Medical Image Analysis. 2018. Vol. 44. P. 115–137. DOI: 10.1016/j.media.2017.12.007.
12. Zhuang X., Li L., Payer C., et al. Evaluation of algorithms for Multi-Modality Whole Heart Segmentation: An open-access grand challenge // Medical Image Analysis. 2019. Vol. 58. Art. 101537. DOI: 10.1016/j.media.2019.101537.
13. Xiong Z., Fedorov A., Li Y., et al. Atrial Segmentation Challenge 2018 (AtriaSeg) // Medical Image Analysis. 2021. Vol. 68. Art. 101948.
14. Leclerc S., Smistad E., Pedrosa J., et al. Deep Learning for Segmentation using an Open Large-Scale Dataset in 2D Echocardiography (CAMUS) // IEEE Trans. Med. Imaging. 2019. Vol. 38, No. 9. P. 2198–2210. DOI: 10.1109/TMI.2019.2900516.
15. Zhuang X. Multisequence CMR Segmentation Challenge (MS-CMRSeg) // Medical Image Analysis. 2020. Vol. 68. Art. 101913.
16. Zhuang X., Li Y., Huang J., et al. Cardiac segmentation for congenital heart disease in CT images // Medical Image Analysis. 2020. Vol. 65. Art. 101743.
17. Campello V.M., Gkontra P., Izquierdo C., et al. Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge // IEEE Trans. Med. Imaging. 2021. Vol. 40, No. 12. P. 3543–3554.
18. Ouyang D., He B., Ghorbani A., et al. Video-based AI for beat-to-beat assessment of cardiac function // Nature. 2020. Vol. 580. P. 252–256. DOI: 10.1038/s41586-020-2145-8.

19. Su B., Peng R., Xue Y., et al. The CMRxMotion challenge: motion artifact reduction and segmentation in cardiac MRI // Medical Image Analysis. 2023. Vol. 88. Art. 102862.
20. Chan T.F., Vese L.A. Active contours without edges // IEEE Trans. Image Process. 2001. Vol. 10, No. 2. P. 266–277.
21. Kaus M.R., von Berg J., Weese J., Niessen W.J., Pekar V. Automated segmentation of the left ventricle in cardiac MRI // Medical Image Analysis. 2004. Vol. 8, No. 3. P. 245–254.
22. Mitchell S.C., Bosch J.G., Lelieveldt B.P.F., van der Geest R.J., Reiber J.H.C., Sonka M. 3-D active appearance models: segmentation of cardiac MR and ultrasound images // IEEE Trans. Med. Imaging. 2002. Vol. 21, No. 9. P. 1167–1178.
23. Paragios N. A level set approach for shape-driven segmentation and tracking of the left ventricle. IEEE Transactions on Medical Imaging. 2003;22(6):773–776. DOI: 10.1109/TMI.2003.814785.
24. Ben Ayed I., Punithakumar K., Li S., Islam A., Chong J., Suri J., et al. Left ventricle segmentation via graph cut distribution matching // MICCAI 2009. LNCS 5762. P. 901–909.
25. Zhuang X. Whole heart segmentation of cardiac MRI using multiple path propagation strategy (MUPPS) // MICCAI 2010. LNCS 6363. P. 109–116.
26. Iglesias J.E., Sabuncu M.R. A unified framework for cross-modality multi-atlas segmentation of brain MRI // Medical Image Analysis. 2013. Vol. 17, No. 8. P. 1181–1191.
27. Vergaro G., Gentile F., Meems L.M.G., et al. NT-proBNP for Risk Prediction in Heart Failure… // JACC: Heart Failure. 2021. Vol. 9, No. 9. P. 653–663. DOI: 10.1016/j.jchf.2021.05.014.
28. Bai W., Shi W., Ledig C., Rueckert D. Multi-atlas segmentation with augmented features for cardiac MR images // Medical Image Analysis. 2015. Vol. 19, No. 1. P. 98–109.
29. Mahapatra D. Cardiac image segmentation from cine cardiac MRI using graph cuts and shape priors // Journal of Digital Imaging. 2013. Vol. 26, No. 4. P. 721–730.
30. Queirós S., Barbosa D., Heyde B., et al. Fast automatic myocardial segmentation in 4D cine CMR datasets // Medical Image Analysis. 2014. Vol. 18, No. 7. P. 1115–1131.
31. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // MICCAI 2015. LNCS 9351. P. 234–241.
32. Çiçek Ö., Abdulkadir A., Lienkamp S.S., Brox T., Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation // MICCAI 2016. LNCS 9901. P. 424–432.
33. Isensee F., Jaeger P.F., Full P.M., et al. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain-specific features // STACOM 2017. LNCS 10663. P. 120–129.
34. Isensee F., Jaeger P.F., Kohl S.A.A., Petersen J., Maier-Hein K.H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation // Nature Methods. 2021. Vol. 18, No. 2. P. 203–211.
35. Chen J., Lu Y., Yu Q., et al. TransUNet: Transformers make strong encoders for medical image segmentation // arXiv:2102.04306, 2021. URL: https://arxiv.org/abs/2102.04306 (дата обращения: 07.10.2025).
36. Hatamizadeh A., Tang Y., Nath V., Yang D., Myronenko A., Landman B., Roth H., Xu D. UNETR: Transformers for 3D Medical Image Segmentation // WACV 2022. P. 574–584. arXiv:2103.10504.
37. Kirillov A., Mintun E., Ravi N., et al. Segment Anything // Proc. ICCV. 2023.
38. Ma J., He Y., Li F., et al. Segment anything in medical images (MedSAM) // Nature Communications. 2024. Vol. 15. Art. 654. DOI: 10.1038/s41467-024-44824-z.
39. Chen Z., Kim S., Ren H., et al. Cine CMR segmentation using temporal-spatial adaptation of prompt-enabled SAM // Journal of Cardiovascular Magnetic Resonance. 2025. Vol. 27, No. 1. Art. 101909. DOI: 10.1016/j.jcmr.2025.101909.
40. Ruan J., Xiang S., Zhang Y., et al. VM-UNet: Vision Mamba UNet for medical image segmentation // arXiv:2402.02491, 2024.
41. Wang H., Suh J.W., Das S.R., Pluta J.B., Altinay M., Yushkevich P.A. Multi-atlas segmentation with joint label fusion and corrective learning // Frontiers in Neuroinformatics. 2013. Vol. 7. Art. 27.
42. Zhou Z., Rahman S., de Souza J., et al. Study of quality of reporting in AI cardiac MRI segmentation studies // Radiology: Cardiothoracic Imaging. 2020. Vol. 2, No. 5. e200193. DOI: 10.1148/ryct.2020200193.

43. Huo G., Dai R., Tang H. SAMba-UNet: synergizing SAM2 and Mamba in UNet with heterogeneous aggregation for cardiac MRI segmentation // arXiv:2505.16304, 2025.
44. Patil S.S., Ramteke M., Verma M., et al. A Domain-Shift Invariant CNN Framework for Cardiac MRI Segmentation Across Unseen Domains // Journal of Digital Imaging. 2023. Vol. 36, No. 5. P. 2148–2163.
45. Li Y., Wang F., Jin J., et al. CMRxRecon: a publicly available k-space dataset and benchmark to advance deep learning for cardiac MRI // Scientific Data. 2024. Vol. 11, No. 1. Art. 687. DOI: 10.1038/s41597-024-03525-4.
46. Liu Y., Wang H., Xiong S., et al. SCANeXt: Enhancing 3D medical image segmentation with dual attention and ConvNeXt // Heliyon. 2024. Vol. 10, No. 5. e26775. DOI: 10.1016/j.heliyon.2024.e26775.
47. Bai W., Sinclair M., Tarroni G., et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks // Journal of Cardiovascular Magnetic Resonance. 2018. Vol. 20, No. 1. Art. 65.
48. Schilling M., Unterberg-Buchwald C., Lotz J., Uecker M., et al. Assessment of deep learning segmentation for real-time free-breathing cardiac MRI at rest and under exercise stress // Scientific Reports. 2024. Vol. 14. Art. 3754.
49. Fu Y., Yi W., Manisty C., et al. CineMA: a foundation model for cine cardiac MRI // arXiv:2506.00679, 2025. URL: https://arxiv.org/abs/2506.00679 (дата обращения: 07.10.2025).
50. Wang Z., Zheng J.-Q., Zhang Y., et al. Mamba-UNet: UNet-like pure visual Mamba for medical image segmentation // arXiv:2402.05079, 2024.
51. Vesal S., Maier A., Ravikumar N. Fully Automated 3D Cardiac MRI Localisation and Segmentation Using Deep Neural Networks // Journal of Imaging. 2020. Vol. 6, No. 7. Art. 65. DOI: 10.3390/jimaging6070065.

О РЕШЕНИИ ЛИНЕЙНЫХ ДИОФАНТОВЫХ УРАВНЕНИЙ В РАМКАХ ПОДХОДА ГРУППОВЫХ ДЕЙСТВИЙ

И.С. Чистов, Л.М. Цыбуля

Московский педагогический государственный университет

В настоящей статье рассмотрены действия групп симметрий правильных *n*-мерных многогранников и группы параллельных переносов на множестве линейных диофантовых уравнений. Раскрыта связь между решениями двух уравнений, одно из которых было получено с помощью некоторого действия из другого. Также были даны рекомендации к применению теории групповых действия на практике при решении определённых задач.

Ключевые слова: действие группы, линейное диофантово уравнение, симметрии, диэдральная группа, параллельный перенос.

1. Введение

Некоторые задачи в алгебре эффективнее решать, применяя групповые действия, а точнее, те закономерности, которые выявляются при действии определённой группы на конкретном множестве.

Целью данной работы является определение взаимосвязей между действиями диэдральной группы D_{2n}, [1], и группы параллельных переносов T_n на множестве всех линейных диофантовых уравнений и на множестве их решений посредством введения и последующего изучения свойств данных действий.

Задачи – задать действия рассматриваемых групп на множестве линейных диофантовых уравнений, исследовать и описать свойства этих действий, установить связь между действиями и решениями линейных диофантовых уравнений.

В первом параграфе вводится и рассматривается действие диэдральной группы на множестве линейных диофантовых уравнений с двумя неизвестными, выявляется связь между векторами общего решения для уравнений, участвующих в действии.

Второй параграф посвящен тому же вопросу, но уже для группы параллельных переносов.

В последнем, третьем параграфе даны рекомендации к применению теории, изложенной в предыдущих параграфах и в [2] на практике.

2. Действия групп симметрий правильных *n*-мерных многогранников на множестве линейных диофантовых уравнений

Определение 1. Линейным диофантовым уравнением (ЛДУ) с n неизвестными будем называть уравнение вида

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

где $a_1, a_2, \dots, a_n, b$ – целые числа, причём хотя бы одно $a_i \neq 0$, и при этом требуется, чтобы неизвестные $x_1, x_2, \dots, x_n$ принимали только целые значения. Числа $a_1, a_2, \dots, a_n$ называются коэффициентами ЛДУ, [3].

Множество всевозможных диофантовых уравнений с n неизвестными будем обозначать так:

$$LDE_n = \{D\colon a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \mid a_i, \quad b \in \mathbb{Z}, i = 1,2, \dots, n\}.$$

Определение 2. Линейное диофантово уравнение назовём совместным, если оно имеет решения, и несовместным в противном случае.

Линейное (диофантово) уравнение с n неизвестными ($n \geq 2$) с геометрической точки зрения задает в пространстве $\mathbb{R}^n$ прямую, плоскость или гиперплоскость в зависимости от n, [4]. Решение ЛДУ представляет собой совокупность точек в целочисленной решетке $\mathbb{Z}^n$, причем

данная совокупность, даже если она пуста, формально целиком покрывается (является подмножеством) прямой, плоскостью, гиперплоскостью в зависимости от n. Таким образом, решения ЛДУ можно рассматривать как части известных геометрических объектов (прямых, плоскостей, гиперплоскостей). При рассмотрении действий групп симметрий правильных многогранников такие геометрические объекты могут подвергаться преобразованиям – поворотам или симметриям.

Рассмотрим для начала случай, когда $n = 2$. В этом случае мы будем иметь дело с группами симметрий правильных многоугольников, которые называются диэдральными группами (группами диэдра).

Так, рассмотрим действие группы диэдра D_8 (с операцией $\circ$ правой композиции отображений) на множестве

$$LDE_2 = \{D\colon a_1x_1 + a_2x_2 = b \mid a_1, a_2, b \in \mathbb{Z}\},$$

содержащее все ЛДУ с двумя переменными и только их.

Устроим отображение

$$d_8\colon D_8 \times LDE_2 \to LDE_2,$$

определённое следующим образом: каждой паре (f, D) поставим в соответствие элемент

$$f \cdot D = f(D),$$

где $f(D)$ есть уравнение прямой, являющейся образом прямой, заданной уравнением D, при движении f.

Утверждение 1. Отображение d_8 есть действие группы D_8 на LDE_2.

Группа D_8 была выбрана неспроста: не все группы диэдра могут таким образом действовать на множестве LDE_n. Для того, чтобы отображение являлось действием, важно, чтобы его результат принадлежал тому же множеству, то есть являлся ЛДУ. Значит, подходят симметрии и повороты только на углы, кратные $\frac{\pi}{2}$ (чтобы коэффициенты нового ЛДУ не содержали дробей или радикалов). Такие преобразования находятся лишь в группе D_8, то есть в группе симметрий квадрата (гексаэдра).

Также специально было выбрано множество LDE_2, так как квадрат, симметрии которого образуют группу D_8, – плоская фигура. Благодаря требованию, что $n = 2$, соответствие d_8 есть именно отображение, поскольку элемент $f(D)$ будет определён для каждой пары (f, D).

В случае $n = 2$ ЛДУ с двумя неизвестными визуально отождествляются с прямыми, которые можно поворачивать относительно точки плоскости и отражать относительно других прямых в прямоугольной декартовой системе координат (ПДСК). При этом будет выбран стандартный базис. В таблице 1 показаны результаты действия d_8 на множестве LDE_2.

Таблица 1. Связь действий a_σ и $a^{(-)}$ с действиями группы D_8

f	**f(D)**	**Интерпретация f(D)**
$\mathrm{R}_0^0 = \mathrm{id}$	$D\colon a_1x_1' + a_2x_2' = b$	D
$\mathrm{R}_0^{\frac{\pi}{2}}$	$-a_2x_1' + a_1x_2' = b$	$a^{(-+)}\left(a_{(12)}(D)\right)$ или $a_{(12)}\left(a^{(+-)}(D)\right)$
R_0^{π}	$-a_1x_1' - a_2x_2' = b$	$a^{(--)}(D)$
$\mathrm{R}_0^{\frac{3\pi}{2}}$	$a_2x_1' - a_1x_2' = b$	$a^{(+-)}\left(a_{(12)}(D)\right)$ или $a_{(12)}\left(a^{(-+)}(D)\right)$
$\mathrm{S}_{\mathrm{y}=0}$	$a_1x_1' - a_2x_2' = b$	$a^{(+-)}(D)$
$\mathrm{S}_{\mathrm{x}=0}$	$-a_1x_1' + a_2x_2' = b$	$a^{(-+)}(D)$
$\mathrm{S}_{\mathrm{y}=\mathrm{x}}$	$a_2x_1' + a_1x_2' = b$	$a_{(12)}(D)$
$\mathrm{S}_{\mathrm{y}=-\mathrm{x}}$	$-a_2x_1' - a_1x_2' = b$	$a^{(--)}\left(a_{(12)}(D)\right)$ или $a_{(12)}\left(a^{(--)}(D)\right)$

Обобщим действие d_8 на случай, когда в ЛДУ число переменных больше двух (или меньше двух) и покажем, какие группы будут подходить для действий в данных случаях. Ясно, что уже в трехмерном случае ситуация с геометрической точки зрения будет видна не так хорошо, как в двумерном. Способ задания таких действий аналогичен d_8. При этом, как и в двумерном случае, стоит выбрать стандартный базис. В таблице 2 показаны подходящие группы для действия на множестве LDE_n.

Таблица 2. Группы, подходящие для действия на LDE_n

n	**Группа, количество элементов в ней**
1	Группа симметрий отрезка (в одномерном пространстве), 2
2	Группа симметрий квадрата (в двумерном пространстве), 8
3	Группа симметрий куба (в трёхмерном пространстве), 48
4	Группа симметрий тессеракта (в четырёхмерном пространстве), 384
5	Группа симметрий пентеракта (в пятимерном пространстве), 3840
6	Группа симметрий гексеракта (в шестимерном пространстве), 46080
…	…
k	Группа симметрий k-мерного куба (в k-мерном пространстве), $2 \cdot 4 \cdot 6 \cdot 8 \cdot \ldots \cdot 2k = k! \cdot 2^k$ элементов

Рассмотрим линейное представление группы D_8, [5]. Для этого гомоморфно инъективно отобразим D_8 в группу $GL_2(\mathbb{Z})$ всех целочисленных матриц с определителем ± 1 с операцией $\cdot$ умножения матриц. То есть имеем

$$h\colon D_8 \to GL_2(\mathbb{Z}).$$

Множество всех матриц, соответствующим движениям из D_8, обозначим через $Im(h)$. Каждый элемент $Im(h)$ можно описать как целочисленную матрицу с определителем ± 1, в каждой строке и в каждом столбце которой лишь один ненулевой элемент, равный ± 1, а остальные элементы нулевые, [6]. Элементы множества $Im(h)$ представлены в таблице 3.

Таблица 3. Матричное представление элементов группы D_8

f	**Матрица f**
$\mathrm{R}_{\mathrm{O}}^{0} = \mathrm{id}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$\mathrm{R}_{\mathrm{O}}^{\frac{\pi}{2}}$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$
$\mathrm{R}_{\mathrm{O}}^{\pi}$	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$
$\mathrm{R}_{\mathrm{O}}^{\frac{3\pi}{2}}$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
$\mathrm{S}_{\mathrm{y=0}}$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$
$\mathrm{S}_{\mathrm{x=0}}$	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$
$\mathrm{S}_{\mathrm{y=x}}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
$\mathrm{S}_{\mathrm{y=-x}}$	$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$

Замечание 1. Будем для удобства воспринимать элементы $\mathbb{Z}^2$ как векторы-строки или как векторы-столбцы в определённых ситуациях.

Теорема 1. Пусть $z_1, z_2 \in \mathbb{Z}^2$ – векторы общего решения совместных ЛДУ $D_1, D_2 \in LDE_2$ соответственно. Пусть имеется преобразование $\mathrm{f} \in D_8$, и при этом $A_f \in Im(h)$ есть матрица f. Тогда

$$\forall\, D_1, D_2 \in LDE_2\ (D_2 = f(D_1) \Longrightarrow z_2 = A_f \cdot z_1).$$

3. Действия групп параллельных переносов на множестве линейных диофантовых уравнений

Для каждого $n \in \mathbb{N}$ рассмотрим T_n – группу параллельных переносов t_m на векторы $m = (m_1, \ldots, m_n) \in \mathbb{Z}^n$, с операцией $\circ$ правой композиции отображений.

Для каждого $n \in \mathbb{N}$ введём отображение

$$\to_n\colon T_n \times LDE_n \to LDE_n,$$

определённое следующим образом: каждой паре (t_m, D), где $m = (m_1, m_2, \ldots, m_n)$ есть вектор длины n параллельного переноса t_m, поставим в соответствие элемент

$$t_m \cdot D = t_m(D),$$

где $t_m(D)$ есть уравнение прямой, являющейся образом прямой, заданной уравнением D, при движении t_m.

Утверждение 2. Отображение $\rightarrow_n$ есть действие группы T_n на LDE_n.

Теорема 2. Пусть $z_1, z_2 \in \mathbb{Z}^n$ – векторы общего решения совместных ЛДУ

$D_1, D_2 \in LDE_n$ соответственно. Пусть $m = (m_1, m_2, \dots, m_n) \in \mathbb{Z}^n$ есть вектор параллельного переноса $t_m \in T_n$. Тогда

$$\forall\, D_1, D_2 \in LDE_n\ (D_2 = t_m(D_1) \Longrightarrow z_2 = z_1 + m).$$

4. Практические рекомендации к применению теории групповых действий при решении линейных диофантовых уравнений

Касаемо вопроса об использовании данной теории на практике при решении конкретных ЛДУ, стоит отметить, что постановка задачи обычно состоит не в том, чтобы получить какое-либо ЛДУ с помощью определённого действия некоторой группы, а в том, чтобы решить ЛДУ (действие группы – не цель, а средство). Используя групповые действия, одно ЛДУ возможно решить с помощью другого ЛДУ, при этом понимая, какое именно действие было совершено, то есть используя признак того или иного действия.

Рассмотрим некоторую задачу. Дано несколько ЛДУ $D_1, D_2, \dots, D_s, s \in \mathbb{N}$ с одинаковым количеством неизвестных n. Требуется решить все данные ЛДУ.

Процесс решения данной задачи с применением теории групповых действий может происходить по следующему алгоритму.

Сначала будет целесообразным привести все ЛДУ к «удобному» виду

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

Также стоит упростить все ЛДУ, например, разделив, если это приемлемо, свободный член на отличный от единицы НОД всех коэффициентов. Именно так проще заметить некоторое сходство между ЛДУ, поскольку при описании действий, упомянутых ранее, использовался именно такой вид ЛДУ. Также для того, чтобы заметить сходство, может быть полезным умножить обе части каких-либо уравнений на -1;

1. Среди данных уравнений вычленить минимум два, которые чем-либо схожи (например, определённая схожесть в коэффициентах);

2. Определить, может ли какое-либо из этих двух ЛДУ быть получено из другого с помощью группового действия, используя признак этого действия (признаки 1-4 ниже);

3. Выяснить, будет ли рациональным использовать групповые действия при решении данной задачи (ответить на вопрос: быстрее/эффективнее ли этот метод решения, чем какой-либо другой?);

4. В случае положительного ответа на вопрос пункта 3 следует решить одно ЛДУ из двух выбранных (то, которое решается проще/быстрее) и применить теорему о связи решений для двух данных ЛДУ (при этом выбрав теорему рационально, если теорем уже несколько. Для рационального выбора теоремы пользоваться нужно теми данными, которые даны в условии).

Ниже сформулированы признаки действия тех или иных групп на множестве линейных диофантовых уравнений с учётом применения к ним шага 0 алгоритма выше. Далее, говоря «ЛДУ-образ» или «ЛДУ-прообраз», будем иметь в виду соответственно ЛДУ, получившееся в результате некоторого действия, и ЛДУ, подвергшееся этому действию.

При действиях S_n в ЛДУ могут меняться местами коэффициенты и/или переменные, при этом формально сами значения коэффициентов остаются теми же, [2].

Признак 1 (действие S_n на LDE). Чтобы понять, что ЛДУ D_2 было получено из ЛДУ D_1 описанными ранее в [2] действиями S_n, нужно обратить внимание на то, что коэффициенты D_2 ровно такие же, как и у D_1, только, быть может, в другом порядке. При этом, свободный член не изменен. (Ранее было показано, что все рассмотренные действия S_n сводятся к перестановке коэффициентов).

При действиях $Aut(\mathbb{Z})$ в ЛДУ может меняться знак у определенных коэффициентов, [2].

Признак 2 (действие $Aut(\mathbb{Z})$) на LDE). Чтобы понять, что ЛДУ D_2 было получено из ЛДУ D_1 описанными ранее в [2] действиями $Aut(\mathbb{Z})$, нужно обратить внимание на то, что абсолютные значения коэффициентов D_2 ровно такие же, как и у D_1, порядок коэффициентов не изме-

нен, но, только, быть может, у определенных коэффициентов противоположные знаки. При этом, абсолютная величина свободного члена не изменена, [2].

При действиях группы D_8, было показано в таблице 1, так как это действие есть комбинация действий группы S_n и $Aut(\mathbb{Z})$, то в ЛДУ от двух переменных может поменяться знак у определенных коэффициентов и/или может поменяться порядок коэффициентов.

Признак 3 (действие D_8 на LDE_2). Чтобы понять, что ЛДУ D_2 было получено из ЛДУ D_1 описанным ранее действием D_8), нужно обратить внимание на то, что абсолютные значения коэффициентов D_2 ровно такие же, как и у D_1, но, быть может, у определенных коэффициентов противоположные знаки и/или их порядок изменен. При этом, абсолютная величина свободного члена не изменена.

При действии D_8 при виде D_1 и D_2 не ясно, что из себя представляет f. Однако, не для всякой задачи f нужно, иногда достаточно комбинации двух предыдущих действий. Но если f дано в условии, то можно воспользоваться теоремой 1 про матрицу. Опять же, поднимается вопрос о рациональности использования той или иной теоремы.

При действии группы T_n левые части уравнений прообраза и образа совпадают, а свободные члены отличаются на величину $D(m)$, которая является результатом подстановки в левую часть прообраза координат вектора m вместо переменных, при этом, если прообраз совместен, то образ тоже.

Признак 4 (действие T_n на LDE_n). Если ЛДУ D_1 и D_2 оба совместны и отличаются только свободным членом, то найдется вектор m такой, что $D_2 = t_m(D_1)$. То есть, чтобы понять, что ЛДУ D_2 было получено из ЛДУ D_1 описанным ранее действием T_n, нужно обратить внимание на то, что оба данные ЛДУ совместны и отличаются лишь свободным членом.

Процесс нахождения данного вектора переноса (его координат) состоит, опять же, в нахождении решения ЛДУ, но уже другого: $a_1m_1+\ldots+a_nm_n = b_2 - b_1$, с переменными m_i, при этом достаточно найти (возможно, подбором) хотя бы одно его частное решение, которое и будет являться искомым вектором m. Но не стоит решать его полностью.

К слову, любые два ЛДУ от двух переменных (на плоскости), отличающиеся только на свободный член, задают две параллельные прямые, значит, всегда существует вектор параллельного переноса m такой, что $D_2 = t_m(D_1)$. Однако, если какое-либо из этих уравнений не совместно, то, ясно, $m \notin \mathbb{Z}^2$.

5. Заключение

По ходу работы были получены следующие важные результаты.

1. Действие диэдральной группы на множестве линейных диофантовых уравнений характеризуется комбинацией определенных действий группы подстановок и группы автоморфизмов группы целых чисел на том же множестве, описанных в [2].

2. При параллельном переносе прямой, плоскости или гиперплоскости, заданной ЛДУ, на определённый вектор, решения данного ЛДУ также переносятся на тот же вектор.

3. Были описаны признаки действий и даны рекомендации для реализации подхода групповых действий при решении ЛДУ на практике.

Таким образом, все задачи работы решены, а цель достигнута.

Исследование предполагается расширить в следующих направлениях:

1) рассмотреть, как данный подход работает на нелинейных диофантовых уравнениях;

2) рассмотреть множества действий конкретной группы или нескольких групп на множестве линейных диофантовых уравнений на предмет наличия определённой структуры.

Литература

1. Dummit, David S., Foote, Richard M. Abstract Algebra. 3rd edition, John Wiley & Sons, Inc., 2004. 932 p.
2. Чистов И.С., Цыбуля Л.М. Некоторые зависимости между решениями линейных диофантовых уравнений при действиях группы подстановок и группы автоморфизмов целых чисел. Алгебра, теория чисел, дискретная геометрия и многомасштабное моделирование: Совре-

менные проблемы, приложения и проблемы истории: Материалы XXIII Международной конференции, посвящённой 80-летию профессора Александра Ивановича Галочкина и 75-летию профессора Владимира Григорьевича Чирского. Тула: Тул. гос. пед. ун-т им. Л. Н. Толстого, 2024. С. 38-41.
3. Бухштаб А.А. Теория чисел. Москва: Просвещение, 1966. 383 с.
4. Атанасян Л.С, Базылев В. Т. Геометрия. В 2-х ч. Ч. I. Учеб. пособие для студентов физ.-мат. фак. пед. ин-тув. Москва: Просвещение, 1986. 336 с.
5. Винберг Э.Б. Линейные представления групп. Москва: Наука, Главная редакция физико-математической литературы, 1985. 144 с.
6. Гельфанд И.М. Лекции по линейной алгебре. 4-е изд., доп. Москва: Наука, Главная редакция физико-математической литературы, 1971. 271 с.

ДИНАМИКА И ХАОС В ДВУМЕРНЫХ СИМПЛЕКТИЧЕСКИХ ОТОБРАЖЕНИЯХ

Д.П. Чернышов, А.М. Сатанин

Национальный исследовательский университет «Высшая школа экономики»

В работе исследуется стохастическая динамика двумерных симплектических отображений на примере обобщенного отображения Чирикова, включающего фазовый сдвиг и дополнительное импульсное воздействие. Используя аналитические и численные методы, рассматривается влияние этих модификаций на структуру фазового пространства

Ключевые слова: симплектические отображения, нелинейная динамика, хаос, показатели Ляпунова, численное моделирование

1. Введение

Проблема неинтегрируемости нелинейных дифференциальных уравнений была впервые сформулирована А. Пуанкаре [1]. Согласно его результатам, причина неинтегрируемости заключается в локальной неустойчивости траекторий гамильтоновых динамических систем, обусловленной существованием или возникновением гиперболических точек и сепаратрис. Это, в свою очередь, приводит к разрушению интегралов движения и появлению сложных режимов эволюции системы при изменении её параметров. Дальнейшее развитие этих идей позволило уточнить представления о формировании стохастических слоев вблизи сепаратрис, возникновении и перекрытии резонансов и, как следствие, привело к появлению концепции глобального хаоса [2]. Многие прикладные задачи демонстрируют хаотическое поведение. В двумерном случае к их числу относятся: ограниченная проблема трех тел, движение цепочек атомов в периодическом потенциале (модель Френкеля – Конторовой), описание орбит частиц в ускорителях, ускорение космических частиц (ускорение Ферми) и т. п. Как правило, уравнения, описывающие динамику нелинейных систем, оказываются неинтегрируемыми. Применительно к системам с 3/2 степенями свободы качественная картина возникновения хаоса была развита в работах Б. В. Чирикова [3, 4].

В настоящей работе аналитическими и численными методами исследуется стохастическая динамика двумерной динамической системы на примере обобщенного отображения Чирикова, представляющего собой естественное расширение стандартного отображения за счет введения параметров фазового сдвига и дополнительного внешнего воздействия («бинарный сигнал»). Такая модификация приводит к усложнению структуры областей в фазовом пространстве уже в главном приближении и открывает новые возможности для управления хаотической динамикой, а также для более точного определения границы между регулярными и хаотическими режимами движения в нетривиальных случаях.

2. Обобщенное отображение Чирикова

Рассмотрим модель плоского ротатора, на который действует «бинарная» последовательность импульсов, схематически изображенная на рис. 1. Функция Гамильтона такой системы записывается в виде:

$$H = \frac{I^2}{2} + \cos\theta \left(K \sum_{n=-\infty}^{\infty} \delta(t - nT) - K_\alpha \sum_{n=-\infty}^{\infty} \delta(t - T(n + \tau)) \right),$$

где (θ, I) – соответственно, обобщенная координата (угол) и обобщенный импульс ротатора с единичным моментом инерции; K и K_α – константы, характеризующие воздействие внешней силы на ротатор, T – период внешней силы, τT – сдвиг фазы.

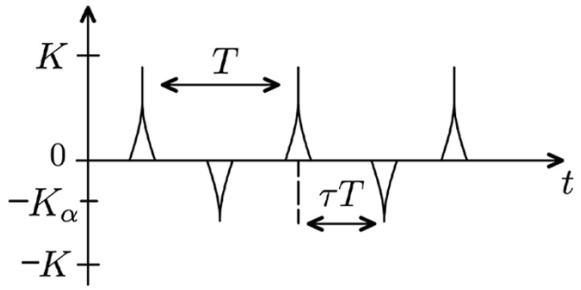


Рис.1. Пример «бинарного сигнала». По вертикальной оси отложена безразмерная амплитуда основного сигнала, равная K, действующего в моменты времени с целыми номерами и амплитуда дополнительного сигнала K_α – сдвинутого по фазе на τT

Поскольку в промежутках между импульсами на ротатор не действует внешняя сила, обобщенный импульс сохраняется, а фаза изменяется с постоянной угловой скоростью. Этот факт позволяет легко получить обобщенное отображение для соответствующих канонических переменных в явном виде. Пусть $I_{\mathrm{n}} \equiv I(nT)$ и $\theta_n \equiv \theta(nT)$ заданы непосредственно перед действием импульса в момент nT. Тогда после изменения $I_{\mathrm{n+\tau}} = I_n + K\sin\theta_n$ под действием «положительного» импульса, приложенного в момент nT, фаза вращается свободно и за время $T\tau$ приобретает приращение $\tau I_{n+\tau}$ (где введены обозначения $I_{\mathrm{n+\tau}} \equiv I(nT+\tau T)$ и $\theta_{n+\tau} \equiv \theta(nT+\tau T)$). Аналогичным образом происходит изменение переменной действия и фазы после действия «отрицательного» импульса в момент $(n+\tau)T$. Следовательно, через период получается комбинированное отображение:

$$\begin{cases} I_{n+\tau} = I_n + K\sin\theta_n, \\ \theta_{n+\tau} = \theta_n + \tau I_{n+\tau}, \\ I_{n+1} = I_{n+\tau} - K_\alpha \sin\theta_{n+\tau}, \\ \theta_{n+1} = \theta_n + (1-\tau) I_{n+\tau}. \end{cases}$$

Отметим, что мы перемасштабировали переменные согласно: $I_n T \to I_n$, $KT \to K$ и $K_\alpha T \to K_\alpha$.

Исключая промежуточные переменные $I_{n+\tau}$ и $\theta_{n+\tau}$, получаем:

$$\begin{cases} I_{n+1} = I_n + K\sin\theta_n - K_\alpha \sin(\theta_n + \tau(I_n + K\sin\theta_n)), \\ \theta_{n+1} = \theta_n + \tau(I_n + K\sin\theta_n) + (1-\tau) I_{n+1}. \end{cases} \quad (1)$$

Очевидно, что при $K_\alpha = 0$ отображение (1) переходит в отображение Чирикова, а в случае $K = 0$, $K_\alpha \neq 0$ – оно снова (после сдвига фазы) сводится к отображению Чирикова.

Отметим, что благодаря периодичности функции $\sin\theta$ динамику системы удобно рассматривать на торе, взяв $[(\theta, I) \bmod 2\pi]$. Математически стандартное отображение представляет собой двумерное гамильтоново (симплектическое) отображение, сохраняющее фазовый объём (якобиан равен 1 в каждой точке фазового пространства).

Дополнительную информацию о характере динамики и структуре фазового пространства системы дают показатели Ляпунова – численные характеристики, отражающие скорость экспоненциального расхождения (или сближения) близких траекторий динамической системы при малом возмущении начальных условий. Путь в фазовом пространстве (на торе) выбрана начальная точка с координатами (I_0, θ_0). Итерируя отображение (1) N-раз, получим траекторию: $(I_0, \theta_0), \ldots, (I_n, \theta_n), \ldots, (I_N, \theta_N)$. Если варьировать начальные условия в малой окрестности, то траектория изменится: $(I_n + \delta I_n, \theta_n + \delta\theta_n)$. Старший показатель Ляпунова определяется соотношением $\Lambda = \lim_{N\to\infty} \Lambda_{\mathrm{N}}$, где:

$$\Lambda_N = \frac{1}{N}\sum_{n=1}^{N} \ln\left(\frac{d_n}{d_{n-1}}\right),$$

а $d_n = \sqrt{\delta I_n^2 + \delta\theta_n^2}$, причем d_0 – начальное расстояние между близкими точками в фазовом пространстве. Инфинитезимальные отклонения находятся совместным решением (1) и:

$$\begin{cases} \delta I_{n+1} = \left(1 - \tau K_\alpha \cos(\theta_n + \tau\left(I_n + K\sin\theta_n\right))\right)\delta I_n + \\ + \left(K\cos\theta_n - K_\alpha \cos(\theta_n + \tau\left(I_n + K\sin\theta_n\right)\left(1 + \tau K\cos\theta_n\right)\right)\delta\theta_n, \\ \delta\theta_{n+1} = \left(1 + \tau K\cos\theta_n\right)\delta\theta_n + \tau\delta I_n + (1-\tau)\delta I_{n+1}. \end{cases} \quad (2)$$

Система (2) позволяет исследовать устойчивость неподвижных точек отображения (1). Например, при $K_\alpha = 0$ найдем, что неподвижная точка $(0, \pi)$ является устойчивой эллиптической точкой при $0 \le K < 4$. Для $K \ge 4$ неподвижная точка $(0, \pi)$ теряет устойчивость и становится гиперболической. При этом эллиптическая неподвижная точка $(0, \pi)$, которая была при $K_\alpha = 0$, становится неустойчивой гиперболической при $K = 0$, $K_\alpha \neq 0$. Интересно, что при $K = K_\alpha$ у отображения (1) появляются неподвижные точки, зависящие от фазы τ. Причем, точка $(0, \pi)$ оказывается устойчивой при любом τ.

3. Численное моделирование

Ниже мы обсудим новые сценарии взаимодействия резонансов в обобщенном отображении Чирикова (1) и приведем дополнительное исследование фазового пространства, а также вычислим старшие показатели Ляпунова, характеризующие резонансную картину.

3.1. Новые сценарии взаимодействия резонансов

Ключевым отличием предлагаемого обобщенного отображения Чирикова (1) от стандартного отображения является использование обобщенных δ-функций, чередующихся по знаку. В отличие от стандартного случая, где на ротатор действует последовательность одинаковых по амплитуде и знаку импульсов, в (1) учитываются импульсы противоположного знака. Это приводит к возникновению новых типов резонансных явлений, которые мы обозначаем как антирезонансы.

На рис. 2 представлены примеры построения сечений Пуанкаре для обобщенного отображения Чирикова при различных значениях параметров $(K,\ K_\alpha,\ \tau,)$. На верхней панели показаны несколько сценариев динамики системы: при отсутствии дополнительных δ-импульсов $(K \neq 0,\ K_\alpha = 0)$ – отображение Чирикова, взаимная аннигиляция резонансов и антирезонансов $(K = K_\alpha,\ \tau = 0)$, одновременное существование резонансов и антирезонансов $(K = K_\alpha,\ \tau \neq 0)$. На нижней панели – динамика системы в условиях сильного внешнего возмущения ($|K - K_\alpha| > K_c$, где K_c – критическое значение, при котором отображение Чирикова переходит в хаотический режим [5]) при различных значениях параметра фазового сдвига τ.

Легко заметить, что с увеличением амплитуды внешнего возмущения динамика системы претерпевает существенные изменения. При $|K - K_\alpha| < K_c$ и $K < K_\alpha$ наблюдается поглощение основного резонанса, контролируемого параметром K, антирезонансом, индуцированным дополнительной серией δ-импульсов. По мере дальнейшего роста возмущения формируется стохастический слой между антирезонансами. В предельном случае антирезонансы сталкиваются, что приводит систему к сценарию глобального хаоса: траектории смешиваются в фазовом пространстве, образуя «стохастическое море» с небольшими вкраплениями «островков» регулярного движения.

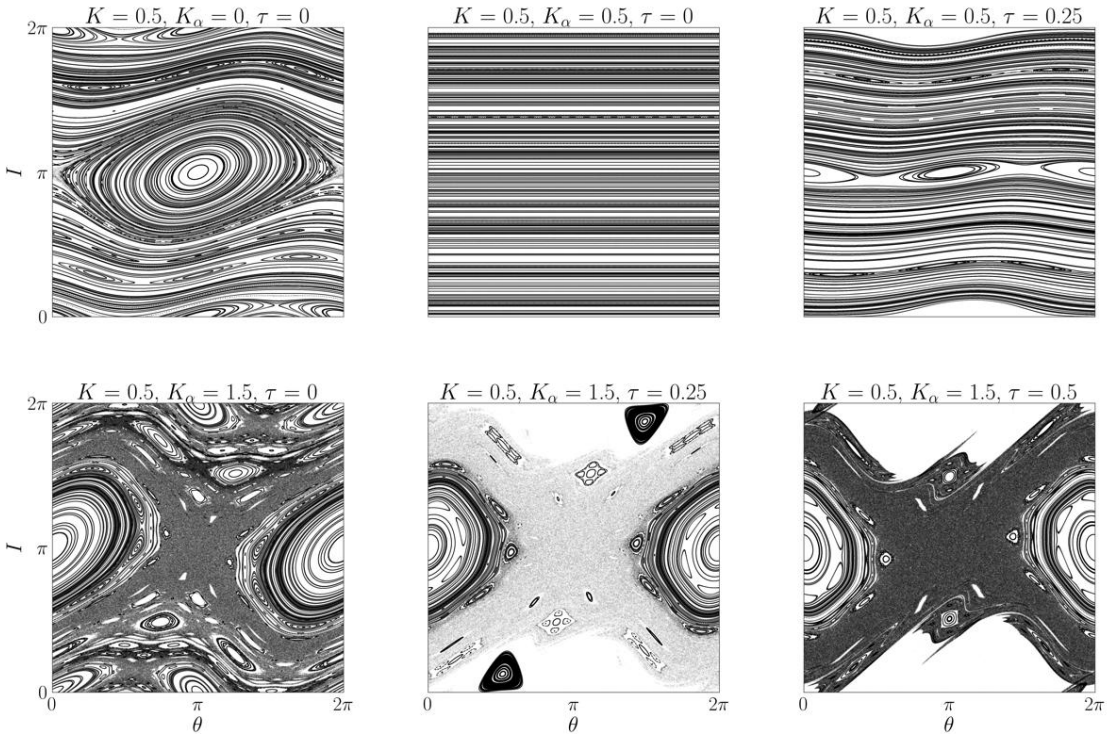


Рис.2. Фазовые портреты обобщенного отображения Чирикова (2)

3.2. Показатели Ляпунова

Вычисление показателей Ляпунова – это нетривиальная задача, особенно для многомерных и гамильтоновых систем. Один из наиболее надежных и теоретически обоснованных подходов к численному определению полного спектра экспонент был предложен в работе [6], где рассмотрены общие принципы вычислений и реализован устойчивый алгоритм на основе ортонормирования возмущений методом Грама–Шмидта.

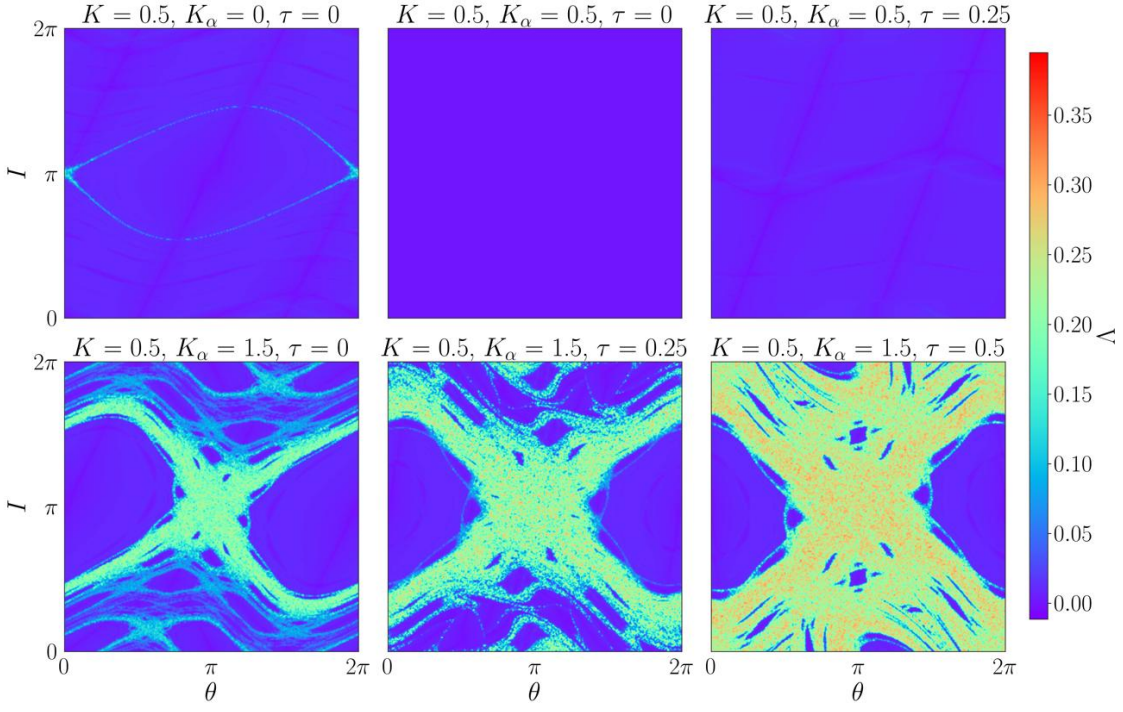


Рис.3. Карты старших показателей Ляпунова Λ для параметров, используемых при построении фазовых портретов обобщенного отображения Чирикова (рис. 2)

Построенные карты старших показателей Ляпунова (рис. 3) позволяют детально исследовать сложную стохастическую динамику обобщенного отображения Чирикова (2). Визуализация значений старших показателей Ляпунова демонстрирует, как на сепаратрисе формируется хаос – экспоненциально узкий стохастический слой, который впоследствии исчезает под воздействием дополнительной внешней силы, компенсирующей основное возмущение ($K_\alpha > K$). Дальнейшая эволюция системы следует сценарию, качественно описанному в пункте 3.1. Отметим, что детальное описание тонкой фрактальной структуры фазового портрета, формируемого отображением Чирикова, до настоящего времени остается недостаточно изученным.

4. Обсуждение результатов

Стандартное отображение реализуется при параметрическом возбуждении нелинейного ротатора или аналогичных нелинейных систем, подвергающихся периодическому воздействию импульсного типа. В работе показано, что обобщенное отображение Чирикова естественным образом возникает при обобщении вида управляющей функции

Для экспериментальной реализации обобщенного отображения Чирикова могут быть использованы высокодобротные джозефсоновские осцилляторы, которые нетрудно сопрячь с волноводными линиями. Такие устройства способны генерировать широкополосный шум, представляющий интерес для применения в сверхпроводящей электронике. Существенным отличием обобщенного отображения от стандартного является возможность совместного существования и взаимодействия резонансов и антирезонансов, что открывает новый сценарий математического управления динамическим хаосом.

Интересным может быть построение многомерного отображения Чирикова и использование технологий машинного обучения для поиска сохраняющихся интегралов движения в системе такого вида [7].

Исследование осуществлено в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики».

Литература

1. Пуанкаре А. Новые методы небесной механики // Избр. труды: В 3 т. М.: Наука, 1971–1974.
2. Lichtenberg A. J., Lieberman M. A. Regular and stochastic motion. Т. 38. – Springer, 1983.
3. Чириков Б. В. Исследования по теории нелинейного резонанса и стохастичности // Препринт ИЯФ. – 1969. – Т. 267.
4. Chirikov B. V. A universal instability of many-dimensional oscillator systems //Physics reports. – 1979. – Т. 52. – №. 5. – С. 263–379.
5. Greene J. M. A method for determining a stochastic transition // Journal of Mathematical Physics. – 1979. – Т. 20. – №. 6. – С. 1183–1201.
6. Benettin G. et al. Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory //Meccanica. – 1980. – Т. 15. – №. 1. – С. 9–20.
7. Chernyshov D., Satanin A., Shchur L. Determining the boundary of dynamical chaos in the generalized Chirikov map via machine learning // arXiv preprint arXiv:2509.11593. – 2025.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПЛОСКИХ СТРУЙНЫХ ТЕЧЕНИЙ ИДЕАЛЬНОЙ ЖИДКОСТИ

М.А. Шадт

Казанский (Приволжский) федеральный университет

Разработан комплексный метод последовательных приближений (КМПП) для построения неизвестных границ в струйных течениях идеальной несжимаемой жидкости, основанный на сочетании комплексного метода граничных элементов (КМГЭ) и принципа баланса расхода (ПБР). Сформулирован и доказан ПБР в естественной системе координат для свободных линий тока и эквипотенциалей.

Ключевые слова: комплексный метод граничных элементов (КМГЭ), принцип баланса расхода (ПБР), комплексный метод последовательных приближений (КМПП), теория струй идеальной жидкости, комплексный потенциал, интеграл типа Коши.

1. Введение

В теории струйных, кавитационных и волновых течений идеальной несжимаемой жидкости наличие неизвестных границ является источником существенных математических трудностей на пути построения и анализа решений. Постановка и решение первых струйных задач были проведены Гельмгольцем и Кирхгофом в середине XIX века [1,2]. Плодотворная идея применения аппарата теории функций комплексного переменного в гидромеханике также принадлежит Кирхгофу. С тех пор данная область гидромеханики активно развивалась трудами многих выдающихся ученых по всему миру. Бурное развитие данной тематики привело к исчерпанию задач, поддающихся аналитическому решению, а стремительное совершенствование вычислительной техники побудило исследователей к созданию соответствующих численных методов [1,3,4,5].

2. Комплексный метод граничных элементов

2.1. Описание КМГЭ

Интегральная формула Коши и интеграл типа Коши составляют основную теоретическую предпосылку для численных исследований краевых задач теории аналитических функций с помощью КМГЭ [3,4,6]. Отличительной чертой многих задач гидромеханики является то, что на каждой части границы течения известна либо вещественная, либо мнимая компонента аналитической функции (комплексного потенциала). Такие задачи называются смешанными краевыми задачами теории аналитических функций.

Комплексный метод граничных элементов позволяет решить следующую задачу.

Пусть на плоскости имеется некоторая область Ω, ограниченная многоугольником (Рис.1). Для простоты будем считать, что область односвязная (хотя в общем случае это необязательно). В каждом узле многоугольника известно значение либо только вещественной части аналитической функции w=w(z), либо только ее мнимой части. Необходимо восстановить недостающие сведения. Иными словами, если в узле известно значение вещественной части, то нужно определить значение мнимой части. Если же в узле известно значение мнимой части, то нужно определить значение вещественной части.

Важно отметить, что сама функция $w = w(z)$ нам неизвестна, мы знаем только, что она аналитическая. КМГЭ позволяет после восстановления недостающих значений аналитической функции восстановить и саму эту функцию (приближенно, но с любой точностью, которая зависит от количества узлов).

В КМГЭ значение аналитической функции обозначается прямой чертой над величиной

$\overline{w}_j = \bar{\varphi}_j + i\bar{\psi}_j$ – значение аналитической функции $w = w(z)$ в узле с номером j, вводится также понятие аппроксимирующей функции, которая обозначается крышкой сверху_$\hat{w}(z_k)$. Обход контура осуществляется так, чтобы область оставалась слева.

Для каждого узла необходимо записать аппроксимирующую функцию:

$$\hat{w}(z_k) = \frac{1}{2\pi i}\overline{w}_k ln\left(\frac{z_{k+1}-z_k}{z_{k-1}-z_k}\right) + \frac{1}{2\pi i}\sum_{\substack{j=1 \\ j\neq k, j\neq k-1}}^{m}\left[\left(\frac{z_k-z_j}{z_{j+1}-z_j}\overline{w}_{j+1} - \frac{z_k-z_{j+1}}{z_{j+1}-z_j}\overline{w}_j\right) ln\left(\frac{z_{j+1}-z_k}{z_j-z_k}\right)\right]$$

Аппроксимирующая функция может быть представлена также в виде, в котором она и используется при программировании метода:

$$\hat{w}(z_k) = \frac{1}{2\pi i}\overline{w}_k ln\left|\frac{z_{k+1}-z_k}{z_{k-1}-z_k}\right| + \frac{\overline{w}_k}{2\pi}\left(arg(z_{k+1}-z_k) - arg(z_{k-1}-z_k)\right) + \frac{1}{2\pi i}\sum_{\substack{j=1 \\ j\neq k, j\neq k-1}}^{m}\left[\left(\frac{z_k-z_j}{z_{j+1}-z_j}\overline{w}_{j+1} - \frac{z_k-z_{j+1}}{z_{j+1}-z_j}\overline{w}_j\right) ln\left(\frac{z_{j+1}-z_k}{z_j-z_k}\right)\right]$$

$$0 \le arg(z_{k+1}-z_k) < 2\pi, \quad 0 \le arg(z_{k-1}-z_k) < 2\pi$$

$$arg(z_{k+1}-z_k) - arg(z_{k-1}-z_k) = \begin{cases} arg(z_{k+1}-z_k) - arg(z_{k-1}-z_k), \\ \text{если } arg(z_{k+1}-z_k) - arg(z_{k-1}-z_k) > 0 \\ 2\pi + arg(z_{k+1}-z_k) - arg(z_{k-1}-z_k), \\ \text{если } arg(z_{k+1}-z_k) - arg(z_{k-1}-z_k) < 0 \end{cases}$$

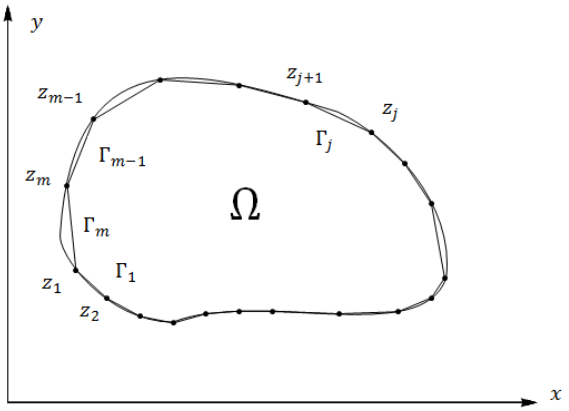


Рис. 1. Разбиение границы области на граничные элементы

Сначала необходимо записать аппроксимирующую функцию для каждого узла границы, оставляя выражение $\overline{w}_j = \bar{\varphi}_j + i\bar{\psi}_j$ в символьном виде, а координаты узлов заменить на их числовые значения. Разделяя теперь действительные и мнимые части, получим систему линейных алгебраических уравнений (СЛАУ) относительно неизвестных составляющих $\overline{w}_j = \bar{\varphi}_j + i\bar{\psi}_j$. Данная система имеет вид $AX + iBX = 0$, причем матирцы A и B – вещественные размера $m \times 2m$ (m строк и $2m$ столбцов), а матрица-столбец $X = X(\bar{\varphi}_1, \bar{\psi}_1, \bar{\varphi}_2, \bar{\psi}_2, \dots, \bar{\varphi}_m, \bar{\psi}_m)$ содержит известные и неизвестные составляющие $\overline{w}_j = \bar{\varphi}_j + i\bar{\psi}_j$. Производя последующее разделение вещественных и мнимых частей и подставляя вместо символа $\overline{w}_j = \bar{\varphi}_j + i\bar{\psi}_j$ его числовое значение, приходим к вещественной СЛАУ, из решения которой определяются недостающие значения $\overline{w}_j = \bar{\varphi}_j + i\bar{\psi}_j$ в узловых точках. Зная теперь все значения $\overline{w}_j = \bar{\varphi}_j + i\bar{\psi}_j$, саму аналитическую функцию (комплексный потенциал) можно восстановить по формуле:

$$w(z) = \frac{1}{2\pi i}\sum_{j=1}^{m}\left[\left(\frac{z-z_j}{z_{j+1}-z_j}\overline{w}_{j+1} - \frac{z-z_{j+1}}{z_{j+1}-z_j}\overline{w}_j\right) ln\left(\frac{z_{j+1}-z}{z_j-z}\right)\right]$$

2.2. Модифицированный КМГЭ

Комплексный метод граничных элементов применим к любой однозначной аналитической функции комплексного переменного. Одной из таких функций в гидродинамике является комплексно-сопряженная скорость $W(z) = v_x(x,y) - iv_y(x,y)$. Специфической особенностью многих задач гидродинамики с неизвестной границей является то, что на одних частях границы известна нормальная компонента скорости v_n, а на других – касательная компонента v_s. Учитывая эту особенность, интегральную формулу Коши можно записать в виде:

$$W(z_0) = \big(v_n(z_0) - iv_s(z_0)\big)e^{-i\alpha(z_0)} = \frac{1}{2\pi i}\int\limits_{\Gamma} \frac{\big(v_n(z) - iv_s(z)\big)e^{-i\alpha(z)}}{z - z_0}\,dz$$

В этой формуле угол α является углом между внешней к области нормалью и осью абсцисс.

Далее все делается по аналогии с вышеизложенным. Для каждого узла области записывается аппроксимирующая функция:

$$\widehat{W}(z_k) = \frac{1}{2\pi i}\bar{W}_k ln\left|\frac{z_{k+1} - z_k}{z_{k-1} - z_k}\right| + \frac{\bar{W}_k}{2\pi}\big(arg(z_{k+1} - z_k) - arg(z_{k-1} - z_k)\big) + \frac{1}{2\pi i}\sum_{\substack{j=1 \\ j\neq k, j\neq k-1}}^{m}\left[\left(\frac{z_k - z_j}{z_{j+1} - z_j}\bar{W}_{j+1} - \frac{z_k - z_{j+1}}{z_{j+1} - z_j}\bar{W}_j\right) ln\left(\frac{z_{j+1} - z_k}{z_j - z_k}\right)\right]$$

$$0 \le arg(z_{k+1} - z_k) < 2\pi, \quad 0 \le arg(z_{k-1} - z_k) < 2\pi$$

$$arg(z_{k+1} - z_k) - arg(z_{k-1} - z_k) = \begin{cases} arg(z_{k+1} - z_k) - arg(z_{k-1} - z_k), \\ \text{если } arg(z_{k+1} - z_k) - arg(z_{k-1} - z_k) > 0 \\ 2\pi + arg(z_{k+1} - z_k) - arg(z_{k-1} - z_k), \\ \text{если } arg(z_{k+1} - z_k) - arg(z_{k-1} - z_k) < 0 \end{cases}$$

$$\bar{W}_j = \bar{v}_{nj}exp\big(-\bar{\alpha}_j\big) - i\bar{v}_{sj}exp\big(-\bar{\alpha}_j\big)$$

$$exp\big(-\bar{\alpha}_j\big) = Cos\big(\bar{\alpha}_j\big) - iSin\big(\bar{\alpha}_j\big)$$

$$Sin\big(\bar{\alpha}_j\big) = Im\left[-i\frac{z_{j+1}-z_j}{|z_{j+1}-z_j|}\right] \qquad Cos\big(\bar{\alpha}_j\big) = Re\left[-i\frac{z_{j+1}-z_j}{|z_{j+1}-z_j|}\right]$$

Разделяя аналогично действительные и мнимые части, получим систему линейных алгебраических уравнений $AX + iBX = 0$. Тут матрица-столбец $X = X(\bar{v}_{n1}, \bar{v}_{s1}, \bar{v}_{n2}, \bar{v}_{s2}, \dots, \bar{v}_{nm}, \bar{v}_{sm})$. Повторное разделение вещественных и мнимых частей позволяет получить матричное уравнение $GY = F$. Матрица G и вектор правой части F получаются по следующему правилу [4] : если в узле z_j задана касательная составляющая скорости $\bar{v}_{sj}$, то берется j-ая строка матрицы B и после выборки элементов строки, соответствующих неизвестным значениям $\bar{v}_s$ или $\bar{v}_n$ во всех остальных узлах, получается j-ая строка матрицы G, j-ый элемент вектора Y будет $\bar{v}_{nj}$, j-ый элемент вектора F – сумма известных значений $\bar{v}_s$ или $\bar{v}_n$, умноженные на соответствующие элементы матрицы B. Если в узле z_j задана нормальная составляющая скорости $\bar{v}_{nj}$, то для построения j-ой строки матрицы G используется матрица A.

Тогда сама комплексно-сопряженная скорость $W(z)$может быть найдена по формуле:

$$W(z) = \frac{1}{2\pi i}\sum_{j=1}^{m}\left[\left(\frac{z - z_j}{z_{j+1} - z_j}\bar{W}_{j+1} - \frac{z - z_{j+1}}{z_{j+1} - z_j}\bar{W}_j\right) ln\left(\frac{z_{j+1} - z}{z_j - z}\right)\right]$$

Комплексный потенциал течения $w(z)$ определяется с точностью до комплексной постоянной, которая физического смысла не имеет и может быть положена равной нулю. Учитывая это, комплексный потенциал течения $w(z)$ находим интегрированием последнего выражения для комплексно-сопряженной скорости $W(z)$: $w(z) = \int_0^z W(z)dz$.

3. Принцип баланса расхода

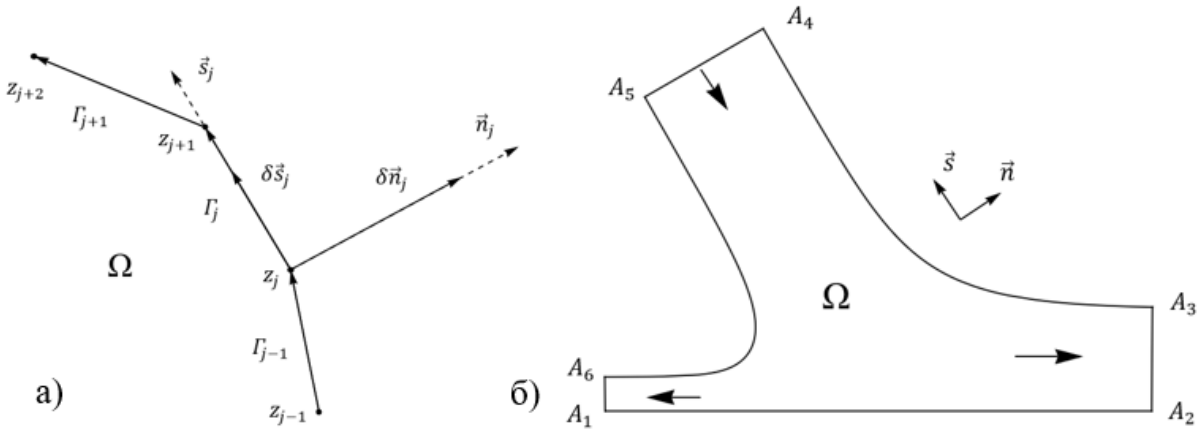


Рис. 2. а) Элемент границы б) Расчетная область

На свободной границе струи (свободные линии тока) A_3A_4 и A_5A_6 постоянен модуль скорости $v_s = \pm v_0 = \pm 1$ (знак зависит от выбора направления осей естественной системы координат), а нормальная компонента скорости равна нулю $v_n = 0$. На линиях равного потенциала A_2A_3 и A_6A_1 касательная компонента скорости равна нулю $v_s = 0$, а нормальная компонента скорости принимает постоянное значение $v_n = \pm 1$. Эта особенность задачи делает ее рассмотрение более удобным в естественной системе координат $(\vec{n}, \vec{s})$. Известно [8], что при принятом здесь введении естественной системы координат (Рис.2) выполняются соотношения (условия Коши–Римана):

$$v_n = \frac{\partial\varphi}{\partial n} = \frac{\partial\psi}{\partial s}; \quad v_s = \frac{\partial\varphi}{\partial s} = -\frac{\partial\psi}{\partial n}$$

Пусть

$$\psi = \psi(n,s) = \psi\big(n(x,y), s(x,y)\big)$$

Тогда, пользуясь свойством инвариантности формы первого дифференциала и условиями Коши–Римана, можно записать:

$$dn = \frac{\partial n}{\partial x}dx + \frac{\partial n}{\partial y}dy; \quad ds = \frac{\partial s}{\partial x}dx + \frac{\partial s}{\partial y}dy$$

$$\frac{\partial\psi}{\partial x} = \frac{\partial\psi}{\partial n}\frac{\partial n}{\partial x} + \frac{\partial\psi}{\partial s}\frac{\partial s}{\partial x}; \quad \frac{\partial\psi}{\partial y} = \frac{\partial\psi}{\partial n}\frac{\partial n}{\partial y} + \frac{\partial\psi}{\partial s}\frac{\partial s}{\partial y}$$

$$d\psi = \frac{\partial\psi}{\partial x}dx + \frac{\partial\psi}{\partial y}dy = \left(\frac{\partial\psi}{\partial n}\frac{\partial n}{\partial x} + \frac{\partial\psi}{\partial s}\frac{\partial s}{\partial x}\right)dx + \left(\frac{\partial\psi}{\partial n}\frac{\partial n}{\partial y} + \frac{\partial\psi}{\partial s}\frac{\partial s}{\partial y}\right)dy =$$

$$= \frac{\partial\psi}{\partial n}\left(\frac{\partial n}{\partial x}dx + \frac{\partial n}{\partial y}dy\right) + \frac{\partial\psi}{\partial s}\left(\frac{\partial s}{\partial x}dx + \frac{\partial s}{\partial y}dy\right) = \frac{\partial\psi}{\partial n}dn + \frac{\partial\psi}{\partial s}ds = -v_s dn + v_n ds$$

Вдоль свободных линии тока A_3A_4 и A_5A_6 функция тока $\psi = \psi(n,s) = Const$ сохраняет постоянное значение и $d\psi = 0$. Это позволяет сформулировать принцип баланса расхода (ПБР) для линии тока: $v_s dn = v_n ds$ или в дискретном виде: $v_{sj}\Delta n_j = v_{nj}\Delta s_j$, где $\Delta s_j = \left|\Delta\vec{s}_j\right| = \left|z_{j+1} - z_j\right| = \left|\Gamma_j\right|$.

Полагая теперь, что

$$\varphi = \varphi(n,s) = \varphi\big(n(x,y), s(x,y)\big)$$

Аналогично предыдущему можно записать:

$$dn = \frac{\partial n}{\partial x}dx + \frac{\partial n}{\partial y}dy; \quad ds = \frac{\partial s}{\partial x}dx + \frac{\partial s}{\partial y}dy$$

$$\frac{\partial\varphi}{\partial x}=\frac{\partial\varphi}{\partial n}\frac{\partial n}{\partial x}+\frac{\partial\varphi}{\partial s}\frac{\partial s}{\partial x};\quad \frac{\partial\varphi}{\partial y}=\frac{\partial\varphi}{\partial n}\frac{\partial n}{\partial y}+\frac{\partial\varphi}{\partial s}\frac{\partial s}{\partial y}$$

$$d\varphi=\frac{\partial\varphi}{\partial x}dx+\frac{\partial\varphi}{\partial y}dy=\left(\frac{\partial\varphi}{\partial n}\frac{\partial n}{\partial x}+\frac{\partial\varphi}{\partial s}\frac{\partial s}{\partial x}\right)dx+\left(\frac{\partial\varphi}{\partial n}\frac{\partial n}{\partial y}+\frac{\partial\varphi}{\partial s}\frac{\partial s}{\partial y}\right)dy=$$
$$=\frac{\partial\varphi}{\partial n}\left(\frac{\partial n}{\partial x}dx+\frac{\partial n}{\partial y}dy\right)+\frac{\partial\varphi}{\partial s}\left(\frac{\partial s}{\partial x}dx+\frac{\partial s}{\partial y}dy\right)=\frac{\partial\varphi}{\partial n}dn+\frac{\partial\varphi}{\partial s}ds=v_n dn+v_s ds$$

Вдоль линий равного потенциала A_2A_3 и A_6A_1 функция $\varphi=\varphi(n,s)=Const$ сохраняет постоянное значение и $d\varphi=0$. Это позволяет сформулировать принцип баланса расхода (ПБР) для эквипотенциали: $v_n dn=-v_s ds$ или в дискретном виде: $v_{nj}\Delta n_j=-v_{sj}\Delta s_j$, где $\Delta s_j=\left|\Delta\vec{s}_j\right|=\left|z_{j+1}-z_j\right|=\left|\Gamma_j\right|$. Неизвестные границы струи сначала задаются произвольно, а потом уточняются по формулам: $x_j'=x_j+\Delta x_j,\ y_j'=y_j+\Delta y_j$, где $x_j',\ y_j'$ – новые координаты, $x_j,\ y_j$ – старые координаты, $\Delta x_j,\ \Delta y_j$ – смещения точек неизвестной границы.

Установим формулы для смещений $\Delta x_j,\ \Delta y_j$. Пусть $\vec{s}_j$ – единичный вектор касательной, проведенный к граничному элементу Γ_j в точке z_j, а $\vec{n}_j$ – единичный вектор нормали, проведенный в той же точке (Рис.2). В этом случае верно равенство:

$$\vec{n}_j=-i\vec{s}_j=-i\frac{z_{j+1}-z_j}{\left|z_{j+1}-z_j\right|}=-i\frac{\left(x_{j+1}+iy_{j+1}\right)-\left(x_j+iy_j\right)}{\left|\Delta\vec{s}_j\right|}=\frac{\left(y_{j+1}-y_j\right)-i\left(x_{j+1}-x_j\right)}{\Delta s_j}$$

Умножим это векторное равенство на число Δn_j:

$$\Delta n_j\vec{n}_j=\delta\vec{n}_j=\frac{\Delta n_j\left(y_{j+1}-y_j\right)-i\Delta n_j\left(x_{j+1}-x_j\right)}{\Delta s_j}$$

Это векторное равенство эквивалентно двум скалярным равенствам:

$$\Delta x_j=\delta n_{jx}=\frac{\Delta n_j}{\Delta s_j}\left(y_{j+1}-y_j\right)\ \text{и}\ \Delta y_j=\delta n_{jy}=-\frac{\Delta n_j}{\Delta s_j}\left(x_{j+1}-x_j\right).$$

Считая, что смещаемые точки принадлежат линии тока, выразим из принципа баланса расхода отношение $\frac{\Delta n_j}{\Delta s_j}=\frac{v_{nj}}{v_{sj}}$ и подставим его в два предыдущих равенства:

$$\Delta x_j=\frac{v_{nj}}{v_{sj}}\left(y_{j+1}-y_j\right)\ \text{и}\ \Delta y_j=-\frac{v_{nj}}{v_{sj}}\left(x_{j+1}-x_j\right).$$

Данные равенства позволяют смещать точки, лежащие на неизвестной линии тока при условии, что скорость направлена по вектору единичной касательной (линия тока A_5A_6 (Рис.2б)). Если же направление скорости и вектора единичной касательной не совпадают, то необходимо использовать формулы(линия тока A_3A_4 (Рис.2б)):

$$\Delta x_j=-\frac{v_{nj}}{v_{sj}}\left(y_{j+1}-y_j\right)\ \text{и}\ \Delta y_j=\frac{v_{nj}}{v_{sj}}\left(x_{j+1}-x_j\right).$$

Точки, лежащие на эквипотенциали, необходимо смещать вдоль касательной.

$$\vec{s}_j=\frac{z_{j+1}-z_j}{\left|z_{j+1}-z_j\right|}=\frac{\left(x_{j+1}+iy_{j+1}\right)-\left(x_j+iy_j\right)}{\left|\Delta\vec{s}_j\right|}=\frac{\left(x_{j+1}-x_j\right)+i\left(y_{j+1}-y_j\right)}{\Delta s_j}$$

Умножим это векторное равенство на число Δn_j: $\Delta n_j\vec{s}_j=\delta\vec{s}_j=\frac{\Delta n_j\left(x_{j+1}-x_j\right)+i\Delta n_j\left(y_{j+1}-y_j\right)}{\Delta s_j}$.

Путем аналогичных рассуждений с использованием принципа баланса расхода для эквипотенциали $\left(\frac{\Delta n_j}{\Delta s_j}=-\frac{v_{sj}}{v_{nj}}\right)$, можно получить формулы:

$$\Delta x_j=-\frac{v_{sj}}{v_{nj}}\left(x_{j+1}-x_j\right)\ \text{и}\ \Delta y_j=-\frac{v_{sj}}{v_{nj}}\left(y_{j+1}-y_j\right).$$

Эти равенства позволяют смещать точки, лежащие на эквипотенциали при условии, что скорость направлена ппротив вектора единичной нормали (эквипотенциаль A_4A_5 (Рис.2б)). Эта эквипотенциаль соответствуют входной границе струи. Однако, данные равенства не находят применения, потому что входная граница струи считается известной. Твердая стенка A_1A_2 также считается заданной.

Для выходных границ (эквипотенциали A_1A_6 и A_2A_3 (Рис.2б)) формулы смещения имеют вид:

$$\Delta x_j=\frac{v_{sj}}{v_{nj}}\left(x_{j+1}-x_j\right)\ \text{и}\ \Delta y_j=\frac{v_{sj}}{v_{nj}}\left(y_{j+1}-y_j\right).$$

4. Результаты расчетов

Модифицированный КМГЭ позволяет задавая в точках на границе области Ω только одну компоненту скорости v_{nj} или v_{sj}, получить недостающие компоненты. Соответствующие формулы смещений, полученные на основании принципа баланса расхода, позволяют занулить нормальную компоненту скорости на линиях тока и касательную компоненту скорости на эквипотенциалях. В условиях данной задачи компоненты скорости задавались следующим образом: на твердой стенке A_1A_2 $v_n = 0$ (условие непротекания), на эквипотенциалях A_2A_3 и A_6A_1 $v_n = 1$ (выходные границы), на линии тока A_3A_4 $v_s = -1$, на линии тока A_5A_6 $v_s = 1$, на эквипотенциали A_4A_5 $v_n = -1$ (входная граница). По найденным значениям компонент скорости можно восстановить функцию $W(z) = v_x(x, y) - iv_y(x, y)$ и далее построить комплексный потенциал течения $w(z) = \int_0^z W(z)dz$. Вычисляя значения функции $w(z)$ в интересующих точках области, можно построить внутренние линии тока и линии равного потенциала по методике [7] (рис. 3).

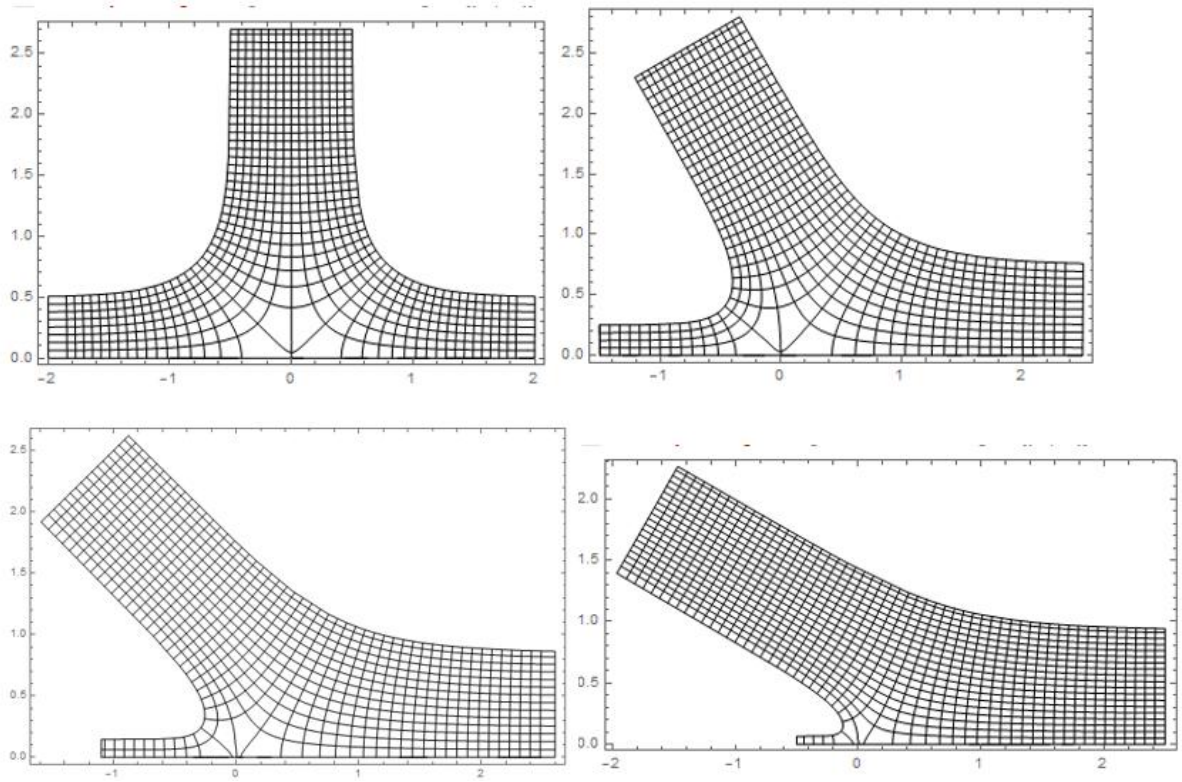


Рис. 3. Линии тока и линии равного потенциала для различных углов натекания струи

5. Заключение

Разработан комплексный вариант метода последовательных приближений для построения неизвестных границ плоских струйных течений идеальной жидкости, основанный на сочетании КМГЭ и ПБР. Сформулирован и доказан ПБР в естественной системе координат для свободных линий тока и эквипотенциалей. Получены формулы для последовательных смещений точек, расположенных на неизвестных границах. Эти формулы удобны тем, что не содержат величин Δn_j и Δs_j, а используют только компоненты скорости. Таким образом, величины Δn_j и Δs_j не вычисляются, что сокращает время работы программы. С помощью предложенного метода построено решение классической задачи о натекании струи на бесконечную стенку для четырех значений углов натекания. В каждом случае построены линии тока и линии равного потенциала. Необходимо отметить, что отличительной чертой метода является возможность построения комплексного потенциала течения.

Литература

1. Kuramoto Y., Battogtokh D. Coexistence of coherence and incoherence in nonlocally coupled phase oscillators // Nonlinear Phenomena in Complex Systems. 2002. Vol. 5, No. 4. P. 380–385. DOI: 10.1007/BF02706196.
1. Гуревич М.И. Теория струй идеальной жидкости. – М.: Наука, 1979. 536 с.
2. Толоконников С. Л. Введение в теорию струйных течений идеальной несжимаемой жидкости (учебное пособие). – М.: Издательство Центра прикладных исследований при механико-математическом факультете МГУ, 2008. 54 с.
3. Громадка Т., Лей Ч. Комплексный метод граничных элементов в инженерных задачах. – М : Мир, 1990. 303 с.
4. Афанасьев К.Е., Стуколов С.В. КМГЭ для решения плоских задач гидродинамики и его реализация на параллельных компьютерах. – Кемерово: КемГУ, 2000. 208 с.
5. Макаров В.В. Численное определение свободных границ в струйных течениях. //ВИНИТИ 2000, №2729–ВОО.
6. Мусхелишвили Н.И. Сингулярные интегральные уравнения. – М.: Наука, 1968. 511 с
7. Шадт М.А. Методика построения внутренних линий тока в задачах теории струй// В сб.: Математическое моделирование и суперкомпьютерные технологии. Труды XXII Международной конференции. Сер. "Суперкомпьютерные дни в России". Нижний Новгород,2022. С 165–167.
8. Голубева О.В. Курс механики сплошных сред. – М.: Высшая школа, 1972. 367 с.

ИССЛЕДОВАНИЕ ПОВЕДЕНИЯ RISC-V GPGPU VORTEX ПРИ ЗАПУСКЕ БИБЛИОТЕКИ РАЗРЕЖЕННОЙ ЛИНЕЙНОЙ АЛГЕБРЫ

Ш.Р. Шаехмитов

Санкт-Петербургский государственный университет

Доклад для молодёжной школы посвящён работе с библиотекой разреженной линейной алгебры на платформе Vortex GPU. Основное внимание уделяется практическим аспектам подготовки окружения, сборки и интеграции библиотек, а также типичным проблемам, возникающим при переносе вычислений на данную архитектуру. В ходе работы были выявлены характерные ошибки совместимости, а также пути их устранения. Доклад будет полезен разработчикам и инженерам, которые планируют использовать Vortex GPU в задачах высокопроизводительных вычислений и хотят заранее учесть возможные технические нюансы.

Ключевые слова: RISC-V, Vortex GPU, разреженная линейная алгебра, ошибки совместимости, анализ графов

1. Введение

Разреженная линейная алгебра является основой многих алгоритмов анализа графов. В частности, стандарт GraphBLAS[2] формализует представление операций на графах через примитивы линейной алгебры над разреженными матрицами, что делает данный подход эффективным и переносимым. Для ускорения вычислений в подобных задачах применяются графические ускорители, обеспечивающие высокую степень параллелизма. Платформа Vortex GPU[3], основанная на архитектуре RISC-V и ориентированная на выполнение вычислений с использованием стандарта OpenCL 1.2, представляет интерес как открытое и гибко настраиваемое решение для высокопроизводительных вычислений. Однако в ходе работы над запуском библиотеки SPLA[1] на Vortex GPU были обнаружены некоторые проблемы в компиляторной инфраструктуре, а также выявлено отсутствие поддержки некоторых операций. Таким образом, изучение особенностей настройки вычислительного окружения, устранения возникающих проблем при работе со сторонними библиотеками на Vortex GPU является актуальной задачей, имеющей практическое значение для расширения применимости данной платформы в высокопроизводительных вычислениях.

2. Цели и задачи работы

Целью работы является анализ практических аспектов использования библиотеки SPLA на платформе Vortex GPU и выявление технических особенностей, возникающих при настройке и запуске вычислений.

Для достижения поставленной цели решаются следующие задачи:

- настройка сборки библиотеки разреженной линейной алгебры для работы с Vortex GPU;
- поиск и исправление ошибок в среде исполнения для Vortex GPU;
- изменение кода библиотеки разреженной линейной алгебры для поддержки работы с памятью устройства;
- изучение поддержки устройством атомарных операций и добавление их частичной поддержки в компилятор PoCL.

3. Выполнение поставленных задач

При выполнении программы, содержащей несколько вычислительных ядер OpenCL, возникла проблема, связанная с корректным переключением между ними при работе на платформе Vortex GPU. В ходе тестирования было обнаружено, что при последовательном вызове различных ядер исполнялось только первое из них, причём его выполнение повторялось несколько раз вместо запуска последующих. Такое поведение указывает на некорректную обработку командной очереди внутри реализации PoCL на Vortex. Проверка параметров вызова ядер, размеров рабочих групп и конфигурации контекста не выявила ошибок в программном коде, что позволило связать проблему непосредственно с особенностями среды исполнения. Анализ кода PoCL позволил выявить, что проблемы с повторным исполнением ядра содержатся в коде среды исполнения, предназначенной для Vortex. Данная проблема потребовала внесения изменений в код PoCL, чтобы избавиться от этого поведения среды исполнения.

В ходе работы с библиотекой возникла проблема, связанная с отсутствием поддержки операции копирования буфера непосредственно внутри устройства. В стандартных реализациях OpenCL такая операция позволяет эффективно перемещать данные между областями памяти GPU без участия центрального процессора. Однако в реализации OpenCL на платформе Vortex GPU соответствующая функциональность не была доступна, что исключало возможность прямого использования команды “copyBuffer” для копирования данных внутри устройства. В результате для переноса данных приходилось использовать последовательное чтение и запись буфера с участием хост-памяти, что усложняло код и увеличивало накладные расходы на обмен данными.

Ещё одной выявленной проблемой стала поддержка атомарных операций. В ходе работы было установлено, что атомарные операции не реализованы на уровне аппаратного описания в SystemVerilog, на котором описано внутреннее устройство Vortex GPU. При этом симулятор simx, используемый для моделирования работы устройства, поддерживает атомарные операции программным образом. Однако компилятор не поддерживал атомарные операции для Vortex GPU, что не позволило исполнять их в simx. Для устранения этой проблемы была проведена модификация компилятора PoCL. Был добавлен код на ассемблере для RISC-V, содержащий в себе реализацию некоторых атомарных операций из стандарта OpencCL 1.2. Такой подход обеспечил возможность корректного выполнения ядер в simx, которые содержали в себе атомарные операции.

4. Выводы

Проведённая работа показала, что использование Vortex GPU для выполнения вычислений с разреженными структурами данных возможно, однако требует учёта ряда особенностей программно-аппаратной среды. Возникающие трудности связаны главным образом не с отсутствием поддержки OpenCL 1.2 или библиотек линейной алгебры, а с деталями реализации устройства. Анализ поведения нескольких вычислительных ядер, организация передачи данных внутри устройства и обеспечение поддержки атомарных операций демонстрируют необходимость доработки этого устройства для использования в более крупных задачах. Полученные результаты и предложенные решения позволяют запускать вычисления, требующие нескольких операций, на платформе Vortex GPU и могут быть использованы как основа для дальнейшего расширения функциональных возможностей Vortex GPU.

Литература

1. Sparse Linear Algebra/spla: An open-source generalized sparse linear algebra library with vendor-agnostic GPUs accelerated computations. URL: https://github.com/SparseLinearAlgebra/spla.
2. GraphBLAS. URL: https://github.com/GraphBLAS.
3. Vortex GPGPU. URL: https://github.com/vortexgpgpu/vortex.

ДОВЕРЕННЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОБНОВЛЕНИИ КЛИНИЧЕСКИХ РЕКОМЕНДАЦИЙ

А.А. Шестова[1], В.Е. Турлапов[1] , В.П. Носов[1,2]

[1]Нижегородский государственный университет им. Н.И. Лобачевского,
[2]Приволжский исследовательский медицинский университет

Ишемическая болезнь сердца (ИБС) остается одной из ведущих причин смертности в Российской Федерации. Эффективность лечения напрямую зависит от соблюдения клинических рекомендаций (КР), однако их внедрение на практике часто затруднено из-за дефицита знаний, вариабельности клинических решений и недостаточной актуализации самих рекомендаций. В работе предложена концепция экспертной системы на основе технологий доверенного искусственного интеллекта (ИИ), обеспечивающей исполнение и обновление КР без искажения их содержания. Представлены алгоритмы диагностики и лечения острого коронарного синдрома, реализующие клинические рекомендации в форме дерева решений с распределением ролей между пациентом, врачом и ИИ.

Ключевые слова: клинические рекомендации, ишемическая болезнь сердца, острый коронарный синдром, доверенный искусственный интеллект, экспертная система, дерево решений, RAG-подсистема.

1. Введение

Клинические рекомендации (КР) - вершина аналитической деятельности в медицине, максимально насыщенный информационный продукт, основанный на концепции доказательной медицины и предназначенный для использования во врачебной практике, а также в образовательном процессе.

Ишемическая болезнь сердца (ИБС) остается одной из ведущих причин смертности в мире и в Российской Федерации. Согласно данным Всемирной организации здравоохранения, сердечно-сосудистые заболевания обусловливают до трети всех случаев смерти взрослого населения [8]. В этих условиях внедрение и соблюдение клинических рекомендаций (КР) приобретает ключевое значение для снижения смертности и повышения качества медицинской помощи. В 2022г в РФ было зарегистрировано > 7,6 млн пациентов с диагнозом ИБС, что охватывает ~5,2% населения страны [1]. В структуре заболеваемости в РФ в 2022 г. удельный вес ИБС от всех болезней составил 3,0%, а от болезней системы кровообращения (БСК) - 20,1%. На хронические формы ИБС (ХИБС) приходится бóльшая часть патологии среди взрослого населения, в структуре заболеваемости всех ИБС в 2022 г. в РФ они составили 63,3%.

Современные клинические рекомендации представляют собой систематизированные документы, основанные на доказательной медицине, призванные стандартизировать диагностику, лечение и профилактику ИБС. Их использование позволяет повысить эффективность терапии, снизить вариабельность медицинских решений и уменьшить риск неблагоприятных исходов. Однако на практике степень их соблюдения остается недостаточной, что напрямую влияет на выживаемость пациентов и эффективность лечения. При этом лишь незначительная часть врачей регулярно обращается к специализированным медицинским журналам, что ведет к низкому уровню осведомленности о действующих клинических рекомендациях [5]. Так, например, в отчете GxP News за 2021 год данные подтверждают низкий интерес к печатным/традиционным журналам: основной источник - онлайн-порталы (50,3%), MedScape (34,4%). Опрос подчеркивает, что только половина респондентов обращает внимание на базы данных (например, PubMed, Scopus) [15]. Реброва О.Ю в 2019 г установлено [2], что врачи, осуществляющие диспансерное наблюдение, в своем большинстве не знают клинических рекомендаций по артериальной гипертонии даже в отношении правил и особенностей клинического измерения АД у взрослых. Средний балл 7,4 ± 5,5 (при максимальном 17), Медиана 5, Мода 3.

Недостаточное знание и применение КР наблюдается не только среди врачей, но и в клинической практике в целом. Так, исследование Ушакова А.А. и др. (2019) выявило, что лишь 9 % пациентов с сахарным диабетом получали лечение в полном соответствии с действующими рекомендациями [6]. Поскольку диабет является одним из ведущих факторов риска ИБС, несоблюдение стандартов лечения этой группы больных существенно повышает риск сердечно-сосудистых осложнений и летальных исходов.

Положительные примеры внедрения рекомендаций демонстрируют, что системные меры могут значительно улучшить ситуацию. Так, в проекте ALIGN (Марцевич С.Ю. и др., 2020) были реализованы комплексные шаги по оценке соответствия терапии стабильной ИБС современным стандартам, что позволило повысить частоту назначения доказанных препаратов и улучшить показатели контроля факторов риска [7]. Эти результаты подтверждают, что внедрение и строгая реализация КР способны оказывать прямое влияние на клинические исходы и продолжительность жизни пациентов.

Таким образом, клинические рекомендации при ИБС являются не только инструментом профессионального стандарта, но и важным фактором сохранения жизни. Их системное внедрение требует комплексных мер: повышения уровня образования врачей, мониторинга соблюдения стандартов и внедрения мотивационных механизмов. Без этого даже самые совершенные рекомендации не смогут реализовать свой потенциал в реальной практике.

Соблюдение клинических рекомендаций – это не формальность, а инструмент спасения жизней. Рост доказательности медицинской практики и внедрение национальных стандартов дают реальную возможность снизить смертность от ИБС, однако только при условии, что рекомендации станут не просто документом, а частью клинического мышления каждого специалиста.

В контексте обозначенных проблем особую актуальность приобретает разработка интеллектуальных систем поддержки принятия врачебных решений, основанных на клинических рекомендациях. Использование технологий доверенного искусственного интеллекта (ИИ) позволяет не только повысить точность и скорость диагностики, но и обеспечить строгое соблюдение алгоритмов лечения, исключая субъективные ошибки и вариативность интерпретаций. Согласно обзору более чем 50 исследований, использование искусственного интеллекта в радиологии (анализ КТ и МРТ) повышает точность диагностики рака легких на 20-30% по сравнению с традиционными методами, применяемыми врачами. Кроме того, ИИ сокращает время обработки медицинских изображений в 5-8 раз. Применение технологий доверенного ИИ, таких как модели с объяснимостью (например, SHAP), позволяет снизить частоту ложноположительных результатов на 15% [9].

Такая экспертная система способна выполнять функции контроля соответствия действий врача утвержденным рекомендациям, напоминания о последовательности этапов ведения пациента и автоматической интерпретации диагностических данных. При этом ключевым принципом остается сохранение решающей роли специалиста: ИИ выполняет вспомогательные, формализованные функции, а окончательное клиническое решение остается за врачом.

2. Диагностика

Важным требованием к алгоритму, реализующему клинические рекомендации, является обеспечение их исполнения без искажения содержания. Для этого предварительно выполнена содержательная формализация клинических рекомендаций по этапам «Диагностика» и «Лечение». На основе этой формализации были распределены роли, источники данных и интерфейсы между основными участниками процесса - пациентом, искусственным интеллектом (ИИ) и врачом. Это позволило построить алгоритм исполнения клинических рекомендаций, обеспечивающий логическую последовательность действий и контроль соблюдения стандартов.

Ключевая идея интеграции клинических рекомендаций в систему ИИ заключается в передаче формализованных функций моделям искусственного интеллекта. Алгоритм при этом выполняет управляющую роль, обеспечивая координацию всех этапов ведения пациента в рамках конкретной нозологии. На рисунке 1 представлена схема алгоритма для этапа “Диагностика” при остром коронарном синдроме (ОКС), демонстрирующая последовательность действий и переход к этапу дифференцированной терапии. В алгоритме выделены функции участников

процесса: пациента, ИИ, бригады скорой медицинской помощи и стационара (или другого лечебно-профилактического учреждения). Левый столбец отражает этапы диагностического процесса, а первые три строки блоков содержат возможные равноуровневые диагнозы, из которых на каждом этапе выбирается только одно направление дальнейших действий.

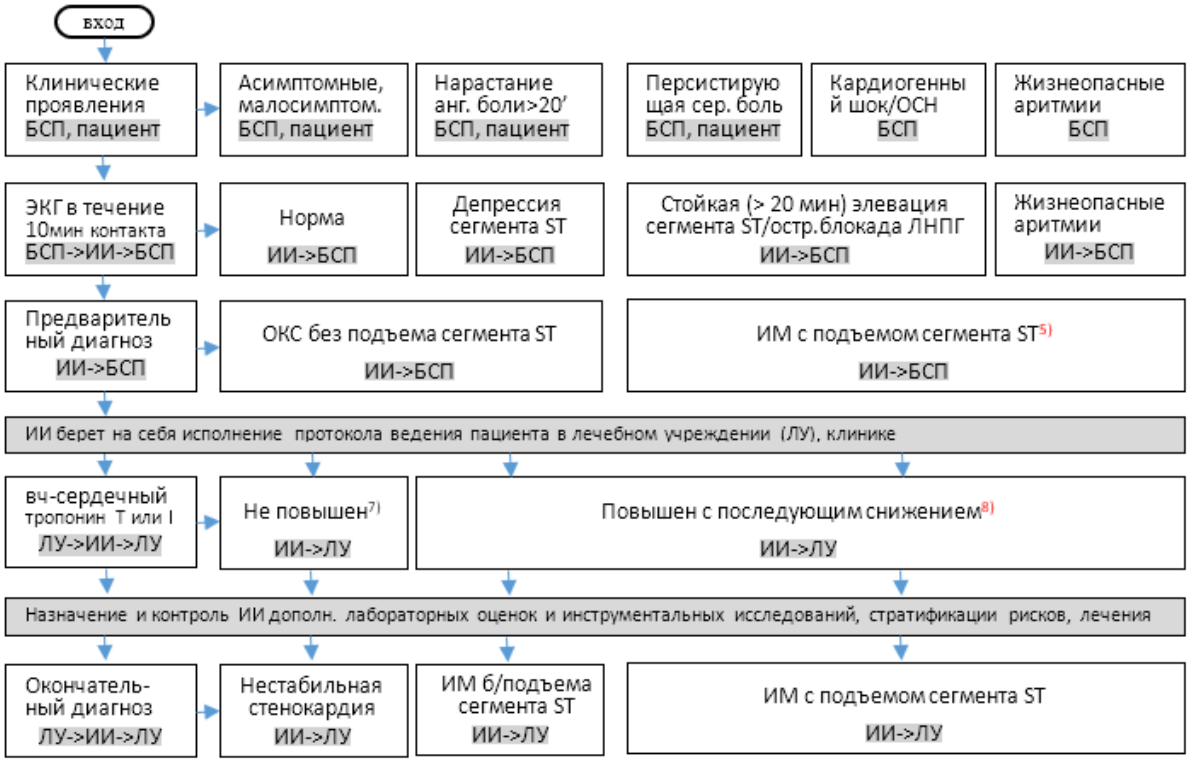


Рис. 1. Алгоритм дерева решений, реализующий клинические рекомендации для ОКС в части диагностики и перехода к дифференцированной терапии

Работоспособность представленной схемы обеспечивается инфраструктурными и программными компонентами серверного комплекса. К ним относятся хранилище данных, где размещены формализованные клинические рекомендации и сопутствующая справочная информация, а также RAG-подсистема (Retrieval-Augmented Generation), которая подгружает тексты клинических рекомендаций, медицинские и фармацевтические справочники, сопроводительные документы скорой помощи, электронные карты пациентов и обученные лингвистические модели (в том числе модели анализа медицинских заключений).

Каждый прямоугольный блок алгоритма отображает участников процесса и направления информационных потоков. При этом итоговое решение остается за медицинским специалистом. Роль ИИ обозначена на этапах предварительной диагностики и сопровождения пациента после госпитализации. Структура блоков алгоритма полностью соответствует структуре формализованных клинических рекомендаций, что обеспечивает их точное воспроизведение в цифровом формате. Отдельные переходы (сноски 5 и 8) инициируют начало лечения в полном соответствии с действующими клиническими рекомендациями по текущему диагнозу.

3. Лечение

На основе клинических рекомендаций возможно построение дерева решений, которое учитывает все параметры пациента: диагноз, сопутствующие заболевания, противопоказания, результаты лабораторных и инструментальных исследований. Такое дерево решений позволяет алгоритму экспертной системы автоматически формировать оптимальные терапевтические стратегии, предлагая врачу варианты лечения, строго соответствующие стандартам. При этом система может анализировать совместимость лекарственных средств, корректировать дозировки с учетом индивидуальных характеристик и сигнализировать о потенциальных рисках. В ре-

зультате обеспечивается персонализированное, но при этом стандартизированное лечение, где каждая рекомендация подкреплена доказательной базой и контролируется механизмами доверенного искусственного интеллекта.

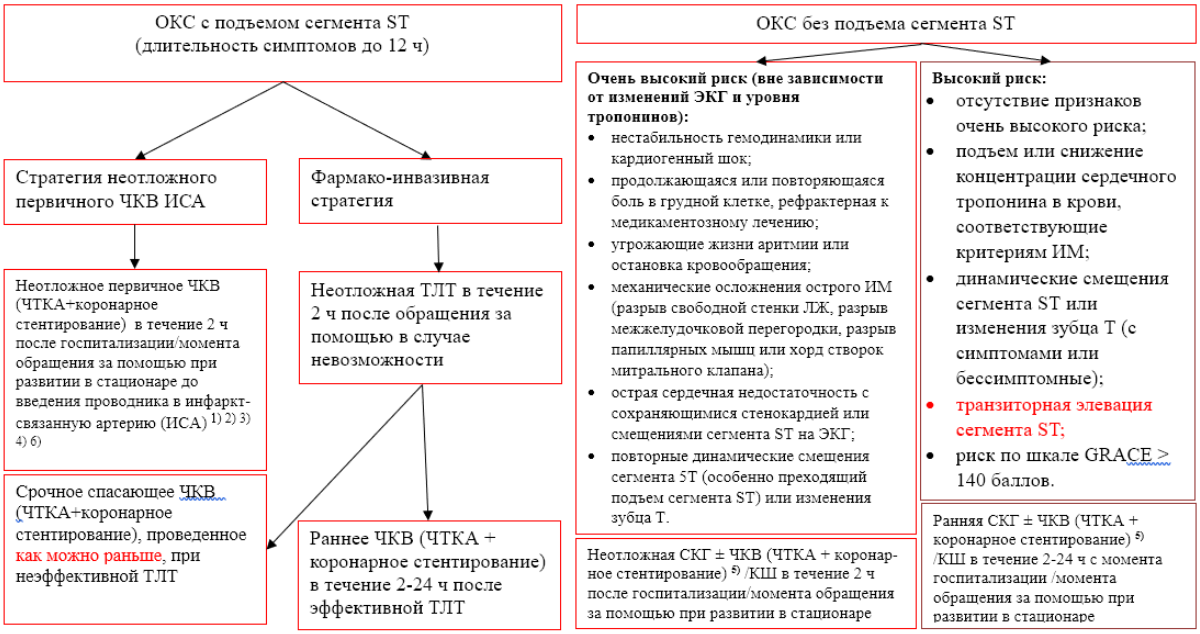


Рис. 2. Дерево решений о виде и сроках инвазивного вмешательства и тромболитической терапии

В частности, для больных с острым коронарным синдромом возможно формирование дерева решений о виде и сроках инвазивного вмешательства и тромболитической терапии (ТЛТ). Ветви такого дерева формируются на основе параметров состояния пациента – времени от начала симптомов, результатов ЭКГ, уровня тропонинов, наличия противопоказаний к тромболизису и технической доступности коронарного вмешательства. Алгоритм обеспечивает выбор оптимальной стратегии: первичное чрескожное коронарное вмешательство (ЧКВ), проведение ТЛТ с последующим ЧКВ или консервативное лечение. Встраивание этих решений в экспертную систему позволяет автоматизировать выбор терапевтической тактики в соответствии с клиническими рекомендациями и минимизировать риски задержки лечения, что особенно важно при жизнеугрожающих состояниях.

4. Адаптация КР

Для сохранения актуальности КР должны регулярно обновляться с учетом новых научных данных, клинических исследований и статистики реальных исходов лечения пациентов. Такой процесс позволяет не только поддерживать соответствие международным стандартам, но и адаптировать рекомендации к условиям национальной системы здравоохранения.

Сбор и анализ статистики об исходах лечения пациентов служит основой для оценки эффективности существующих КР. В реальной клинической практике значительная часть данных остается неструктурированной – в историях болезни, медицинских отчетах и регистрах. Их систематизация позволяет выявлять расхождения между предписанными и фактическими лечебными подходами, оценивать клинические и экономические результаты, а также обосновывать необходимость пересмотра отдельных положений рекомендаций. Таким образом, данные о результатах лечения становятся важнейшим источником информации для динамического обновления КР.

Механизм обновления КР включает в себя сбор данных из медицинских информационных систем, регистров и телемедицинских платформ, где фиксируются фактические показатели эффективности и безопасности медицинских вмешательств. На основе анализа этих данных формируются предложения для экспертных сообществ, которые пересматривают рекомендации в соответствии с выявленными тенденциями. Такой подход обеспечивает переход от статичной

модели к «живым» клиническим рекомендациям, которые отражают текущее состояние науки и практики.

Использование искусственного интеллекта (ИИ) открывает новые возможности для автоматизации процесса обновления КР. Системы на основе машинного обучения способны проводить интеллектуальный поиск и оценку новых публикаций, клинических испытаний и данных из электронных медицинских карт. Применение RAG-подсистем позволяет непрерывно пополнять базу знаний экспертной системы точной и актуальной информацией, извлекая её из надежных источников – международных баз данных, журналов и реестров. Это обеспечивает своевременное выявление новых доказательств, требующих пересмотра действующих рекомендаций.

Интеграция ИИ в процесс поддержки и обновления экспертной системы делает возможным автоматический контроль за соответствием клинических решений последним версиям КР. Для обеспечения высокого качества КР и их соответствия принципам доказательной медицины применяются стандартизированные инструменты оценки, такие как AGREE II и RIGHT. AGREE II – стандартный вопросник “Appraisal of Guidelines for Research and Evaluation II” (AGREE), состоящий из 23 вопросов в 6 аспектах (доменах) (процесс разработки КР; качество представления рекомендаций; полнота описания; клиническая валидность) [11, 12] или его сокращенной версии AGREE Global Rating Scale (AGREE GRS) [10]. RIGHT – контрольный список (Reporting Items for practice Guidelines in HealThcare), включающего 22 вопроса в 7 доменах (основная информация, введение, доказательства, рекомендации, рецензирование и обеспечение качества, финансирование, декларирование и управление интересами, другая информация) [13, 14]. Интеграция этих инструментов в системы доверенного искусственного интеллекта позволяет автоматизировать анализ КР, выявлять недостатки и предлагать корректировки, обеспечивая их актуальность и надежность для клинической практики. В результате экспертная система становится самоподдерживающейся структурой, способной не только обеспечивать выполнение клинических рекомендаций, но и адаптировать их на основе накопленного опыта и реальных клинических исходов, сохраняя при этом прозрачность, достоверность и доверие со стороны медицинского сообщества.

Литература

1. Woolf SH, Grol R, Hutchinson A, et al. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ. 1999;318(7182):527-530. https://doi.org/10.1136/bmj.318.7182.527.
2. Реброва О.Ю. О трех способах подготовки клинических рекомендаций. Инструменты ADAPTE и AGREE GRS. Проблемы Эндокринологии. 2019;65(3):197-203. https://doi.org/10.14341/probl10100.
3. Котова Е. Г., Кобякова О.С., Стародубов В.И. и др. Заболеваемость всего населения России в 2022 году: статистические материалы. М.: ФГБУ "ЦНИИОИЗ" Минздрава России, 2023. 146 с. ISBN: 978-5-94116-072-3.
4. А.Г. Ластовецкий, В.В. Кочубей, Я.П. Сандаков, М.А. Иванова. Проблемы стандартизации в здравоохранении. – 2019. – № 9-10. – С. 18-22. – DOI 10.26347/1607-2502201909-10018-022. – EDN ZLYGMF.
5. Герасимов, А.А. Оценка отношения врачей кардиологов и терапевтов к клиническим рекомендациям по профилактике и лечению инфаркта миокарда / А.А. Герасимов, Р.В. Полибин // Профилактическая медицина. – 2019. – Т. 22, № 5. – С. 37-44. – DOI 10.17116/profmed20192205137. – EDN LHZKID.
6. Ушаков, А.А. Соответствие медицинской помощи клиническим рекомендациям по данным анкетирования пациентов / А. А. Ушаков, А. В. Войтенко, Ю. О. Назаренко // 73-я Итоговая научная конференция студентов Ростовского государственного медицинского университета : Сборник материалов, Ростов-на-Дону, 12 апреля 2019 года. – Ростов-на-Дону: Ростовский государственный медицинский университет, 2019. – С. 132-133. – EDN IUXYPM.
7. Попытка привести качество терапии больных со стабильно протекающей ишемической болезнью сердца в соответствие с современными клиническими рекомендациями (исследование ALIGN): дизайн и первые результаты / С.Ю. Марцевич, Е.Д. Жаркова, Н.П. Кутишенко

[и др.] // Рациональная фармакотерапия в кардиологии. – 2020. – Т. 16, № 1. – С. 75-81. – DOI 10.20996/1819-6446-2020-02-08. – EDN DKXZWC.
8. Fact Sheet "Cardiovascular diseases (CVDs)" от Всемирной организации здравоохранения (ВОЗ). url: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
9. Soleimani A, Soltani P, Karimi H, Mirzaei M, Esfahanian F, Yavari M, Esfahani MP. The effect of moderate-intensity aerobic exercise on non-proliferative diabetic retinopathy in type II diabetes mellitus patients: A clinical trial. Microvasc Res. 2023 Sep;149:104556. doi: 10.1016/j.mvr.2023.104556. Epub 2023 Jun 2. PMID: 37269942.
10. Brouwers MC, Kho ME, Browman GP, et al. The global rating scale complements the AGREE II in advancing the quality of practice guidelines. J Clin Epidemiol. 2012;65(5):526-534. https://doi.org/10.1016/j.jclinepi.2011.10.008.
11. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in healthcare. CMAJ. 2010;182(18):E839-842. https://doi.org/10.1503/cmaj.090449.
12. Опросник по экспертизе и аттестации руководств. The AGREE Collaboration. September 2001. – 26 с. [Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument. The AGREE Collaboration. September 2001. 26 p. (In Russ.)]. https://www.agreetrust.org/wp-content/uploads/2013/06/AGREE_Instrument_Russian.pdf.
13. RIGHT Statement. [cited 2025, Aug 30].. http://www.right-statement.org/.
14. Chen Y., Yang K., Marušié A. et al.; for the RIGHT (Reporting Items for Practice Guidelines in Healthcare) Working Group. A Reporting Tool for Practice Guidelines in Health Care: The RIGHT Statement. Ann Intern Med. 2017;166:128–132. https://doi.org/10.7326/M16-1565
15. Результаты опроса: какие медицинские ресурсы предпочитают врачи // Медвестник. – 2021. – 17 ноября. – URL: https://www.medvestnik.ru/content/news/Rezultaty-oprosa-kakie-medicinskie-resursy-predpochitayut-vrachi.html (дата обращения: 17.10.2025).

КОГЕРЕНТНО-НЕКОГЕРЕНТНО-ГРАДИЕНТНЫЙ РЕЖИМ В СИСТЕМЕ ЧАСТИЦ С ЗАДАННЫМ ДВИЖЕНИЕМ И ВНУТРЕННЕЙ СТЕПЕНЬЮ СВОБОДЫ[1*]

П.А. Щербаков, Л.А. Смирнов, М.И. Болотов, В.А. Костин, Г.В. Осипов

Нижегородский государственный университет им. Н.И. Лобачевского

Рассматривается система нелокально связанных фазовых осцилляторов на кольце с периодическими граничными условиями. В отличие от известных динамических режимов, формирующихся для движущихся осцилляторов с равномерным распределением положений во всей среде, мы локализуем движущиеся элементы среди стационарных и получаем динамический режим, состоящий из трёх пространственных областей: когерентная, некогерентная и градиентная. При помощи кросс-корреляционного анализа мы демонстрируем, что некогерентный участок возникает вследствие разброса амплитуд движущихся элементов. Также, мы аппроксимируем локальное поле в термодинамическом пределе в предположении малых амплитуд движения и показываем, что формирование градиентных областей является эффектом асимметрии эффективного ядра нелокальной связи.

Ключевые слова: химерный режим, связанные осцилляторы, нелокальная связь.

1. Введение

С момента своего первого описания [1] химерные режимы продолжают оставаться в фокусе активных междисциплинарных исследований [2–4]. Классически, химерный режим рассматривается для системы нелокально связанных фазовых осцилляторов на эквидистантной пространственной сетке, где уравнения Курамото-Баттогтоха демонстрируют бистабильность между полностью синхронным и химерным состоянием.

При этом, случай неэквидистантной или эволюционирующей во времени пространственной сетки исследован значительно слабее. В работе [5] был проведён анализ системы со случайным движением элементов; с помощью полуаналитических методов было показано, что химерное состояние устойчиво возникает как при диффузионном, так и при баллистическом движении. При статическом случайном распределении элементов вероятность его возникновения также существенно выше, чем у полностью синхронного режима.

В то время как в исследовании [6] рассматривалось неквазистатическое частичное движение, закономерно приводящее к разделению системы на химеры на быстро движущихся и неподвижных частицах, в настоящей работе исследуется противоположный сценарий – адиабатическое движение локализованного подансмабля. Мы рассматриваем систему частиц, обладающих внутренней степенью свободы, описываемой фазой, причём фазовые переменные не влияют на пространственное положение частиц.

2. Модель

Рассмотрим систему из N осцилляторов, расположенных на кольце длины L с периодическими граничными условиями. Позиции осцилляторов характеризуются координатами $x_n \in [0; L)(n = 1,2, \dots, N)$. Состояние каждого осциллятора описывается фазой $\varphi_n \in [-\pi; \pi)$. Фазы удовлетворяют классическим уравнениям Курамото и Баттогтоха [1]:

$$\dot{\varphi}_n = \frac{1}{N}\sum_{j=1}^{N} G\left(x_j(t) - x_n(t)\right) \cdot \sin(\varphi_j - \varphi_n - \alpha)\,,$$

[1*] Работа поддержана Министерством науки и высшего образования Российской Федерации, контракт # FSWR-2020-0036.

где α – фазовый сдвиг. В качестве ядра нелокальной связи мы используем вариант из работы [5]:

$$G(x) = \kappa\ \cosh(\kappa(|x| - L/2))/2\sinh(\kappa L/2)\,,$$

где κ – коэффициент нелокального взаимодействия.

Координаты частиц осциллируют на кольце:

$$x_n(t) = x_n^{(0)} + a_n \sin(\Omega_n t + \Psi_n).$$

Для простоты мы рассматриваем случай $\Omega_n = \Omega = const, \Psi_n \in \{0, \pi\}$. Обозначим $A_n = a_n \cos\Psi_n$. Начальные положения частиц эквидистантны: $x_n^{(0)} = nL/N$. Выделим в кольце симметричный подотрезок вокруг точки $x = L/2$: пусть частицы движутся только если их начальное положение находится в подотрезке:

$$A_n = \begin{cases} b_n, |n - N/2| \le \sigma N; \\ 0, |n - N/2| > \sigma N. \end{cases}$$

Здесь b_n – реализация случайной величины с равномерным распределением на $[c - a; c + a]$. Выборка случайных величин генерируется в начальный момент времени и дальше фиксируется. Параметр σ определяет отношение между движущимися и покоящимися частицами.

Фазовые уравнения интегрируются численно методом Рунге-Кутты 4 порядка. Рассматриваются вспомогательные метрики: модуль глобального параметра порядка:

$$|R(t)| = |\frac{1}{N}\sum_{j=1}^{N} e^{i\varphi_j(t)}|$$

и модуль локального поля:

$$|H_n(t)| = \left|\frac{1}{N}\sum_{j=1}^{N} G\left(x_j(t) - x_n(t)\right) \cdot e^{i\varphi_j(t)}\right|.$$

Начальные фазы $\varphi_n(0)$ имеют равномерное распределение на $[-\pi; \pi)$. В численных симуляциях используются следующие значения параметров: сила связи $\kappa = 4.0$, длина среды $L = 1.0$, фазовый сдвиг $\alpha = 1.457$. Также рассматривается адиабатическое значение частоты $\Omega = 0.01$.

3. Результаты

При значениях параметров $a = c = 0.06, \sigma = 0.1, N = 512$ был получен динамический режим смешанного типа (Рис. 1), одновременно имеющий черты химерного режима и неравномерно градиентного состояния. Режим имеет три пространственные области: когерентную, некогерентную и градиентную, причём с течением времени когерентная и градиентная области периодически меняются местами.

В докладе будет представлено полуаналитическое обоснование формирования некогерентного участка путём подсчёта локальных корреляций фазового профиля движущихся частиц для больших N, а также аналитическое подтверждение образования градиентных областей путём анализа эффективного ядра нелокального взаимодействия в термодинамическом пределе в предположении малых амплитуд.

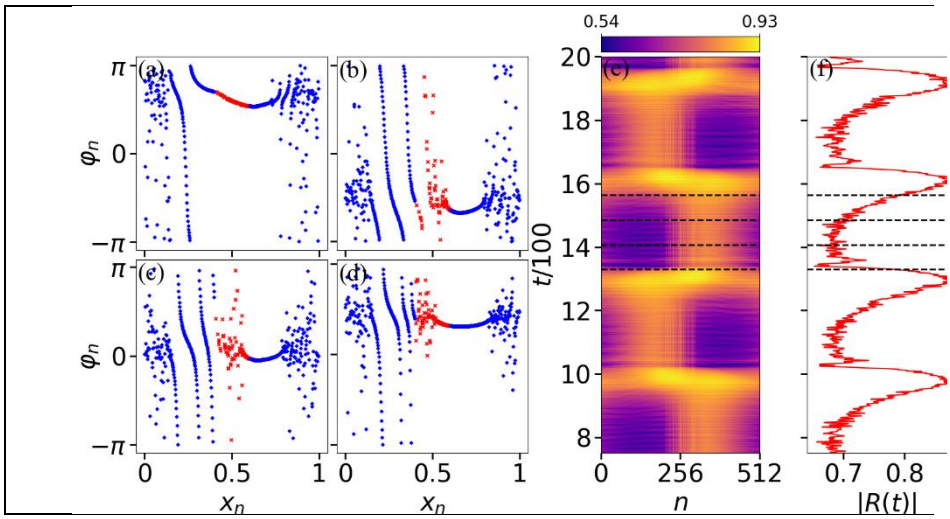


Рис. 1. Когерентно-некогерентно-градиентный режим. (a-d) мгновенные фазовые снимки, красным цветом показаны фазы движущихся частиц. (e) модуль локального поля $|H_n(t)|$. (f) модуль глобального параметра порядка $|R(t)|$

Литература

1. Kuramoto Y., Battogtokh D. Coexistence of coherence and incoherence in nonlocally coupled phase oscillators // Nonlinear Phenomena in Complex Systems. 2002. Vol. 5, No. 4. P. 380–385. DOI: 10.1007/BF02706196.
2. Omel'chenko O.E. Traveling chimera states // Journal of Physics A: Mathematical and Theoretical. 2019. Vol. 52, No. 10. P. 104001. DOI: 10.1088/1751-8121/ab0043.
3. Yao N., Zheng Z. Chimera states in spatiotemporal systems: Theory and applications // International Journal of Modern Physics B. 2016. Vol. 30, No. 7. P. 1630002. DOI: 10.1142/S0217979216300021.
4. Wolfrum M., Omel'chenko O.E. Chimera states are chaotic transients // Physical Review E. 2011. Vol. 84, No. 1. P. 015201. DOI: 10.1103/PhysRevE.84.015201.
5. Smirnov L.A., Bolotov M.I., Osipov G.V., Pikovsky A. Disorder fosters chimera in an array of motile particles // Physical Review E. 2021. Vol. 104, No. 3. P. 034205. DOI: 10.1103/PhysRevE.104.034205.
6. Bolotov M.I., Smirnov L.A., Osipov G.V. Chimera states in a system of stationary and flying-through deterministic particles with an internal degree of freedom // Chaos: An Interdisciplinary Journal of Nonlinear Science. 2025. Vol. 35, No. 1. DOI: 10.1063/5.0235078.

СОДЕРЖАНИЕ

E.A. Kubrakova
The impact of news sentiment on the quality of predictive analytics using a Long-Short Term Memory (LSTM) model 4

Н.Ф. Алексиадис
О задаче распознавания *A*-функции в функциональной системе рациональных функций с рациональными коэффицентами 10

А.А. Бердюгин, В.Е. Турлапов
Гибридный метод временного суперразрешения с нейросетевой постобработкой кадров в реальном времени 15

М.И. Болотов, Л.А. Смирнов, В.О. Муняев, Г.В. Осипов, И.В. Белых
Циклопные состояния в сетях неидентичных осцилляторов со связью по высшим модам 20

А.А. Воденеева, В.Д. Волокитин, М.А. Загрядсков, К.И. Ковалев, Е.А. Козинов, А.И. Кулик, Д.Д. Литвяков, А.Ю. Пирова, А.В. Устинов, И.Б. Мееров
Оптимизация алгоритмов умножения разреженной матрицы на вектор для процессоров архитектуры RISC-V 22

В.Д. Волокитин
Моделирование производительности метода квантовых траекторий для оптимизации параметров алгоритма 27

Д. С. Жариков, М.В. Пискунов, Д.С. Коломенский
Метод комплексной валидации численной модели соударения капли с поверхностью с использованием комбинации динамического и кинематического параметров движения 31

А.С. Зайцев, К.А. Баркалов
Об алгоритме поиска глобального минимума для одного класса многомерных многоэкстремальных функций 36

А.О. Захаров, Ю.В. Захарова
Алгоритмическая сложность задач составления расписаний в системах с учетом статической и динамической составляющих расхода энергии 42

В.И. Звонилов
Камеры и стенки в пространстве вещественных алгебраических кривых бистепени (4,3) на гиперболоиде 44

Д.К. Кармаев, Л.Г. Афраймович
Решение задачи о назначениях фиксированной размерности при помощи нейронных сетей 49

Е.В. Кондрашова
Принятие решений при оптимизации управления образовательным процессом с использованием моделей теории игр 52

С.Ю. Коннов, Е.А. Козинов
Об опыте применения алгоритма многокритериальной оптимизации для настройки гиперпараметров моделей машинного обучения 54

Р.А. Кононов, О.В. Масленников, В.И. Некоркин
Итеративное уточнение признаков в сверточных нейронных сетях с помощью рекуррентных блоков 59

М.И. Кузнецов, А.В. Кондратьева
Фундаментальные представления ортогональной алгебры Ли и новые простые подалгебры неальтернирующих гамильтоновых алгебр Ли 64

М.И. Кузнецов, М.М. Рабиа
Вычисление локальных деформаций расширений простых алгебр Ли типа А с помощью внешнего дифференцирования над полем характеристики 2 66

С.А. Лавренков, А.В. Васюков
Моделирование тонких композитных пластин под вибрационной нагрузкой с уточнением параметров материала 68

О.В. Любимцев, А.А. Туганбаев
Расширения центрально существенных колец 72

А.А. Маркелов, А.С. Дмитричев, В.И. Некоркин
Бифуркации предельных циклов и торов в системе взаимосвязанных осцилляторов Баутина 75

И.Б. Мееров, С.А. Борисов, В.Д. Волокитин, А.В. Горшков, М.А. Козлов, Е.А. Панова, А.Ю. Пирова, А.В. Линев
Об опыте разработки переносимых программ для научных вычислений 79

Моршинин А.В.
Гибридный эволюционный алгоритм для задачи 2-CORRELATION CLUSTERING 83

Д.А. Никитенко
О трех опорных задачах суперкомпьютерного кодизайна 87

А.Г. Никольская, Д.А. Рыболовлев
Особенности состязательных атак с ограничениями против систем обнаружения вторжений, основанных на применении методов машинного обучения 95

В.П. Носов, А.А. Шестова, В.Е. Турлапов
Интеграция доверенного искусственного интеллекта в инфраструктуру здравоохранения 100

Е.А. Панкратова, И.Т. Бубукин, И.В. Ракуть, М.И. Агафонов, А.А. Яблоков, В.И. Бубукин, А.Л. Панкратов
Обработка и анализ данных поглощения атмосферы на полигоне «Кара-Даг» 108

Е.В. Пройдакова, А.С. Бабанова
Построение прогностической модели выбора специализации студентами бакалавриата института ИТММ на примере реальных данных 111

Е.В. Пыряев, К.А. Беклемышева
Численное моделирование биофабрикации разветвленных кровеносных сосудов акустической левитации 116

Ю.А. Родимков
Использование циклического буфера для повышения точности восстановления искажений волнового фронта в оптических системах 119

С.В. Русаков, В.А. Старостин, И.И. Хейфец
Имитационная модель горизонтально расположенного цилиндрического резервуара для оценки погрешности градуировочной процедуры объёмного метода 122

В.П. Савельев, Н.И. Сутягина
Зоны иммунитета малого бизнеса в условиях внешних воздействий 127

А. Сарвари
Некоторые эндоморфизмы абелевых групп 132

М.Ю. Сахно
Алгоритм списочного типа для размещения виртуальных машин на сервера с учетом NUMA-архитектуры 135

С.В. Сидоров, Т.Г. Смирнова
О количестве базисов в классе булевых функций..... 137

С.В. Сидоров, Г.В. Уткин
Классы подобия целочисленных нильпотентных матриц 4-го порядка с минимальным многочленом степени не выше 3 141

И.Е. Смирнов, А.В. Васюков
Численное моделирование воздействия динамической нагрузки на конструкции из композиционных материалов..... 143

Л.А. Смирнов, В.О. Муняев, М.И. Болотов, И.В. Белых
Редукция высшего порядка модели Курамото с задержкой по времени 146

И.С. Созонов, Т.С. Бородина, А.А. Акобян, Л.А. Лейфер
Применение нейронных сетей для анализа локальных рынков недвижимости..... 148

А.А. Сорокин, С.М. Титов, А.Ю. Нестеров, А.А. Оболенский
Коллективная разработка библиотеки для вывода результатов сверточных нейронных сетей..... 152

А.В. Степанов
Построение обобщенного H∞ управления для системы Матье на конечном интервале по априорным и экспериментальным данным 158

А.А. Тимофеев, М.А. Кривов
Графический ускоритель GeForce 3090 и метод Монте-Карло как универсальное решение одной классической обратной задачи 167

Е.Б. Титова, И.Н. Алпутов, О.А. Петров, Р.Д. Махнёв
Структура левого нуль-пространства матрицы ограничений четырёхиндексной планарной транспортной задачи..... 173

М.А. Толстых
Прямой экстремальный подход для идентификации параметров в виде вектора функций в диффузионной модели..... 176

М.М. Хамков, Л.А. Смирнов, М.И. Болотов
Ротобризерные циклопные состояния в ансамбле фазовых осцилляторов со связью по высшим модам..... 178

А.И. Хоменко, В.Е. Турлапов
Сегментация сердца методами глубокого обучения: обзор публикаций..... 180

И.С. Чистов, Л.М. Цыбуля
О решении линейных диофантовых уравнений в рамках подхода групповых действий 189

Д.П. Чернышов, А.М. Сатанин
Динамика и хаос в двумерных симплектических отображениях 195

М.А. Шадт
Математическое моделирование плоских струйных течений идеальной жидкости 200

Ш.Р. Шаехмитов
Исследование поведения RISC-V GPGPU Vortex при запуске библиотеки разреженной линейной алгебры 207

А.А. Шестова, В.Е. Турлапов , В.П. Носов
Доверенный искусственный интеллект в обновлении клинических рекомендаций 209

П.А. Щербаков, Л.А. Смирнов, М.И. Болотов, В.А. Костин, Г.В. Осипов
Когерентно-некогерентно-градиентный режим в системе частиц с заданным движением и внутренней степенью свободы..... 215

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И СУПЕРКОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ

Труды XXV Международной конференции

Нижний Новгород, 17–19 ноября 2025 г.

Под ред. проф. Д.В. Баландина

Отв. за выпуск К.А. Баркалов

Публикуется в авторской редакции

Издательство Нижегородского госуниверситета им. Н.И. Лобачевского.
603950, Н. Новгород, пр. Гагарина, 23.