

Министерство образования и науки Российской Федерации
Ярославский государственный университет им. П. Г. Демидова

МОДЕЛИРОВАНИЕ И АНАЛИЗ ИНФОРМАЦИОННЫХ СИСТЕМ

Том 23 № 4(64) 2016

Основан в 1999 году
Выходит 6 раз в год

Главный редактор

В.А. Соколов,

доктор физико-математических наук, профессор, Россия

Редакционная коллегия

С.М. Абрамов, д-р физ.-мат. наук, чл.-корр. РАН, Россия; **В.С. Афраймович**, проф.-исследователь, Мексика; **О.Л. Бандман**, д-р техн. наук, Россия; **В.Н. Белых**, д-р физ.-мат. наук, проф., Россия; **В.А. Бондаренко**, д-р физ.-мат. наук, проф., Россия; **С.Д. Глызин**, д-р физ.-мат. наук, проф., Россия (зам. гл. ред.); **А. Дехтярь**, проф., США; **М.Г. Дмитриев**, д-р физ.-мат. наук, проф., Россия; **В.Л. Дольников**, д-р физ.-мат. наук, проф., Россия; **В.Г. Дурнев**, д-р физ.-мат. наук, проф., Россия; **В.А. Захаров**, д-р физ.-мат. наук, проф., Россия; **Л.С. Казарин**, д-р физ.-мат. наук, проф., Россия; **Ю.Г. Карпов**, д-р техн. наук, проф., Россия; **С.А. Кащенко**, д-р физ.-мат. наук, проф., Россия; **А.Ю. Колесов**, д-р физ.-мат. наук, проф., Россия; **Н.А. Кудряшов**, д-р физ.-мат. наук, проф., Заслуженный деятель науки РФ, Россия; **О. Кушнаренок**, проф., Франция; **И.А. Ломазова**, д-р физ.-мат. наук, проф., Россия; **Г.Г. Малинецкий**, д-р физ.-мат. наук, проф., Россия; **В.Э. Малышкин**, д-р техн. наук, проф., Россия; **А.В. Михайлов**, д-р физ.-мат. наук, проф., Великобритания; **В.А. Непомнящий**, канд. физ.-мат. наук, Россия; **Н.Х. Розов**, д-р физ.-мат. наук, проф., чл.-корр. РАО, Россия; **Н. Сидорова**, д-р наук, Нидерланды; **Р.Л. Смелянский**, д-р физ.-мат. наук, проф., член-корр. РАН, академик РАЕН, Россия; **Е.А. Тимофеев**, д-р физ.-мат. наук, проф., Россия (зам. гл. ред.); **М.Б. Трахтенброт**, д-р комп. наук, Израиль; **Д.В. Тураев**, проф., Великобритания; **Х. Файт**, д-р наук, проф., Австрия; **Ф. Шнеблен**, проф., Франция

Ответственный секретарь **Е. В. Кузьмин**, д-р физ.-мат. наук, проф., Россия

Адрес редакции: ЯрГУ, ул. Советская, 14, г. Ярославль, 150000, Россия
Website: <http://mais-journal.ru>, e-mail: mais@uniyar.ac.ru; телефон (4852) 79-77-73

Научные статьи в журнал принимаются по электронной почте. Статьи должны содержать УДК, аннотации на русском и английском языках и сопровождаться набором текста в редакторе LaTeX . Плата с аспирантов за публикацию рукописей не взимается.

СОДЕРЖАНИЕ

Моделирование и анализ информационных систем. Т. 23, №4. 2016

Сравнение алгоритмов удвоения размера изображения <i>Ваганов С. Е., Хашии С. И.</i>	389
Об оптимизации и распараллеливании алгоритма Литтла для решения задачи коммивояжера <i>Васильчиков В. В.</i>	401
Облачный сервис для интерактивного моделирования межтерриториальной торговли <i>Величко А. С., Грибова В. В., Федорищев Л. А.</i>	412
Компьютерное моделирование "Интеллектуального здания" <i>Марьясин О. Ю., Колоджина А. С., Огарков А. А.</i>	427
Об алгебраических циклах на расслоенных произведениях неизотривиальных семейств регулярных поверхностей с геометрическим родом 1 <i>Никольская О. В.</i>	440
Сетевая модель для задачи целочисленного сбалансирования четырехмерной матрицы <i>Смирнов А. В.</i>	466
Резервное копирование с хранением в базе данных <i>Таранин С. М.</i>	479
Вероятностный анализ систем организации турниров <i>Цирлин А. М., Ахременков Ал. А.</i>	492

Свидетельство о регистрации СМИ ПИ №ФС77-66186 от 20.06.2016 выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций. Учредитель – Федеральное государственное бюджетное образовательное учреждение высшего образования "Ярославский государственный университет им. П. Г. Демидова". Подписной индекс – 31907 в Объединенном каталоге "Пресса России". Редактор, корректор А.А. Аладьева. Редактор перевода Э.И. Соколова. Подписано в печать 22.08.2016. Дата выхода в свет 31.08.2016. Формат 60x84¹/₈. Усл. печ. л. 14,41. Уч.-изд. л. 12,0. Объем 123 с. Тираж 50 экз. Свободная цена. Заказ 051/016. Адрес типографии: ул. Советская, 14, оф. 109, г. Ярославль, 150003 Россия. Адрес издателя: Ярославский государственный университет им. П. Г. Демидова, ул. Советская, 14, г. Ярославль, 150003 Россия.

ISSN 1818–1015 (Print)
ISSN 2313–5417 (Online)

P.G. Demidov Yaroslavl State University

MODELING AND ANALYSIS OF INFORMATION SYSTEMS

Volume 23 No 4(64) 2016

Founded in 1999
6 issues per year

Editor-in-Chief

V. A. Sokolov,

Doctor of Sciences in Mathematics, Professor, Russia

Editorial Board

S.M. Abramov, Prof., Dr. Sci., Corr. Member of RAS, Russia; **V. Afraimovich**, Prof.-researcher, Mexico; **O.L. Bandman**, Prof., Dr. Sci., Russia; **V.N. Belykh**, Prof., Dr. Sci., Russia; **V.A. Bondarenko**, Prof., Dr. Sci., Russia; **S.D. Glyzin**, Prof., Dr. Sci., Russia (*Deputy Editor-in-Chief*); **A. Dekhtyar**, Prof., USA; **M.G. Dmitriev**, Prof., Dr. Sci., Russia; **V.L. Dol'nikov**, Prof., Dr. Sci., Russia; **V.G. Durnev**, Prof., Dr. Sci., Russia; **L.S. Kazarin**, Prof., Dr. Sci., Russia; **Yu.G. Karpov**, Prof., Dr. Sci., Russia; **S.A. Kashchenko**, Prof., Dr. Sci., Russia; **A.Yu. Kolesov**, Prof., Dr. Sci., Russia; **N.A. Kudryashov**, Dr. Sci., Prof., Russia; **O. Kouchnarenko**, Prof., France; **I.A. Lomazova**, Prof., Dr. Sci., Russia; **G.G. Malinetsky**, Prof., Dr. Sci., Russia; **V.E. Malyshkin**, Prof., Dr. Sci., Russia; **A.V. Mikhailov**, Prof., Dr. Sci., Great Britain; **V.A. Nepomniaschy**, PhD, Russia; **N.H. Rozov**, Prof., Dr. Sci., Corr. Member of RAE, Russia; **Ph. Schnoebelen**, Senior Researcher, France; **N. Sidorova**, Dr., Assistant Prof., Netherlands; **R.L. Smeliansky**, Prof., Dr. Sci., Corr. Member of RAS, Russia; **E.A. Timofeev**, Prof., Dr. Sci., Russia (*Deputy Editor-in-Chief*); **M. Trakhtenbrot**, Dr., Israel; **D. Turaev**, Prof., Great Britain; **H. Veith**, Prof., Dr. Sci., Austria; **V.A. Zakharov**, Prof., Dr. Sci., Russia

Responsible Secretary **E. V. Kuzmin**, Prof., Dr. Sci., Russia

Editorial Office Address: P.G. Demidov Yaroslavl State University,
Sovetskaya str., 14, Yaroslavl, 150000, Russia
Website: <http://mais-journal.ru>, e-mail: mais@uniyar.ac.ru

© P.G. Demidov Yaroslavl State University, 2016

Contents

Modeling and Analysis of Information Systems. Vol. 23, No 4. 2016

Comparison of Doubling the Size of Image Algorithms <i>Vaganov S. E., Khashin S. I.</i>	389
On the Optimization and Parallelizing Little Algorithm for Solving the Traveling Salesman Problem <i>Vasilchikov V. V.</i>	401
Cloud Service for Interactive Simulation of Interregional Trade <i>Velichko A. S., Gribova V. V., Fedorishchev L. A.</i>	412
Computer Modelling "Smart Building" <i>Maryasin O. Yu., Kolodkina A. S., Ogarkov A. A.</i>	427
On Algebraic Cycles on Fibre Products of Non-isotrivial Families of Regular Surfaces with Geometric Genus 1 <i>Nikol'skaya O. V.</i>	440
Network Model for The Problem of Integer Balancing of a Four-dimensional Matrix <i>Smirnov A. V.</i>	466
Backup with Storage in a Database <i>Taranin S. M.</i>	479
Probabilistic Analysis of Tournament Organization Systems <i>Tsirlin A. M., Akhremenkov Al. A.</i>	492

©Ваганов С. Е., Хашин С.И., 2016

DOI: 10.18255/1818-1015-2016-4-389-400

УДК 519.67

Сравнение алгоритмов удвоения размера изображения

Ваганов С. Е., Хашин С.И.

получена 18 апреля 2016

Аннотация. В данной работе произведен сравнительный анализ качества некоторых интерполяционных неадаптивных методов увеличения размера изображения в два раза. В качестве оценки точности (качества) аппроксимации использовалась величина среднеквадратичного отклонения. Артефакты (алиасинг, эффект Гиббса, размытие и др.), вносимые интерполяционными методами, не рассматривались. Приведено описание интерполяционных алгоритмов удвоения, таких как: метод ближайшего соседа, линейная и кубическая интерполяции, интерполяция сверткой с ядром Ланцоша (при $a=1, 2, 3$), а также метод 17-точечной интерполяции. Для каждого метода удвоения были найдены оптимальные коэффициенты ядер свертки для различных алгоритмов уменьшения размера вдвое. Рассмотрены различные методы уменьшения размера вдвое: усреднение по 4-м ближайшим точкам и взвешенное 16-ти ближайших точек с оптимальными коэффициентами. Оптимальные коэффициенты найдены для каждого рассматриваемого в работе метода удвоения, они подбирались таким образом, чтобы минимизировать величину среднеквадратичного отклонения точного значения от приближения.

В работе предлагается простой метод дополнительной корректировки произвольного алгоритма удвоения размера. Этот метод показывает хорошие результаты на простых интерполяционных алгоритмах. Однако при использовании наиболее сложных алгоритмов (17-точечный, Ланцош $a = 3$) улучшения оказываются незначительными. По результатам численных экспериментов самым точным среди рассмотренных алгоритмов является метод 17-точечной интерполяции, немного хуже – интерполяция посредством свертки с ядром Ланцоша с параметром $a = 3$ (см. таблицу в конце работы).

Ключевые слова: интерполяция, свертка функций, фильтр Ланцоша, 17-точечная интерполяция

Для цитирования: Ваганов С. Е., Хашин С.И., "Сравнение алгоритмов удвоения размера изображения", *Моделирование и анализ информационных систем*, **23:4** (2016), 389–400.

Об авторах:

Ваганов Сергей Евгеньевич, orcid.org/0000-0002-8642-8555, аспирант, Ивановский государственный университет, ул. Ермака, 39, г. Иваново, 153025 Россия, e-mail: pro100-pioner@mail.ru

Хашин Сергей Иванович, orcid.org/0000-0002-6508-663X, канд. физ.-мат. наук, доцент, Ивановский государственный университет, ул. Ермака, 39, г. Иваново, 153025 Россия, e-mail: khash2@mail.ru

Благодарности:

Работа выполнена при финансовой поддержке проекта 162 госзадания на НИР №2014/40

Введение

Масштабирование изображений является одной из классических задач компьютерной графики [2, 5, 10]. Частным случаем данной задачи является уменьшение и увеличение размера изображения в два раза. Особую роль данные методы играют при сжатии растровых изображений [1, 7, 8, 10], где от качества аппроксимации зависит общая эффективность кодирования.

В работе произведен сравнительный анализ погрешности некоторых интерполяционных методов увеличения размера изображения вдвое. В качестве оценки точности алгоритма использовалось среднеквадратичное отклонение. Бралась во внимание только неадаптивные методы, вносимые артефакты (алиасинг, размытие и др.) не рассматривались. Предложен простой способ «коррекции» (п. 1.5.), дающий улучшение аппроксимации на простых алгоритмах.

Пусть $a[x, y]$ – матрица яркости серого изображения либо одна из цветовых компонент полноцветного, $m_x \times m_y$ – её размеры. Для удобства будем полагать, что значения $a[x, y]$ определены для всех пар целых чисел (x, y) по формуле

$$a[x, y] = a[x', y'],$$

где

$$x' = \begin{cases} 0 & \text{если } 0 < x, \\ x & \text{если } 0 \leq x < m_x, \\ m_x - 1 & \text{если } x \geq m_x, \end{cases}$$

$$y' = \begin{cases} 0 & \text{если } 0 < y, \\ y & \text{если } 0 \leq y < m_y, \\ m_y - 1 & \text{если } y \geq m_y. \end{cases}$$

Кроме того, можно рассматривать матрицу $a[x, y]$ как функцию $a(x, y)$ от двух целых переменных и распространить её значения на всю действительную плоскость с помощью некоторой интерполяционной формулы. Таким образом, функция $a : \mathbb{R}^2 \rightarrow \mathbb{R}$ определена на всей действительной плоскости.

Для получения изображения с матрицей $b[x, y]$ вдвое меньшего размера $\frac{m_x}{2} \times \frac{m_y}{2}$ можно использовать один из двух основных методов.

1. Усреднение:

$$b[x, y] = \frac{a[2x, 2y] + a[2x + 1, 2y] + a[2x, 2y + 1] + a[2x + 1, 2y + 1]}{4}. \quad (1)$$

2. Четные точки:

$$b[x, y] = a[2x, 2y]. \quad (2)$$

В настоящей работе мы ограничимся только первым методом.

В некоторых задачах (например при сжатии изображений) алгоритм уменьшения размера может использовать большее число смежных точек. Возникает необходимость вместо (1) рассматривать формулу более общего вида:

$$b[x, y] = \sum_{i,j} a[2x + i, 2y + j] \cdot \beta(i, j), \quad (3)$$

где $\beta(i, j)$ – некоторая весовая функция.

По результатам измерений можно отметить, что переход от усреднения по четырем точкам к взвешенному среднему по 16 показывает заметное улучшение точности аппроксимации. Взвешивание по 36 дает ещё некоторое улучшение, но дальнейшее увеличение области практически не сказывается на результатах интерполяции. Поэтому в настоящей работе мы ограничились случаем 36 ненулевых коэффициентов функции $\beta(i, j)$, среди которых только 6 являются различными. Они могут быть расположены следующим образом:

$$\begin{array}{cccccc} \beta_5 & \beta_4 & \beta_3 & \beta_3 & \beta_4 & \beta_5 \\ \beta_4 & \beta_2 & \beta_1 & \beta_1 & \beta_2 & \beta_4 \\ \beta_3 & \beta_1 & \beta_0 & \beta_0 & \beta_1 & \beta_3 \\ \beta_3 & \beta_1 & \beta_0 & \beta_0 & \beta_1 & \beta_3 \\ \beta_4 & \beta_2 & \beta_1 & \beta_1 & \beta_2 & \beta_4 \\ \beta_5 & \beta_4 & \beta_3 & \beta_3 & \beta_4 & \beta_5 \end{array}$$

причем их сумма должна быть равна 1:

$$4\beta_0 + 8\beta_1 + 4\beta_2 + 8\beta_3 + 8\beta_4 + 4\beta_5 = 1.$$

Работа состоит из четырех частей. В первом разделе приведено описание некоторых алгоритмов линейного удвоения размера изображений (метод ближайшего соседа, линейная и кубическая интерполяция, интерполяция посредством свертки с ядром Ланцоша), их свойства, а также описание алгоритма коррекции и метода 17-точечной интерполяции. Второй раздел включает описание методов оптимального уменьшения размера изображений для рассмотренных в первом разделе методов удвоения. В третьем разделе содержатся результаты численных оценок качества (точности) рассмотренных в работе методов удвоения. В последнем разделе подведены итоги проведенной работы.

1. Методы линейного удвоения размера изображения

Если матрица $b[x, y]$ половинного размера была получена по формуле (1) из матрицы $a[x, y]$, то

$$a(x, y) \approx b(0.5x - 0.25, 0.5y - 0.25), \quad (4)$$

а если по формуле (2),

$$a(x, y) \approx b(0.5x, 0.5y). \quad (5)$$

Для приближения значения функции в точке с нецелыми координатами будем использовать свертку с некоторым ядром $R(x, y)$:

$$f(x, y) = \sum_{i, j \in \mathbb{Z}} f(i, j) R(x - i, y - j).$$

Обычно предполагается, что ядро удовлетворяет условиям:

- $R(x, y) = 0$, если $|x| \geq Q, |y| \geq Q$ для некоторой константы Q ;
- $R(x, y) = R(|x|, |y|)$;
- $R(x, y) = R(y, x)$;
- $R(0, 0) = 1$;
- $R(i, j) = 0$ при целых (i, j) не равных нулю одновременно.

К данным условиям часто добавляют ещё одно, которое формализуется как

$$\sum_{i, j \in \mathbb{Z}} R(x - i, y - j) = 1$$

для всех действительных (x, y) . Очевидно, что достаточно проверить это равенство при $0 \leq x, y \leq 1$.

Будем называть ядро $R(x, y)$ *разложимым*, если

$$R(x, y) = R(x) \cdot R(y).$$

В случае применения формулы (4) нам требуются значения функции-ядра лишь в точках вида

$$(i + 0.25, j + 0.25), (i + 0.75, j + 0.25), (i + 0.25, j + 0.75), (i + 0.75, j + 0.75)$$

для целых (i, j) . Во всех далее рассматриваемых примерах константа Q не превышает 3, поэтому в случае разложимого ядра достаточно знать значения функции $R(x)$ в точках

$$0.25, \quad 0.75, \quad 1.25, \quad 1.75, \quad 2.25, \quad 2.75, \quad (6)$$

то есть всего шесть значений.

1.1. Тривиальная, линейная и кубическая интерполяции

Наиболее популярны следующие методы интерполяции значений яркости (см., например, [2], стр. 406):

- Тривиальная интерполяция (метод ближайшего соседа), то есть по ближайшей целой точке. Ядро интерполяции:

$$R_0(x) = \begin{cases} 1 & 0 \leq |x| < 0.5 \\ 0 & 0.5 \leq |x| \end{cases}$$

- Линейная интерполяция, ядро

$$R_1(x) = \begin{cases} 1 - |x| & 0 \leq |x| < 1 \\ 0 & 1 \leq |x| \end{cases}$$

- Кубическая интерполяция, её (одномерное) ядро можно записать в виде

$$R_3(x) = \begin{cases} (1 - |x|) \cdot (1 + |x|/2 - |x|^2/2) & 0 \leq |x| \leq 1 \\ (1 - |x|)(2 - |x|)(3 - |x|)/6 & 1 \leq |x| \leq 2 \\ 0 & |x| > 2 \end{cases}$$

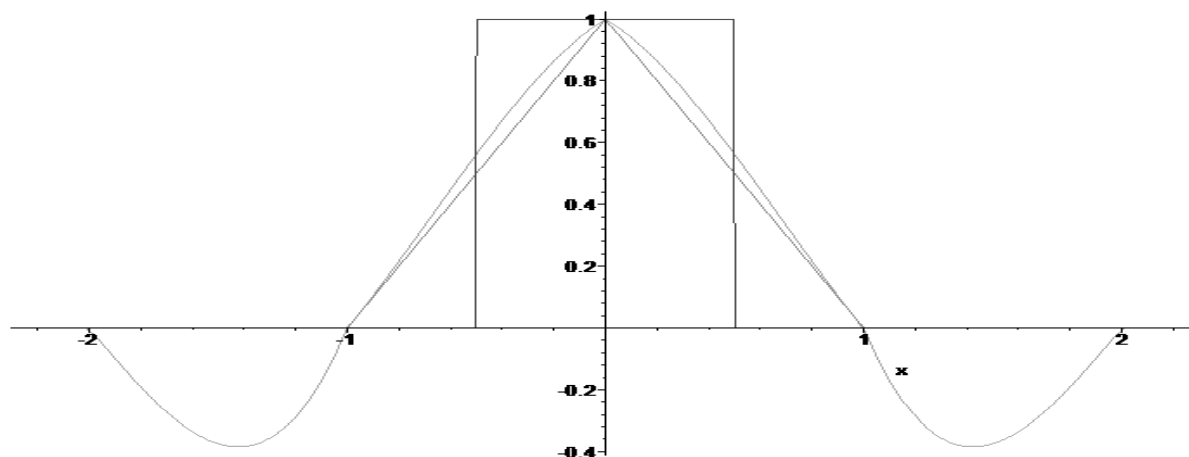


Рис. 1. Функции $R_0(x)$, $R_1(x)$, $R_3(x)$

Fig. 1. Functions $R_0(x)$, $R_1(x)$, $R_3(x)$

1.2. Ядро Ланцоша

Ядра Ланцоша считаются одними из наиболее эффективных при интерполяции значения яркости изображений, см., например, [6], стр. 231–232.

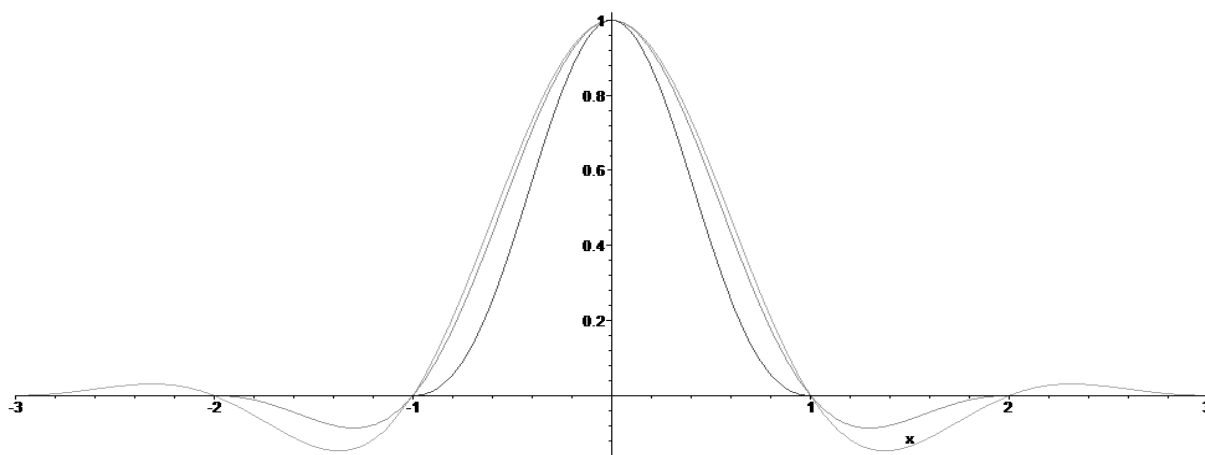


Рис. 2. Функции $L_1(x)$, $L_2(x)$, $L_3(x)$

Fig. 2. Functions $L_1(x)$, $L_2(x)$, $L_3(x)$

Пусть sinc – кардинальный синус, то есть

$$\text{sinc}(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x} & x \neq 0 \\ 1 & x = 0 \end{cases}$$

Функцией Ланцоша с натуральным параметром a (размером окна) называется функция

$$L(x) = \begin{cases} C \cdot \text{sinc}(x) \text{sinc}(\frac{x}{a}) & |x| < a \\ 0 & |x| \geq a \end{cases}$$

где константа C подбирается таким образом, чтобы

$$\int L(x) = 1.$$

a	C
1	1.1076364143517988223
2	0.9903050711891360020
3	1.0029533506415752856

При использовании ядра Ланцоша следует иметь в виду, что из-за обнуления его значений при $|x| > a$, условие $\sum_i R(x-i) \equiv 1$ будет выполняться лишь приближенно. Поэтому весовые коэффициенты ядра необходимо пронормировать таким образом, чтобы их сумма равнялась 1.

1.3. Разложимые ядра

Как уже говорилось выше, в случае разложимого ядра для нахождения матрицы удвоенного размера нам нужна не вся функция $R(x)$, а лишь шесть её значений:

$$R(0.25), R(0.75), R(1.25), R(1.75), R(2.25), R(2.75),$$

то есть, фактически, вектор из шести чисел. Выпишем явно этот вектор во всех интересующих нас случаях.

- Тривиальная интерполяция, то есть по ближайшей целой точке. Значения разложимой функции R_0 в точках (6) таковы:

$$(1, 0, 0, 0, 0, 0).$$

- Линейная интерполяция, значения функции в шести точках:

$$(3/4, 1/4, 0, 0, 0, 0).$$

- Кубическая интерполяция:

$$(105, 35, -7, -5, 0, 0)/128.$$

- Интерполяция 5-й степени:

$$(3465, 1155, -693, -495, 77, 63)/8192.$$

- Ядро Ланцоша с параметром $a = 1$

$$(0.9, 0.1, 0, 0, 0, 0).$$

- Ядро Ланцоша с параметром $a = 2$

$$(0.8686065442, 0.2330001887, -0.08388006799, -0.01772666415, 0, 0),$$

или, с округлением до $1/256$:

$$(222, 60, -21, -5, 0, 0)/256.$$

- Ядро Ланцоша с параметром $a = 3$

$(0.89277077, 0.27101057, -0.13327464, -0.067997263, 0.030112286, 0.0073782709),$

или, с округлением до $1/256$:

$$(229, 69, -34, -18, 8, 2)/256.$$

Замечание 1. Коэффициенты в ядрах Ланцоша нормируются так, чтобы их сумма равнялась единице.

Эти коэффициенты можно подобрать так, чтобы минимизировать погрешность на реальных файлах. Такие наборы будем называть оптимальными.

- Оптимальная линейная интерполяция

$$(7/8, 1/8, 0, 0, 0, 0).$$

- Оптимальная кубическая интерполяция

$$(254, 48, -38, -8, 0, 0)/256.$$

- Оптимальная пятой степени

$$(256, 37, -36, -6, 4, 1)/256.$$

1.4. 17-точечная интерполяция

Во всех рассмотренных выше случаях ядро $R(x, y)$ имело вид произведения: $R(x, y) = R(x)R(y)$. В общем случае это не обязательно так. В работе [4] предложена 17-точечная интерполяционная формула, ядро которой не может быть записано в таком виде.

Если мы хотим использовать функцию-ядро лишь для удвоения размера изображения, достаточно будет определить значения функции R в точках вида $R(k \pm 0.25, l \pm 0.25)$ где k, l – целые числа. Поэтому с учетом симметричности ($R(x, y) = R(|x|, |y|)$) нужные нам значения функции $R(x, y)$ можно задать таблицей:

R_{17}	$1/4$	$3/4$	$5/4$	$7/4$	$9/4$
$1/4$	r_0	r_1	r_3	r_4	r_9
$3/4$	r_1	r_2	r_5	r_7	
$5/4$	r_3	r_5	r_6	r_8	
$7/4$	r_4	r_7	r_8		
$9/4$	r_9				

Таким образом, функция $R(x, y)$ определяется десятью коэффициентами. 17-точечной она названа потому, что интерполированное значение вычисляется как взвешенная сумма значений исходной функции в 17 ближайших целых точках.

Большое количество численных экспериментов позволило найти формулу близкую к оптимальной в этом классе:

$$\{r_0, \dots, r_9\} = \{256, 45, 13, -46, -7, -9, 3, -2, 1, 10\}/256. \quad (7)$$

Добавление дополнительных ненулевых коэффициентов к ядру позволяет улучшить качество алгоритма в лучшем случае на сотые процента и вряд ли может быть признано целесообразным.

1.5. Коррекция алгоритма удвоения

Пусть B — изображение, уменьшенное в два раза с помощью (1). Допустим, что у нас имеется некоторый алгоритм удвоения размера изображения $\psi : B \rightarrow C$. Если увеличенное изображение C уменьшить вдвое, то мы не обязательно получим B при его увеличении.

Поэтому рассмотрим следующий алгоритм улучшения восстановленного изображения C . Пусть B' — аппроксимация изображения B , полученная усреднением увеличенного изображения C . Тогда, в силу линейности (1) коррекция значений C определяется следующим соотношением:

$$\begin{aligned} C(2x, 2y) &= C(2x, 2y) + e(x, y), \\ C(2x + 1, 2y) &= C(2x + 1, 2y) + e(x, y), \\ C(2x, 2y + 1) &= C(2x, 2y + 1) + e(x, y), \\ C(2x + 1, 2y + 1) &= C(2x + 1, 2y + 1) + e(x, y), \end{aligned} \quad (8)$$

где $e(x, y) = B(x, y) - B'(x, y)$ — величина отклонения истинного значения от аппроксимации.

Приведенный метод коррекции позволяет улучшить качество аппроксимации некоторых линейных методов удвоения (см. таблицу далее).

2. Оптимальное уменьшение размера изображения

Если при уменьшении размера изображения вдвое мы перейдем от формулы (1) к более сложной (3), то для всех рассмотренных выше методов можно подобрать оптимальные значения шести коэффициентов

$$(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5) \quad (9)$$

таким образом, чтобы погрешность получалась минимальной. Рассмотрим эти методы более детально.

Линейная интерполяция. Оптимальные значения коэффициентов β_i :

$$(109, 0, -3, -30, 5, 8)/256.$$

Кубическая интерполяция:

$$(82, 8, -6, -20, 5, 2)/256.$$

Пятой степени:

$$(74, 8, -1, -17, 3, 3)/256.$$

Ланцош $a = 1$:

$$(77, 6, -2, -16, 4, 1)/256.$$

Ланцош $a = 2$:

$$(75, 7, -2, -16, 4, 1)/256.$$

Ланцош $a = 3$:

$$(64, 12, -3, -14, 3, 1)/256.$$

Оптимальная линейная формула. Будем одновременно подбирать два коэффициента линейной интерполяционной формулы $(R(1/4), R(3/4))$ и шесть коэффициентов β_i . Оптимальным следует признать набор

$$(R(1/4), R(3/4)) = (200, 56)/256$$

и

$$(\beta_0, \dots, \beta_5) = (103, 0, -1, -26, 2, 10)/256.$$

Оптимальная кубическая формула. Будем одновременно подбирать четыре коэффициента линейной интерполяционной формулы $(R(1/4), R(3/4), R(5/4), R(7/4))$ и β_i . Оптимальным следует признать набор

$$(R(\dots)) = (235, 47, -23, -3)/256,$$

и

$$(\beta_0, \dots, \beta_5) = (70, 7, 0, -12, 0, 4)/256.$$

Оптимальная формула 5-й степени. Будем одновременно подбирать шесть коэффициентов линейной интерполяционной формулы

$$(R(1/4), R(3/4), R(5/4), R(7/4), R(9/4), R(11/4))$$

и β_i .

Оптимальным следует признать набор

$$(R(\dots)) = (236, 54, -49, -7, 21, 1)/256$$

и

$$(\beta_0, \dots, \beta_5) = (62, 14, 3, -13, -2, 1)/256.$$

И, наконец, **17-точечная формула.** Оптимальные коэффициенты β_i оказались следующими:

$$\{\beta_0, \dots, \beta_5\} = \{72, 15, 2, -17, -4, 2\}/256,$$

а соответствующие коэффициенты $r_0, \dots, r_9$ 17-точечной формулы:

$$\{190, 45, 14, -34, -4, -2, 6, 1, 1, 16\}/256.$$

3. Результаты расчетов

При сравнении алгоритмов было проведено тестирование на большом количестве изображений различного происхождения и качества. В таблице, представленной ниже, даны результаты для 13 небольших стандартных изображений, часто используемых в такого рода экспериментах:

Airplane.bmp, Car.bmp, Couple.bmp, Girl.bmp, House.bmp, Lake.bmp, Lena.bmp, Mandrill.bmp, Peppers.bmp, Splash.bmp, Tiffany.bmp, Tree.bmp, Woman.bmp.

Путь B – изображение, уменьшенное вдвое, и C_T – изображение B , увеличенное тривиальным методом. Тогда оценку качества алгоритма удвоения можно представить соотношением:

$$D(A, C) = MSE(A, C)/MSE(A, C_T), \quad (10)$$

где A – исходное изображение, а C – увеличенное вдвое рассматриваемым методом изображение B .

Для каждого метода в таблице представлены усредненные значения (10) по всему набору файлов в четырех вариантах:

- 1) исходный метод;
- 2) с коррекцией $corr$;
- 3) с оптимальным уменьшением ods ;
- 4) с оптимальным уменьшением и коррекцией $ods + corr$.

Например, для метода «Ланцош $a = 3$ » погрешность составляет 88.289% от погрешности тривиального метода, а при условии оптимального уменьшения и последующей коррекции — 83.809% от погрешности тривиального метода.

Метод	D	$corr$	ods	$ods + corr$
Тривиальный	10.10817			
Линейный	0.91519	0.82115	0.79500	0.78666
Кубический	0.84574	0.79950	0.78953	0.77392
5-й степени	0.82708	0.79283	0.78304	0.76908
Ланцош $a=1$	0.92254	0.90744	0.90265	0.90601
Ланцош $a=2$	0.82749	0.80067	0.78471	0.77886
Ланцош $a=3$	0.80676	0.79032	0.77608	0.76897
Лин. оптим.	0.91224	0.88835	0.80943	0.78512
Куб. оптим.	0.79730	0.79565	0.79589	0.77739
5-й ст.опт.	0.80825	0.80825	0.78501	0.77580
17 точек	0.79032	0.79032	0.76898	0.76890

Данные на других изображениях, конечно же, отличаются от представленных в таблице, однако не слишком сильно. И, самое главное, соотношение между различными методами остается тем же самым.

4. Выводы

- Если уменьшение изображения проводится путем усреднения по 4-м соседним точкам, то наилучшие результаты среди рассмотренных линейных формул дает 17-точечная интерполяция с ядром (7). Дальнейшее усложнение формулы улучшает результат на сотые процента и вряд ли может быть признано целесообразным.
- Коррекция удвоенного изображения (п. 1.5) – предложенный способ повышения точности аппроксимации. Дает отличные результаты для простых методов (например линейная интерполяция). Однако при использовании наиболее сложных алгоритмов (17-точечный, Ланцош $a = 3$) улучшения оказываются незначительными.

- Если имеется возможность выбирать метод уменьшения изображения вдвое или известен алгоритм уменьшения, посредством которого было получено исходное изображение A , то алгоритм удвоения можно несколько улучшить.
- Дальнейшее улучшение точности возможно лишь при использовании нелинейных или адаптивных методов.

Список литературы / References

- [1] Ватолин Д. и др., *Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео*, Диалог – МИФИ, М., 2002, ISBN: 5-86404-170-X, 384 с.; [Vatolin D. et al., *Metody szhatiya dannyykh. Ustroystvo arkhivatorov, szhatie izobrazheniy i video*, Dialog – MIFI, M., 2002, 384 pp., (in Russian).]
- [2] Гонсалес Р., Вудс Р., *Цифровая обработка изображений*, Техносфера, М., 2012, ISBN: 978-5-94836-331-8, 1104 с.; [Gonsales R., Vuds R., *Tsifrovaya obrabotka izobrazheniy*, Tekhnosfera, M., 2012, 1104 pp., (in Russian).]
- [3] Прэтт У., *Цифровая обработка изображений*, 1, Мир, М., 1982, 312 с.; [Prett U., *Tsifrovaya obrabotka izobrazheniy*, 1, Mir, M., 1982, 312 pp., (in Russian).]
- [4] Хашин С.И., “Семнадцатиточечная интерполяционная формула от 2 переменных”, *Вестник ИвГУ*, 2003, №3, 133–137; [Khashin S.I., “Semnadsatitocheynaya interpolatsionnaya formula ot 2 peremennykh”, *Ivanovo State University Bulletin*, 2003, №3, 133–137, (in Russian).]
- [5] Яне Б., *Цифровая обработка изображений*, Техносфера, М., 2007, 583 с.; [Yane B., *Tsifrovaya obrabotka izobrazheniy*, Tekhnosfera, M., 2007, 583 pp., (in Russian).]
- [6] Wilhelm Burger, Mark J. Burge, *Principles of digital image processing: core algorithms*, Springer-Verlag, London, 2009, ISBN: 978-1-84800-194-7, 327 pp.
- [7] David S. Taubman, Michael W. Marcellin, *JPEG2000: image compression fundamentals, standards, and practice*, Springer Science+Business Media, LLC, 2002, ISBN: 978-1-4613-5245-7, 773 pp.
- [8] Wallace G. K., “The JPEG still picture compression standard”, *IEEE Transactions on Consumer Electronics*, **38**:1 (1992), xviii – xxxiv.
- [9] Robert G. Keys, “Cubic Convolution Interpolation for Digital Image”, *Processing IEEE transaction on acoustics, speech, and signal processing*, **29**:6 (1981), 1153–1160.
- [10] Klette R., *Concise Computer Vision. An Introduction into Theory and Algorithms*, Springer-Verlag, London, 2014, ISBN: 978-1-4471-6320-6, 429 pp.
- [11] Xin Li, Michael T. Orchard, “New Edge-Directed Interpolation”, *IEEE Transactions on Image Processing*, **10**:10 (2001), 1521–1527.

Vaganov S. E., Khashin S. I., "Comparison of Doubling the Size of Image Algorithms", *Modeling and Analysis of Information Systems*, **23**:4 (2016), 389–400.

DOI: 10.18255/1818-1015-2016-4-389-400

Abstract. In this paper the comparative analysis for quality of some interpolation non-adaptive methods of doubling the image size is carried out. We used the value of a mean square error for estimation accuracy (quality) approximation. Artifacts (aliasing, Gibbs effect (ringing), blurring, etc.) introduced by interpolation methods were not considered. The description of the doubling interpolation upscale algorithms are presented, such as: the nearest neighbor method, linear and cubic interpolation, Lanczos convolution interpolation (with $a=1,2,3$), and 17-point interpolation method. For each method

of upscaling to twice optimal coefficients of kernel convolutions for different down-scale to twice algorithms were found. Various methods for reducing the image size by half were considered the mean value over 4 nearest points and the weighted value of 16 nearest points with optimal coefficients. The optimal weights were calculated for each method of doubling described in this paper. The optimal weights were chosen in such a way as to minimize the value of mean square error between the accurate value and the found approximation.

A simple method performing correction for approximation of any algorithm of doubling size is offered. The proposed correction method shows good results for simple interpolation algorithms. However, these improvements are insignificant for complex algorithms (17-point interpolation, Lanczos $a=3$). According to the results of numerical experiments, the most accurate among the reviewed algorithms is the 17-point interpolation method, slightly worse is Lanczos convolution interpolation with the parameter $a=3$ (see the table at the end).

Keywords: interpolation, convolution of function, Lanczos filter, 17-point interpolation

On the authors:

Vaganov Sergey Evgenevich, orcid.org/0000-0002-8642-8555, graduate student,
Ivanovo State University, Ermaka str., 39, Ivanovo, 153025, Russia,
e-mail: pro100-pioner@mail.ru

Khashin Sergey Ivanovich, orcid.org/0000-0002-6508-663X, PhD,
P.G. Ivanovo State University,
Ermaka str., 39, Ivanovo, 153025, Russia, e-mail: khash2@mail.ru

Acknowledgments:

This work was supported by the Federal targeted Program "Scientific and Scientific-pedagogical Personnel of Innovative Russia" under the Grant No 162 (2014/40).

©Васильчиков В. В., 2016

DOI: 10.18255/1818-1015-2016-4-401-411

УДК 519.688: 519.85

Об оптимизации и распараллеливании алгоритма Литтла для решения задачи коммивояжера

Васильчиков В. В.

получена 31 мая 2016

Аннотация.

В данной работе рассматриваются способы ускорения решения NP-полной задачи коммивояжера. Классический алгоритм Литтла, относящийся к категории "методов ветвей и границ", позволяет ее решать как для ориентированных, так и для неориентированных графов. Однако для неориентированных графов его работу можно ускорить за счет исключения рассмотрения фактически ранее рассмотренных вариантов. В работе предлагаются изменения, которые следует внести в ключевые операции алгоритма для ускорения его работы. Приводятся результаты численного эксперимента, показавшего значительное ускорение решения задачи с использованием усовершенствованного алгоритма. Другой ресурс для ускорения – это разработка параллельного алгоритма. Для задач подобного рода весьма сложно сразу разбить вычисления на достаточное количество сравнимых по трудоемкости подзадач. Параллелизм у них выявляется динамически во время вычислений. Для таких задач разумным представляется использование рекурсивно-параллельной организации вычислений. В нашем случае хорошим выбором оказалась разработанная автором библиотека RPM_ParLib, позволяющая создавать эффективные параллельные программы для вычислений на локальной сети в среде .NET Framework на любом поддерживаемом ею языке программирования. Мы при разработке программы использовали язык C#. Были написаны параллельные программы для реализации как исходного, так и модифицированного алгоритмов, проведено их сравнение. Эксперименты проводились для графов с количеством вершин до 45 с количеством компьютеров в сети до 16. Дополнительно исследовалось ускорение, которого можно достичь за счет распараллеливания базового алгоритма Литтла для ориентированных графов. Результаты этих серий экспериментов также приводятся в работе.

Ключевые слова: задача коммивояжера, алгоритм Литтла, параллельный алгоритм, рекурсия, .NET

Для цитирования: Васильчиков В. В., "Об оптимизации и распараллеливании алгоритма Литтла для решения задачи коммивояжера", *Моделирование и анализ информационных систем*, **23:4** (2016), 401–411.

Об авторах:

Васильчиков Владимир Васильевич, orcid.org/0000-0001-7882-8906, канд. техн. наук, зав. кафедрой вычислительных и программных систем,
Ярославский государственный университет им. П.Г. Демидова,
ул. Советская, 14, г. Ярославль, 150000 Россия, e-mail: vvv193@mail.ru

Благодарности:

Работа выполнена в рамках инициативной НИР ВИП-004 (номер госрегистрации АААА-А16-116070610022-6).

Введение

Задача коммивояжера является одной из самых известных задач дискретной оптимизации. Она относится к категории NP-полных задач, для которых не найдено алгоритмов точного решения за полиномиальное от размера задачи время [1]. Для нахождения точного решения некоторых задач такого рода достаточно эффективно могут применяться алгоритмы, относящиеся к категории "метод ветвей и границ". Метод впервые был предложен в [2] применительно к задаче целочисленного линейного программирования. Для решения задачи коммивояжера с успехом применяется алгоритм Литтла [3–6], который относится именно к этой категории методов. К сожалению, для него не существует приемлемой оценки трудоемкости, и время его работы очень сильно зависит от исходных данных.

Алгоритм Литтла работает и для ориентированных и для неориентированных графов. Однако в ходе экспериментов на компьютере было отмечено, что для случая симметричной матрицы расстояний (неориентированный граф) алгоритм выполняет повторное рассмотрение фактически уже исследованных ветвей вычислений. В работе предлагается модификация алгоритма, позволяющая устранить это явление и ускорить получение результата. Следует отметить, что и ранее разными авторами предлагались различные модификации базового алгоритма Литтла с целью ускорить его работу, в качестве примера можно назвать, скажем, работы [7, 8], однако наше предложение носит принципиально иной характер.

Другая возможность ускорить решение задачи – это распараллеливание. Поскольку количество вычислений и объем необходимой памяти при решении задачи коммивояжера с увеличением ее размера быстро возрастают, наличие возможности организации вычислений в параллельном режиме становится очень актуальным. Исследования в этом направлении могут быть основаны на различных парадигмах и инструментальных средствах параллельного программирования [9]. Автор же данной работы выбрал в качестве основы парадигму рекурсивно-параллельного (РП) программирования [10] и соответствующие библиотеки поддержки [11, 12].

В [10] изложены основные принципы организации рекурсивно-параллельных вычислений и описаны основные алгоритмы и механизмы поддержки этого стиля программирования. Библиотеки [11, 12] позволяют относительно легко создавать, отлаживать и эксплуатировать РП-приложения в среде .NET Framework. В [13] подробно описаны функциональные возможности упомянутых библиотек, а также процесс разработки и исследования РП-алгоритма решения NP-полной задачи о клике.

1. Задача коммивояжера

Напомним формулировку задачи.

Задано конечное множество $C = \{c_0, c_1, \dots, c_{n-1}\}$ "городов" и для каждой пары $c_i, c_j \in C$ "расстояние" между ними $d_{ij} = d(c_i, c_j) \in Z^+$ (здесь Z^+ означает положительные целые числа). Матрица расстояний $D = \|d_{ij}\|$ не обязательно должна быть симметричной, также мы не требуем выполнения неравенства треугольника, то есть рассматриваем задачу в самом общем виде.

Коммивояжер, выехав из города c_0 , должен объехать все остальные города, побывав в каждом по одному разу, и вернуться в город c_0 . Обозначим соответствующий

циклический маршрут $r = \{0, i_1, i_2, \dots, i_{n-1}, 0\}$, здесь мы без ограничения общности положили $i_0 = i_n = 0$. Очевидно, длина пути определяется как

$$l(r) = \sum_{k=0}^{n-1} d_{i_k i_{k+1}}.$$

Пусть R – множество всех таких маршрутов. Требуется найти цикл $r_0 \in R$, такой что

$$l(r_0) = \min_{r \in R} l(r).$$

Очевидно, что количество всех допустимых маршрутов равно $(n-1)!$ и полный перебор крайне неэффективен.

2. Алгоритм Литтла

Напомним, как организованы вычисления при использовании базовой реализации алгоритма Литтла.

Выберем произвольное положительное значение ∞ , заведомо большее любого из d_{ij} . Мы будем присваивать его элементам матрицы D , которые соответствуют дугам, не подлежащим включению в искомый маршрут. Сразу же пропишем это значение во всех элементах главной диагонали, чтобы исключить наличие петель. Также присвоим переменной *Record*, предназначенной для хранения длины кратчайшего на данный момент цикла, заведомо неконкурентоспособное значение. Далее алгоритм носит рекурсивный характер, его основной шаг выглядит следующим образом:

1. Осуществляем редукцию матрицы D по строкам путем вычитания из всех элементов каждой строки минимального в строке значения. Очевидно, что кратчайший маршрут после такой операции так и останется кратчайшим.
2. Аналогичным образом осуществляем приведение матрицы D по столбцам.
3. Сумму вычтенных на предыдущих двух этапах минимальных значений назовем константой редукции. Она имеет смысл нижней оценки длины кратчайшего маршрута. Если бы теперь можно было построить гамильтонов цикл только из ребер нулевой длины, это и был бы оптимальный вариант.
4. Для каждого ребра нулевой длины в модифицированной матрице D получим так называемую "оценку нуля" v_{ij} . Для элемента в позиции (i, j) она равна сумме минимальных элементов в i -й строке и j -м столбце (не считая сам этот нуль).
5. Выбираем нулевой элемент с максимальной оценкой, пусть это элемент, расположенный в i -й строке и j -м столбце. Теперь все допустимые маршруты можно разбить на два класса: содержащие ребро (i, j) и не содержащие. Для последних нижняя оценка длины оптимального маршрута увеличивается на величину v_{ij} .

6. Для продолжения вычислений организуем рекурсивное разделение процесса. Для маршрутов первого класса (с ребром (i, j)) включаем его в искомый маршрут, убедившись, что при этом не образуется цикла, проходящего менее чем через n вершин, далее исключаем из матрицы D i -ю строку и j -й столбец (или заполняем их значениями ∞), затем переходим к пункту 1. Для маршрутов без ребра (i, j) просто полагаем значение d_{ij} равным ∞ и переходим к пункту 1.

Рекурсивный процесс завершается в следующих случаях:

- В маршрут включено $n - 1$ ребро, в этом случае маршрут очевидным образом достраивается, и если его длина меньше, чем значение $Record$, маршрут запоминается как лучший на текущий момент.
- Если полученная на этапе 5 оценка длины маршрута больше или равна $Record$, ветвь считается неперспективной, и вычисления на ней завершаются немедленно.

Отметим, что для эффективной организации работы алгоритма необходимо хорошо продумать способ хранения информации, относящейся к каждой ветви вычислений. Ограниченность объема оперативной памяти сразу вынуждает нас несколько отступить от изложенной выше последовательности действий. На этапе 5 наиболее выгодным было бы выбирать для продолжения вычислений нулевой элемент с глобально максимальной оценкой, так, чтобы вычисления продолжались во всех возможных направлениях. Однако рост количества незаконченных ветвей носит в этом случае экспоненциальный характер, причем для каждой требуется хранить свой вариант редуцированной матрицы, что для программной реализации неприемлемо. Поэтому, когда мы говорим о выборе нулевого элемента с максимальной оценкой, мы имеем в виду максимум *для текущей ветви*.

Следует также отметить необходимость эффективной организации проверки на отсутствие коротких циклов, которую требуется выполнять на каждой итерации. При наличии подходящей структуры данных это можно сделать за $O(n)$ операций.

Насколько автору известно, в настоящее время не существует сколько-нибудь информативных оценок трудоемкости алгоритма Литтла – это все-таки перебор, хотя и хорошо оптимизированный. Скорость решения очень зависит от исходных данных, здесь мы имеем в виду не размер задачи, а матрицу расстояний. Для случайных графов с одинаковым количеством вершин и одинаковым распределением длин ребер, как показывает эксперимент, скорость получения результата может отличаться в десятки и даже сотни раз. На нее влияет и начальное значение $Record$, и даже порядок нумерации вершин, правда не так радикально, как расстояния между вершинами.

3. Модификация алгоритма Литтла для симметричной матрицы расстояний

Как было ранее отмечено, алгоритм работает и для ориентированных, и для неориентированных графов. Последний случай отличается только тем, что матрица рас-

стояний симметрична относительно главной диагонали. Кстати, в ходе экспериментов было отмечено, что для неориентированных графов алгоритм, как правило, работает значительно дольше. Это обусловлено многими причинами. Одна из них заключается в том, что, исключив на каком-то этапе из рассмотрения ребро (a, b) , мы в дальнейшем все равно вынуждены рассматривать варианты с включением и исключением ребра (b, a) . При продолжении ветви это делается не один раз. В итоге мы выполняем большое количество операций, которых можно было бы избежать, если сразу иметь в виду, что ребро можно пройти в обоих направлениях, а значит, исключив ребро (a, b) , мы могли бы сразу же исключить и ребро (b, a) .

Однако алгоритм для этого потребует значительно изменить. Во-первых, редукцию нужно осуществлять так, чтобы текущая матрица расстояний постоянно оставалась симметричной. Во-вторых, требуется изменить способ вычисления "оценки нуля", чтобы учесть возможность прохождения ребра в разных направлениях. Наконец, модификацию матрицы при включении ребра в маршрут также следует производить иначе.

Ниже перечислены предлагаемые модификации. Их смысл легче понять, если представить, что мы строим не один цикл, а два: в прямом и обратном направлении.

- *Редукция.* Сразу после редукции i -й строки проделываем эту же операцию с i -м столбцом. Фактически достаточно просто скопировать значения элементов строки в элементы столбца, причем в константе редукции следует вычтенное значение учесть и для строки, и для столбца. Кроме того, если вершина еще не включена в строящийся цикл, в константе редукции она учитывается с весом 2, а если один раз уже включена, то с весом 1. Элементы, равные ∞ , не редуцируются.
- *"Оценка нуля".* Для оценки нуля в позиции (i, j) возьмем сумму оценок для i -й и j -й вершин. Оценка i -й вершины вычисляется следующим образом. Если вершина i встречается в уже построенном участке цикла один раз, то оценкой является минимальный элемент в строке (или столбце) i , исключая сам нуль в позиции (i, j) . Если вершина еще не встречалась, то оценка равна сумме *двух самых маленьких элементов* в строке. Два раза или более вершина при выполнении данной операции, разумеется, встретиться не может.
- *Изменение матрицы при включении или исключении ребра.* Если мы на данной ветви исключили ребро (i, j) , то в матрице на позициях (i, j) и (j, i) записываем значение ∞ . При добавлении ребра (i, j) выполняем следующие преобразования матрицы расстояний. Если вершина i включается в строящийся цикл повторно (в нее и вошли, и вышли), всем элементам i -й строки и i -го столбца присваивается значение ∞ . Аналогично поступаем с вершиной j .

Для оценки качества предлагаемого варианта алгоритма был проведен численный эксперимент. Сразу отметим, что приводимые ниже оценки имеют чисто качественный характер. Во-первых, как было отмечено выше, количество итераций для получения результата непредсказуемо. Во-вторых, предложенная модификация радикально изменяет порядок рассмотрения вариантов построения цикла в базовом и оптимизированном алгоритме, что делает не вполне корректным их сравнение даже на одних и тех же исходных данных.

Однако несколько десятков проведенных экспериментов позволяют утверждать, что модифицированный алгоритм работает лучше почти всегда. Слово "почти" означает только то, что ни в одном из экспериментов он не работал дольше базового алгоритма, хотя непредсказуемость порядка выбора ребер для построения цикла не позволяет утверждать, что так будет всегда. При этом на достаточно больших примерах (35, 40 и более вершин) скорость работы модифицированного алгоритма зачастую была выше в десять и более раз.

Для сравнения скорости работы двух алгоритмов было следующим образом сгенерировано несколько десятков неориентированных графов. Вершины случайным образом равномерно размещались в квадрате размером 3000*3000. Расстояния округлялись до ближайшего целого. Количество вершин бралось в пределах от 20 до 40. При количестве вершин, равном 45, базовый алгоритм обычно работал слишком долго, поэтому мы ограничились только этими значениями.

На этом этапе для сбора статистики использовался компьютер на базе двухъядерного процессора AMD с тактовой частотой 1.90 GHz и 6 GB оперативной памяти, работающий под управлением 64-разрядной ОС Windows 7. Ниже приводится таблица 1, содержащая результаты экспериментов на 10 графах, выбранных случайным образом.

№ п/п	Кол-во вершин	Базовый алгоритм		Оптимизированный алгоритм		Полученное ускорение	
		Итераций	Время (мс)	Итераций	Время (мс)	Итерации	Время
1.	35	107 230	13 418	1 602	262	66.94	51.21
2.	35	847 722	115 643	14 727	2 199	57.56	52.59
3.	35	45 487 366	6 488 281	7 528 783	899 248	6.04	7.22
4.	40	31 832 080	6 694 203	242 609	4 1706	131.21	160.51
5.	40	11 242 224	2 093 984	1 362 560	161 358	8.25	12.98
6.	40	25 721 496	4 736 154	1 601 237	256 790	16.06	18.44
7.	40	1 144 205	190 076	492 012	63 070	2.33	3.01
8.	40	2 864 366	480 548	1 316 074	189 683	2.18	2.53
9.	40	673 835	129 462	9 320	1 597	72.30	81.07
10.	40	3 670 605	649 980	255 015	34 266	14.39	18.97

Таблица 1. Сравнение скорости работы базового и оптимизированного алгоритмов

Несмотря на существенный разброс результатов сравнения двух алгоритмов, обусловленный вышеупомянутыми причинами, по-видимому, есть основания отдать предпочтение оптимизированному варианту.

4. Параллельный вариант алгоритма Литтла

Для таких трудоемких задач, как задача коммивояжера, совершенно естественным выглядит желание ускорить получение решения за счет распараллеливания алгоритма. Например, в [9] описан опыт распараллеливания алгоритма на суперкомпьютере MVS-15000 ВМ с применением такого подхода, как динамическое назначение подзадач на обработку. Задача коммивояжера решалась для графов с количеством вершин от 21 до 40, при этом на 12 процессорах достигалось ускорение до 6-7 раз.

Следует отметить, что способ разбиения данной задачи на ветви, допускающие параллельную обработку отнюдь не очевиден, а от него в основном и зависит достигнутое ускорение. За основу мы взяли рекурсивно-параллельный подход к построению алгоритма [10], язык программирования C#, а в качестве основного инструмента организации параллельных вычислений использовали библиотеку RPM_ParLib [11–13].

Такой подход естественным образом вытекает из рекурсивной организации алгоритма Литтла. На каждой ветви после выбора очередного ребра (a, b) ("нулевой элемент" с максимальной оценкой) вычисления могут быть разбиты на подзадачу, содержащую это ребро, и подзадачу, его не содержащую. Первую процессорный модуль (ПМ) оставляет себе, вторую оформляет как потенциально мигрирующий процесс, который при необходимости может быть выполнен на любом модуле. Из соображений эффективности такое рекурсивное дробление нецелесообразно продолжать до предела, поэтому мы заложили ограничение на уровень вложенности рекурсии и на количество уже включенных ребер. При превышении этих значений подзадача решается с использованием последовательного алгоритма. В наших экспериментах оба ограничения были равны 16, этого было достаточно, чтобы механизм динамической балансировки нагрузки, реализованный в библиотеке RPM_ParLib, мог достаточно равномерно распределить работу.

При передаче подзадачи на другой модуль достаточно передавать только накопленные для нее текущую оценку и векторы редукции, поскольку исходная матрица расстояний в самом начале работы рассылалась на все процессорные модули. Как только на одной из веток улучшался текущий рекорд, полученное значение немедленно рассылалось на все модули и использовалось ими в процессе вычислений для отсека неперспективных ветвей.

		Вар.1	Вар.2	Вар.3	Вар.4	Вар.5	Вар.6
Последов. алгоритм	Итераций	655 505	7 563 800	16 195 164	5 517 076	1 078 332	14 133 682
	Время (с.)	76	690	1343	527	127	1176
1 ПМ	Время (с.)	80	722	1391	535	143	1282
	Ускорение	0.94	0.96	0.97	0.98	0.89	0.92
2 ПМ	Время (с.)	32	265	603	84	49	405
	Ускорение	2.34	2.60	2.23	6.24	2.62	2.90
4 ПМ	Время (с.)	23	183	232	46	37	261
	Ускорение	3.35	3.77	5.78	11.46	3.48	4.50
6 ПМ	Время (с.)	22	70	145	35	35	208
	Ускорение	3.49	9.85	9.28	15.16	3.61	5.66
8 ПМ	Время (с.)	20	54	136	30	34	203
	Ускорение	3.84	12.81	9.90	17.63	3.73	5.78
12 ПМ	Время (с.)	19	38	124	30	32	148
	Ускорение	3.96	18.26	10.86	17.45	4.03	7.94
16 ПМ	Время (с.)	16	33	93	27	28	121
	Ускорение	4.66	20.90	14.39	19.80	4.59	9.73

Таблица 2. Ускорение параллельного оптимизированного алгоритма на графах с 45 вершинами

В процессе тестирования использовались компьютеры на базе четырехъядерного процессора Intel Core i3 с тактовой частотой 3.07 GHz и 4 GB оперативной памяти, работающие под управлением 64-разрядной ОС Windows 7. Пропускная способность сети равнялась 100 Mb/s.

В таблице 2 приведены результаты вычислений, позволяющие оценить ускорение, полученное при распараллеливании оптимизированного алгоритма на 6 случайных графах с 45 вершинами. Способ их генерации описан выше. Отметим, что это ускорение для оптимизированного алгоритма, в котором, как отмечалось выше, уже исключены лишние вычисления. Строка "ускорение" для одного ПМ содержит отношение времени работы параллельного алгоритма на одном модуле ко времени работы алгоритма последовательного. Это своего рода накладные расходы, и, как видно из таблицы, они не слишком высоки. В остальных случаях ускорение считалось также относительно времени работы именно последовательного варианта.

Отметим, что ускорение нередко превышало количество исполнителей. Для рекурсивно-параллельной организации алгоритмов "ветвей и границ" это вполне естественно, поскольку при поиске на решения "с разных сторон" больше шансов быстро найти хорошее решение и за счет этого отсеять большее количество перспективных ветвей. Такое явление наблюдалось и при решении задачи о клике [13].

		Базовый алгоритм Литтла			Оптимизированный алгоритм Литтла		
		Вар.1	Вар.2	Вар.3	Вар.1	Вар.2	Вар.3
Посл. алг.	Итераций	31 832 080	11 242 224	25 721 496	242 609	1 362 560	1 601 237
	Время (с.)	3331	1119	2528	25	79	143
1 ПМ	Время (с.)	3423	1150	2617	29	85	156
	Ускорение	0.97	0.97	0.97	0.87	0.93	0.92
	Отн. баз.				115.30	13.16	16.23
2 ПМ	Время (с.)	1091	284	489	21	23	54
	Ускорение	3.05	3.95	5.17	1.22	3.41	2.67
	Отн. баз.				161.94	48.20	47.11
4 ПМ	Время (с.)	312	122	366	17	17	39
	Ускорение	10.68	9.16	6.90	1.44	4.55	3.70
	Отн. баз.				191.07	64.41	65.22
8 ПМ	Время (с.)	189	76	199	16	15	31
	Ускорение	17.66	14.82	12.67	1.54	5.30	4.59
	Отн. баз.				203.87	74.92	81.00
12 ПМ	Время (с.)	119	54	153	14	16	25
	Ускорение	27.93	20.75	16.49	1.77	5.00	5.66
	Отн. баз.				234.07	70.73	99.93
16 ПМ	Время (с.)	98	51	141	15	12	24
	Ускорение	34.01	22.04	17.92	1.66	6.43	6.00
	Отн. баз.				219.20	90.91	105.93

Таблица 3. Ускорение параллельной версии для базового и оптимизированного алгоритмов на графах с 40 вершинами

Для того чтобы оценить ускорение одновременно и за счет оптимизации, и за счет распараллеливания, для нескольких примеров были проведены эксперимен-

ты одновременно как для базового, так и для оптимизированного алгоритма. Мы ограничились здесь графами с 40 вершинами, поскольку, как было выше отмечено, при большем числе вершин базовый алгоритм обычно работает недопустимо долго. Из числа сгенерированных случайных графов мы выбрали три с наибольшим количеством итераций, потребовавшихся для решения задачи.

Из таблицы 3 видно, что параллельный вариант оптимизированного алгоритма относительно своего же последовательного варианта обеспечивает меньший выигрыш, чем базовый параллельный относительно базового последовательного, что, если честно, неудивительно. Однако он все равно работает намного быстрее базового варианта. И с ростом размера задачи этот разрыв, по-видимому, будет только расти, достаточно сравнить эти результаты с результатами для графов с 45 вершинами. Можно отметить, что описанный подход к распараллеливанию алгоритма для базового варианта работает также очень хорошо.

5. Результаты использования РП-алгоритма для ориентированных графов

Все упомянутые выше результаты относились к неориентированным графам, для них, собственно, и разрабатывался оптимизированный алгоритм. Однако нас интересовали также и возможности рекурсивного распараллеливания алгоритма Литтла для ориентированных графов, разумеется, уже без описанной выше оптимизации.

Для эксперимента мы генерировали случайные графы, длина дуг в которых была равномерно распределена в промежутке от 1 до 5000. Разумеется, применительно к исходной формулировке задачи такие примеры выглядят очень уж неестественно, однако нас в первую очередь интересовало качество получающегося параллельного алгоритма и возможности библиотеки RPM_ParLib.

		Вар.1	Вар.2	Вар.3	Вар.4	Вар.5	Вар.6
Последов. алгоритм	Итераций	2714275	1774096	1586139	2006961	2723686	1394583
	Время (с.)	801	553	485	590	853	414
1 ПМ	Время (с.)	817	593	489	594	878	425
	Ускорение	0.98	0.93	0.99	0.99	0.97	0.97
2 ПМ	Время (с.)	240	149	182	233	228	118
	Ускорение	3.34	3.72	2.67	2.53	3.75	3.51
4 ПМ	Время (с.)	156	106	164	130	137	85
	Ускорение	5.15	5.22	2.96	4.53	6.23	4.90
8 ПМ	Время (с.)	145	71	124	94	88	78
	Ускорение	5.53	7.83	3.91	6.30	9.75	5.29
12 ПМ	Время (с.)	106	67	108	63	69	75
	Ускорение	7.52	8.26	4.51	9.43	12.42	5.50
16 ПМ	Время (с.)	109	65	99	53	59	75
	Ускорение	7.34	8.56	4.90	11.18	14.49	5.54

Таблица 4. Результаты вычислений для ориентированных графов с 70 вершинами

Поскольку для графов, построенных таким образом, отсечение неперспективных направлений поиска кратчайшего цикла происходит намного быстрее, за приемлемое время мы можем решать задачи большего размера. В таблице 4 приведены результаты некоторых вычислений для графов с 70 вершинами. Для этого эксперимента мы сгенерировали 15 вариантов исходных данных и выбрали из них 6 самых трудоемких.

Рассмотрение полученных результатов приводит нас к выводу о том, что и в этом случае алгоритм Литтла обладает хорошим потенциалом для организации вычислений в рекурсивно-параллельном стиле. При этом, как правило, на более трудоемких задачах удастся достичь большего ускорения.

Список литературы / References

- [1] Гэри М. Джонсон Д., *Вычислительные машины и труднорешаемые задачи*, Мир, Москва, 1982; [Garey M. R., Johnson D. S., *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Co, San Francisco, 1979.]
- [2] Land A. H. Doig A. G., “An Automatic Method of Solving Discrete Programming Problems”, *Econometrica*, **28** (1960), 497–520.
- [3] John D. C. Little, Katta G. Murty, Dura W. Sweeney, Caroline Karel, “An Algorithm for the Traveling Salesman Problem”, *Operations Research*, **11:6** (1963), 972–989.
- [4] Кофман А., *Введение в прикладную комбинаторику*, Наука, Москва, 1975; [Kaufmann A., *Intraduction a la combinatorique en vue des applications*, Dumond, Paris, 1968.]
- [5] Рейнгольд Э. Нивергельт Ю. Део Н., *Комбинаторные алгоритмы. Теория и практика*, Мир, Москва, 1980; [Reingold E. Nievergelt Ju. Deo N., *Combinatorial Algorithms: Theory and Practice*, Prentice Hall College, 1977.]
- [6] Кормен Т. Лейзерсон Ч. Ривест Р. Штайн К., *Алгоритмы: построение и анализ*, Вильямс, Москва, 2006; [Cormen T. Leiserson Ch. Rivest R. Stein C., *Introduction to Algorithms*, The MIT Press, 2001.]
- [7] Петрунин С. В., “Использование метода последовательной сепарации для решения задачи коммивояжера”, *Научный вестник МТГУ ГА*, 2009, № 146, 105–108; [Petrovnine S. V., “Employment of permanent separation method for solution of traveling salesman problems”, *Scientifik bulletin of the MSTU CA*, 2009, № 146, 105–108, (in Russian).]
- [8] Костюк Ю. Л., “Эффективная реализация алгоритма решения задачи коммивояжера методом ветвей и границ”, *Прикладная дискретная математика*, 2013, № 2, 78–90; [Kostyuk Yu. L., “Effective implementation of algorithm for solving the travelling salesman problem by branch-and-bound method”, *Prikl. Diskr. Mat.*, 2013, № 2, 78–90, (in Russian).]
- [9] Сигал И. Х., Бабинская Я. Л., Посыпкин М. А., “Параллельная реализация метода ветвей и границ в задаче коммивояжера на базе библиотеки BNB-Solver”, *Труды ИСА РАН*, **25** (2006), 26–36; [Sigal I. Kh., Babinskaya Ya. L., Posypkin M. A., “Parallelnaya realizaciya metoda vetvey i granic v zadache kommivoyazhera na baze biblioteki BNB-Solver”, *Trudy ISA RAN*, **25** (2006), 26–36, (in Russian).]
- [10] Васильчиков В. В., *Средства параллельного программирования для вычислительных систем с динамической балансировкой загрузки*, ЯрГУ, Ярославль, 2001; [Vasilchikov V. V., *Sredstva parallelnogo programmirovaniya dlya vychislitelnykh sistem s dinamicheskoy balansirovkoj zagruzki*, YarGU, Yaroslavl, 2001, (in Russian).]
- [11] Васильчиков В. В., *Коммуникационный модуль для организации полносвязного соединения компьютеров в локальной сети с использованием .NET Framework*, Свидетельство о государственной регистрации программы для ЭВМ № 2013619925,

- 2013; [Vasilchikov V. V., *Kommunikatsionnyy modul dlya organizatsii polnosvyaznogo soedineniya kompyuterov v lokalnoy seti s ispolzovaniem .NET Framework*, Svidetelstvo o gosudarstvennoy registratsii programmy dlya EVM № 2013619925, 2013, (in Russian).]
- [12] Васильчиков В. В., *Библиотека поддержки рекурсивно-параллельного программирования для .NET Framework*, Свидетельство о государственной регистрации программы для ЭВМ № 2013619926, 2013; [Vasilchikov V. V., Svidetelstvo o gosudarstvennoy registratsii programmy dlya EVM № 2013619926, 2013, (in Russian).]
- [13] Васильчиков В. В., “О поддержке рекурсивно-параллельного программирования в .NET Framework”, *Модел. и анализ информ. систем*, **21**:2 (2014), 15–25; [Vasilchikov V. V., “On the Recursive-Parallel Programming for the .NET Framework”, *Modeling and Analysis of Information Systems*, **21**:2 (2014), 15–25, (in Russian).]

Vasilchikov V. V., "On the Optimization and Parallelizing Little Algorithm for Solving the Traveling Salesman Problem", *Modeling and Analysis of Information Systems*, **23**:4 (2016), 401–411.

DOI: 10.18255/1818-1015-2016-4-401-411

Abstract. The paper describes some ways to accelerate solving the NP-complete Traveling Salesman Problem. The classic Little algorithm belonging to the category of "branch and bound methods" can solve it both for directed and undirected graphs. However, for undirected graphs its operation can be accelerated by eliminating the consideration of branches examined earlier. The paper proposes changes to be made in the key operations of the algorithm to speed up its execution. It also describes the results of an experiment that demonstrated a significant acceleration of solving the problem by using an advanced algorithm. Another way to speed up the work is to parallelize the algorithm. For problems of this kind it is difficult to break the task into a sufficient number of subtasks having comparable complexity. Their parallelism arises dynamically during the execution. For such problems, it seems reasonable to use parallel-recursive algorithms. In our case the use of the library RPM.ParLib developed by the author was a good choice. It allows us to develop effective applications for parallel computing on a local network using any .NET-compatible programming language. We used C# to develop the programs. Parallel applications were developed as for basic and modified algorithms, the comparing of their speed was made. Experiments were performed for the graphs with the number of vertexes up to 45 and with the number of network computers up to 16. We also investigated the acceleration that can be achieved by parallelizing the basic Little algorithm for directed graphs. The results of these experiments are also presented in the paper.

Keywords: Traveling Salesman Problem, Little algorithm, parallel algorithm, recursion, .NET

On the authors:

Vasilchikov Vladimir Vasilyevich, orcid.org/0000-0001-7882-8906, PhD,
P.G. Demidov Yaroslavl State University,
Sovetskaya str., 14, Yaroslavl, 150000, Russia, e-mail: vvv193@mail.ru

Acknowledgments:

This work was supported by initiative program VIP-004 (state registration number AAAA-A16-116070610022-6).

©Величко А. С., Грибова В. В., Федорищев Л. А., 2016

DOI: 10.18255/1818-1015-2016-4-412-426

УДК 519.8, 004.92

Облачный сервис для интерактивного моделирования межтерриториальной торговли

Величко А. С., Грибова В. В., Федорищев Л. А.

получена 14 апреля 2016

Аннотация. В работе описана математическая модель торговых потоков между территориями региона или страны в транспортной сети, представленной одним агрегированным или разнообразными видами морского и наземного транспорта. Используется подход моделирования сложных коммуникационных систем для определения наиболее вероятных значений потоков в условиях неполноты информации о рассматриваемой системе. Транспортные издержки между территориями моделируются в рамках гравитационной модели. Плата за перевозку зависит от расстояния между регионами, оцениваемого как длина кратчайшего пути между пунктами в заданной транспортной сети или определяемого как географическое расстояние. Математическая постановка задачи относится к классу выпуклых задач математического программирования и предполагает численное решение нелинейной оптимизационной задачи с линейными ограничениями. На основе модели реализовано программное средство как облачный сервис на гетерогенных вычислительных архитектурах: модуль моделирования выполнен на высокопроизводительной серверной платформе, модули управления и визуализации – на облачной платформе IACPaas. Связь между платформами осуществляется через асинхронные http-запросы. Для обмена информацией между модулями разработана и реализована декларативная модель в формате JSON для объектов, рассматриваемых в математической модели: продуктов, зон и коммуникаций. Модуль визуализации позволяет наглядно представить исходные и получаемые в результате моделирования матричные данные, интерактивно изменять входные параметры модели. В работе продемонстрировано использование программного комплекса для моделирования межрегиональных грузопотоков в транспортной сети регионов Дальнего Востока России на основе входных данных, взятых из открытых статистических источников.

Ключевые слова: межрегиональный, транспорт, сеть, мультимодальный, гравитационный, энтропийный, равновесие, модель

Для цитирования: Величко А. С., Грибова В. В., Федорищев Л. А., "Облачный сервис для интерактивного моделирования межтерриториальной торговли", *Моделирование и анализ информационных систем*, **23**:4 (2016), 412–426.

Об авторах:

Величко Андрей Сергеевич, orcid.org/0000-0001-8760-316X, канд. физ.-мат. наук, научный сотрудник, Институт автоматизации и процессов управления ДВО РАН
ул. Радио, 5, г. Владивосток, 690041 Россия, e-mail: vandre@dvo.ru

Грибова Валерия Викторовна, orcid.org/0000-0001-9393-351X, д-р техн. наук, зав. лабораторией, Институт автоматизации и процессов управления ДВО РАН
ул. Радио, 5, г. Владивосток, 690041 Россия, e-mail: gribova@iacp.dvo.ru

Федорищев Леонид Александрович, orcid.org/0000-0002-2049-2570, канд. техн. наук, научный сотрудник, Институт автоматизации и процессов управления ДВО РАН
ул. Радио, 5, г. Владивосток, 690041 Россия, e-mail: fleo1987@mail.ru

Благодарности:

Работа выполнена при частичной финансовой поддержке РФФИ, №16-07-00340-а.

Введение

Работа транспортной системы, связывающей регионы страны, является необходимой для безопасного и устойчивого функционирования регионов как экономической системы, обеспечивающей процессы производства и потребления. Поскольку создание такой системы – трудоемкая задача, необходимым условием является предварительный анализ и оценка проектных и управленческих решений до их применения в реальных условиях. Такие решения могут быть направлены на модернизацию и строительство отдельных участков транспортной сети региона, прежде всего на подходах к портам, крупным промышленным районам, крупным городским агломерациям и новым месторождениям.

Транспортные системы в современных условиях характеризуются независимым некооперативным поведением большого количества разнородных экономических субъектов, реализующих свои индивидуальные интересы, совместно использующих сеть автомобильного, воздушного, трубопроводного, морского, железнодорожного транспорта как для грузовых, так и для пассажирских перевозок.

Моделирование межрегиональных потоков в межотраслевой постановке рассматривалось В. В. Леонтьевым [1], с точки зрения общего экономического равновесия и «гравитационного» подхода – в работе [2]. В работах [3, 4] моделировались коммуникационные системы в условиях неполноты информации. В настоящее время эти подходы активно используются для моделирования пассажиро- и грузопотоков в транспортных системах как в России, так и за рубежом [5–10]. Однако в построении математической модели торговых потоков наблюдаются следующие недостатки: не учитываются экспортные и импортные потоки за пределы системы рассматриваемых регионов; перевозки могут быть как многопродуктовыми, так и мультимодальными, то есть могут осуществляться по разным путям и/или различными видами транспорта: по автомобильным дорогам, железнодорожным, морским и авиационным видами транспорта; пункты ввоза и вывоза, моделируемые процессами спроса и предложения товаров, могут быть отнесены не только к одинаковым, но и к различным регионам. Помимо этого входные данные могут быть заданы неточно, в этом случае возможна модификация модели с интервальным заданием входных данных, что обсуждалось в статье [11]. Не менее важной задачей для использования моделей в практической деятельности является создание интерактивных систем моделирования со средствами визуализации полученных результатов в удобном для пользователя виде, прежде всего в виде графов.

Целью данной работы является описание программного комплекса, реализованного на гетерогенных вычислительных архитектурах, для интерактивного моделирования межтерриториальной торговли.

1. Требования к программному комплексу и основные принципы его реализации

Моделирование экономических связей территорий используется при принятии решений о региональной социально-экономической политике, направленной на снижение негативных проявлений межрегиональной неоднородности, в целях предот-

вращения образования депрессивных, неконкурентоспособных, бедствующих территорий, отстающих от регионов-лидеров, и в целом – в целях интеграции регионов страны.

Программное средство моделирования межтерриториальной торговли предназначено для специалистов различных министерств и ведомств, занимающихся проблемой оптимизации и планирования межтерриториальных потоков продуктов отраслей экономики с учетом их многопродуктовости в многомодальных транспортных сетях.

Основными требованиями к программному комплексу являются:

1. Реализация комплекса как облачного сервиса для обеспечения как можно более широкой доступности разрабатываемого приложения для заинтересованного сообщества. Использование приложения как облачного сервиса не накладывает дополнительных ограничений на платформу, установленную на компьютере клиента, оперативную память, другие характеристики компьютера, а также не требует навыков инсталляции приложения [12, 13].

2. Возможность в реальном режиме времени моделировать торговые потоки. Учитывая, что моделирование торговых потоков требует больших вычислительных ресурсов, блок моделирования должен быть реализован на высокопроизводительных вычислительных платформах для обеспечения эффекта «реального времени» во время процесса моделирования.

3. Система должна иметь средства ввода исходных данных и визуализации результатов моделирования в удобном для пользователя виде.

Учитывая ключевые требования к программному комплексу, в качестве средства реализации системы визуализации и интерактивного моделирования выбрана облачная платформа IASPaas [14], которая представляет собой программно-информационный интернет-комплекс для обеспечения поддержки разработки, управления и удаленного использования прикладных и инструментальных мультиагентных облачных сервисов и их компонентов. Для реализации блока моделирования используется высокопроизводительный вычислительный кластер. Связь между платформой IASPaas и вычислительным кластером осуществляется через асинхронные http-запросы.

На рисунке 1 представлена обобщенная архитектура облачного сервиса. В общем случае сервис моделирования состоит из трех основных модулей: модуля моделирования торговых потоков, модуля управления и модуля визуализации.

2. Модуль моделирования торговых потоков

В данном разделе представлена математическая модель торговых потоков и ее программная реализация.

2.1. Математическая модель торговых потоков

Рассмотрим модель экономики k регионов, в каждом из которой имеется n продуктов. Пусть z_{ij}^{rm} – неизвестное, моделируемое количество продукта r -го типа, $r = 1, \dots, n$, поставляемого из i -го региона в j -й, $i, j = 1, \dots, k$, перевозимое видом

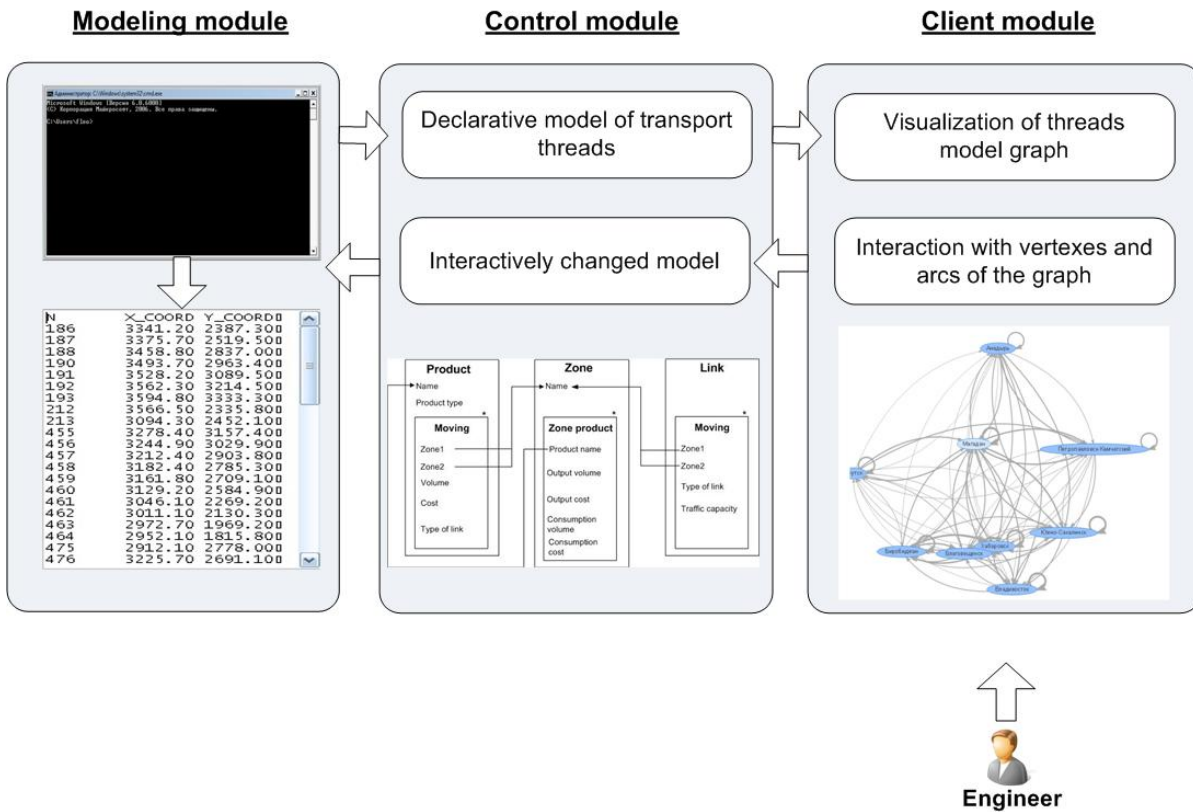


Рис. 1. Архитектура взаимодействия программных модулей сервиса интерактивного моделирования торговых потоков

Fig. 1. Interaction architecture of software modules for interactive trade flows service

транспорта $m = 1, 2, \dots, M$, где M – количество различных типов транспорта, используемых для перевозок. Здесь и далее верхние индексы соответствуют типам продукта и виду транспорта, нижние – регионам. Заметим, что поток z_{ii}^{rm} не обязательно предполагается равным нулю, что соответствует потреблению регионом продукта, перевозка которого не осуществляется.

Суммарный поток продукта r в j -й регион («полное потребление» продукта r в регионе j) различными видами транспорта неизвестен и равен $\sum_{m=1}^M \sum_{i=1}^k z_{ij}^{rm}$, что равно V_j^r – известному совокупному ввозу продукта r в регион j . Суммарный вывоз продукта r из региона i в остальные регионы («полное производство» продукта r в регионе i) с использованием всех видов транспорта равен $\sum_{m=1}^M \sum_{j=1}^k z_{ij}^{rm}$, что в свою очередь является известной из статистических данных величиной W_i^r .

Определяемое таким образом производство и потребление для каждого продукта $r = 1, \dots, n$ подчиняется очевидному для замкнутой системы уравнению баланса $\sum_{m=1}^M \sum_{i=1}^k \sum_{j=1}^k z_{ij}^{rm} = \sum_i V_j^r = \sum_j W_i^r$, что накладывает дополнительное ограничение на величины V_j^r и W_i^r .

Однако, поскольку система регионов не замкнута, то указанный выше баланс,

формируемый на основе статистических данных, не будет соблюдаться. Это означает, что существуют потоки продуктов между рассматриваемыми k регионами и некоторыми в общем случае неизвестными, «внешними» регионами по отношению к рассматриваемой системе. Задача осложняется тем, что и суммарный ввоз, и вывоз для таких «внешних» регионов являются неизвестными. Очевидно, что модель в этом случае требует модификации.

Для решения этой проблемы агрегируем подобные «внешние» регионы в регион $k + 1$ и будем рассматривать дополнительные неизвестные потоки $z_{i\ k+1}^{rm}$ и $z_{k+1\ j}^{rm}$.

Торговые потоки осуществляются экономическими агентами под воздействием платы за перевозку, которая во многом зависит от географического расстояния между регионами. Рассмотрим гравитационную модель транспортных издержек, которая может быть представлена соотношением $v_{ij}^{rm} = \exp(-d^{rm}T_{ij})$, где v_{ij}^{rm} – априорно заданный поток продуктов из i -го региона в j -й, T_{ij} – оценка географического расстояния между регионами i и j , а d^{rm} – параметры, отвечающие за чувствительность потока к расстоянию для данного продукта и вида используемого транспорта для его перевозки. Параметры d^{rm} неотрицательны, что означает меньшую величину потока между регионами i и j при большем значении расстояния между ними. В данном случае предполагается, что $v_{ii}^{rm} = 0$ и $T_{ii} = 0$ и $v_{ij}^{rm} = v_{ji}^{rm}$.

Калибровка неотрицательных параметров d^{rm} на фактических статистических данных представляет собой отдельную задачу прикладной статистики. Такая оценка проводится методами, применяемыми к модели регрессии, которая представима линейной по параметрам моделью $\ln v_{ij}^{rm} = \alpha - d^{rm}T_{ij} + \delta^{rm}$ для всех $i, j = 1, \dots, k$ и $i > j$, где δ^{rm} – нормально распределенные остатки регрессий для всех r и m .

В работах [3, 4, 10] рассматривается подход моделирования потоков в коммуникационной сети, соответствующий принципу наиболее вероятного значения распределения потоков системы в условиях неполноты информации, когда заданы лишь балансовые соотношения на потоки. Адаптация данного подхода для рассматриваемой нами модели торговых межрегиональных потоков приводит к необходимости минимизации нелинейной функции вида $\sum_{i,j=1,i \neq j}^{k+1} z_{ij}^r \ln(z_{ij}^{rm}/v_{ij}^{rm})$ по совокупности неизвестных потоков z_{ij}^{rm} .

Наличие такой функции приводит к необходимости задания строго положительных торговых потоков z_{ij}^{rm} , что моделируется заданием для потоков ограничений снизу малым наперед заданным параметром $\varepsilon > 0$.

Таким образом, для моделирования торговых потоков необходимо решать оптимизационную задачу с нелинейной целевой функцией и линейными ограничениями:

$$\sum_{i,j=1,i \neq j}^{k+1} z_{ij}^r \ln(z_{ij}^{rm}/v_{ij}^{rm}) \rightarrow \min_{\{z_{ij}^{rm}\}},$$

где $v_{ij}^{rm} = \exp(-\hat{d}^{rm}T_{ij})$, $\hat{d}^{rm}$ – известные оценки параметров и ограничения задачи имеют вид

$$\sum_{m=1}^M \sum_{i=1}^{k+1} z_{ij}^{rm} = V_j^r \text{ для всех } j = 1, 2, \dots, k \text{ и } r = 1, 2, \dots, n,$$

$$\sum_{m=1}^M \sum_{j=1}^{k+1} z_{ij}^{rm} = W_i^r \text{ для всех } i = 1, 2, \dots, k \text{ и } r = 1, 2, \dots, n,$$

$$z_{ij}^{rm} \geq \varepsilon > 0 \text{ для всех } i, j = 1, 2, \dots, k+1, r = 1, 2, \dots, n, m = 1, 2, \dots, M.$$

2.2. Программная реализация математической модели торговых потоков

В соответствии с описанной математической моделью реализовано программное средство, реализующее описанную модель. Программа [16] предназначена для нахождения равновесных межрегиональных грузопотоков в транспортной сети системы регионов на основе вышеописанной математической модели. Ее реализация выполнена на языке MPL [17] и Octave [18] на серверной платформе SuperMicro SuperServer 6037R-72RFT+. Данное программное средство позволяет рассчитать равновесные значения объемов грузопотоков между узлами транспортной сети.

Для применения программы используется следующая входная информация: статистические данные по совокупному ввозу и вывозу продуктов для регионов, расстояния между основными транспортными узлами регионов при осуществлении перевозок железнодорожным, автомобильным и/или морским транспортом.

Результатом выполнения программы являются рассчитанные равновесные значения объемов грузопотоков между транспортными узлами транспортной сети, матрицы межрегиональных потоков и матрицы «связанности» регионов как в разрезе отдельных разделов ОКВЭД, так и в целом по всем моделируемым разделам ОКВЭД.

Показателем «связанности» регионов i и j по определенному виду продукции называется величина $L_{ij} = \frac{2(x_{ij}+x_{ji})}{E_i+E_j+I_i+I_j}$, где x_{ij} – торговый поток из региона i в регион j , E_i – суммарный вывоз продукции из региона i во все остальные регионы, E_j – аналогично определяемая величина для региона j . Величины I_i, I_j – суммарный ввоз продукции в регион i и j соответственно из всех остальных регионов.

Программа обеспечивает вывод матриц вычисленных межрегиональных потоков и матриц показателей коэффициентов связности регионов по каждому товару и в агрегатном представлении по всем товарам.

3. Модуль управления

Для обеспечения удобного анализа и интерактивной обработки полученных результатов математических вычислений необходимо выполнить визуализацию полученных результатов и предоставить пользователю интерфейс для интерактивного моделирования различных ситуаций. Модуль моделирования торговых потоков и модуль визуализации являются полностью независимыми и внешними по отношению друг к другу программными средствами, находящимися на разных серверах. Для объединения этих модулей необходима дополнительная программа-посредник – модуль управления взаимодействием независимых внешних модулей. Поэтому вычисленные с помощью модуля моделирования матрицы передаются в модуль управления, расположенный на облачной платформе IASPaas.

Главная задача модуля управления – получать и передавать информацию от модуля моделирования и модуля визуализации в заранее определенном формате, понятном обеим сторонам взаимодействия. Формат данных, с которым «удобно»

работать модулю моделирования, – матричное представление; формат данных для визуализации – графовая структура. Оба эти формата являются узкоспециализированными и отражают специфику решаемых задач. Для обмена информацией между ними необходим промежуточный формат, который будет легко преобразовываться в специальный и наоборот. Для этой цели был разработан формат декларативной модели торговых потоков [19], состоящий из трех основных типов объектов: продукты, зоны и коммуникации. Структура каждого типа объекта и связи между ними представлены на рисунке 2.

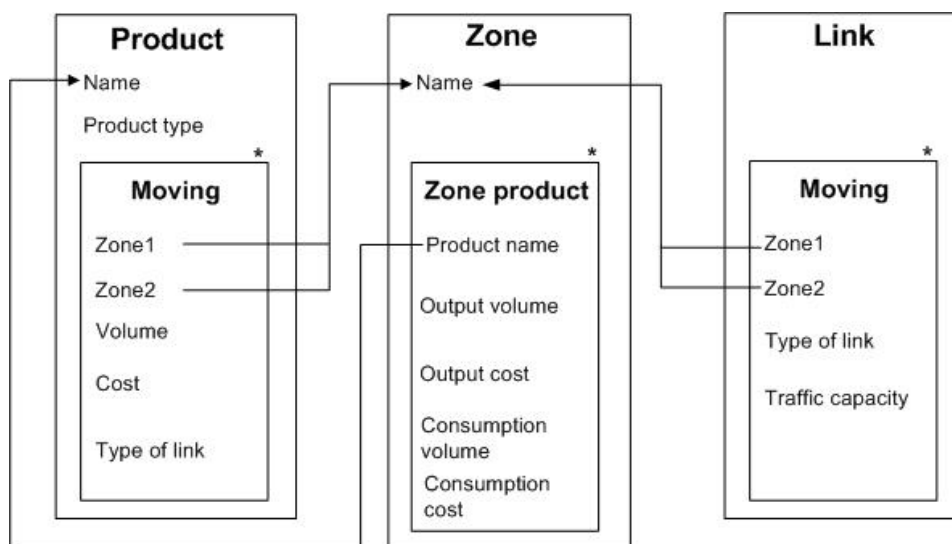


Рис. 2. Декларативная модель торговых потоков
 Fig. 2. The declarative model of trade flows

Формальное представление объекта «Продукты» выглядит следующим образом.

```

Продукты: [
{ // Продукт <i>
Имя: <Строковое значение>,
Тип продукции отрасли: <Строковое значение>
Движение: [
{ // Движение <i>
Зона1: <Строковое значение>,
Зона2: <Строковое значение>,
Объем нагрузки: <Числовое значение>,
Стоимость: <Числовое значение>,
Тип коммуникации: <Качественное значение>
},
},...
]
},
{ },...
],
    
```

Модуль управления работает на основе стандартного http-протокола и асинхронных запросов. Модуль визуализации отправляет асинхронный http-запрос модулю управления на получение входных данных от модуля моделирования. Модуль управления инициирует http-запрос на получение или расчет данных о состоянии торговых потоков. Модуль моделирования (произведя необходимые расчеты, если необходимо) формирует необходимые данные в декларативном представлении в вышеуказанном формате и затем отправляет ответ модулю визуализации через модуль управления. По полученным данным модуль визуализации строит граф торговых потоков. В процессе работы с сервисом модуль визуализации может послать новые запросы, содержащие измененную информацию о модели на сервер модуля моделирования.

4. Модуль визуализации

Модуль визуализации позволяет наглядно (графически) представить сложные матричные данные о возможных торговых потоках в виде ориентированного графа, в котором дуги являются непосредственно торговыми потоками, а вершины представляют точки назначения этих потоков.

Визуализированная транспортная сеть позволяет пользователю не только увидеть результаты математического моделирования, но и интерактивно изменять их, меняя параметры вершин и дуг с помощью программного интерфейса.

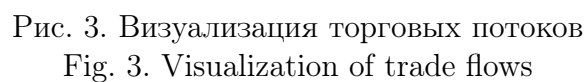
Модуль визуализации позволяет отображать различные варианты ориентированного графа в зависимости от нужных параметров. Для этого на основе декларативной модели автоматически формируются элементы управления графом торговых потоков [20]. С помощью элементов управления пользователь может указать, какой товар и какой тип коммуникации необходимо отобразить в необходимой выборке. Параметрическая выборка и визуализация только части информации имеет большое значение для наглядности получаемых результатов, особенно если части этих результатов являются полностью (частично) независимыми друг от друга (рисунок 3).

В визуализации графа торговых потоков используются дополнительные графические параметры: цвет и толщина дуг. Цвет характеризует тип коммуникации потоков. Толщина дуг характеризует объем загруженности потоков.

Интерактивное редактирование графа предполагает изменение параметров как вершин графа (рисунок 4), так и его дуг. При нажатии на соответствующий элемент графа пользователь получает структурированную информацию об объекте, которую может изменить. В соответствии с внесенными изменениями на сервер модуля моделирования через модуль управления будет отправлен динамический асинхронный запрос, в результате которого будет получена новая декларативная модель торговых потоков, и визуализация графа изменится.

Реализация модуля визуализации как облачного сервиса открывает удобную возможность использования и демонстрации математических результатов моделирования для многих пользователей через Интернет.

Для использования сервиса достаточно зайти на облачную платформу и получить доступ на его использование у администратора без установки дополнительных



модулей на компьютере пользователя. Важно, что размещение сервиса на облачной платформе дает возможность пользоваться им на любом компьютере независимо от его местоположения.

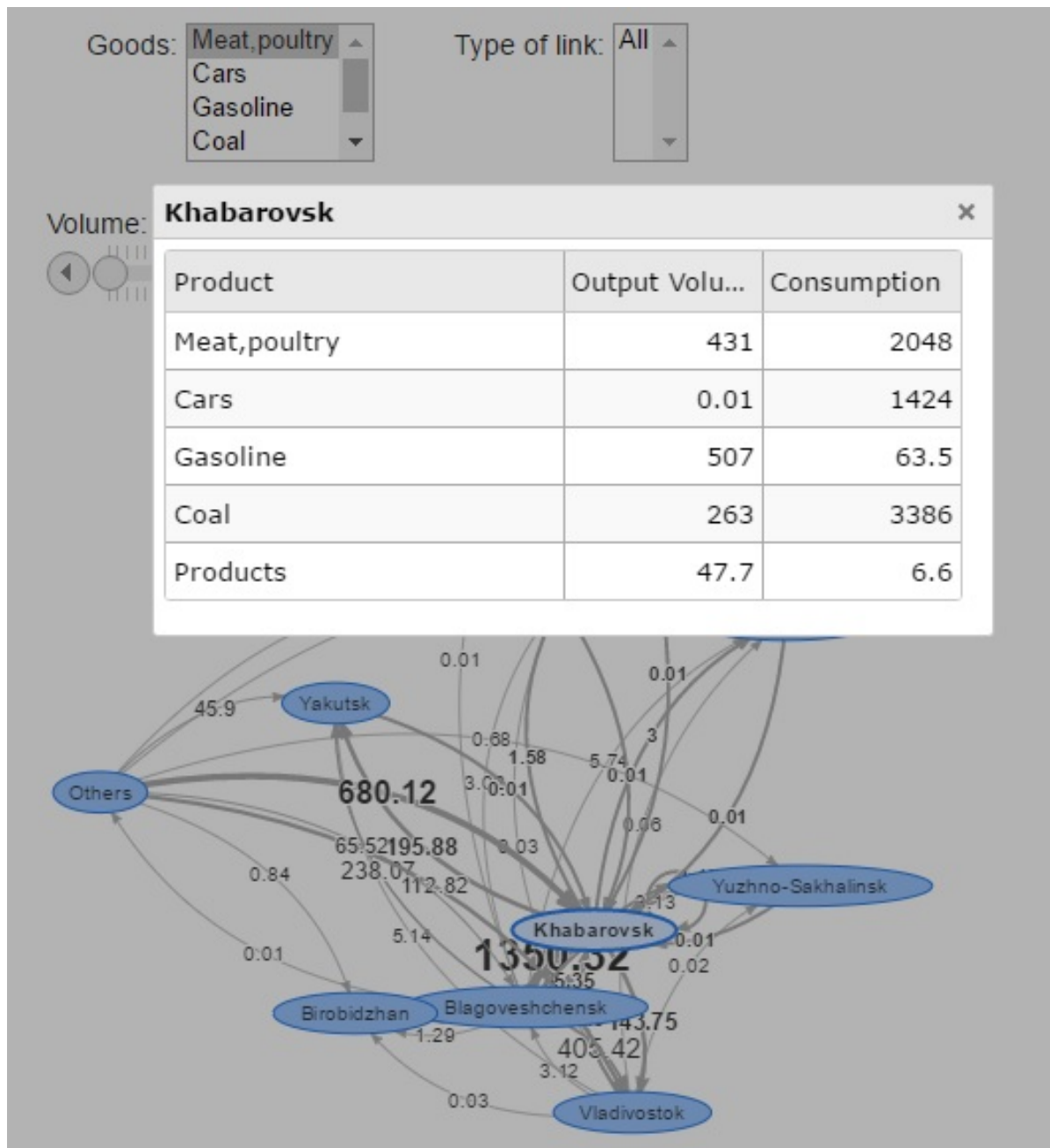


Рис. 4. Редактирование параметров модели
Fig. 4. Editing the parameters of the model

5. Визуализация и анализ решения

Программный комплекс для интерактивного моделирования межтерриториальной торговли применяется для определения равновесных межрегиональных грузопотоков в транспортной сети железнодорожного, автомобильного и морского транспорта регионов Дальнего Востока России. В качестве входных данных используется информация Росстата из официальных статистических справочников разных лет «Ре-

гионы России. Социально-экономические показатели» о межрегиональной торговле основными видами продукции (товаров): продовольственными товарами, топливом, товарами технического назначения и др.

В качестве пунктов ввоза и вывоза продукции рассматриваются административные центры 9 регионов Дальнего Востока: Приморский край (г. Владивосток), Хабаровский край (г. Хабаровск), Амурская область (г. Благовещенск), Еврейская автономная область (г. Биробиджан), Республика Саха - Якутия (г. Якутск), Магаданская область (г. Магадан), Сахалинская область (г. Южно-Сахалинск), Камчатский край (г. Петропавловск-Камчатский), Чукотский автономный округ (г. Анадырь). Оценки расстояний между регионами приведены в таблице 1, они соответствуют кратчайшим путям между административными центрами рассматриваемых регионов в транспортной сети железнодорожного, автомобильного и морского транспорта Дальнего Востока России.

Таблица 1. Оценки расстояний между регионами Дальнего Востока России, км
 Table 1. Estimates of the distances between the regions of the Russian Far East, km

Регионы Regions	1	2	3	4	5	6	7	8	9
1	0	762	1410	938	3314	2490	990	2490	4490
2	762	0	779	176	2552	2534	1034	2534	4534
3	1410	779	0	603	2035	3182	1813	3182	5182
4	938	176	603	0	2376	2710	1210	2710	4710
5	3314	2552	2035	2376	0	1736	3586	2736	4736
6	2490	2534	3182	2710	1736	0	1500	1000	3000
7	990	1034	1813	1210	3586	1500	0	1500	3500
8	2490	2534	3182	2710	2736	1000	1500	0	2000
9	4490	4534	5182	4710	4736	3000	3500	2000	0

Примечание:

Приморский край – 1, Хабаровский край – 2, Амурская область – 3, Еврейская автономная область – 4, Республика Саха (Якутия) – 5, Магаданская область – 6, Сахалинская область – 7, Камчатский край – 8, Чукотский автономный округ – 9.

В таблице 2 приведен результат моделирования перевозок угля, как решение задачи определения наиболее вероятных перевозок товаров в системе дальневосточных регионов и их торговли с другими регионами. Внешние по отношению к системе рассматриваемых регионов территории агрегированы в регион, который в таблице указан последней строкой (столбцом) под номером «10».

Таблица 2. Результаты моделирования. Перевозки угля, тыс. тонн
Table 2. The simulation results. Transportation of coal, thous. tons

Регионы Regions	1	2	3	4	5	6	7	8	9	10
1	0	5,53	2,29	0,78	0,72	1,45	0	0,78	0,36	0
2	100,69	0	44,63	5,18	19,6	52,13	0	27,98	12,79	0
3	172,49	184,94	0	16,44	14,47	60,61	0	32,53	13,53	0
4	0	0	0	0	0	0	0	0	0	0
5	1133	1693,23	301,64	180,99	0	92,41	0,01	78,16	34,55	0,01
6	20,95	41,37	11,61	5,08	0,85	0	0	0,7	0,54	0
7	12,86	26,08	10,21	3,5	2,71	3,03	0	1,63	0,97	0
8	0	0,01	0	0	0	0	0	0	0	0
9	33,11	64,88	16,57	7,74	2,03	3,44	0	1,23	0	0
10	705,91	1369,96	306,05	157,29	41,03	109,91	0	58,99	15,06	0

Примечание:

Приморский край – 1, Хабаровский край – 2, Амурская область – 3, Еврейская автономная область – 4, Республика Саха (Якутия) – 5, Магаданская область – 6, Сахалинская область – 7, Камчатский край – 8, Чукотский автономный округ – 9, другие регионы – 10.

Визуальное представление данных таблицы 2 приведено на рисунке 5.

В настоящее время моделирование производится по 25 товарам различного назначения: продовольственные товары, топливо, товары технического назначения и другие. Результат моделирования определяет вероятные перевозки товаров между регионами Дальнего Востока России. В условиях неполноты статистических данных использование математического моделирования позволило получить интерпретируемые результаты, имеющие экономический смысл и прогноз наиболее вероятного распределения потоков товаров между регионами Дальнего Востока России.

Представленный подход позволяет выяснить характер взаимосвязей экономик регионов Дальнего Востока России в условиях существующей пространственной структуры размещения производств. Эти взаимосвязи сопровождаются пространственной неравномерностью и сильной дифференциацией по продуктам. Новая структура взаимосвязей экономик регионов возможна при изменении параметров экономического пространства (транспортная, энергетическая инфраструктура; пространственная структура производства и расходования валового регионального продукта).

6. Заключение

В работе описан программный комплекс, который предназначен для интерактивного моделирования межтерриториальной торговли. Сервис моделирования является облачным и состоит из трех основных модулей: модуля моделирования торговых потоков, модуля управления и модуля визуализации. В основе модуля моделирования торговых потоков лежит математическая модель торговых потоков. Учитывая

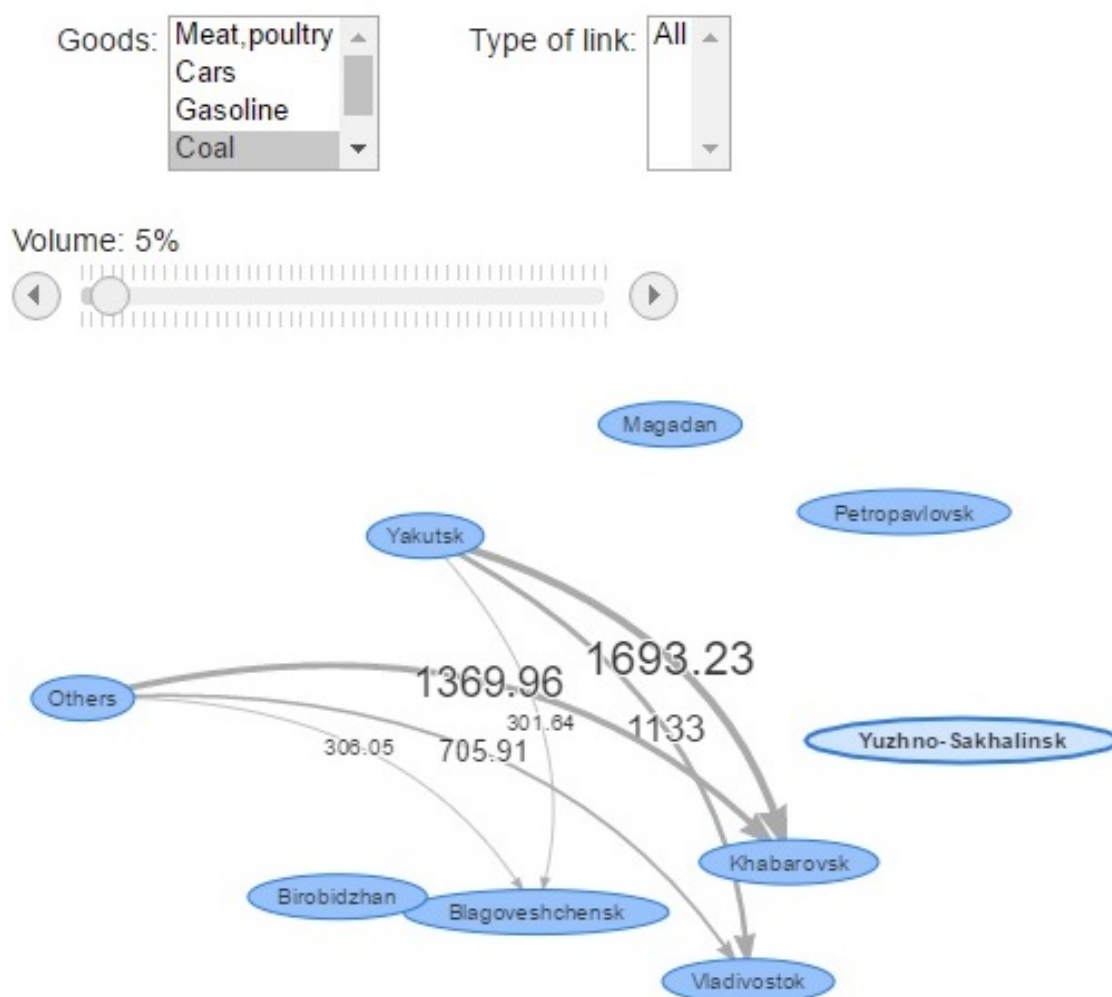


Рис. 5. Торговые потоки угля
 Fig. 5. Trade flows of coal

большую вычислительную сложность процесса моделирования, модуль моделирования реализован на высокопроизводительной вычислительной архитектуре, модули управления и визуализации реализованы на облачной платформе IASPaas с использованием мультиагентного подхода. Взаимодействие между вычислительным кластером и платформой IASPaas осуществляется по http-протоколу на основе динамических асинхронных вызовов.

Программный комплекс предназначен для специалистов различных министерств и ведомств, занимающихся проблемой оптимизации и планирования межтерриториальных потоков продуктов отраслей экономики с учетом их многопродуктовости в многомодальных транспортных сетях, и может быть использован для моделирования межтерриториальной торговли в различных регионах России.

Список литературы / References

- [1] Леонтьев В. В., *Избранные произведения в 3-х тт*, **1**, Общеэкономические проблемы межотраслевого анализа, Экономика, М., 2006; [Leont'ev V. V., *Izbrannyye proizvedeniya v 3-kh tt*, **1**, Obshcheekonomicheskie problemy mezhotraslevogo analiza, Ekonomika, М., 2006, (in Russian).]
- [2] Anderson J. E., Wincoop E., "Gravity with gravitas: a solution to the border puzzle", *The American Economic Review*, **93**:1 (2003), 170–192.
- [3] Вильсон А. Дж., *Энтропийные методы моделирования сложных систем*, Наука, М., 1978; [Vil'son A. Dzh., *Entropiynye metody modelirovaniya slozhnykh sistem*, Nauka, М., 1978, (in Russian).]
- [4] Fang S. C., Rajasekara J. R., Tsao H. S. J., *Entropy Optimization and Mathematical Programming*, Kluwer Academic, Dordrecht, 1997.
- [5] Batten D. F., Boyce D. E., "Spatial interaction and interregional commodity flow models", *Handbook on regional and urban economics*, **1**, 1987, 357–406.
- [6] Ham H., Kim T. J., Boyce D., "Implementation and estimation of a combined model of interregional, multimodal commodity shipments and transportation network flows", *Transportation Research. Part B*, 2005, № 39, 65–79.
- [7] Ham H., Kim T. J., Boyce D., "Assessment of economic impacts from unexpected events with an interregional commodity flow and multimodal transportation network model", *Transportation Research. Part A*, 2005, № 39, 849–860.
- [8] Васильева Е. М., Левит Б. Ю., Лившиц В. Н., *Нелинейные транспортные задачи на сетях*, Транспорт, М., 1981; [Vasil'eva E. M., Levit B. Yu., Livshits V. N., *Nelineynye transportnye zadachi na setyakh*, Transport, М., 1981, (in Russian).]
- [9] Васильева Е. М., Игудин Р. В., Лившиц В. Н., *Оптимизация планирования и управления транспортными системами*, Транспорт, М., 1987; [Vasil'eva E. M., Igudin R. V., Livshits V. N., *Optimizatsiya planirovaniya i upravleniya transportnymi sistemami*, Transport, М., 1987, (in Russian).]
- [10] Попков Ю. С. и др., *Системный анализ и проблемы развития городов*, Наука, М., 1983; [Popkov Yu. S. et al., *Sistemnyy analiz i problemy razvitiya gorodov*, Nauka, М., 1983, (in Russian).]
- [11] Величко А. С., Давыдов Д. В., "Интервальная энтропийная модель межрегионального производственного баланса", *Пространственная экономика*, **3**:3 (2009), 20–35; [Velichko A. S., Davydov D. V., "Intervalnaya entropiynaya model mezhregionalnogo proizvodstvennogo balansa", *Prostranstvennaya ekonomika*, **3**:3 (2009), 20–35, (in Russian).]
- [12] Zhang Q., Cheng L., Boutaba R., "Cloud computing: state-of-the-art and research challenges", *J. Internet Serv. Appl.*, **1** (2010), 7–18.
- [13] Sun L. et al., "Cloud service selection: State-of-the-art and future research directions", *Journal of Network and Computer Applications*, **45** (2014), 134–150.
- [14] Gribova V. et al., "A Cloud Computing Platform for Lifecycle Support of Intelligent Multi-agent Internet-services", International Conference on Power Electronics and Energy Engineering (PEEE) (Hong Kong, China, April 19–20, 2015), 231–235.
- [15] Грибова В. В. и др., "Базовая технология разработки интеллектуальных сервисов на облачной платформе IACPaaS. Часть 1. Разработка базы знаний и решателя задач", *Программная инженерия*, **12** (2015), 3–11; [Gribova V. V. et al., "A Base Technology for Development of Intelligent Services with the Use of IACPaaS Cloud Platform. Part 1. A Development of Knowledge Base and Problem Solver", *Software Engineering*, **12** (2015), 3–11, (in Russian).]
- [16] Величко А. С., "Вычисление равновесных межрегиональных потоков товаров в условиях интервальной неопределенности с использованием метода перекрестной энтропии", свидетельство о госрегистрации ПрЭВМ РФ №2011613757 от 13.05.11, *RU ОБПБТ*, **3(76)** (2011), 256–257; [Velichko A. S., "Vychislenie ravnovesnykh

mezhhregionalnykh potokov tovarov v usloviyakh intervalnoy neopredelennosti s ispolzovaniem metoda perekrestnoy entropii”, sv. o gos. reg. prog. dlya EVM Ross. Fed. №2011613757 ot 13.05.11, *RU OBPBT*, **3(76)** (2011), 256–257, (in Russian).]

- [17] *AMPL: A Modeling Language for Mathematical Programming*, <http://www.ampl.com>.
- [18] *GNU Octave: High-level language for numerical computations*, <https://www.gnu.org/software/octave/>.
- [19] Gribova V. V. et al., “A Two-level Model of Information Units with Complex Structure that Correspond to the Questioning Metaphor”, *Automatic Documentation and Mathematical Linguistics*, **49:5** (2015), 172–181.
- [20] Gribova V., Fedorishchev L., “Cloud service for development of virtual interactive environments”, International Conference on Advanced ICT for Education (ICAICTE) (Dalian, China, August 16–17, 2014), 20–23.

Velichko A. S., Gribova V. V., Fedorishchev L. A., "Cloud Service for Interactive Simulation of Interregional Trade", *Modeling and Analysis of Information Systems*, **23:4** (2016), 412–426.

DOI: 10.18255/1818-1015-2016-4-412-426

Abstract. The paper describes a mathematical model of trade flows between the territories of a region or a country in a transport network having one or more different types of marine or ground transportation. We use the approach of modeling complex communication systems to determine the most probable values of flows in case of incomplete information about the system. Transport costs between the territories are modeled within the framework of the gravity model. The payment for transportation depends on the distance between regions, the distance is estimated as the shortest way length in a given transport network or geographical distance. The mathematical formulation of the problem belongs to the class of convex mathematical programming problems and assumes the numerical solution of nonlinear optimization problem with linear constraints. Based on the model, the software is implemented as a cloud service on heterogeneous computing architectures: the simulation module is made on a high-performance server platform, management and visualization modules are produced with IACPaaS cloud platform. Communication between the platforms is established via asynchronous http-queries. For information exchange between the modules the declarative model with JSON format is developed and implemented for the objects considered in the mathematical model which are products, areas and communications. Visualization module allows to present graphically the original and the resulting matrix data and to modify the input parameters of the model interactively. The paper demonstrates the use of software for the simulation of inter-regional freight traffic of the Russian Far East region based on input data provided by open statistics sources.

Keywords: interregional, transportation, network, multimodal, gravity, entropy, equilibrium, model

On the authors:

Velichko Andrei Sergeevich, orcid.org/0000-0001-8760-316X, PhD, researcher, Institute for Automation and Control Processes FEB RAS, 5, Radio Str., Vladivostok, 690041, Russia, e-mail: vandre@dvo.ru

Gribova Valeriya Viktorovna, orcid.org/0009-0203-2514-7755, DSc, laboratory head, Institute for Automation and Control Processes FEB RAS, 5, Radio Str., Vladivostok, 690041, Russia, e-mail: gribova@iacp.dvo.ru

Fedorishchev Leonid Aleksandrovich, orcid.org/0009-0203-2514-7755, PhD, researcher, Institute for Automation and Control Processes FEB RAS, 5, Radio Str., Vladivostok, 690041, Russia, e-mail: fleo1987@mail.ru

Acknowledgments:

The work is partially supported by the RFBR grant No.16-07-00340-a.

©Марьясин О.Ю., Колодкина А.С., Огарков А.А., 2016

DOI: 10.18255/1818-1015-2016-4-427-439

УДК 628.87

Компьютерное моделирование "Интеллектуального здания"

Марьясин О.Ю., Колодкина А.С., Огарков А.А.

получена 15 марта 2016

Аннотация.

В настоящее время в развитых странах активно развивается технология "Интеллектуальное здание", или "Умный дом" ("Smart building"). Концепция "Интеллектуального здания", или "Умного дома", состоит в том, что система должна уметь распознавать конкретные ситуации, происходящие в доме, и соответствующим образом на них реагировать. Для автоматизированного контроля и управления, а также организации взаимодействия между отдельными системами инженерного оборудования организуется автоматизированная система управления зданием, в которую, в виде отдельных составляющих, входят подсистемы автоматизации того или иного инженерного оборудования. Для проведения исследований различных режимов функционирования работающих в здании инженерных подсистем и всей системы в целом необходимо использовать математическое и компьютерное моделирование. С точки зрения математического описания "Интеллектуальное здание" представляет собой непрерывно-дискретную или гибридную систему, состоящую из взаимодействующих элементов различной природы, поведение которых описывается как непрерывными, так и дискретными процессами. В данной статье авторы предлагают компьютерную модель "Интеллектуального здания", позволяющую моделировать работу основных инженерных подсистем здания и алгоритмов управления этими подсистемами. Модель построена в системе Simulink пакета Matlab, с использованием библиотеки "физического моделирования" Simscape и библиотеки Stateflow. Особенность данной модели заключается в использовании специализированных алгоритмов контроля и управления, позволяющих обеспечить скоординированное взаимодействие подсистем и оптимизировать энергопотребление.

Ключевые слова: "Интеллектуальное здание", АСУЗ, гибридная система, компьютерная модель, Simulink, Stateflow

Для цитирования: Марьясин О.Ю., Колодкина А.С., Огарков А.А., "Компьютерное моделирование «Интеллектуального здания»", *Моделирование и анализ информационных систем*, **23**:4 (2016), 427–439.

Об авторах:

Марьясин Олег Юрьевич, orcid.org/0000-0001-7449-518X, канд. тех. наук, доцент,
Ярославский государственный технический университет,
Московский пр., 88, г. Ярославль, 150023 Россия, e-mail: maryasinou@ystu.ru

Колодкина Анна Сергеевна, orcid.org/0000-0001-7736-5372, преподаватель,
Ярославский промышленно-экономический колледж,
ул. Гагарина, 8, г. Ярославль, 150023 Россия, e-mail: anuta_25@mail.ru

Огарков Андрей Алексеевич, orcid.org/0000-0002-4151-0379, инженер,
Компания «Интегратор»,
ул. Чехова, 2, г. Ярославль, 150054 Россия, e-mail: drivemox@mail.ru

Введение

Современное здание, независимо от своего типа – жилой ли это дом, государственное или общественное учреждение, производственный корпус – содержит большой объем инженерного оборудования. Причем доля инженерного оборудования в общей стоимости здания непрерывно растет. Все это происходит по той причине, что с каждым годом неуклонно повышаются представления о комфортности и безопасности пребывания человека в здании.

В настоящее время поддержанием в здании требуемых санитарно-гигиенических условий, обеспечением его безопасности и защищенности от нештатных ситуаций занимаются множество подсистем инженерного оборудования, каждая из которых характеризуется достаточно большим набором контролируемых технологических параметров и сигналов управления. Основными из них являются:

- система управления отоплением, включая автоматизированный тепловой пункт;
- система управления вентиляцией и кондиционированием;
- система управления энергоснабжением, включая систему освещения;
- система охранно-пожарной сигнализации;
- система контроля доступа;
- телефонная и информационная сеть здания.

В совокупности все они образуют то, что называется системой жизнеобеспечения здания.

Традиционные решения инженерного оборудования здания представляют собой совокупность отдельных, не взаимодействующих между собой (автономных) систем. Здание, в котором эти системы объединены в единый интегрированный комплекс, "имеет право" называться "интеллектуальным" [1]. Понятие "Интеллектуальное здание" ("Умный дом") родилось в США в начале 1980-х годов и очень быстро стало модным. Если отвлечься от рекламных лозунгов, то "интеллектуальным" следует называть здание, оснащенное средствами автоматического контроля над всеми системами жизнеобеспечения [2]. Концепция "Интеллектуального здания", или "Умного дома", состоит в том, что система должна уметь распознавать конкретные ситуации, происходящие в доме, и соответствующим образом на них реагировать [3].

Для автоматизированного контроля и управления, а также организации взаимодействия между отдельными системами инженерного оборудования, организуется автоматизированная система управления зданием (АСУЗ), в которую, в виде отдельных составляющих, входят подсистемы автоматизации того или иного инженерного оборудования. Система способна за счет полной информации от всех эксплуатируемых подсистем, будь то охранно-пожарная, электропитание, водоснабжение, отопление, кондиционирование и т. д., принять правильное решение и выполнить соответствующее действие, проинформировать жильцов или соответствующую службу о событии.

Определение допустимой или тем более оптимальной структуры, состава и значений характеристик такой сложной системы, как "Интеллектуальное здание", невозможно без проведения исследований различных режимов функционирования входящих в ее состав подсистем и всей системы в целом. Так как проведение натурных экспериментов на действующей системе является весьма затратным, а часто вообще невозможным, то для этого необходимо использовать математическое и компьютерное моделирование. В настоящее время существуют компьютерные системы и программы, позволяющие моделировать или отдельные, или принципиально схожие инженерные подсистемы и их системы управления. Например, для моделирования систем отопления, вентиляции и кондиционирования (ОВК, или HVAC, как их называют на западе) используются такие специализированные пакеты, как TRNSYS [4], ESP-r [5], IDA ICE [6] и др. Однако эти системы не позволяют проводить моделирование функционирования всего комплекса основных инженерных подсистем здания.

В данной статье авторы предлагают компьютерную модель "Интеллектуального здания", позволяющую моделировать работу основных инженерных подсистем здания и алгоритмов управления этими подсистемами. Особенность данной модели заключается в использовании специализированных алгоритмов контроля и управления, позволяющих обеспечить скоординированное взаимодействие подсистем и оптимизировать энергопотребление.

1. Гибридные и реактивные системы и средства для их моделирования

С точки зрения математического описания "Интеллектуальное здание" представляет собой непрерывно-дискретную или гибридную систему, состоящую из взаимодействующих элементов различной природы, поведение которых описывается как непрерывными, так и дискретными процессами. Общее поведение гибридной системы описывается последовательностью локальных поведений отдельных элементов, смена которых происходит под воздействием событий. Наступление того или иного события может зависеть от поступления внешних сигналов, может быть привязано ко времени или зависеть от значений различных параметров и, следовательно, от функций локальных поведений. Одно событие может порождать другие, и сам процесс функционирования системы описывается в общем случае сложным, нетривиальным алгоритмом [7].

Примерами непрерывно-дискретных систем могут служить транспортные системы, системы управления физическими процессами с аварийными ситуациями, системы, состоящие из параллельных взаимодействующих динамических подсистем.

Поведение гибридной системы можно, таким образом, представить бесконечной последовательностью сменяющих друг друга или протекающих параллельно непрерывных и мгновенных локальных процессов. Эти процессы могут быть независимыми друг от друга или конкурировать за общие ресурсы и/или взаимодействовать между собой, выполняя общую задачу. Динамические системы, имеющие такое поведение, еще называют реактивными системами [7].

К классу реактивных систем относят операционные системы, мультипрограмм-

ные среды, параллельные и распределенные дискретные системы, дискретные системы реального времени, вычислительные сети.

Реактивную систему и ее окружение удобно представлять в виде параллельной системы, в которой компоненты функционируют параллельно, взаимодействуя друг с другом. Глобальное поведение параллельной дискретной системы представляется множеством последовательностей смены ее глобальных состояний во времени, которые можно привязать либо к событиям, вызывающим смену состояний, либо к тактам времени.

Гибридной системой в теории реактивных систем называется система, которая в классических задачах моделирования и анализа называлась бы непрерывно-дискретной или событийно-управляемой.

Специфика подобных систем заключается, с одной стороны, в сложности динамики каждого элемента системы, а с другой – в возможности возникновения разного рода событий, которые по-разному в зависимости от момента их наступления мгновенно изменяют динамику элементов и структуру системы в целом.

В настоящее время для моделирования дискретной динамики реактивных систем широко используется предложенный Д. Харелом визуальный формализм – Statechart (диаграммы состояний и переходов) [8]. Основные неграфические компоненты таких диаграмм – это событие и действие, основные графические компоненты – состояние и переход [9].

Событие – нечто, происходящее вне рассматриваемой системы, возможно требуя некоторых ответных действий. События могут быть вызваны поступлением некоторых данных или некоторых задающих сигналов со стороны человека или некоторой другой части системы. События считаются мгновенными (для выбранного уровня абстрагирования).

Действия – это реакции моделируемой системы на события. Подобно событиям, действия принято считать мгновенными.

Состояние – условия, в которых моделируемая система пребывает некоторое время, в течение которого она ведет себя одинаковым образом. В диаграмме состояния часто представляются кругами или прямоугольными полями со скругленными углами.

Переход – изменение состояния, обычно вызываемое некоторым значительным событием. Как правило, состояние соответствует промежутку времени между двумя такими событиями. Переходы показываются в диаграммах линиями со стрелками, указывающими направление перехода.

Каждому переходу могут быть сопоставлены условия, при выполнении которых переход осуществляется.

С каждым переходом и каждым состоянием могут быть соотнесены некоторые действия. Действия могут дополнительно обозначаться как действия, выполняемые однократно при входе в состояние; действия, выполняемые многократно внутри некоторого состояния; действия, выполняемые однократно при выходе из состояния т.д.

Диаграммы состояний и переходов в настоящее время широко используются для моделировании сложных систем. Достаточно упомянуть унифицированный язык моделирования (Unified Modeling Language (UML)) [10], одним из элементов кото-

рого являются диаграммы состояний и язык "Графсет" [11], который используется при программировании логических контроллеров систем автоматизации.

Существенно повышает степень наглядности модели использование анимации, которая отображает изменения в системе, сопровождающиеся переходами от одного состояния к другому. Построение таких имитационных моделей возможно с использованием различных современных инструментальных средств, одним из которых является программный пакет AnyLogic [12].

Система AnyLogic – это среда компьютерного моделирования общего назначения, охватывающая основные направления моделирования: дискретно-событийное, системной динамики, агентное. Графическая среда моделирования AnyLogic поддерживает проектирование, разработку и выполнение компьютерных экспериментов с моделью, включая различные виды анализа [13]. AnyLogic также включает в себя возможность создания интерактивной анимации для улучшения наглядности моделей. Основным недостатком системы AnyLogic является отсутствие компонентов для построения непрерывных систем, описываемых в виде передаточных функций или уравнений состояния, и отсутствие компонентов для выполнения "физического моделирования".

Построение моделей гибридных, реактивных систем возможно также с помощью перспективного языка Modelica и систем "физического" моделирования, поддерживающих данный язык [14]. Среди них есть как коммерческие системы, такие как Dymola [15], MapleSim [16], Wolfram SystemModeler [17], так и свободно-доступные OpenModelica [18], JModelica [19] и др. На основе языка Modelica разработано большое число свободно-доступных библиотек компонентов, одной из которых является библиотека Modelica_StateGraph2, включающая компоненты для моделирования дискретных событий, реактивных и гибридных систем с помощью иерархических диаграмм состояния [20]. Основным недостатком систем на базе языка Modelica является отсутствие достаточной информации, в особенности на русском языке, как по системам моделирования, поддерживающим данный язык, так и по отдельным библиотекам компонентов.

Другим инструментом моделирования реактивных систем является пакет MATLAB [21]. MATLAB обеспечивает доступ к различным типам данных, высокоуровневому программированию и включает систему Simulink, которая поддерживает проектирование непрерывных и дискретных динамических систем в графической среде (в виде блок-схем). Система Simulink в свою очередь включает библиотеку Stateflow, которая является интерактивным инструментом проектирования сложных, событийно-управляемых систем [22]. Stateflow-диаграммы, использующие визуальный формализм Д. Харела, включаются в Simulink-модели, чтобы обеспечить возможность моделирования процессов, управляемых событиями. Stateflow обеспечивает ясное описание поведения сложных систем, используя диаграммы состояний и переходов. Дополнительная возможность следить в режиме реального времени за процессом выполнения диаграммы путем включения режима анимации делает процесс моделирования реактивных систем по-настоящему наглядным.

Комбинация MATLAB-Simulink-Stateflow является мощным универсальным инструментом моделирования реактивных систем. Поэтому она была выбрана авторами для разработки компьютерной модели "Интеллектуального здания".

2. Компьютерная модель системы

Разработанная авторами компьютерная модель "Интеллектуального здания", включает в себя следующие подсистемы: сбора и обработки информации от датчиков и выдачи сигналов исполнительным устройствам, автоматического управления освещением, автоматического управления микроклиматом в помещении, управления системами безопасности (системами охранной, пожарной сигнализации, обнаружения протечки воды и утечки газа), управления электроснабжением, учета энергоресурсов.

Модель построена в системе Simulink пакета Matlab, с использованием библиотеки "физического моделирования" Simscape [23] и библиотеки Stateflow. Использование Simscape позволяет создавать модели, состоящие из блоков, моделей физических компонентов, таких как электродвигатели, насосы, гидравлические клапаны, емкости и т.д. Применение компонентов Simscape по сравнению с обычными компонентами системы Simulink дает возможность сделать модель более компактной, легко читаемой, контролировать размерности сигналов. Использование Stateflow дает возможность моделировать поведение гибридных или сложных динамических, событийно-управляемых (реактивных) систем, базируясь на картах состояния Хarelла.

В состав разработанной модели входят блоки Simulink и Simscape, диаграммы и таблицы истинности Stateflow. Таблицы истинности позволяют задавать логические функции в наглядной табличной форме. Сложные модели отдельных подсистем представляются в виде маскированных подсистем блок-диаграммы Simulink, что повышает наглядность модели, удобство в работе и позволяет придать моделям более простую, иерархическую структуру.

Система Simulink обладает большими возможностями как при моделировании различных возмущений (детерминированных или случайных с различными законами распределения), действующих на систему, так и при моделировании всевозможных штатных или нештатных ситуаций, возникающих в реальной жизни.

Например, можно, смоделировать ситуацию, когда при возникновении пожара "срабатывают" извещатели системы пожарной сигнализации. При этом формируются сигналы для управления другими системами интегрированного комплекса:

- система управления микроклиматом выключает приточную систему вентиляции, обслуживающую данную зону, чтобы предотвратить поступление свежего воздуха к очагу возгорания. Для удаления дыма из коридоров, холлов, лестниц (вдоль маршрутов эвакуации) включается соответствующая подсистема дымоудаления (открываются заслонки, включаются вентиляторы);
- система управления электроснабжением отключает цепи электропитания вблизи зоны пожара;
- система управления освещением включает аварийное освещение;
- система управления доступом разблокирует двери для беспрепятственной эвакуации людей и т.д.

В рамках одной статьи трудно описать все подсистемы реализованной компьютерной модели. Поэтому далее мы ограничимся описанием только двух подсистем для

одного помещения здания. На основе этих двух подсистем мы покажем заложенные в них алгоритмы управления и возможности координации с другими инженерными подсистемами.

3. Модель управления электрическим освещением помещения

Рассмотрим фрагмент модели системы "Интеллектуального здания", включающий систему управления электрическим освещением помещения, представленный на рис. 1.

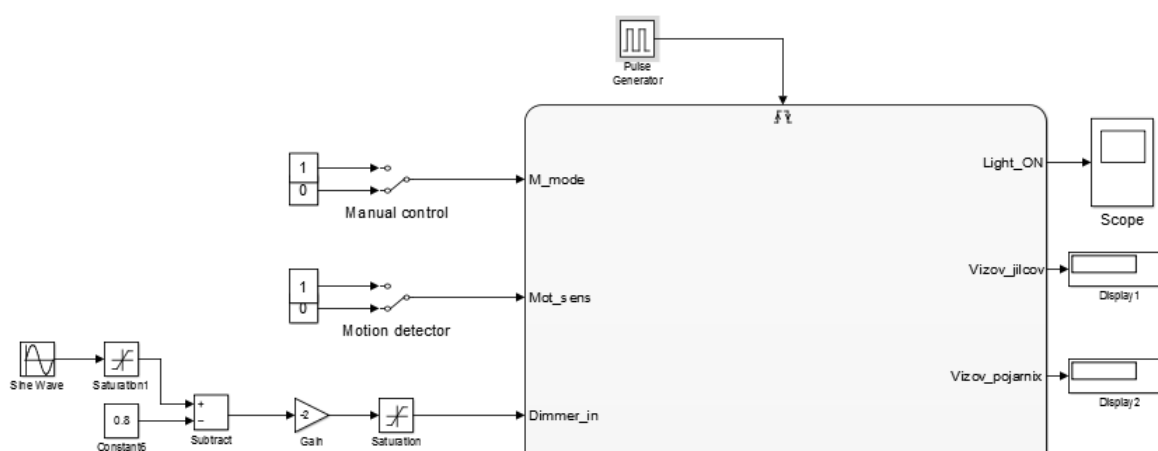


Рис. 1. Фрагмент модели, включающий систему управления освещением помещения
Fig. 1. Fragment of the model, including lighting facilities management system

На вход диаграммы состояний подаются различные сигналы с датчиков, установленных в доме. Для системы управления электрическим освещением это сигнал с датчика движения, сигнал с датчика освещенности, а также сигнал с кнопки ручного включения освещения. В зависимости от состояния этих сигналов, а также за счет наличия в системе устройств регулирования электрической мощности светильников (диммеров), на выходе имеем регулируемую яркость ламп. Поэтому освещенность всегда поддерживается на комфортном для человека уровне при минимизации энергозатрат.

Алгоритм, в соответствии с которым работает система, закладывается в диаграмме состояний, фрагмент которой представлен на рис. 2.

При появлении людей в помещении свет включается автоматически, а при их отсутствии гаснет. Для исключения частых включений и отключений ламп реализована настраиваемая задержка по времени отключения после исчезновения движения, поэтому срок службы лампы не уменьшается. Данная функция особенно полезна для проходных помещений, таких как коридоры, лестничные пролеты.

Световой поток от лампы регулируется пропорционально уровню естественной освещенности, в темное время суток он максимален, в светлое время суток минимален, а при достаточной естественной освещенности лампы полностью отключаются.

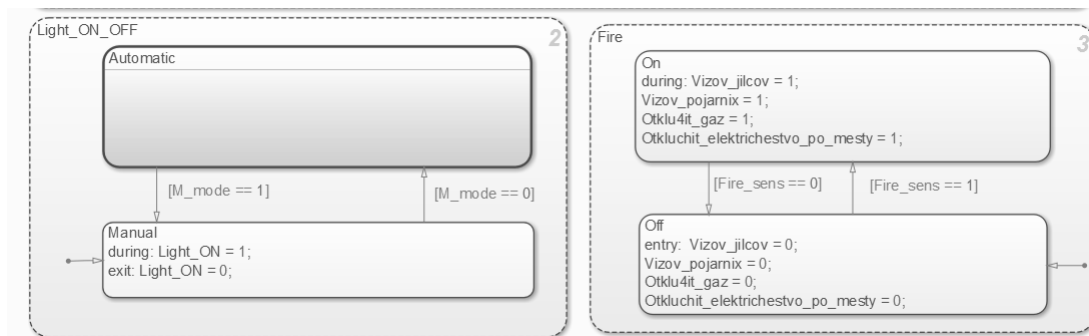


Рис. 2. Фрагмент диаграммы состояний управления электрическим освещением

Fig. 2. Fragment diagram of electric lighting control states

4. Модель системы управления микроклиматом в помещении

На рис. 3 показана модель системы управления микроклиматом в помещении, интегрированная в общую модель системы.

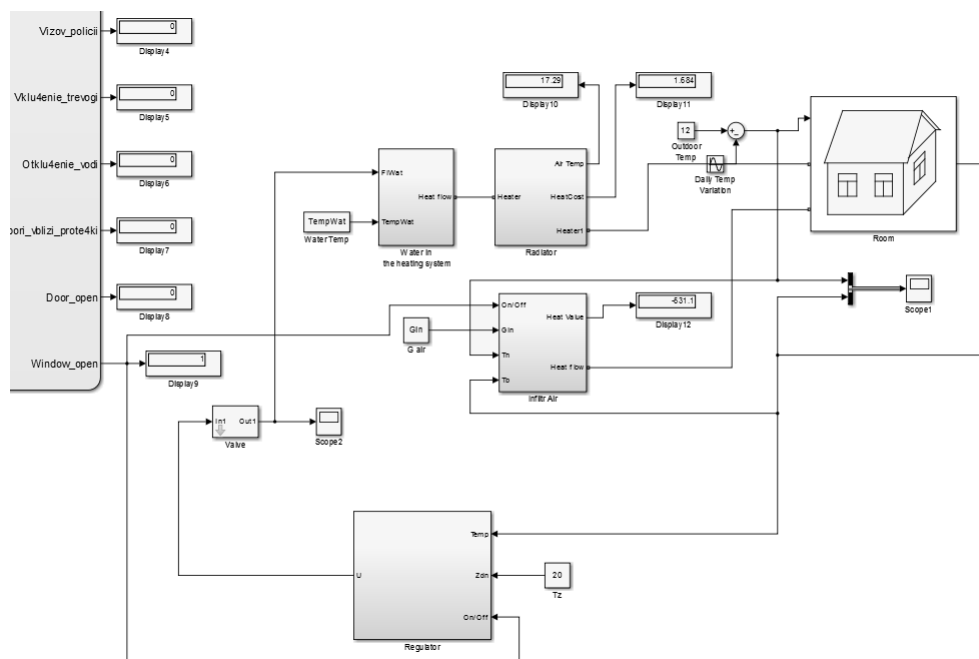


Рис. 3. Модель системы управления микроклиматом в помещении

Fig. 3. Model systems for climate control in a room

Подсистема Room содержит блок-диаграмму модели теплового режима помещения, построенную с использованием теплотехнических компонентов Thermal библиотеки Simscape и показанную на рис. 4.

Основу модели составляют три параллельные цепочки элементов, представляющие передачу тепла через стены, окна и потолок (крышу). При необходимости можно добавить подобные цепочки для моделирования теплопередачи через пол

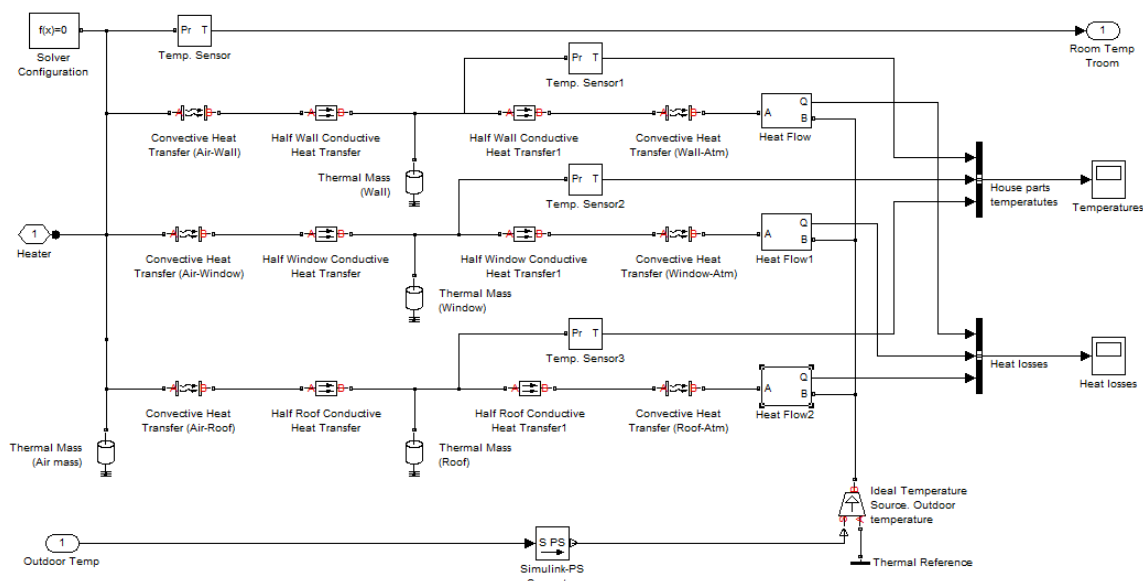


Рис. 4. Модель теплового режима помещения
Fig. 4. Model thermal regime premises

и двери. Каждая цепочка включает в себя конвективный теплообмен между наружным воздухом и ограждающей поверхностью, передачу тепла теплопроводностью через ограждающую поверхность и конвективный теплообмен между ограждающей поверхностью и внутренним воздухом помещения. В случае использования многослойных ограждающих поверхностей каждый слой должен быть представлен на модели отдельно. Термическая масса всех ограждающих поверхностей, а также внутреннего воздуха моделируется соответствующими элементами Thermal Mass.

При необходимости, модель теплового режима помещения позволяет также учитывать инфильтрацию воздуха через ограждающие конструкции помещения или окна и двери, влияние естественной или принудительной вентиляции, наличие сплит-системы, включение внутренних электрических нагревательных приборов и другие факторы.

Подсистема Room имеет входной порт, куда подается сигнал о температуре наружного воздуха, и выходной порт для снятия температуры внутреннего воздуха помещения.

Подсистема Radiator моделирует передачу тепла от теплоносителя к воздуху внутри помещения. Блок Radiator имеет выход, который может использоваться для определения суммарного количества тепла, затраченного на обогрев помещения.

Подсистема Water in the heating system моделирует нагрев теплоносителя в системе отопления. Подсистема имеет два входа для задания расхода и разности температуры воды в прямом и обратном трубопроводах отопления и один выход для передачи теплового потока в блок Radiator.

Подсистема Regulator моделирует перестраиваемый регулятор и содержит модель, показанную на рис. 5.

Подсистема Regulator включает ПИД-регулятор и Stateflow-блок Thermostat, управляющий работой регулятора и показанный на рис. 6.

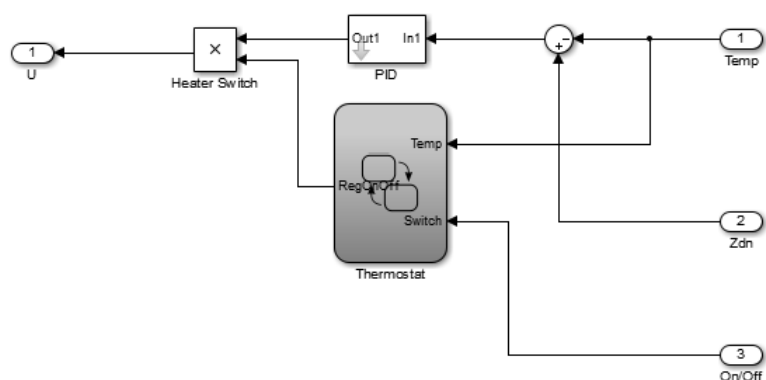


Рис. 5. Подсистема Regulator
 Fig. 5. Subsystem Regulator

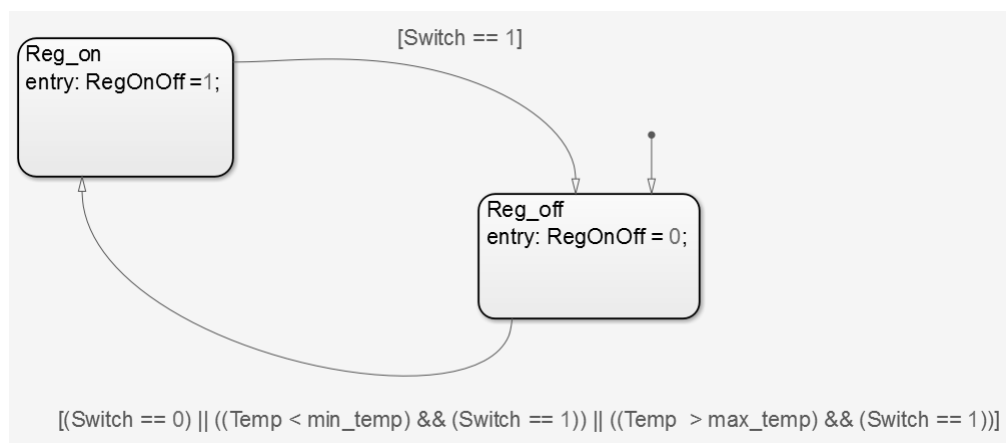


Рис. 6. Stateflow-блок Thermostat
 Fig. 6. Stateflow-block Thermostat

При открытии окна в помещении, например, с целью его проветривания, сигнал от датчика открытия окна передается из системы охранно-пожарной сигнализации в систему управления микроклиматом в помещении. При этом регулятор временно отключается и включается только при достижении минимально допустимой или максимально допустимой температуры. При закрытии окна регулятор переходит в штатный режим работы.

Переходный процесс в замкнутой системе регулирования с перестраиваемым регулятором (верхний график), в течение сорока восьми часов (двух суток), показан на рис. 7. Здесь предполагается, что примерно через 10 часов после начала эксперимента открывается окно, которое остается открытым еще примерно 10 часов, а потом закрывается. Во время открытия окна температура в помещении снижается и поддерживается на минимально допустимом уровне. Это позволяет избежать ненужных потерь тепла на обогрев помещения.

Перестраиваемый регулятор позволяет также учитывать наличие человека в помещении, определяемое по датчику движения. В случае отсутствия людей в помеще-

нии минимально допустимые и максимально допустимые температуры могут иметь более широкие значения.

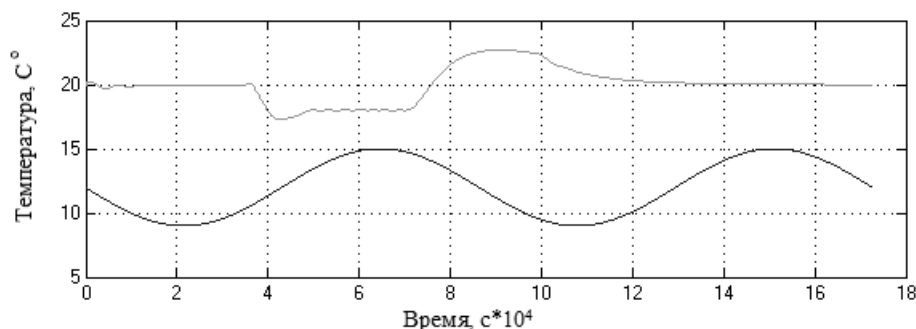


Рис. 7. Переходный процесс в замкнутой системе регулирования с перестраиваемым регулятором

Fig. 7. The transition process in a closed system of regulation with adjustable regulator

5. Заключение

С построенной компьютерной моделью "Интеллектуального здания" были проведены численные эксперименты с целью проверки непротиворечивости системы логических условий, отсутствия не реальных, не допустимых и тупиковых (deadlock) ситуаций, оценки качества динамических режимов, определения объемов энергопотребления в различных режимах и выявления возможностей экономии энергоресурсов. На основании этих экспериментов получен непротиворечивый набор логических правил, который может использоваться при разработке алгоритмов управления "Интеллектуальным зданием".

Для генерации программного С-кода сложных гибридных Simulink и Stateflow моделей для конкретных аппаратных платформ может применяться Simulink Coder [24] совместно с Embedded Coder [25]. Полученный С-код может использоваться при разработке программного обеспечения микроконтроллеров АСУЗ, управляющих инженерным оборудованием и для проведения совместного программно-аппаратного тестирования в связке модель-устройство. Среда генерации кода позволяет гибко настраивать получаемый С-код.

Список литературы / References

- [1] Репин В., "Интеллектуальное здание. Концепция крупным планом", *СтройПРО-Фильм*, 2001, №8; [Repin V., "Intellectualnoe zdanie. Konserpcija krupnym planom", *StrojPROFil*, 2001, №8, (in Russian).]
- [2] Табунщиков Ю.А., "Интеллектуальные здания", *АВОК*, 2001, №3, 6–14; [Tabunshnikov Ju. A., "Intellectualnye zdaniya", *AVOK*, 2001, №3, 6–14, (in Russian).]
- [3] Харке В., *Умный дом. Объединение в сеть бытовой техники и систем коммуникаций в жилищном строительстве*, Техносфера, М, 2006, 288 с.; [Harke V., *Umnyj dom. Obedinenie v set bytovoj tehniki i sistem kommunikacij v zhilishhnom stroitelstve*, Tehnosfera, М, 2006, 288 pp., (in Russian).]

- [4] “TRNSYS”, <http://sel.me.wisc.edu/trnsys/index.html>.
- [5] “ESP-r”, <http://www.esru.strath.ac.uk/Programs/ESP-r.htm>.
- [6] “IDA ICE”, <http://www.equa.se/en/ida-ice>.
- [7] Парийская Е. Ю., “Применение методов теории реактивных систем в задачах моделирования и качественного анализа непрерывно-дискретных систем”, *Дифференциальные уравнения и процессы управления*, 1998, № 1, 367–417; [Parijskaia E. Ju., “The Comparative Analysis of Mathematical Models and Simulating Approaches for Continuous-discrete Systems”, *Differential Equations and Control Processes*, 1998, № 1, 367–417, (in Russian).]
- [8] Harel D., “Statecharts: a Visual Formalism for Complex Systems”, *Sci. Comput. Prog.*, **8** (1987), 231–274.
- [9] “Stateflow”, <http://matlab.exponenta.ru/stateflow>.
- [10] Буч Г. и др., *Язык UML. Руководство пользователя*, ДМК Пресс, М, 2006, 496 с.; [Buch G. et al., *Jazyk UML. Rukovodstvo polzovatelja*, DMK Press, M, 2006, 496 pp., (in Russian).]
- [11] Шалыто А. А., *Switch-технология. Алгоритмизация и программирование задач логического управления*, Наука, СПб, 1998, 628 с.; [Shalyto A. A., *Switch-tehnologija. Algoritimizacija i programmirovanie zadach logicheskogo upravlenija*, Nauka, SPb, 1998, 628 pp., (in Russian).]
- [12] “AnyLogic – инструмент многоподходного имитационного моделирования”, <http://matlab.exponenta.ru/stateflow>; [“AnyLogic – instrument mnogopodhodnogo imitacionnogo modelirovanija”, <http://matlab.exponenta.ru/stateflow>, (in Russian).]
- [13] Боев В. Д. и др., *Компьютерное моделирование: Пособие для курсового и дипломного проектирования*, ВАС, СПб, 2011, 348 с.; [Boev V.D. et al., *Kompjuternoe modelirovanie: Posobie dlja kursovogo i diplomnogo proektirovanija*, VAS, SPb, 2011, 348] pp.
- [14] “Modelica and the Modelica Association”, <https://modelica.org>.
- [15] “Dymola”, <http://www.3ds.com/products-services/catia/products/dymola>.
- [16] “MapleSim”, <http://www.maplesoft.com/products/maplesim/index.aspx>.
- [17] “Wolfram SystemModeler”, <http://www.wolfram.com/system-modeler/features/>.
- [18] “Welcome to OpenModelica”, www.openmodelica.org/index.php.
- [19] “JModelica”, www.jmodelica.org.
- [20] “Modelica_StateGraph2”, https://github.com/modelica/Modelica_StateGraph2.
- [21] “MATLAB and Simulink”, <http://matlab.ru/products>.
- [22] The MathWorks, Inc., “Stateflow User’s Guide”, 2012.
- [23] The MathWorks, Inc., “Simscape User’s Guide”, 2012.
- [24] “Simulink Coder.”, <http://matlab.ru/products/simulink-coder>.
- [25] “Embedded Coder.”, <http://matlab.ru/products/embedded-coder>.

Maryasin O.Yu., Kolodkina A.S., Ogarkov A.A., "Computer Modelling «Smart Building», *Modeling and Analysis of Information Systems*, **23:4** (2016), 427–439.

DOI: 10.18255/1818-1015-2016-4-427-439

Abstract. Currently "Smart building" or "Smart house" technology is developing actively in industrialized countries. The main idea of "smart building" or "smart house" is to have a system which is able to identify definite situations happening in house and respond accordingly. Automated house management system is made for automated control and management and also for organization of interaction between separated systems of engineering equipment. This system includes automation subsystems of one or another engineering equipment as separated components. In order to perform study of different functioning modes of engineering subsystems and the whole system, mathematical and computer modeling needs to be used. From mathematical point of view description of "Smart building" is a continuous-discrete or hybrid system consisting of interacting elements of different nature, whose behavior is described by continuous and discrete processes. In the article the authors present a computer model "Smart building" which allows to model the work of main engineering subsystems and management algorithms. The model is created in Simulink Matlab system with "physical modeling" library Simscape and Stateflow library. The peculiarity of this model is the use of specialized management and control algorithms which allow providing coordinated interaction of subsystems and optimizing power consumption.

Keywords: "Smart building", ABMS, hybrid system, computer model, Simulink, Stateflow

On the authors:

Maryasin Oleg Yurievich, orcid.org/0000-0001-7449-518X, PhD,
Yaroslavl State Technical University,
Moskovsky ave., 88, Yaroslavl, 150023, Russia, e-mail: maryasinou@ystu.ru

Kolodkina Anna Sergeevna, orcid.org/0000-0001-7736-5372, teacher,
Yaroslavl Industrial and Economical College,
Gagarina str., 8, Yaroslavl, 150023, Russia, e-mail: anuta_25@mail.ru

Ogarkov Andrei Alekseevich, orcid.org/0000-0002-4151-0379, engineer,
"Integrator" Company, Chehova str., 2, Yaroslavl, 150054, Russia, e-mail: drivemox@mail.ru

©Никольская О. В., 2016

DOI: 10.18255/1818-1015-2016-4-440-465

УДК 512.7

Об алгебраических циклах на расслоенных произведениях неизотривиальных семейств регулярных поверхностей с геометрическим родом 1

Никольская О. В.

получена 7 июня 2016

Аннотация. Пусть $\pi_k : X_k \rightarrow C$ ($k = 1, 2$) – проективное семейство поверхностей (возможно, с вырождениями) над гладкой проективной кривой C . Предположим, что дискриминантные локусы $\Delta_k = \{\delta \in C \mid \text{Sing}(X_{k\delta}) \neq \emptyset\}$ ($k = 1, 2$) не пересекаются, причем $h^{2,0}(X_{ks}) = 1$, $h^{1,0}(X_{ks}) = 0$ для любого гладкого слоя X_{ks} и отображение периодов, ассоциированное с вариацией структур Ходжа $R^2\pi'_{k*}\mathbb{Q}$ (где $\pi'_k : X'_k \rightarrow C \setminus \Delta_k$ – гладкая часть морфизма π_k), является непостоянным. Если для общих геометрических слоев X_{1s} и X_{2s} выполнены следующие условия:

- (i) $b_2(X_{1s}) - \text{rank NS}(X_{1s})$ является нечетным числом;
- (ii) $b_2(X_{1s}) - \text{rank NS}(X_{1s}) \neq b_2(X_{2s}) - \text{rank NS}(X_{2s})$,

то для любой гладкой проективной модели X расслоенного произведения $X_1 \times_C X_2$ верна гипотеза Ходжа об алгебраических циклах.

Если, кроме того, морфизмы π_k гладкие, $p_k = b_2(X_{ks}) - \text{rank NS}(X_{ks})$ ($k = 1, 2$) – нечетные простые числа и $p_1 \neq p_2$, то для $X_1 \times_C X_2$ и для расслоенного квадрата $X_1 \times_C X_1$ верны гипотеза Ходжа и стандартная гипотеза Гротендика об алгебраичности операторов $*$ и Λ теории Ходжа.

Этот результат доставляет новые примеры гладких проективных 5-мерных многообразий, для которых верны гипотезы Ходжа и Гротендика, потому что в качестве гладких слоев морфизма $\pi_k : X_k \rightarrow C$ могут быть $K3$ -поверхности, а также минимальные регулярные поверхности основного типа (размерности Кодаиры $\kappa = 2$) с геометрическим родом 1, принадлежащие одному из следующих типов: (а) поверхности с $K^2 \leq 2$; (б) поверхности с $3 \leq K^2 \leq 8$, модули которых лежат в одной компоненте модулей с поверхностью Тодорова; (с) поверхности с $K^2 = 3$ с кручением группы Пикара $\mathbb{Z}/3\mathbb{Z}$.

Ключевые слова: гипотеза Ходжа, стандартная гипотеза, расслоенное произведение, группа Ходжа, цикл Пуанкаре

Для цитирования: Никольская О. В., "Об алгебраических циклах на расслоенных произведениях неизотривиальных семейств регулярных поверхностей с геометрическим родом 1", *Моделирование и анализ информационных систем*, **23:4** (2016), 440–465.

Об авторах:

Никольская Ольга Владимировна, orcid.org/0000-0002-6742-8453, канд. физ.-мат. наук, доцент, Владимирский государственный университет им. А.Г. и Н.Г. Столетовых, ул. Горького, 87, г. Владимир, 600000, Россия, e-mail: rapichonok@yandex.ru

Благодарности:

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант N 16-31-00266).

Введение

Пусть X — гладкое проективное комплексное многообразие, $r \in \mathbb{Z}$, $r \geq 0$,

$$H^{2r}(X, \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C} = \bigoplus_{p+q=2r} H^{p,q}(X, \mathbb{C})$$

— разложение Ходжа. Гипотеза Ходжа [1] утверждает, что $\mathbb{Q}$ -пространство

$$H^{2r}(X, \mathbb{Q}) \cap H^{r,r}(X, \mathbb{C})$$

порождается классами когомологий алгебраических циклов коразмерности r на X . Эта гипотеза верна для всех многообразий размерности ≤ 3 , для всех простых абелевых многообразий простой размерности [2], а также для некоторых других типов абелевых многообразий, перечисленных в обзоре [3]. Справедливость гипотезы Ходжа не установлена даже для произведения двух $K3$ -поверхностей. Тем не менее эта гипотеза верна для гладких моделей расслоенных произведений некоторых неизотривиальных семейств $K3$ -поверхностей над гладкой проективной кривой [4], [5], [6].

С другой стороны, стандартная гипотеза Гротендика $B(X)$ типа Лефшеца утверждает [7], что классические операторы $*$ и Λ теории Ходжа, рассматриваемые как соответствия на декартовом квадрате $X \times X$, представлены алгебраическими классами когомологий на $X \times X$. Эта гипотеза верна для всех гладких комплексных проективных кривых, поверхностей, абелевых многообразий [8] и трехмерных многообразий размерности Кодaira $\kappa(X) < 3$ [9]. Кроме того, $B(X)$ выполняется для гиперкэлеровых многообразий, являющихся деформациями точечных схем Гильберта $K3$ -поверхностей [10], а также для расслоенного произведения $X_1 \times_C X_2$ двух проективных неизотривиальных гладких семейств $\pi_k : X_k \rightarrow C$ ($k = 1, 2$) $K3$ -поверхностей над гладкой проективной кривой C при условии, что ранги решеток трансцендентных циклов на общих геометрических слоях X_{ks} ($k = 1, 2$) являются различными простыми нечетными числами [4]. Другие известные примеры справедливости стандартной гипотезы содержатся в [11].

В данной работе гипотеза Ходжа и стандартная гипотеза Гротендика доказываются для расслоенных произведений семейств регулярных поверхностей с геометрическим родом 1 при некоторых естественных условиях на ранги решеток трансцендентных циклов общих геометрических слоев. Это доставляет в том числе примеры 5-мерных гладких проективных многообразий размерности Кодaira 5, для которых верны обе гипотезы.

Пусть $\pi_k : X_k \rightarrow C$ ($k = 1, 2$) — проективное семейство поверхностей (возможно, с вырождениями) над гладкой проективной кривой C . Предположим, что дискриминантные локусы $\Delta_k = \{\delta \in C \mid \text{Sing}(X_{k\delta}) \neq \emptyset\}$ ($k = 1, 2$) не пересекаются, причем $h^{2,0}(X_{ks}) = 1$, $h^{1,0}(X_{ks}) = 0$ для любого гладкого слоя X_{ks} и отображение периодов, ассоциированное с вариацией структур Ходжа $R^2 \pi'_{k*} \mathbb{Q}$ (где $\pi'_k : X'_k \rightarrow C \setminus \Delta_k$ — гладкая часть морфизма π_k), является непостоянным. Если для общих геометрических слоев X_{1s} и X_{2s} выполнены следующие условия:

- (i) $b_2(X_{1s}) - \text{rank NS}(X_{1s})$ является нечетным числом;
- (ii) $b_2(X_{1s}) - \text{rank NS}(X_{1s}) \neq b_2(X_{2s}) - \text{rank NS}(X_{2s})$,

то для любой гладкой проективной модели X расслоенного произведения $X_1 \times_C X_2$ верна гипотеза Ходжа об алгебраических циклах.

Если, кроме того, морфизмы π_k гладкие, $p_k = b_2(X_{ks}) - \text{rank NS}(X_{ks})$ ($k = 1, 2$) – нечетные простые числа и $p_1 \neq p_2$, то для $X_1 \times_C X_2$ и для расслоенного квадрата $X_1 \times_C X_1$ верны гипотеза Ходжа и стандартная гипотеза Гротендика об алгебраичности операторов $*$ и Λ теории Ходжа.

Заметим, что в качестве гладких слоев морфизма $\pi_k : X_k \rightarrow C$ могут быть $K3$ -поверхности, а также минимальные регулярные поверхности основного типа (размерности Кодайры $\kappa = 2$) с геометрическим родом 1, принадлежащие одному из следующих типов: (а) поверхности с $K^2 \leq 2$; (б) поверхности с $3 \leq K^2 \leq 8$, модули которых лежат в одной компоненте модулей с поверхностью Тодорова; (с) поверхности с $K^2 = 3$ с кручением группы Пикара $\mathbb{Z}/3\mathbb{Z}$.

1. Отображение периодов и группы Ходжа

1.1. Отображение периодов и монодромия

По определению, отображение периодов сопоставляет точке s базы семейства гладких проективных многообразий над полем $\mathbb{C}$ комплексных чисел когомологии слоя над этой точкой, снабженные структурой Ходжа. Полученная при этом структура Ходжа рассматривается как точка в многообразии модулей структур Ходжа данного типа.

Пусть $\{\mathcal{X}_s\}_{s \in \mathcal{S}}$ – семейство слоев гладкого проективного морфизма $f : \mathcal{X} \rightarrow \mathcal{S}$, где $\mathcal{S}$ – гладкое алгебраическое многообразие. Тогда когомологии $H^p(\mathcal{X}_s, \mathbb{Z}) = V_{\mathbb{Z}}$ снабжены чистой поляризованной структурой Ходжа, которая задается морфизмом вещественных алгебраических групп $h : \text{Res}_{\mathbb{C}/\mathbb{R}}(G_m) \rightarrow G_{\mathbb{R}}$, где $\text{Res}_{\mathbb{C}/\mathbb{R}}(G_m)$ – мультипликативная группа $\mathbb{C}^{\times}$ поля комплексных чисел, рассматриваемая как вещественная алгебраическая группа,

$$G = \{g \in \text{GL}(V) \mid \psi(gx, gy) = \lambda(g)\psi(x, y)\}$$

– алгебраическая группа линейных автоморфизмов пространства V , умножающих невырожденную (симметрическую или кососимметрическую) билинейную форму поляризации ψ на скалярный множитель, причем автоморфизм $\text{Ad } h(i)$ группы $G_{\mathbb{R}}$ является инволюцией Картана и $h(\mathbb{R}^{\times})$ лежит в центре группы $G_{\mathbb{R}}$.

Множество X_G морфизмов групп $h : \text{Res}_{\mathbb{C}/\mathbb{R}}(G_m) \rightarrow G_{\mathbb{R}}$, обладающих этими свойствами, естественным образом снабжено $G_{\mathbb{R}}$ -инвариантной структурой однородного кэлерова многообразия, а фактор $M_G = X_G/G_{\mathbb{Z}}$ является пространством модулей структур Ходжа.

Голоморфное отображение в X_G или M_G называется *горизонтальным*, если образ его касательного отображения лежит в горизонтальном подрасслоении. Хорошо известно, что отображение периодов $\Phi : \mathcal{S} \rightarrow M_G$ горизонтально [12], [13, п. (2.4)]. Образ группы монодромии

$$\Phi_*(\pi_1(\mathcal{S}, s)) \subset G_{\mathbb{Z}}$$

полупрост во всяком рациональном представлении группы G [13, теорема (3.3)], а преобразования обхода T вокруг дивизора с нормальными пересечениями $\bar{\mathcal{S}} \setminus \mathcal{S}$ в гладкой компактификации $\bar{\mathcal{S}}$ многообразия $\mathcal{S}$ (преобразования Пикара – Лефшеца)

порождают квазиунипотентные (другими словами, имеющие в качестве собственных чисел корни из 1) элементы $\Phi_*(T) \in G_{\mathbb{Z}}$ [13, теорема (3.1)]. Группа монодромии важна в силу теоремы жесткости [13, теорема (5.4)], [14], [15]: если над $\mathcal{S}$ имеются два семейства гладких проективных многообразий, то соответствующие отображения периодов Φ_1 и Φ_2 из $\mathcal{S}$ в M_G совпадают тогда и только тогда, когда $\Phi_1(s_0) = \Phi_2(s_0)$ в некоторой точке s_0 и морфизмы групп $\Phi_{i*} : (\pi_1(\mathcal{S}, s_0)) \rightarrow G_{\mathbb{Z}}$, $i = 1, 2$ совпадают.

1.2. Поверхность Тодорова

Поверхностью Тодорова называется любая проективная комплексная поверхность Z с эйлеровой характеристикой структурного пучка $\chi(\mathcal{O}_Z) = 2$, двойными рациональными особыми точками и обильным каноническим классом при условии, что ее биканонический образ является $K3$ -поверхностью S с двойными рациональными особыми точками (в частности, $q(S) = 0$, $K_S = 0$) [16, § 2].

1.3. Типы гладких слоев семейств поверхностей

Пусть V – гладкая проективная поверхность основного типа над конечнопорожденным полем $k \hookrightarrow \mathbb{C}$ с $h^{2,0}(V \otimes_k \mathbb{C}) = 1$. Если минимальная модель поверхности $V \otimes_k \mathbb{C}$ принадлежит одному из следующих типов:

- (а) поверхности с $q = 0$ и $K^2 \leq 2$;
- (б) поверхности с $q = 0$ и $3 \leq K^2 \leq 8$, модули которых лежат в одной компоненте модулей с поверхностью Тодорова;
- (с) поверхности с $q = 0$ и $K^2 = 3$ с кручением группы Пикара $\mathbb{Z}/3\mathbb{Z}$;
- (д) поверхности с $q = 1$ и $K^2 = 2$;
- (е) поверхности с $q = 1$ и $K^2 = 3$ и общим слоем отображения Альбанезе рода 3;
- (ф) поверхности с $q = 1$ и $K^2 = 4$ в любой из восьми компонент модулей, описанных в работе Пигнателли [17],

то можно считать, что существует такой гладкий проективный k -морфизм $f : \mathcal{X} \rightarrow \mathcal{S}$ над некоторой гладкой связной базой $\mathcal{S}$, что отображение периодов, ассоциированное с вариацией структур Ходжа $R^2 f_{\mathbb{C}*} \mathbb{Q}$, непостоянно, причем многообразие V является слоем морфизма f над некоторой точкой $s \in \mathcal{S}(k)$, гипотеза Тэйта для дивизоров V и гипотеза Мамфорда – Тэйта для когомологий степени 2 верны [18, теорема 9.3].

1.4. Второе число Бетти гладкого слоя

Используя формулу для арифметического рода поверхности

$$1 - q + p_g = \frac{K^2 + \chi}{12},$$

где $\chi = 2 - 2b_1 + b_2$ – эйлерова характеристика, мы видим, что в рассматриваемых случаях

$$b_2 = \chi - 2 + 2b_1 = 12(1 - q + p_g) - K^2 - 2 + 4q = 10 - 8q + 12p_g - K^2 = 22 - 8q - K^2 \geq 10.$$

1.5. Определение группы Ходжа

Пусть S – гладкая проективная поверхность с $h^{2,0}(S) = 1$. Рассмотрим разложение Ходжа

$$H^2(S, \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C} = H^{2,0}(S, \mathbb{C}) \oplus H^{1,1}(S, \mathbb{C}) \oplus H^{0,2}(S, \mathbb{C}),$$

где $H^{p,q}(S, \mathbb{C})$ – пространство гармонических форм типа (p, q) . Известно, что $H^{2,0}(S, \mathbb{C}) \xrightarrow{\sim} H^0(S, \Omega_S^2)$, $H^{0,2}(S, \mathbb{C}) \xrightarrow{\sim} H^2(S, \mathcal{O}_S)$ являются одномерными пространствами над $\mathbb{C}$.

Пусть $U^1 = \{e^{i\theta} | \theta \in \mathbb{R}\}$ – единичная окружность. Определим ее действие в $H^2(S, \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C}$ следующим образом: $e^{i\theta}$ действует на $H^{p,q}(S, \mathbb{C})$ как умножение на число $e^{i\theta(p-q)}$. В итоге мы получаем морфизм групп

$$U^1 \xrightarrow{h} \mathrm{GL}(H^2(S, \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{R}),$$

где $h(e^{i\theta})(w_{2,0} + w_{1,1} + w_{0,2}) = e^{2i\theta}w_{2,0} + w_{1,1} + e^{-2i\theta}w_{0,2}$ на пространстве $H^2(S, \mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{C}$.

По определению, *группой Ходжа* структуры Ходжа $H^2(S, \mathbb{Q})$ называется наименьшая алгебраическая $\mathbb{Q}$ -подгруппа $\mathrm{Hg}(H^2(S, \mathbb{Q})) \hookrightarrow \mathrm{GL}(H^2(S, \mathbb{Q}))$, группа $\mathbb{R}$ -точек которой содержит $h(U^1)$.

По теореме Лефшеца о дивизорах имеем:

$$H^2(S, \mathbb{Q})^{\mathrm{Hg}(H^2(S, \mathbb{Q}))} = H^2(S, \mathbb{Q}) \cap H^{1,1}(S, \mathbb{C}) = \mathrm{NS}(S) \otimes_{\mathbb{Z}} \mathbb{Q} \stackrel{\mathrm{def}}{=} \mathrm{NS}_{\mathbb{Q}}(S),$$

где $\mathrm{NS}(S)$ группа Нерона–Севери поверхности S .

1.6. Некоторые спаривания

Пусть $\mathrm{NS}_{\mathbb{Q}}(S)^{\perp}$ – ортогональное дополнение к $\mathrm{NS}_{\mathbb{Q}}(S)$ в $H^2(S, \mathbb{Q})$ относительно билинейного спаривания

$$\langle \cdot, \cdot \rangle : H^2(S, \mathbb{Q}) \times H^2(S, \mathbb{Q}) \xrightarrow{x \times y \mapsto x \smile y} H^4(S, \mathbb{Q}) \xrightarrow{\sim} \mathbb{Q}(-2) \xrightarrow{(2\pi i)^2} \mathbb{Q}.$$

Из описания Ю.Г. Зархиным группы $\mathrm{Hg}(H^2(S, \mathbb{Q}))$ известно, что $\mathrm{Hg}(H^2(S, \mathbb{Q}))$ -модуль $\mathrm{NS}_{\mathbb{Q}}(S)^{\perp}$ простой и $E = E(S) \stackrel{\mathrm{def}}{=} \mathrm{End}_{\mathrm{Hg}(H^2(S, \mathbb{Q}))} \mathrm{NS}_{\mathbb{Q}}(S)^{\perp}$ – вполне вещественное поле $E_0 = E_0(S)$ или мнимое квадратичное расширение вполне вещественного поля E_0 [19, теоремы 1.4.1, 1.6, 1.5.1]. Пусть

$$\Phi : \mathrm{NS}_{\mathbb{Q}}(S)^{\perp} \times \mathrm{NS}_{\mathbb{Q}}(S)^{\perp} \xrightarrow{x \times y \mapsto \alpha} E$$

– спаривание, определяемое формулой

$$\langle ex, y \rangle = \mathrm{tr}_{E/\mathbb{Q}}(e\alpha) \quad \text{для всех } e \in E.$$

1.7. Структура группы Ходжа в случае вполне вещественного поля

Если $E = E_0$, то группа $\mathrm{Hg}(H^2(S, \mathbb{Q}))$ полупростая [19, замечание 1.5.3.b],

$$\mathrm{Hg}(H^2(S, \mathbb{Q})) = \mathrm{Res}_{E_0/\mathbb{Q}}(\mathrm{SO}(\mathrm{NS}_{\mathbb{Q}}(S)^{\perp}, \Phi)),$$

где $\text{Res}_{E_0/\mathbb{Q}}(\text{SO}(\text{NS}_{\mathbb{Q}}(S)^{\perp}, \Phi))$ получается из E_0 -группы $\text{SO}(\text{NS}_{\mathbb{Q}}(S)^{\perp}, \Phi)$ ограничением поля скаляров до $\mathbb{Q}$ [19, теорема 2.2.1]. Лемма Шура и равенство

$$E \otimes \overline{\mathbb{Q}} = \text{End}_{\text{Hg}(H^2(S, \mathbb{Q}))} \text{NS}_{\mathbb{Q}}(S)^{\perp} \otimes \overline{\mathbb{Q}}$$

дают разложение

$$\text{NS}_{\mathbb{Q}}(S)^{\perp} \otimes \overline{\mathbb{Q}} = V_1 \oplus \cdots \oplus V_e,$$

где $e = [E : \mathbb{Q}]$, $V_i = V_i(S)$ ($i = 1, \dots, e$) – неприводимые попарно неизоморфные ортогональные $\text{Hg}(H^2(S, \mathbb{Q})) \otimes \overline{\mathbb{Q}}$ -модули, образ канонического морфизма

$$\text{Hg}(H^2(S, \mathbb{Q})) \otimes \overline{\mathbb{Q}} \rightarrow \text{GL}(V_i)$$

совпадает с $\text{SO}(V_i)$, группа Галуа $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ транзитивно переставляет $V_1, \dots, V_e$. Следовательно,

$$\text{Hg}(H^2(S, \mathbb{Q})) \otimes \overline{\mathbb{Q}} \simeq \prod_{i=1}^e \text{SO}(V_i).$$

Пара

$$(\text{тип } \text{Lie Hg}(H^2(S, \mathbb{Q})) \otimes \overline{\mathbb{Q}}, \text{NS}_{\mathbb{Q}}(S)^{\perp} \otimes \overline{\mathbb{Q}})$$

принимает одно из следующих значений:

$$\left(A_1 \times \cdots \times A_1 = A_1^e, \quad E(2\omega_1^{(1)}) \oplus \cdots \oplus E(2\omega_1^{(e)}) \right),$$

если $\dim V_i = 3$; (1.1)

$$\left(A_1 \times \cdots \times A_1 = A_1^{2e}, \quad E(\omega_1^{(1)} + \omega_1^{(2)}) \oplus \cdots \oplus E(\omega_1^{(2e-1)} + \omega_1^{(2e)}) \right),$$

если $\dim V_i = 4$; (1.2)

$$\left(B_n \times \cdots \times B_n = B_n^e, \quad E(\omega_1^{(1)}) \oplus \cdots \oplus E(\omega_1^{(e)}) \right),$$

если $\dim V_i = 2n + 1, \quad n \geq 2$; (1.3)

$$\left(D_n \times \cdots \times D_n = D_n^e, \quad E(\omega_1^{(1)}) \oplus \cdots \oplus E(\omega_1^{(e)}) \right),$$

если $\dim V_i = 2n, \quad n \geq 3$, (1.4)

где через $E(\omega_1^{(i)})$ обозначается стандартное неприводимое представление со старшим весом $\omega_1^{(i)}$ (в обозначениях Н. Бурбаки [20]) i -го простого фактора типа A_1, B_n или D_n полупростой алгебры Ли $\text{Lie Hg}(H^2(S, \mathbb{Q})) \otimes \overline{\mathbb{Q}}$.

1.8. Лемма Мустафина

Лемма Мустафина [21, § 4, лемма 3]. Пусть $\rho : g \rightarrow \text{End}_{\mathbb{Q}} V$ – точное $\mathbb{Q}$ -неприводимое представление $\mathbb{Q}$ -полупростой алгебры Ли g , и пусть $g_{\mathbb{C}} = g_1 \times \cdots \times g_e$ – разложение на простые факторы комплексификации g , где $e = \dim_{\mathbb{Q}}(Z(\text{End}_g V))$, $Z(\text{End}_g V)$ – центр алгебры $\text{End}_g V$. Тогда алгебра Ли g $\mathbb{Q}$ -простая.

Согласно лемме Мустафина, алгебра Ли $\text{Lie Hg}(H^2(S, \mathbb{Q})S)$ является $\mathbb{Q}$ -простой в случаях (1.1), (1.3) – (1.4) (другими словами, если $\dim V_i \neq 4$).

1.9. Структура группы Ходжа в случае мнимого квадратичного расширения вполне вещественного поля

Предположим, что $E \neq E_0$. Тогда

$$\mathrm{Hg}(H^2(S, \mathbb{Q})) = \mathrm{Res}_{E_0/\mathbb{Q}}(\mathrm{U}(\mathrm{NS}_{\mathbb{Q}}(S)^{\perp}, \Phi)),$$

где $\mathrm{U}(\mathrm{NS}_{\mathbb{Q}}(S)^{\perp}, \Phi)$ – унитарная группа E -векторного пространства $\mathrm{NS}_{\mathbb{Q}}(S)^{\perp}$ относительно эрмитовой формы Φ [19, теорема 2.3.1].

В этом случае легко проверить, что полупростая часть $\mathrm{Lie} \, \mathrm{Hg}(H^2(S, \mathbb{Q}))^{\mathrm{ss}} \otimes \overline{\mathbb{Q}}$ редуктивной алгебры Ли $\mathrm{Lie} \, \mathrm{Hg}(H^2(S, \mathbb{Q})) \otimes \overline{\mathbb{Q}}$ является произведением $e_0 = e/2$ экземпляров простой алгебры Ли типа A_n ($n \geq 1$) и $\mathrm{Lie} \, \mathrm{Hg}(H^2(S, \mathbb{Q}))^{\mathrm{ss}} \otimes \overline{\mathbb{Q}}$ -модуль $\mathrm{NS}_{\mathbb{Q}}(S)^{\perp} \otimes \overline{\mathbb{Q}}$ допускает разложение

$$\mathrm{NS}_{\mathbb{Q}}(S)^{\perp} \otimes \overline{\mathbb{Q}} = E(\omega_1^{(1)}) \oplus E(\omega_n^{(1)}) \oplus \cdots \oplus E(\omega_1^{(e_0)}) \oplus E(\omega_n^{(e_0)}), \quad (1.5)$$

где $E(\omega_1^{(i)})$ – стандартное неприводимое представление алгебры Ли типа A_n в $(n+1)$ -мерном пространстве над $\overline{\mathbb{Q}}$ и $E(\omega_n^{(i)}) = E(\omega_1^{(i)})^{\vee}$ – представление, двойственное стандартному (оно изоморфно n -й внешней степени стандартного представления).

Действительно, поскольку E_0 – вполне вещественное поле, то алгебра $E_0 \otimes_{\mathbb{Q}} \mathbb{R}$ – прямая сумма полей $\mathbb{R}_{\sigma} = \mathbb{R}$, индексированных вложениями $\sigma : E_0 \hookrightarrow \mathbb{R}$, где

$$\mathbb{R}_{\sigma} = \{a \in E_0 \otimes_{\mathbb{Q}} \mathbb{R} \mid ea = \sigma(e)a \text{ для } e \in E_0\} = E_0 \otimes_{E_0, \sigma} \mathbb{R}.$$

Поэтому $E_0 \otimes_{\mathbb{Q}} \mathbb{R}$ -модуль $\mathrm{NS}_{\mathbb{Q}}(S)^{\perp} \otimes_{\mathbb{Q}} \mathbb{R} = \mathrm{NS}_{\mathbb{Q}}(S)^{\perp} \otimes_{E_0} (E_0 \otimes_{\mathbb{Q}} \mathbb{R})$ является прямой суммой таких $\mathbb{R}_{\sigma}$ -векторных пространств W_{σ} , что

$$W_{\sigma} = \mathrm{NS}_{\mathbb{Q}}(S)^{\perp} \otimes_{E_0, \sigma} \mathbb{R}$$

и $\dim_{\mathbb{R}} W_{\sigma} = \dim_{E_0} \mathrm{NS}_{\mathbb{Q}}(S)^{\perp}$ для любого σ . Аналогично $E_0 \otimes_{\mathbb{Q}} \mathbb{R}$ -алгебра Ли $\mathrm{Lie} \, \mathrm{Hg}(H^2(S, \mathbb{Q})) \otimes_{\mathbb{Q}} \mathbb{R}$ является прямой суммой таких $\mathbb{R}_{\sigma}$ -алгебр Ли $hg_{\sigma} \subset \mathrm{End}_{\mathbb{R}} W_{\sigma}$, что

$$hg_{\sigma} = \mathrm{Lie} \, \mathrm{Hg}(H^2(S, \mathbb{Q})) \otimes_{E_0, \sigma} \mathbb{R}$$

и $\dim_{\mathbb{R}} hg_{\sigma} = \dim_{E_0} \mathrm{Lie} \, \mathrm{Hg}(H^2(S, \mathbb{Q}))$. Более того, $\mathrm{Lie} \, \mathrm{Hg}(H^2(S, \mathbb{Q})) \otimes_{\mathbb{Q}} \mathbb{R}$ -модуль $\mathrm{NS}_{\mathbb{Q}}(S)^{\perp} \otimes_{\mathbb{Q}} \mathbb{R}$ – прямая сумма точных hg_{σ} -модулей W_{σ} и

$$\mathrm{End}_{hg_{\sigma}} W_{\sigma} = \mathbb{C}. \quad (1.6)$$

В частности, hg_{σ} – неприводимая подалгебра в $\mathrm{End}_{\mathbb{R}} W_{\sigma}$ [19, замечание 1.9.4]. Поэтому остается проверить, что существует разложение

$$W_{\sigma} \otimes_{\mathbb{R}} \mathbb{C} = V_{\sigma} \oplus \bigvee V_{\sigma}, \quad (1.7)$$

где V_{σ} – неприводимый $hg_{\sigma} \otimes_{\mathbb{R}} \mathbb{C}$ -модуль.

В силу (1.6) и леммы Шура имеется разложение $W_{\sigma} \otimes_{\mathbb{R}} \mathbb{C} = V_{\sigma} \oplus U_{\sigma}$, где V_{σ} и U_{σ} – неизоморфные неприводимые $hg_{\sigma} \otimes_{\mathbb{R}} \mathbb{C}$ -модули. Введем обозначения

$$V_{\sigma} = E(\omega_{\sigma}^{(1)})(\chi_{\sigma}^{(1)}), \quad U_{\sigma} = E(\omega_{\sigma}^{(2)})(\chi_{\sigma}^{(2)}),$$

где $\omega_\sigma^{(1)}$ – старший вес пространства V_σ , рассматриваемого как модуль над полупростой частью $[hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}]^{\text{ss}}$ алгебры Ли $hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}$, и $\chi_\sigma^{(1)}$ – вес $\text{Cent}(hg_\sigma \otimes_{\mathbb{R}} \mathbb{C})$ -модуля V_σ ; аналогичные обозначения используются для U_σ . Хорошо известно, что для невырожденной эрмитовой формы Ψ на $(n+1)$ -мерном комплексном векторном пространстве $V_{\mathbb{C}}$ существует каноническая точная последовательность вещественных групп Ли [22, гл. IX, § 6, предложение 3]:

$$1 \rightarrow \text{SU}(\Psi) \rightarrow \text{U}(\Psi) \rightarrow \{e^{i\theta} \mid \theta \in \mathbb{R}\} \rightarrow 1,$$

где $\text{SU}(\Psi)$ – компактная вещественная форма комплексной группы Ли $\text{SL}(V_{\mathbb{C}})$ типа A_n [23, гл. IX. § 4, таблица I]. Следовательно, $\chi_\sigma^{(1)} \neq 0$ или $\chi_\sigma^{(2)} \neq 0$. Мы можем (и будем) считать, что $\chi_\sigma^{(1)} \neq 0$. Тогда

$$[\text{Sym}^2(V_\sigma)]^{hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}} = [[\text{Sym}^2(E(\omega_\sigma^{(1)}))](2\chi_\sigma^{(1)})]^{hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}} = 0. \quad (1.8)$$

Поскольку ограничение формы поляризации на неприводимую $\mathbb{R}$ -структуру Ходжа W_σ веса 2 невырождено и hg_σ -инвариантно [19, п. 0.3.1.1 - 0.3.1.2, 0.3.2], то разложение

$$\begin{aligned} [\text{Sym}^2(W_\sigma \otimes_{\mathbb{R}} \mathbb{C})]^{hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}} &= [\text{Sym}^2(V_\sigma \oplus U_\sigma)]^{hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}} = \\ &= [\text{Sym}^2(V_\sigma)]^{hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}} \oplus [V_\sigma \otimes_{\mathbb{C}} U_\sigma]^{hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}} \oplus [\text{Sym}^2(U_\sigma)]^{hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}}, \end{aligned}$$

лемма Шура и (1.8) показывают, что имеется ненулевой элемент $\mathbb{C}$ -пространств $[V_\sigma \otimes_{\mathbb{C}} U_\sigma]^{hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}}$, определяющий некоторый изоморфизм $hg_\sigma \otimes_{\mathbb{R}} \mathbb{C}$ -модулей $U_\sigma \xrightarrow{\sim} \check{V}_\sigma$. Формула (1.7) доказана.

В дальнейшем мы будем обозначать i -е слагаемое в разложении (1.5) через $V_i = V_i(S)$.

1.10. Точки, общие в смысле Ходжа

Лемма. Пусть $f : \mathcal{X} \rightarrow \mathcal{S}$ – гладкий проективный морфизм над гладкой связной комплексной кривой $\mathcal{S}$, слоями которого являются поверхности $\mathcal{X}_s$ с $h^{2,0}(\mathcal{X}_s) = 1$. Если отображение периодов, ассоциированное с вариацией структур Ходжа $R^2 f_* \mathbb{Q}$, непостоянно, то существует такое счётное подмножество $\Delta_{\text{countable}} \subset \mathcal{S}$, что функция $s \mapsto \text{rank NS}(\mathcal{X}_s)$ постоянна на множестве $\mathcal{S} \setminus \Delta_{\text{countable}}$.

Доказательство. Используя замену базы, мы можем (и будем считать), что замыкание $G^{(2)}$ образа представления монодромии $\pi_1(\mathcal{S}, s) \rightarrow \text{GL}(H^2(\mathcal{X}_s, \mathbb{Q}))$ связное в топологии Зариского. По теореме Делиня, для любой точки $s \in \mathcal{S}$ вне некоторого счётного подмножества $\Delta_{\text{countable}}$, группа $G^{(2)}$ – нормальная подгруппа в группе Ходжа $\text{Hg}(H^2(\mathcal{X}_s, \mathbb{Q}))$ структуры Ходжа $H^2(\mathcal{X}_s, \mathbb{Q})$ [24, теорема 7.3]. Для таких точек s , называемых *общими в смысле Ходжа*, имеем согласно предложениям 1.2 и 1.3 в [21] и теореме Лефшеца о дивизорах:

$$\begin{aligned} H^2(\mathcal{X}_s, \mathbb{Q})^{\pi_1(\mathcal{S}, s)} &= H^2(\mathcal{X}_s, \mathbb{Q})^{G^{(2)}} \hookrightarrow \text{NS}_{\mathbb{Q}}(\mathcal{X}_s) \\ &= H^2(\mathcal{X}_s, \mathbb{Q})^{\text{Hg}(H^2(\mathcal{X}_s, \mathbb{Q}))} \hookrightarrow H^2(\mathcal{X}_s, \mathbb{Q})^{G^{(2)}}, \end{aligned}$$

поэтому $H^2(\mathcal{X}_s, \mathbb{Q})^{\pi_1(\mathcal{S}, s)} = \text{NS}_{\mathbb{Q}}(\mathcal{X}_s)$ и, следовательно,

$$\text{rank NS}(\mathcal{X}_s) = \dim_{\mathbb{Q}} H^0(\mathcal{S}, R^2 f_* \mathbb{Q})$$

не зависит от выбора точки $s \in \mathcal{S} \setminus \Delta_{\text{countable}}$.

Эта лемма позволяет заменять выражение "точка, общая в смысле Ходжа" на выражение "общая геометрическая точка", где "общность" точки означает ее принадлежность множеству $\mathcal{S} \setminus \Delta_{\text{countable}}$ для некоторого счётного подмножества $\Delta_{\text{countable}} \hookrightarrow \mathcal{S}$.

1.11. Непостоянное отображение периодов и бесконечная группа монодромии

Предложение [21, §1, предложение 1.2]. Пусть $f : \mathcal{X} \rightarrow \mathcal{S}$ – гладкое проективное семейство поверхностей с $h^{2,0} = 1$. Тогда следующие условия эквивалентны:

а) отображение периодов голоморфной 2-формы не является отображением $\mathcal{S}$ в точку (другими словами, отображение периодов, ассоциированное с вариацией структур Ходжа $R^2 f_* \mathbb{Q}$, непостоянно);

б) образ представления монодромии $\pi_1(\mathcal{S}, s) \rightarrow \text{GL}(H^2(\mathcal{X}_s, \mathbb{Q}))$ бесконечен.

2. Доказательство основной теоремы об алгебраических циклах на расслоенном произведении

2.1. Формулировка основной теоремы

Теорема. Пусть $\pi_k : X_k \rightarrow C$ ($k = 1, 2$) – проективное семейство поверхностей (возможно, с вырождениями) над гладкой проективной кривой C . Предположим, что множества $\Delta_k = \{\delta \in C \mid \text{Sing}(X_{k\delta}) \neq \emptyset\}$ ($k = 1, 2$) не пересекаются, причём $h^{2,0}(X_{ks}) = 1$, $h^{1,0}(X_{ks}) = 0$ для любого гладкого слоя X_{ks} и отображение периодов, ассоциированное с вариацией структур Ходжа $R^2 \pi'_{k*} \mathbb{Q}$, непостоянно (где $\pi'_k : X_k \setminus \pi_k^{-1}(\Delta_k) \rightarrow C \setminus \Delta_k$ – гладкая часть морфизма π_k).

Если для общих геометрических слоев X_{1s} и X_{2s} выполнены следующие условия:

- (i) $b_2(X_{1s}) - \text{rank NS}(X_{1s})$ является нечётным числом;
- (ii) $b_2(X_{1s}) - \text{rank NS}(X_{1s}) \neq b_2(X_{2s}) - \text{rank NS}(X_{2s})$,

то для любой гладкой проективной модели X расслоенного произведения $X_1 \times_C X_2$ верна гипотеза Ходжа об алгебраических циклах.

Если, кроме того, морфизмы π_1 и π_2 гладкие, $p_k = b_2(X_{ks}) - \text{rank NS}(X_{ks})$ ($k = 1, 2$) – нечётные простые числа и $p_1 \neq p_2$, то для $X_1 \times_C X_2$ верна стандартная гипотеза Гротендика об алгебраичности операторов $*$ и Λ теории Ходжа.

Здесь общность точки $s \in C$ означает, что она принадлежит множеству $C \setminus \Delta_{\text{countable}}$, где $\Delta_{\text{countable}}$ – счётное подмножество, зависящее от семейств π_k ; в силу леммы п. 1.10 мы можем также предполагать, что функции $s \mapsto \text{rank NS}(X_{ks})$ ($k = 1, 2$) постоянны на множестве $C \setminus \Delta_{\text{countable}}$.

2.2. Некоторые обозначения и дополнительные условия

Мы начинаем доказательство этой теоремы.

Прежде всего, если 5-мерные гладкие проективные многообразия X и $\tilde{X}$ бирационально эквивалентны, то гипотезы Ходжа для X и $\tilde{X}$ эквивалентны в силу следствия 1.8 в [25]. С другой стороны, если $f : X \rightarrow Y$ - сюръективный морфизм гладких проективных многообразий, то из гипотезы Ходжа для X следует гипотеза Ходжа для Y [26, следствие 1.2]. Поэтому в силу теоремы Мамфорда о полустабильных редукциях [27, с. 53-54] мы можем (и будем) предполагать, что любой особый слой морфизма $\pi_k : X_k \rightarrow C$ является объединением гладких поверхностей кратности 1 с нормальными пересечениями, и $X = X_1 \times_C X_2$ (очевидно, что условие $\Delta_1 \cap \Delta_2 = \emptyset$ сохраняется при замене базы и многообразие X не имеет особенностей).

Пусть

$$\begin{aligned}\Delta &= \Delta_1 \cup \Delta_2, \\ C' &= C \setminus \Delta, \quad X'_k = \pi_k^{-1}(C'), \quad \pi'_k = \pi_k|_{X'_k} : X'_k \rightarrow C', \\ X' &= X'_1 \times_{C'} X'_2, \quad \pi' = \pi'_1 \times_{C'} \pi'_2 : X' \rightarrow C',\end{aligned}$$

кроме того, пусть $\pi = \pi_1 \times_C \pi_2 : X \rightarrow C$ - структурный морфизм.

Фиксируем точку $s \in C'$. Тогда X_{1s} и X_{2s} - поверхности с $h^{2,0} = 1$.

Пусть

$$\rho^{(i)} : \pi_1(C', s) \longrightarrow \mathrm{GL}(H^i(X_s, \mathbb{Q}))$$

-представление монодромии, $G^{(i)} = \overline{\rho^{(i)}(\pi_1(C', s))} \subset \mathrm{GL}(H^i(X_s, \mathbb{Q}))$ - замыкание группы $\rho^{(i)}(\pi_1(C', s))$ в топологии Зариского группы $\mathrm{GL}(H^i(X_s, \mathbb{Q}))$. Аналогично определяются группы $G_k^{(i)}$, связанные с представлениями $\rho_k^{(i)} : \pi_1(C', s) \longrightarrow \mathrm{GL}(H^i(X_{ks}, \mathbb{Q}))$.

Пусть $(G^{(i)})^0$ - связная компонента единицы группы $G^{(i)}$. Рассматривая в случае необходимости конечное разветвлённое накрытие $\tilde{C} \rightarrow C$ и $\tilde{X} = X \times_C \tilde{C} \rightarrow \tilde{C}$, мы можем (и будем) считать, что $G^{(i)} = (G^{(i)})^0$, потому что $B(\tilde{X}) \Rightarrow B(X)$ [28, теорема 1.6] и из гипотезы Ходжа для $\tilde{X}$ следует гипотеза Ходжа для X [26, следствие 1.2]. По теореме Делиня о полупростоте представления монодромии группа $G^{(i)}$ полупростая [29, следствие 4.2.9] и не зависит от выбора точки $s \in C'$.

Согласно одной теореме Делиня [24, теорема 7.3] можно считать, что существует такое счётное подмножество $\Delta_{\mathrm{countable}} \subset C$, что для любой точки $s \in C \setminus \Delta_{\mathrm{countable}}$ группа $G_k^{(2)}$ является нормальной подгруппой в группе Ходжа $\mathrm{Hg}(X_{ks})$ ($k = 1, 2$). Точки $s \in C \setminus \Delta_{\mathrm{countable}}$ мы будем называть *общими* в смысле Ходжа. Для таких точек s алгебра Ли $\mathrm{Lie} G_k^{(2)}$ является *идеалом* в $\mathrm{Lie} \mathrm{Hg}(X_{ks})$.

2.3. Морфизмы трансцендентных частей и представления монодромии

Лемма. При выполнении условий теоремы п. 2.1 для общих геометрических слоёв X_{1s} , X_{2s} имеем:

$$\mathrm{Hom}_{\pi_1(C', s)}(\mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp}, \mathrm{NS}_{\mathbb{Q}}(X_{2s})^{\perp}) = 0.$$

Доказательство. Рассматривая, в случае необходимости, подходящее разветвленное накрытие $\tilde{C} \rightarrow C$, мы можем считать, что группы $G_k^{(2)}$ связны ($k = 1, 2$).

Из условия (i) следует, что для поверхности X_{1s} выполнено равенство $E(X_{1s}) = E_0(X_{1s})$, причем случай (1.2) невозможен. Поэтому, в силу леммы п. 1.8, алгебра Ли $\text{Lie Hg}(X_{1s})$ является $\mathbb{Q}$ -простой. С другой стороны, точка $s \in C$ является общей в смысле Ходжа, поэтому $G_1^{(2)}$ является нетривиальной (в силу предложения п. 1.11 нормальной подгруппой $\mathbb{Q}$ -простой связной алгебраической группы $\text{Hg}(X_{1s})$). Поэтому

$$G_1^{(2)} = \text{Hg}(X_{1s}) \quad (2.1)$$

и, следовательно, $\pi_1(C', s)$ -модуль $\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}$ является простым, так как он является простым $\text{Hg}(X_{1s})$ -модулем по теореме Ю.Г. Зархина [19, теорема 1.6].

Будем предполагать в дальнейшем, что $\text{Hom}_{\pi_1(C', s)}(\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}, \text{NS}_{\mathbb{Q}}(X_{2s})^{\perp}) \neq 0$. Тогда простой $\pi_1(C', s)$ -модуль $\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}$ является прямым слагаемым полупростого $\pi_1(C', s)$ -модуля $\text{NS}_{\mathbb{Q}}(X_{2s})^{\perp}$ [29, теорема 4.2.6].

Если $E(X_{2s}) = E_0(X_{2s})$ и для поверхности X_{2s} выполнено одно из условий (1.1), (1.3), (1.4), то по лемме Мустафина алгебра Ли $\text{Lie Hg}(X_{2s})$ является $\mathbb{Q}$ -простой и, следовательно, $\pi_1(C', s)$ -модуль $\text{NS}_{\mathbb{Q}}(X_{2s})^{\perp}$ является простым, причем он содержит простой $\pi_1(C', s)$ -модуль $\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}$. Поэтому $\dim \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp} = \dim \text{NS}_{\mathbb{Q}}(X_{2s})^{\perp}$, что противоречит условию (ii). Если для поверхности X_{2s} выполнено условие (1.2), то $V_i(X_{2s})$ является неприводимым $\pi_1(C, s)$ -модулем степени 4 или прямой суммой $\pi_1(C, s)$ -модулей степеней 1 или 2 (потому что $G_2^{(2)}$ является связной нормальной подгруппой в группе Ходжа $\text{Hg}(X_{2s})$), в то время как $V_i(X_{1s})$ является неприводимым $\pi_1(C, s)$ -модулем степени 3, $2n + 1$ ($n \geq 2$), $2n$ ($n \geq 3$). В этом случае утверждение леммы следует из леммы Шура.

Если $E(X_{2s}) \neq E_0(X_{2s})$, то из (1.5) следует, что $V_i(X_{2s})$ является прямой суммой 1-мерных модулей (с тривиальным действием $\pi_1(C', s)$) или неприводимым $\pi_1(C', s)$ -модулем степени $n + 1$ (потому что $G_2^{(2)}$ является связной нормальной подгруппой в группе Ходжа $\text{Hg}(X_{2s})$ с полупростой частью $\text{Hg}(X_{2s})^{\text{ss}}$ типа $A_n \times \cdots \times A_n = A_n^{e_0}$ ($n \geq 1$)). Поскольку для поверхности X_{1s} выполнено одно из условий (1.1), (1.3) – (1.4) и включение $\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \hookrightarrow \text{NS}_{\mathbb{Q}}(X_{2s})^{\perp}$ индуцирует сюръективный морфизм алгебраических $\mathbb{Q}$ -групп $G_2^{(2)} \rightarrow G_1^{(2)}$, то из (2.1) следует, что простые факторы алгебры Ли $\text{Lie Hg}(X_{1s}) \otimes \overline{\mathbb{Q}}$ имеют тип A_n ($n \geq 2$), причем неприводимые представления $V_i(X_{1s})$ не являются автодуальными (другими словами, $V_i(X_{1s})^{\vee}$ не изоморфно $V_i(X_{1s})$ над алгеброй Ли $\text{Lie Hg}(X_{1s}) \otimes \overline{\mathbb{Q}}$). Это противоречит автодуальности $V_i(X_{1s})$ при выполнении одного из условий (1.1), (1.3) – (1.4) для поверхности X_{1s} . Лемма доказана.

2.4. Точная последовательность рациональных структур Ходжа

Спектральная последовательность Лере $E_2^{p,q} = H^p(C, R^q \pi_{k*} \mathbb{Q}) \Rightarrow H^*(X_k, \mathbb{Q})$ ($k = 1, 2$) вырождена [30, следствие (15.15)], причем краевое отображение

$$H^{2i}(X_k, \mathbb{Q}) \rightarrow H^0(C, R^{2i} \pi_{k*} \mathbb{Q})$$

сюръективно [30, следствие (15.14)], поэтому имеется точная последовательность $\mathbb{Q}$ -структур Ходжа [30, теорема (15.11), следствие (15.15), теорема (15.16)]

$$\begin{aligned} 0 \rightarrow H^2(C, R^{2i-2}\pi_{k*}\mathbb{Q}) &\rightarrow \text{Ker}[H^{2i}(X_k, \mathbb{Q}) \rightarrow H^0(C, R^{2i}\pi_{k*}\mathbb{Q})] \\ &\rightarrow H^1(C, R^{2i-1}\pi_{k*}\mathbb{Q}) \rightarrow 0. \end{aligned} \quad (2.2)$$

Так как $H^1(X_{ks}, \mathbb{Q}) = 0$ для всех $s \in C'$, то для $s \in C'$ имеем:

$$H^1(X_{ks}, \mathbb{Q}) = H^3(X_{ks}, \mathbb{Q}) = 0.$$

Поэтому пучок $R^{2i-1}\pi_{k*}\mathbb{Q}$ сосредоточен на конечном множестве $\Delta = C \setminus C'$ и, следовательно, $H^1(C, R^{2i-1}\pi_{k*}\mathbb{Q}) = 0$. Из (2.2) следует точность последовательности

$$0 \rightarrow H^2(C, R^{2i-2}\pi_{k*}\mathbb{Q}) \rightarrow H^{2i}(X_k, \mathbb{Q}) \rightarrow H^0(C, R^{2i}\pi_{k*}\mathbb{Q}) \rightarrow 0. \quad (2.3)$$

В случае $i = 1$ получаем из (2.3) точную последовательность $\mathbb{Q}$ -структур Ходжа

$$0 \rightarrow H^2(C, \mathbb{Q}) \rightarrow H^2(X_k, \mathbb{Q}) \rightarrow H^0(C, R^2\pi_{k*}\mathbb{Q}) \rightarrow 0, \quad (2.4)$$

где $H^2(C, \mathbb{Q})$ порождается алгебраическим классом $\text{cl}_{X_k}(X_{ks})$ слоя X_{ks} (и, следовательно, имеет тип Ходжа $(1, 1)$).

2.5. Вычисление типа одной рациональной структуры Ходжа

Лемма. *Рациональная структура Ходжа $H^0(C, R^2\pi_{k*}\mathbb{Q})$ имеет тип $(1, 1)$.*

Доказательство. Как показано в п. 1.10, мы имеем:

$$H^0(C', R^2\pi'_{k*}\mathbb{Q}) \xrightarrow{\sim} H^2(X_{ks}, \mathbb{Q})^{\pi_1(C', s)} \hookrightarrow \text{NS}_{\mathbb{Q}}(X_{ks}). \quad (2.5)$$

Используя алгоритм п. 1.4 в [9], легко получить из теории смешанных структур Ходжа, что

$$\text{Coker} [H^1(C, R^1\pi_{k*}\mathbb{Q}) \hookrightarrow H^1(C', R^1\pi'_{k*}\mathbb{Q}) \cap \text{Im}[H^2(X_k, \mathbb{Q}) \rightarrow H^2(X'_k, \mathbb{Q})]]$$

– чистая $\mathbb{Q}$ -структура Ходжа типа $(1, 1)$. Поскольку из (2.5) следует, что $H^0(C', R^2\pi'_{k*}\mathbb{Q})$ – рациональная структура Ходжа типа $(1, 1)$ то $H^0(C, R^2\pi_{k*}\mathbb{Q})$ является $\mathbb{Q}$ -структурой Ходжа типа $(1, 1)$ в силу результатов п. 2.3 в [31]. Лемма доказана.

2.6. Рациональные когомологии степени 2 и группы Нерона – Севери

Лемма. *Имеем: $H^2(X_k, \mathbb{Q}) = \text{NS}_{\mathbb{Q}}(X_k)$, $H^2(X, \mathbb{Q}) = \text{NS}_{\mathbb{Q}}(X)$.*

Доказательство. Первое равенство следует из (2.4), леммы п. 2.5 и теоремы Лефшеца о дивизорах. Спектральная последовательность Лере

$$E_2^{p,q} = H^p(C, R^q\pi_*\mathbb{Q}) \Rightarrow H^*(X, \mathbb{Q})$$

вырождена [30, следствие (15.15)], причем краевое отображение

$$H^{2i}(X, \mathbb{Q}) \rightarrow H^0(C, R^{2i}\pi_*\mathbb{Q})$$

сюръективно [30, следствие (15.14)], поэтому имеется точная последовательность $\mathbb{Q}$ -структур Ходжа [30, теорема (15.11), следствие (15.15), теорема (15.16)]

$$\begin{aligned} 0 \rightarrow H^2(C, R^{2i-2}\pi_*\mathbb{Q}) \rightarrow \text{Ker}[H^{2i}(X, \mathbb{Q}) \rightarrow H^0(C, R^{2i}\pi_*\mathbb{Q})] \\ \rightarrow H^1(C, R^{2i-1}\pi_*\mathbb{Q}) \rightarrow 0. \end{aligned} \quad (2.6)$$

Так как $H^1(X_{ks}, \mathbb{Q}) = 0$ для всех $s \in C \setminus \Delta_k$, то пучок $R^{2j-1}\pi_{k*}\mathbb{Q}$ сосредоточен на конечном множестве Δ_k и, следовательно,

$$H^1(C, R^{2i-1}\pi_*\mathbb{Q}) = 0 \quad (2.7)$$

в силу формулы Кюннета

$$R^{2i-1}\pi_*\mathbb{Q} = \bigoplus_{p+q=2i-1} R^p\pi_{1*}\mathbb{Q} \otimes R^q\pi_{2*}\mathbb{Q}.$$

Из (2.7) следует точность последовательности

$$0 \rightarrow H^2(C, R^{2i-2}\pi_*\mathbb{Q}) \rightarrow H^{2i}(X, \mathbb{Q}) \rightarrow H^0(C, R^{2i}\pi_*\mathbb{Q}) \rightarrow 0. \quad (2.8)$$

В случае $i = 1$ получаем из (2.8) точную последовательность $\mathbb{Q}$ -структур Ходжа

$$0 \rightarrow H^2(C, \mathbb{Q}) \rightarrow H^2(X, \mathbb{Q}) \rightarrow H^0(C, R^2\pi_*\mathbb{Q}) \rightarrow 0, \quad (2.9)$$

где в силу условия $\Delta_1 \cap \Delta_2 = \emptyset$ имеем $R^1\pi_{1*}\mathbb{Q} \otimes R^1\pi_{2*}\mathbb{Q} = 0$, поэтому из формулы Кюннета следует, что

$$R^2\pi_*\mathbb{Q} = R^2\pi_{1*}\mathbb{Q} \oplus [R^1\pi_{1*}\mathbb{Q} \otimes R^1\pi_{2*}\mathbb{Q}] \oplus R^2\pi_{2*}\mathbb{Q} = R^2\pi_{1*}\mathbb{Q} \oplus R^2\pi_{2*}\mathbb{Q}. \quad (2.10)$$

Остается применить (2.9), лемму п. 2.5 и теорему Лефшеца о дивизорах. Лемма доказана.

2.7. Алгебраичность рациональных когомологий степени 4

Согласно сильной теореме Лефшеца $H^4(X_k, \mathbb{Q}) = H^2(X_k, \mathbb{Q}) \smile \text{cl}_{X_k}(H_k)$, где H_k – гиперплоское сечение многообразия X_k . Поэтому из леммы п. 2.6 следует, что пространство $H^4(X_k, \mathbb{Q})$ порождается алгебраическими классами когомологий.

2.8. Доказательство гипотезы Ходжа для расслоенного произведения

Пусть

$$f: Z \rightarrow \pi^{-1}(\Delta)$$

– нормализация схемы $\pi^{-1}(\Delta)$. Тогда Z – несвязное объединение гладких неприводимых компонент дивизора $\pi^{-1}(\Delta)$. Поскольку $f: Z \rightarrow \pi^{-1}(\Delta)$ – разрешение

особенностей замкнутой подсхемы $i_\Delta : \pi^{-1}(\Delta) \hookrightarrow X$, то имеется точная последовательность смешанных $\mathbb{Q}$ -структур Ходжа [32, следствие (8.2.8)]:

$$H^2(Z, \mathbb{Q}) \xrightarrow{(i_\Delta f)_*} H^4(X, \mathbb{Q}) \rightarrow H^4(X', \mathbb{Q}),$$

где $(i_\Delta f)_*$ – морфизм типа $(1, 1)$ рациональных структур Ходжа. Поэтому

$$(i_\Delta f)_* H^2(Z, \mathbb{Q}) = \text{Ker} [H^4(X, \mathbb{Q}) \rightarrow H^4(X', \mathbb{Q})]. \quad (2.11)$$

Хорошо известно, что имеется канонический сюръективный морфизм $\mathbb{Q}$ -структур Ходжа $H^4(X, \mathbb{Q}) \rightarrow H^0(C, R^4\pi_*\mathbb{Q})$ [30, следствие (15.14)], включенный в коммутативную диаграмму с точными строками [30, диаграмма (15.1) для случая $m = 3$]

$$\begin{array}{ccccccc} 0 & \rightarrow & L_4 & \rightarrow & H^4(X, \mathbb{Q}) & \rightarrow & H^0(C, R^4\pi_*\mathbb{Q}) \rightarrow 0 \\ & & \downarrow \alpha_4 & & \downarrow & & \downarrow \\ & & H^1(C, R^3\pi_*\mathbb{Q}) & & \downarrow & & \downarrow \\ & & \simeq \downarrow \gamma_4 & & \downarrow & & \downarrow \\ & & H^1(C, j_*R^3\pi'_*\mathbb{Q}) & & \downarrow \varphi_4 & & \downarrow \psi_4 \\ & & \cap & & & & \\ 0 & \rightarrow & H^1(C', R^3\pi'_*\mathbb{Q}) & \xrightarrow{\beta_4} & H^4(X', \mathbb{Q}) & \rightarrow & H^0(C', R^4\pi'_*\mathbb{Q}) \rightarrow 0, \end{array} \quad (2.12)$$

где $\gamma_4 \circ \alpha_4$ – морфизм $\mathbb{Q}$ -структур Ходжа и β_4 – морфизм смешанных структур Ходжа [30, теоремы (15.3), (15.4)].

По теореме Делиня каноническое отображение $H^4(X, \mathbb{Q}) \rightarrow H^0(C', R^4\pi'_*\mathbb{Q})$ является сюръективным морфизмом рациональных структур Ходжа [29, п. 4.1.1 – 4.1.2]. Поэтому диаграмма (2.12) даёт коммутативную диаграмму с точными строками

$$\begin{array}{ccccccc} 0 & \rightarrow & L_4 & \rightarrow & H^4(X, \mathbb{Q}) & \rightarrow & H^0(C, R^4\pi_*\mathbb{Q}) \rightarrow 0 \\ & & \downarrow \alpha_4 & & \downarrow & & \downarrow \\ & & H^1(C, R^3\pi_*\mathbb{Q}) & & \downarrow \varphi_4 & & \downarrow \psi_4 \\ & & \cap & & & & \\ 0 & \rightarrow & H^1(C', R^3\pi'_*\mathbb{Q}) \cap \text{Im}(\varphi_4) & \rightarrow & \text{Im}(\varphi_4) & \rightarrow & H^0(C', R^4\pi'_*\mathbb{Q}) \rightarrow 0. \end{array} \quad (2.13)$$

Спектральная последовательность Лере $E_2^{p,q} = H^p(C, R^q\pi_*\mathbb{Q})$ вырождается и даёт точную последовательность $\mathbb{Q}$ -структур Ходжа [30, теорема (15.11), предложение (15.12), следствие (15.15), теорема (15.16)]

$$0 \rightarrow H^2(C, R^2\pi_*\mathbb{Q}) \rightarrow \text{Ker}[H^4(X, \mathbb{Q}) \rightarrow H^0(C, R^4\pi_*\mathbb{Q})] \xrightarrow{\alpha_4} H^1(C, R^3\pi_*\mathbb{Q}) \rightarrow 0. \quad (2.14)$$

Поэтому, принимая во внимание (2.11) и присоединяя канонические отображения ядер и коядер к диаграмме (2.13), мы получим коммутативную диаграмму с точными строками и столбцами

$$\begin{array}{ccccccc} & & 0 & & 0 & & 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \rightarrow & H^2(C, R^2\pi_*\mathbb{Q}) & \rightarrow & (i_\Delta f)_* H^2(Z, \mathbb{Q}) & \rightarrow & \text{Ker}(\psi_4) \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \rightarrow & L_4 & \rightarrow & H^4(X, \mathbb{Q}) & \rightarrow & H^0(C, R^4\pi_*\mathbb{Q}) \rightarrow 0 \\ & & \downarrow \alpha_4 & & \downarrow \varphi_4 & & \downarrow \psi_4 \\ 0 & \rightarrow & H^1(C', R^3\pi'_*\mathbb{Q}) \cap \text{Im}(\varphi_4) & \rightarrow & \text{Im}(\varphi_4) & \rightarrow & H^0(C', R^4\pi'_*\mathbb{Q}) \rightarrow 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ & & \frac{H^1(C', R^3\pi'_*\mathbb{Q}) \cap \text{Im}(\varphi_4)}{H^1(C, R^3\pi_*\mathbb{Q})} & \rightarrow & 0 & \rightarrow & 0. \end{array} \quad (2.15)$$

Поскольку $R^3\pi'_*\mathbb{Q} = 0$ в силу формулы Кюннета, то диаграмма (2.15) даёт точную последовательность рациональных структур Ходжа

$$0 \rightarrow (i_{\Delta}f)_* H^2(Z, \mathbb{Q}) \rightarrow H^4(X, \mathbb{Q}) \rightarrow H^0(C', R^4\pi'_*\mathbb{Q}) \rightarrow 0. \quad (2.16)$$

С другой стороны,

$$R^4\pi'_*\mathbb{Q} = R^4\pi'_{1*}\mathbb{Q} \oplus [R^2\pi'_{1*}\mathbb{Q} \otimes R^2\pi'_{2*}\mathbb{Q}] \oplus R^4\pi'_{2*}\mathbb{Q}.$$

Пусть $s \in C'$ — общая точка в смысле Ходжа для семейств $\pi'_k : X'_k \rightarrow C'$ ($k = 1, 2$). Тогда можно считать (переходя в случае необходимости к разветвлённому накрытию $\tilde{C} \rightarrow C$ и подходящей гладкой модели многообразия $X_k \times_C \tilde{C}$ над $\tilde{C}$), что $G_k^{(2)} \hookrightarrow \text{Hg}(X_{ks})$ ($k = 1, 2$) — нормальная связная подгруппа. В этом случае

$$\text{NS}_{\mathbb{Q}}(X_{ks}) = H^2(X_{ks}, \mathbb{Q})^{\text{Hg}(X_{ks})} = \text{NS}_{\mathbb{Q}}(X_{ks})^{\pi_1(C', s)} \quad (2.17)$$

по теореме Лефшеца о дивизорах. Имеем в силу (2.5):

$$[\text{NS}_{\mathbb{Q}}(X_{ks}) \oplus \text{NS}_{\mathbb{Q}}(X_{ks})^{\perp}]^{\pi_1(C', s)} \hookrightarrow \text{NS}_{\mathbb{Q}}(X_{ks}),$$

поэтому

$$[\text{NS}_{\mathbb{Q}}(X_{ks})^{\perp}]^{\pi_1(C', s)} = 0. \quad (2.18)$$

Из (2.17) — (2.18) следует, что

$$\begin{aligned} & H^0(C', R^4\pi'_*\mathbb{Q}) \xrightarrow{\sim} H^4(X_{1s} \times X_{2s}, \mathbb{Q})^{\pi_1(C', s)} \\ & = H^4(X_{1s}, \mathbb{Q})^{\pi_1(C', s)} \oplus [H^2(X_{1s}, \mathbb{Q}) \otimes H^2(X_{2s}, \mathbb{Q})]^{\pi_1(C', s)} \oplus H^4(X_{2s}, \mathbb{Q})^{\pi_1(C', s)}, \end{aligned} \quad (2.19)$$

где

$$\begin{aligned} & [H^2(X_{1s}, \mathbb{Q}) \otimes H^2(X_{2s}, \mathbb{Q})]^{\pi_1(C', s)} \\ & = [[\text{NS}_{\mathbb{Q}}(X_{1s}) \oplus \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}] \otimes [\text{NS}_{\mathbb{Q}}(X_{2s}) \oplus \text{NS}_{\mathbb{Q}}(X_{2s})^{\perp}]]^{\pi_1(C', s)} \\ & = [\text{NS}_{\mathbb{Q}}(X_{1s}) \otimes \text{NS}_{\mathbb{Q}}(X_{2s}) \oplus [\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \text{NS}_{\mathbb{Q}}(X_{2s})^{\perp}]]^{\pi_1(C', s)} \\ & \xrightarrow{\sim} [\text{NS}_{\mathbb{Q}}(X_{1s}) \otimes \text{NS}_{\mathbb{Q}}(X_{2s}) \oplus \text{Hom}_{\pi_1(C', s)}[\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}, \text{NS}_{\mathbb{Q}}(X_{2s})^{\perp}]] = \\ & \quad \text{NS}_{\mathbb{Q}}(X_{1s}) \otimes \text{NS}_{\mathbb{Q}}(X_{2s}) \end{aligned} \quad (2.20)$$

согласно лемме п. 2.3. В силу леммы п. 2.6, (2.17) — (2.18) и теоремы Делиния [29, теорема 4.1.1], существует каноническое сюръективное отображение ограничения

$$\text{NS}_{\mathbb{Q}}(X_k) = H^2(X_k, \mathbb{Q}) \rightarrow H^0(C', R^2\pi'_{k*}\mathbb{Q}) \xrightarrow{\sim} H^2(X_{ks}, \mathbb{Q})^{\pi_1(C', s)} = \text{NS}_{\mathbb{Q}}(X_{ks}).$$

Поэтому элементы из $\text{NS}_{\mathbb{Q}}(X_{ks})$ поднимаются до алгебраических классов из пространства $H^2(X_k, \mathbb{Q})$. С другой стороны, элементы пространства $H^4(X_{ks}, \mathbb{Q})^{\pi_1(C', s)}$ поднимаются до алгебраических классов из $H^4(X_k, \mathbb{Q})$, потому что каноническое отображение ограничения $H^4(X_k, \mathbb{Q}) \rightarrow H^4(X_{ks}, \mathbb{Q})^{\pi_1(C', s)}$ сюръективно по теореме Делиния [29, теорема 4.1.1]. Согласно (2.16), (2.19) — (2.20) пространство $H^0(C', R^4\pi'_*\mathbb{Q})$ порождается образами алгебраических классов в $H^4(X, \mathbb{Q})$, лежащих в $\text{NS}_{\mathbb{Q}}(X_1) \cup$

$\mathrm{NS}_{\mathbb{Q}}(X_2)$, $H^4(X_1, \mathbb{Q})$, $H^4(X_2, \mathbb{Q})$, где $\mathrm{NS}_{\mathbb{Q}}(X_k)$ отождествляется с образом $\mathrm{NS}_{\mathbb{Q}}(X_k)$ при каноническом отображении $q_k^* : H^2(X_k, \mathbb{Q}) \hookrightarrow H^2(X, \mathbb{Q})$, определённом проекцией $X = X_1 \times_C X_2 \xrightarrow{q_k} X_k$ и $H^4(X_k, \mathbb{Q})$ отождествляется с образом $H^4(X_k, \mathbb{Q})$ при каноническом отображении $q_k^* : H^4(X_k, \mathbb{Q}) \hookrightarrow H^4(X, \mathbb{Q})$. В частности, точная последовательность (2.16) рациональных структур Ходжа даёт точную последовательность $\mathbb{Q}$ -структур Ходжа

$$0 \rightarrow (i_{\Delta}f)_* [H^2(Z, \mathbb{Q}) \cap H^{1,1}(Z, \mathbb{C})] \rightarrow [H^4(X, \mathbb{Q}) \cap H^{2,2}(X, \mathbb{C})] \rightarrow H^0(C', R^4\pi'_*\mathbb{Q}) \rightarrow 0. \quad (2.21)$$

Теорема Лефшеца о дивизорах на гладком (возможно, несвязном) многообразии Z показывает, что (2.21) имеет вид

$$0 \rightarrow (i_{\Delta}f)_* \mathrm{NS}_{\mathbb{Q}}(Z) \rightarrow [H^4(X, \mathbb{Q}) \cap H^{2,2}(X, \mathbb{C})] \rightarrow H^0(C', R^4\pi'_*\mathbb{Q}) \rightarrow 0.$$

Поскольку $H^0(C', R^4\pi'_*\mathbb{Q})$ порождается образами алгебраических классов на X , то $H^4(X, \mathbb{Q}) \cap H^{2,2}(X, \mathbb{C})$ порождается классами когомологий алгебраических циклов коразмерности 2 на X . Значит, для X верна гипотеза Ходжа об алгебраических циклах.

2.9. Доказательство стандартной гипотезы для расслоенного произведения

Будем предполагать в дальнейшем, что морфизмы π_1 и π_2 гладкие, $p_k = b_2(X_{ks}) - \mathrm{rank} \mathrm{NS}(X_{ks})$ ($k = 1, 2$) – нечётные простые числа и $p_1 \neq p_2$. Мы будем использовать основные свойства циклов Пуанкаре [31, §1], ассоциированных с $\mathbb{Q}$ -структурами Ходжа.

Для любой точки $t \in C$ 4-мерное проективное многообразие $X_t = X_{1t} \times X_{2t}$ снабжено поляризацией. Она задаёт невырожденную билинейную форму поляризации

$$\psi_t : H^2(X_t, \mathbb{Q}) \times H^2(X_t, \mathbb{Q}) \rightarrow \mathbb{Q}$$

на пространстве $H^2(X_t, \mathbb{Q})$. Очевидно, что

$$\dim_{\mathbb{Q}} H^2(X_t, \mathbb{Q}) \neq 2.$$

Поэтому корректно определен цикл Пуанкаре $\wp(H^2(X_t, \mathbb{Q}))$ [31, § 1], являющийся образующей 1-мерного пространства

$$[H^2(X_t, \mathbb{Q}) \otimes_{\mathbb{Q}} H^2(X_t, \mathbb{Q})]^{SO(H^2(X_t, \mathbb{Q}), \psi_t)},$$

где алгебраическая $\mathbb{Q}$ -группа $SO(H^2(X_t, \mathbb{Q}), \psi_t)$ действует диагонально на тензорном произведении $H^2(X_t, \mathbb{Q}) \otimes_{\mathbb{Q}} H^2(X_t, \mathbb{Q})$. Это действие задается формулой

$$\sigma(x \otimes y) = \sigma(x) \otimes \sigma(y) \quad (\sigma \in SO(H^2(X_t, \mathbb{Q}), \psi_t), \quad x, y \in H^2(X_t, \mathbb{Q})).$$

Хорошо известно, что $\wp(H^2(X_t, \mathbb{Q})) \in [H^2(X_t, \mathbb{Q}) \otimes H^2(X_t, \mathbb{Q})] \cap H^{2,2}(X_t \times X_t, \mathbb{C})$ является *циклом Ходжа* [31, § 1].

Для общей в смысле Ходжа точки $t \in C \setminus \Delta_{\text{countable}}$ пространство $\text{NS}_{\mathbb{Q}}(X_{kt})^{\perp}$ является неприводимым $\text{Hg}(H^2(X_{kt}, \mathbb{Q}))$ -модулем простой нечетной степени p_k , поэтому из классификации (1.1) – (1.4) следует, что $\text{Hg}(H^2(X_{kt}, \mathbb{Q}))$ – простая алгебраическая группа типа $B_{\frac{p_k-1}{2}}$. Из неравенства $p_1 \neq p_2$ следует, что в силу существования включения

$$\text{Hg}(H^2(X_{1t} \times X_{2t}, \mathbb{Q})) \hookrightarrow \text{Hg}(H^2(X_{1t}, \mathbb{Q})) \times \text{Hg}(H^2(X_{2t}, \mathbb{Q}))$$

и сюръективности канонических морфизмов

$$\text{Hg}(H^2(X_{1t} \times X_{2t}, \mathbb{Q})) \rightarrow \text{Hg}(H^2(X_{kt}, \mathbb{Q})) \quad (k = 1, 2)$$

группа

$$\text{Hg}(H^2(X_t, \mathbb{Q})) = \text{Hg}(H^2(X_{1t} \times X_{2t}, \mathbb{Q})) = \text{Hg}(H^2(X_{1t}, \mathbb{Q}) \oplus H^2(X_{2t}, \mathbb{Q}))$$

имеет тип $B_{\frac{p_1-1}{2}} \times B_{\frac{p_2-1}{2}}$. В частности,

$$\text{Hg}(H^2(X_t, \mathbb{Q})) = \text{Hg}(H^2(X_{1t}, \mathbb{Q}) \times \text{Hg}(H^2(X_{2t}, \mathbb{Q})).$$

Используя алгоритм доказательства теоремы 0.8 в [26], легко проверить, что пространство

$$H^4(X_{kt} \times X_{kt}, \mathbb{Q}) \cap H^{2,2}(X_{kt} \times X_{kt}, \mathbb{C})$$

порождается классами пересечений дивизоров и классом диагонали $\Delta_{X_{kt}} \hookrightarrow X_{kt} \times X_{kt}$ (в частности, гипотеза Ходжа выполнена для многообразия $X_{kt} \times X_{kt}$).

Действительно, разложение

$$H^2(X_{kt}, \mathbb{Q}) = \text{NS}_{\mathbb{Q}}(X_{kt}) \oplus \text{NS}_{\mathbb{Q}}(X_{kt})^{\perp}$$

и теорема Лефшеца о дивизорах дают равенства

$$H^2(X_{kt}, \mathbb{Q})^{\text{Hg}(H^2(X_{kt}, \mathbb{Q}))} = \text{NS}_{\mathbb{Q}}(X_{kt}), \quad [\text{NS}_{\mathbb{Q}}(X_{kt})^{\perp}]^{\text{Hg}(H^2(X_{kt}, \mathbb{Q}))} = 0,$$

поэтому из хорошо известного равенства [3, следствие B55]

$$H^4(X_{kt} \times X_{kt}, \mathbb{Q}) \cap H^{2,2}(X_{kt} \times X_{kt}, \mathbb{C}) = [H^4(X_{kt} \times X_{kt}, \mathbb{Q})]^{\text{Hg}(H^2(X_{kt}, \mathbb{Q}))},$$

теоремы Кюннета и леммы Шура следует, что это пространство порождается классами пересечений дивизоров и 1-мерным пространством

$$[\text{NS}_{\mathbb{Q}}(X_{kt})^{\perp} \otimes \text{NS}_{\mathbb{Q}}(X_{kt})^{\perp}]^{\text{Hg}(H^2(X_{kt}, \mathbb{Q}))} \xrightarrow{\sim} \text{End}_{\text{Hg}(H^2(X_{kt}, \mathbb{Q}))} \text{NS}_{\mathbb{Q}}(X_{kt})^{\perp}.$$

Из равенства $h^{1,0}(X_{kt}) = 0$ и результатов Окамото [33] следует, что алгебраический класс диагонали не является линейной комбинацией классов пересечений дивизоров на квадрате $X_{kt} \times X_{kt}$ поверхности X_{kt} , поэтому для многообразия $X_{kt} \times X_{kt}$ верна гипотеза Ходжа.

Очевидно, что

$$\begin{aligned}
 & [\mathrm{NS}_{\mathbb{Q}}(X_{1t}) \otimes \mathrm{NS}_{\mathbb{Q}}(X_{2t})^{\perp}]^{\mathrm{Hg}(H^2(X_{1t}, \mathbb{Q})) \times \mathrm{Hg}(H^2(X_{2t}, \mathbb{Q}))} \\
 &= \mathrm{NS}_{\mathbb{Q}}(X_{1t}) \otimes [\mathrm{NS}_{\mathbb{Q}}(X_{2t})^{\perp}]^{\mathrm{Hg}(H^2(X_{2t}, \mathbb{Q}))} = 0; \\
 & [\mathrm{NS}_{\mathbb{Q}}(X_{1t})^{\perp} \otimes \mathrm{NS}_{\mathbb{Q}}(X_{2t})^{\perp}]^{\mathrm{Hg}(H^2(X_{1t}, \mathbb{Q})) \times \mathrm{Hg}(H^2(X_{2t}, \mathbb{Q}))} \\
 &= [\mathrm{NS}_{\mathbb{Q}}(X_{1t})^{\perp}]^{\mathrm{Hg}(H^2(X_{1t}, \mathbb{Q}))} \otimes [\mathrm{NS}_{\mathbb{Q}}(X_{2t})^{\perp}]^{\mathrm{Hg}(H^2(X_{2t}, \mathbb{Q}))} = 0; \\
 & [\mathrm{NS}_{\mathbb{Q}}(X_{1t})^{\perp} \otimes \mathrm{NS}_{\mathbb{Q}}(X_{2t})]^{\mathrm{Hg}(H^2(X_{1t}, \mathbb{Q})) \times \mathrm{Hg}(H^2(X_{2t}, \mathbb{Q}))} \\
 &= [\mathrm{NS}_{\mathbb{Q}}(X_{1t})^{\perp}]^{\mathrm{Hg}(H^2(X_{1t}, \mathbb{Q}))} \otimes \mathrm{NS}_{\mathbb{Q}}(X_{2t}) = 0,
 \end{aligned}$$

поэтому для многообразия

$$X_t \times X_t = X_{1t} \times X_{2t} \times X_{1t} \times X_{2t} \xrightarrow{\sim} X_{1t} \times X_{1t} \times X_{2t} \times X_{2t}$$

пространство

$$\begin{aligned}
 & H^4(X_{1t} \times X_{1t} \times X_{2t} \times X_{2t}, \mathbb{Q}) \cap H^{2,2}(X_{1t} \times X_{1t} \times X_{2t} \times X_{2t}, \mathbb{C}) \\
 &= [H^4(X_{1t} \times X_{1t} \times X_{2t} \times X_{2t}, \mathbb{Q})]^{\mathrm{Hg}(H^2(X_{1t}, \mathbb{Q})) \times \mathrm{Hg}(H^2(X_{2t}, \mathbb{Q}))} \\
 &= [H^4(X_{1t} \times X_{1t}, \mathbb{Q})]^{\mathrm{Hg}(H^2(X_{1t}, \mathbb{Q}))} \\
 &\oplus [H^2(X_{1t} \times X_{1t}, \mathbb{Q}) \otimes H^2(X_{2t} \times X_{2t}, \mathbb{Q})]^{\mathrm{Hg}(H^2(X_{1t}, \mathbb{Q})) \times \mathrm{Hg}(H^2(X_{2t}, \mathbb{Q}))} \\
 &\oplus [H^4(X_{2t} \times X_{2t}, \mathbb{Q})]^{\mathrm{Hg}(H^2(X_{2t}, \mathbb{Q}))}
 \end{aligned}$$

порождается алгебраическими классами когомологий. В частности, класс Пуанкаре

$$\wp(H^2(X_t, \mathbb{Q})) \in [H^2(X_t, \mathbb{Q}) \otimes H^2(X_t, \mathbb{Q})] \cap H^{2,2}(X_t \times X_t, \mathbb{C})$$

алгебраический для всех точек $t \in C \setminus \Delta_{\text{countable}}$.

Если $\eta \in C$ – общая схемная точка кривой C , то можно считать (используя в случае необходимости разветвлённое накрытие $\tilde{C} \rightarrow C$ и замену базы $X \times_C \tilde{C} \rightarrow \tilde{C}$), что цикл Пуанкаре $\wp(H^2(X_{\bar{\eta}}, \mathbb{Q}))$ на общем геометрическом слое $X_{\bar{\eta}}$ определён над полем $\kappa(\eta) = k(C)$ рациональных функций кривой C . Пусть Z_{η} – цикл Пуанкаре на $X_{\eta} \times_{\kappa(\eta)} X_{\eta}$. Замыкание его компонент в топологии Зариского многообразия $X \times_C X$ даёт алгебраический цикл Z на $X \times_C X$, определяющий алгебраические циклы Z_t на $X_t \times X_t$ для всех $t \in C$. Каждый цикл Z_t определяет изоморфизм $H^6(X_t, \mathbb{Q}) \xrightarrow{\sim} H^2(X_t, \mathbb{Q})$ (это следствие теории редукции (специализации) алгебраических циклов [34] и леммы 1.5 в [31]); в нашем случае рассматриваются алгебраические циклы с коэффициентами из поля $\mathbb{Q}$.

Используя идею доказательства теоремы 8.3 в [31], можно легко показать, что существует алгебраический изоморфизм

$$H^1(C, R^6 \pi_* \mathbb{Q}) \xrightarrow{\sim} H^1(C, R^2 \pi_* \mathbb{Q}), \quad (2.22)$$

определённый алгебраическим классом $\iota_*(\mathrm{cl}_{X \times_C X}(Z))$, где $\iota : X \times_C X \hookrightarrow X \times X$ – каноническое вложение.

Действительно, алгебраический цикл Z_t коразмерности 2 на $X_t \times X_t$ определяет изоморфизм $H^6(X_t, \mathbb{Q}) \xrightarrow{\sim} H^2(X_t, \mathbb{Q})$, являющийся композицией

$$\begin{aligned} H^6(X_t, \mathbb{Q}) &\xrightarrow{\text{pr}_{1t}^*} H^6(X_t, \mathbb{Q}) \otimes H^0(X_t, \mathbb{Q}) \xrightarrow{\sim \text{cl}_{X_t \times X_t}(Z_t)} \\ &H^8(X_t, \mathbb{Q}) \otimes H^2(X_t, \mathbb{Q}) \xrightarrow{\text{pr}_{2t*}} H^2(X_t, \mathbb{Q}), \end{aligned} \quad (2.23)$$

где pr_{1t} и pr_{2t} – канонические проекции $X_t \times X_t \rightarrow X_t$.

С другой стороны, цикл Z имеет коразмерность 2 на $X \times_C X$. Последовательность (2.23) допускает естественную глобализацию

$$\begin{aligned} R^6 \pi_* \mathbb{Q} &\xrightarrow{q_1^*} R^6 \pi_* \mathbb{Q} \otimes \pi_* \mathbb{Q} \xrightarrow{\sim \text{cl}_{X \times_C X}(Z)} R^{10} \tau_* \mathbb{Q} = \\ &\bigoplus_{i=0}^{10} R^i \pi_* \mathbb{Q} \otimes R^{10-i} \pi_* \mathbb{Q} \longrightarrow R^8 \pi_* \mathbb{Q} \otimes R^2 \pi_* \mathbb{Q} \xrightarrow{q_{2*}} R^2 \pi_* \mathbb{Q}, \end{aligned}$$

где q_1, q_2 – канонические проекции $X \times_C X \rightarrow X$ и $\tau = \pi q_i : X \times_C X \rightarrow C$ – структурный морфизм. Учитывая, что $q_{2*}(R^i \pi_* \mathbb{Q} \otimes R^{10-i} \pi_* \mathbb{Q}) = 0$ для всех $i \neq 8$ [31, формула (1.2)], мы видим, что определён изоморфизм

$$H^1(C, R^6 \pi_* \mathbb{Q}) \xrightarrow[\sim]{x \mapsto q_{2*}(q_1^*(x) \sim \text{cl}_{X \times_C X}(Z))} H^1(C, R^2 \pi_* \mathbb{Q}).$$

Заметим, что $q_k = \text{pr}_k \iota$, где $\text{pr}_k : X \times X \rightarrow X$ – каноническая проекция и $\iota : X \times_C X \hookrightarrow X \times X$ – каноническое вложение. В силу формулы проекции имеем:

$$\begin{aligned} q_{2*}(q_1^*(x) \sim \text{cl}_{X \times_C X}(Z)) &= [\text{pr}_2 \iota]_*((\text{pr}_1 \iota)^* x \sim \text{cl}_{X \times_C X}(Z)) = \\ &= \text{pr}_{2*} \iota_*(\iota^* \text{pr}_1^* x \sim \text{cl}_{X \times_C X}(Z)) = \text{pr}_{2*}(\text{pr}_1^* x \sim \iota_*(\text{cl}_{X \times_C X}(Z))), \end{aligned}$$

поэтому алгебраический класс $\iota_*(\text{cl}_{X \times_C X}(Z))$ определяет алгебраический изоморфизм

$$H^1(C, R^6 \pi_* \mathbb{Q}) \xrightarrow[\sim]{x \mapsto q_{2*}(q_1^*(x) \sim \text{cl}_{X \times_C X}(Z))} H^1(C, R^2 \pi_* \mathbb{Q}).$$

С другой стороны, вырожденная последовательность Лере $E_2^{p,q} = H^p(C, R^q \pi_* \mathbb{Q}) \Rightarrow H^*(X, \mathbb{Q})$ даёт точные последовательности

$$\begin{aligned} 0 \rightarrow H^2(C, R^1 \pi_* \mathbb{Q}) \rightarrow \text{Ker}[H^3(X, \mathbb{Q}) \rightarrow H^0(C, R^3 \pi_* \mathbb{Q})] \rightarrow H^1(C, R^2 \pi_* \mathbb{Q}) \rightarrow 0, \\ 0 \rightarrow H^2(C, R^5 \pi_* \mathbb{Q}) \rightarrow \text{Ker}[H^7(X, \mathbb{Q}) \rightarrow H^0(C, R^7 \pi_* \mathbb{Q})] \rightarrow H^1(C, R^6 \pi_* \mathbb{Q}) \rightarrow 0. \end{aligned}$$

Поскольку $R^1 \pi_* \mathbb{Q} = R^3 \pi_* \mathbb{Q} = R^7 \pi_* \mathbb{Q} = 0$, то мы получаем канонические изоморфизмы

$$H^3(X, \mathbb{Q}) \xrightarrow{\sim} H^1(C, R^2 \pi_* \mathbb{Q}), \quad H^7(X, \mathbb{Q}) \xrightarrow{\sim} H^1(C, R^6 \pi_* \mathbb{Q}).$$

Значит, из (2.22) следует, что алгебраический класс $\iota_*(\text{cl}_{X \times_C X}(Z))$ даёт алгебраический изоморфизм

$$H^7(X, \mathbb{Q}) \xrightarrow{\sim} H^3(X, \mathbb{Q}).$$

Кроме того, в силу леммы 1.5 в [31], существуют алгебраические изоморфизмы

$$\begin{aligned} H^{10}(X, \mathbb{Q}) &\xrightarrow{\sim} H^0(X, \mathbb{Q}), & H^9(X, \mathbb{Q}) &\xrightarrow{\sim} H^1(X, \mathbb{Q}), \\ H^8(X, \mathbb{Q}) &\xrightarrow{\sim} H^2(X, \mathbb{Q}), & H^6(X, \mathbb{Q}) &\xrightarrow{\sim} H^4(X, \mathbb{Q}), \end{aligned}$$

определённые соответственно алгебраическими циклами Пуанкаре

$$\begin{aligned} \wp(H^0(X, \mathbb{Q})) &\in H^0(X, \mathbb{Q}) \otimes H^0(X, \mathbb{Q}), \\ \wp(H^1(X, \mathbb{Q})) &\in H^1(X, \mathbb{Q}) \otimes H^1(X, \mathbb{Q}) \end{aligned}$$

(цикл $\wp(H^1(X, \mathbb{Q}))$ алгебраический, потому что

$$\wp(H^1(X, \mathbb{Q})) \in [H^1(X, \mathbb{Q}) \otimes H^1(X, \mathbb{Q})] \cap H^{1,1}(X \times X, \mathbb{C})$$

является циклом Ходжа [31, п. 1.4] и

$$[H^1(X, \mathbb{Q}) \otimes H^1(X, \mathbb{Q})] \cap H^{1,1}(X \times X, \mathbb{C}) \hookrightarrow \text{NS}_{\mathbb{Q}}(X \times X)$$

по теореме Лефшеца для дивизоров на многообразии $X \times X$);

$$\begin{aligned} \wp(H^2(X, \mathbb{Q})) &\in H^2(X, \mathbb{Q}) \otimes H^2(X, \mathbb{Q}), \\ \wp(H^4(X, \mathbb{Q})) &\in H^4(X, \mathbb{Q}) \otimes H^4(X, \mathbb{Q}) \end{aligned}$$

(напомним, что в рассматриваемом случае $Z = \emptyset$, поэтому в силу формулы (2.16) имеем равенство $H^4(X, \mathbb{Q}) = H^0(C', R^4\pi'_*\mathbb{Q})$, где по доказанному выше пространство $H^0(C', R^4\pi'_*\mathbb{Q})$ порождается образами алгебраических классов в $H^4(X, \mathbb{Q})$, лежащих в $\text{NS}_{\mathbb{Q}}(X_1) \cup \text{NS}_{\mathbb{Q}}(X_2)$, $H^4(X_1, \mathbb{Q})$, $H^4(X_2, \mathbb{Q})$, причем $\text{NS}_{\mathbb{Q}}(X_k)$ отождествляется с образом $\text{NS}_{\mathbb{Q}}(X_k)$ при каноническом отображении $q_k^* : H^2(X_k, \mathbb{Q}) \hookrightarrow H^2(X, \mathbb{Q})$, определенном проекцией $X = X_1 \times_C X_2 \xrightarrow{q_k} X_k$, и $H^4(X_k, \mathbb{Q})$ отождествляется с образом $H^4(X_k, \mathbb{Q})$ при каноническом отображении $q_k^* : H^4(X_k, \mathbb{Q}) \hookrightarrow H^4(X, \mathbb{Q})$; в итоге из результатов п. 2.7 и леммы п. 2.6 следует, что пространства $H^2(X, \mathbb{Q})$ и $H^4(X, \mathbb{Q})$ порождаются классами алгебраических циклов). Значит $B(X)$ верна в силу теоремы п. 2.9 в [35]. Теорема п. 2.1 доказана.

3. Об алгебраических циклах на расслоенном квадрате

3.1. Формулировка теоремы о циклах на расслоенном квадрате

Теорема. Пусть C - гладкая проективная кривая над полем комплексных чисел, $\pi_1 : X_1 \rightarrow C$ - гладкое проективное семейство поверхностей с $h^{2,0}(X_{1s}) = 1$, $h^{1,0}(X_{1s}) = 0$ и непостоянным отображением периодов, ассоциированным с вариацией структур Ходжа $R^2\pi_{k*}\mathbb{Q}$, причем для общего геометрического слоя X_{1s} число $b_2(X_{1s}) - \text{rank NS}(X_{1s}) = p_1$ является нечётным простым. Тогда для расслоенного квадрата $X = X_1 \times_C X_1$ верны гипотеза Ходжа и стандартная гипотеза Гротендика $B(X)$ типа Лефшеца об алгебраичности операторов $*$ и Λ теории Ходжа.

3.2. Доказательство стандартной гипотезы для расслоенного квадрата

Гипотеза $B(X)$ легко проверяется с помощью метода из п. 2.9, потому что в рассматриваемом случае для общей в смысле Ходжа точки $s \in C \setminus \Delta_{\text{countable}}$ группа Ходжа $\text{Hg}(H^2(X_{1s}, \mathbb{Q}))$ имеет тип $B_{\frac{p_1-1}{2}}$ и, следовательно, $\mathbb{Q}$ -пространство

$$\begin{aligned} H^4(X_{1s} \times X_{1s} \times X_{1s} \times X_{1s}, \mathbb{Q}) \cap H^{2,2}(X_{1s} \times X_{1s} \times X_{1s} \times X_{1s}, \mathbb{C}) \\ = H^4(X_{1s} \times X_{1s} \times X_{1s} \times X_{1s}, \mathbb{Q})^{\text{Hg}(H^2(X_{1s}, \mathbb{Q}))} \end{aligned}$$

порождается классами алгебраических циклов, происходящих из пересечений дивизоров на произведениях вида $X_{1s} \times X_{1s}$ и диагоналей $\Delta_{X_{1s}} \hookrightarrow X_{1s} \times X_{1s}$; в частности, класс Пуанкаре

$$\wp(H^2(X_s, \mathbb{Q})) \in [H^2(X_s, \mathbb{Q}) \otimes H^2(X_s, \mathbb{Q})] \cap H^{2,2}(X_s \times X_s, \mathbb{C})$$

алгебраический для всех точек $s \in C \setminus \Delta_{\text{countable}}$.

Поэтому остается доказать гипотезу Ходжа для $X = X_1 \times_C X_1$.

В силу леммы п. 2.6 и результатов п. 2.7 – 2.8 имеем: $H^2(X_1, \mathbb{Q}) = \text{NS}_{\mathbb{Q}}(X_1)$, $H^4(X_1, \mathbb{Q})$, $H^2(C, R^2\pi_{1*}\mathbb{Q})$ и $H^2(X, \mathbb{Q})$ порождены классами алгебраических циклов.

3.3. Доказательство гипотезы Ходжа для расслоенного квадрата

В силу (2.18) имеем:

$$[\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\pi_1(C,s)} = 0. \quad (3.1)$$

С другой стороны, $\pi_1(C, s)$ оставляет неподвижными элементы из $\text{NS}_{\mathbb{Q}}(X_{1s})$ в силу включения $G_1^{(2)} \hookrightarrow \text{Hg}(X_{1s})$. Значит, из (3.1) следует, что

$$\begin{aligned} H^0(C, R^2\pi_{1*}\mathbb{Q} \otimes R^2\pi_{1*}\mathbb{Q}) &\xrightarrow{\sim} [H^2(X_{1s}, \mathbb{Q}) \otimes H^2(X_{1s}, \mathbb{Q})]^{\pi_1(C,s)} = \\ &= [[\text{NS}_{\mathbb{Q}}(X_{1s}) \oplus \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\otimes 2}]^{\pi_1(C,s)} = \\ &= [\text{NS}_{\mathbb{Q}}(X_{1s}) \otimes \text{NS}_{\mathbb{Q}}(X_{1s})]^{\pi_1(C,s)} \oplus [\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \text{NS}_{\mathbb{Q}}(X_{1s})]^{\pi_1(C,s)} \oplus \\ &\oplus [\text{NS}_{\mathbb{Q}}(X_{1s}) \otimes \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\pi_1(C,s)} \oplus [\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\pi_1(C,s)} = \\ &= [\text{NS}_{\mathbb{Q}}(X_{1s}) \otimes \text{NS}_{\mathbb{Q}}(X_{1s})] \oplus [\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\pi_1(C,s)}. \end{aligned} \quad (3.2)$$

Очевидно, что

$$[\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\pi_1(C,s)} \xrightarrow{\sim} \text{Hom}_{\pi_1(C,s)}(\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}, \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}).$$

Поскольку $\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}$ является абсолютно неприводимым $\pi_1(C, s)$ -модулем степени p_1 , то в силу леммы Шура имеем

$$\text{Hom}_{\pi_1(C,s)}(\text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}, \text{NS}_{\mathbb{Q}}(X_{1s})^{\perp}) = \mathbb{Q}.$$

В итоге

$$\dim_{\mathbb{Q}}[\mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\pi_1(C,s)} = 1. \quad (3.3)$$

Нормальная подгруппа $G_1^{(2)}$ либо совпадает с простой группой Ходжа $\mathrm{Hg}(X_{1s})$, либо равна 1 (второй вариант невозможен в силу предложения п. 1.11, так как семейство $\pi_1 : X_1 \rightarrow C$ определяет непостоянное отображение периодов, ассоциированное с вариацией структур Ходжа $R^2\pi_{1*}\mathbb{Q}$). Значит,

$$\begin{aligned} G_1^{(2)} &= \mathrm{Hg}(X_{1s}), \\ H^0(C, R^2\pi_{1*}\mathbb{Q}) &\xrightarrow{\sim} H^2(X_{1s}, \mathbb{Q})^{\pi_1(C,s)} = H^2(X_{1s}, \mathbb{Q})^{\mathrm{Hg}(X_{1s})} = \mathrm{NS}_{\mathbb{Q}}(X_{1s}), \end{aligned} \quad (3.4)$$

поэтому

$$\begin{aligned} [\mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\pi_1(C,s)} &= [\mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{G_1^{(2)}} = \\ &= [\mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp} \otimes \mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp}]^{\mathrm{Hg}(X_{1s})} \hookrightarrow [H^2(X_{1s}, \mathbb{Q}) \otimes H^2(X_{1s}, \mathbb{Q})]^{\mathrm{Hg}(X_{1s})} \\ &\hookrightarrow H^4(X_{1s} \times X_{1s}, \mathbb{Q}) \cap H^{2,2}(X_{1s} \times X_{1s}, \mathbb{C}). \end{aligned}$$

Как показано в п. 2.9, пространство $H^4(X_{1s} \times X_{1s}, \mathbb{Q}) \cap H^{2,2}(X_{1s} \times X_{1s}, \mathbb{C})$ порождается классами пересечений дивизоров и классом диагонали $\Delta_{X_{1s}} \hookrightarrow X_{1s} \times X_{1s}$, потому что для *общей в смысле Ходжа* точки $s \in C$, пространство $\mathrm{NS}_{\mathbb{Q}}(X_{1s})^{\perp}$ является абсолютно неприводимым $\mathrm{Hg}(X_{1s})$ -модулем.

Заметим, что в силу результатов Окамото ([33, следствие 2.5]) класс

$$\mathrm{cl}_{X_{1s} \times X_{1s}}(\Delta_{X_{1s}}) \in H^4(X_{1s} \times X_{1s}, \mathbb{Q})$$

не когомологичен сумме пересечений дивизоров с коэффициентами из $\mathbb{Q}$. Значит, образ

$$\mathrm{cl}_{X_1 \times_C X_1}(\Delta_{X_1}) \in H^4(X_1 \times_C X_1, \mathbb{Q}) = H^4(X, \mathbb{Q})$$

при каноническом сюръективном морфизме $H^4(X, \mathbb{Q}) \rightarrow H^0(C, R^4\pi_*\mathbb{Q})$ не когомологичен сумме пересечений дивизоров, потому что

$$H^0(C, R^4\pi_*\mathbb{Q}) \xrightarrow{\sim} H^4(X_{1s} \times X_{1s}, \mathbb{Q})^{\pi_1(C,s)} \subset H^4(X_{1s} \times X_{1s}, \mathbb{Q})$$

является образом $H^4(X, \mathbb{Q})$ при каноническом морфизме ограничения

$$H^4(X, \mathbb{Q}) \rightarrow H^4(X_{1s} \times X_{1s}, \mathbb{Q}).$$

Обозначим через Δ_{X_1} диагональ в $X_1 \times_C X_1 = X$. Тогда алгебраический класс $\mathrm{cl}_X(\Delta_{X_1})$ принадлежит $H^4(X, \mathbb{Q})$, потому что Δ_{X_1} имеет коразмерность 2 в X .

У нас есть точная последовательность $\mathbb{Q}$ -структур Ходжа

$$0 \rightarrow H^2(C, R^2\pi_*\mathbb{Q}) \rightarrow \mathrm{Ker}[H^4(X, \mathbb{Q}) \rightarrow H^0(C, R^4\pi_*\mathbb{Q})] \rightarrow H^1(C, R^3\pi_*\mathbb{Q}) \rightarrow 0$$

(формула (2.6) при $i = 2$). Поэтому

$$H^2(C, R^2\pi_*\mathbb{Q}) = \mathrm{Ker}[H^4(X, \mathbb{Q}) \rightarrow H^0(C, R^4\pi_*\mathbb{Q})]$$

и имеется точная последовательность

$$0 \rightarrow H^2(C, R^2\pi_*\mathbb{Q}) \rightarrow H^4(X, \mathbb{Q}) \rightarrow H^0(C, R^4\pi_*\mathbb{Q}) \rightarrow 0.$$

Отображение $H^4(X, \mathbb{Q}) \rightarrow H^0(C, R^4\pi_*\mathbb{Q})$ сюръективно по теореме Делиня [29, теорема 4.1.1]. Кроме того,

$$H^0(C, R^4\pi_*\mathbb{Q}) = H^0(C, R^4\pi_{1*}\mathbb{Q}) \oplus H^0(C, R^2\pi_{1*}\mathbb{Q} \otimes R^2\pi_{1*}\mathbb{Q}) \oplus H^0(C, R^4\pi_{1*}\mathbb{Q});$$

при этом $H^0(C, R^4\pi_{1*}\mathbb{Q})$ входит в точную последовательность

$$0 \rightarrow H^2(C, R^2\pi_{1*}\mathbb{Q}) \rightarrow H^4(X_1, \mathbb{Q}) \rightarrow H^0(C, R^4\pi_{1*}\mathbb{Q}) \rightarrow 0$$

(формула (2.3) при $i = 2$), пространство $H^4(X_1, \mathbb{Q})$ порождено алгебраическими циклами и даже классами пересечений дивизоров (п. 2.7). В частности, $H^0(C, R^4\pi_{1*}\mathbb{Q})$ порождается классами пересечений дивизоров на X_1 и, следовательно, на X .

Из (3.2) – (3.4) следует, что $H^0(C, R^4\pi_*\mathbb{Q})$ порождено пересечениями дивизоров на X и образом класса $\text{cl}_{X_1 \times_C X_1}(\Delta_{X_1}) = \text{cl}_X(\Delta_{X_1})$ в $H^0(C, R^4\pi_*\mathbb{Q})$.

Осталось заметить, что по формуле Кюннета

$$R^2\pi_*\mathbb{Q} = R^2\pi_{1*}\mathbb{Q} \oplus R^2\pi_{1*}\mathbb{Q},$$

поэтому $H^2(C, R^2\pi_*\mathbb{Q}) = H^2(C, R^2\pi_{1*}\mathbb{Q}) \oplus H^2(C, R^2\pi_{1*}\mathbb{Q})$ порождается классами пересечений дивизоров на множителях расслоенного произведения $X = X_1 \times_C X_1$ и, следовательно, на X . Теорема п. 3.1 доказана.

Список литературы / References

- [1] Hodge W. V.D., “The topological invariants of algebraic varieties”, *Proceedings of International Congress of Mathematicians*, **1** (1952), 182–192.
- [2] Танкеев С. Г., “Циклы на простых абелевых многообразиях простой размерности”, *Изв. АН СССР. Сер. матем.*, **46**:1 (1982), 155–170; English transl.: Tankeev S. G., “Cycles on simple abelian varieties of prime dimension”, *Mathematics of the USSR-Izvestiya*, **20**:1 (1983), 157–171.
- [3] Gordon B. B., “A survey of the Hodge conjecture for Abelian varieties, Appendix in: J.D. Lewis”, *A survey of the Hodge conjecture*, **10**, Second edition, CRM Monograph Series, American Mathematical Society, Providence, RI, 1999, 297–356.
- [4] Никольская О. В., “Об алгебраических циклах на расслоенном произведении семейств K3 поверхностей”, *Изв. РАН. Сер. матем.*, **77**:1 (2013), 145–164; English transl.: Nikolskaya O. V., “On algebraic cycles on a fibre product of families of K3 surfaces”, *Izv. Math.*, **77**:1 (2013), 143–162.
- [5] Никольская О. В., “О геометрии гладкой модели расслоенного произведения семейств K3 поверхностей”, *Матем. сб.*, **205**:2 (2014), 123–130; English transl.: Nikolskaya O. V., “On the geometry of a smooth model of a fibre product of families of K3 surfaces”, *Sbornik: Mathematics*, **205**:2 (2014), 269–276.
- [6] Никольская О. В., “Об алгебраических классах когомологий на гладкой модели расслоенного произведения семейств K3 поверхностей”, *Матем. заметки*, **96**:5 (2014), 738–746; English transl.: Nikolskaya O. V., “On algebraic cohomology classes on a smooth model of a fiber product of families of K3 surfaces”, *Math. Notes*, **96**:5 (2014), 745–752.
- [7] Grothendieck A., “Standard conjectures on algebraic cycles”, *Algebraic Geometry*, Oxford University Press, London, 1969, 193–199, International Colloquium (Bombay, 1968).

- [8] Lieberman D. I., “Numerical and homological equivalence of algebraic cycles on Hodge manifolds”, *Amer. J. Math.*, **90**:2 (1968), 366–374.
- [9] Танкеев С. Г., “О стандартной гипотезе типа Лефшеца для комплексных проективных трехмерных многообразий. II”, *Изв. РАН. Сер. матем.*, **75**:5 (2011), 177–194; English transl.: Tankeev S. G., “On the standard conjecture of Lefschetz type for complex projective threefolds. II”, *Izv. Math.*, **75**:5 (2011), 1047–1062.
- [10] Charles F., Markman E., “The standard conjectures for holomorphic symplectic varieties deformation equivalent to Hilbert schemes of K3 surfaces”, *Compos. Math.*, **149**:3 (2013), 481–494.
- [11] Танкеев С. Г., “О стандартной гипотезе для комплексных 4-мерных эллиптических многообразий и компактификаций минимальных моделей Нерона”, *Изв. РАН. Сер. матем.*, **78**:1 (2014), 181–214; English transl.: Tankeev S. G., “On the standard conjecture for complex 4-dimensional elliptic varieties and compactifications of Neron minimal models”, *Izv. Math.*, **78**:1 (2014), 169–200.
- [12] Deligne P., “Travaux de Griffiths”, *Séminaire Bourbaki 1969/70, Exposé 376*, Heidelberg, Benjamin, New York, 1971, 213–235.
- [13] Гриффитс Ф. А., “Периоды интегралов на алгебраических многообразиях: обзор основных результатов и обсуждение открытых проблем”, *УМН*, **25**:3 (1970), 175–234; English transl.: Griffiths Ph. A., “Periods of integrals on algebraic manifolds: summary of main results and discussion of open problems”, *Bull. Amer. Math. Soc.*, **76**:2 (1970), 228–296.
- [14] Griffiths Ph. A., “A transcendental method in algebraic geometry”, *Actes du Congrès international des mathématiciens (Nice, 1970)*, **1** (1971), 113–119.
- [15] Schmid W., “Variation of Hodge structure: the singularities of the period mapping”, *Invent. math.*, **22** (1973), 211–319.
- [16] Morrison D. R., “On the moduli of Todorov surfaces”, *Algebraic geometry and commutative algebra in honor of Masayoshi Nagata*, Kinokuniya C. Ltd., 1988, 313–356.
- [17] Pignatelli R., “Some (big) irreducible components of the moduli space of minimal surfaces of general type with $p_g = q = 1$ and $K^2 = 4$ ”, *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, **20**:3 (2009), 207–226.
- [18] Moonen B., “On the Tate and Mumford - Tate conjectures in codimension one for varieties with $h^{2,0} = 1$ ”, *arXiv*: 1504.05406v1, 21 Apr 2015, 1–45.
- [19] Zarhin Yu. G., “Hodge groups of K3 surfaces”, *Journal für die reine und angewandte Mathematik*, **341** (1983), 193–220.
- [20] Бурбаки Н., *Группы и алгебры Ли*, Мир, Москва, гл. 1–3 (1976); гл. 4–6 (1972); гл. 7–8 (1978); French transl.: Bourbaki N., *Groupes et algèbres de Lie, Chaps. 1–8*, Actualités Sci. Indust., nos. 1285, 1349, 1337, 1364, Hermann, Paris, 1971, 1972, 1968, 1975.
- [21] Мустафин Г. А., “Семейства алгебраических многообразий и инвариантные циклы”, *Изв. АН СССР. Сер. матем.*, **49**:5 (1985), 948–978; English transl.: Mustafin G. A., “Families of algebraic varieties and invariant cycles”, *Mathematics of the USSR-Izvestiya*, **27**:2 (1986), 251–278.
- [22] Бурбаки Н., *Элементы математики. Алгебра. Гл. VII – IX. Модули, кольца, формы*, Наука, Москва, 1966; French transl.: Bourbaki N., *Éléments de mathématique. Algèbre, Ch. IX, Formes sesquilineaires et formes quadratiques*, Hermann, Paris, 1959.
- [23] Хелгасон С., *Дифференциальная геометрия и симметрические пространства*, Мир, Москва, 1964; English transl.: Helgason S., *Differential geometry and symmetric spaces*, Academic Press, New York and London, 1962.
- [24] Зархин Ю. Г., “Веса простых алгебр Ли в когомологиях алгебраических многообразий”, *Изв. АН СССР. Сер. матем.*, **48**:2 (1984), 264–304; English transl.: Zarhin Yu. G., “Weights of simple Lie algebras in the cohomology of algebraic varieties”, *Math. USSR-Izv.*, **24**:2 (1985), 245–281.

- [25] Танкеев С. Г., “Моноидальные преобразования и гипотезы об алгебраических циклах”, *Изв. РАН. Сер. матем.*, **71**:3 (2007), 197–224; English transl.: Tankeev S. G., “Monoidal transformation and conjectures on algebraic cycles”, *Izv. Math.*, **71**:3 (2007), 629–655.
- [26] Танкеев С. Г., “Об арифметике и геометрии общего гиперповерхностного сечения”, *Изв. РАН. Сер. матем.*, **66**:2 (2002), 173–204; English transl.: Tankeev S. G., “The arithmetic and geometry of a generic hypersurface section”, *Izv. Math.*, **66**:2 (2002), 393–424.
- [27] Kempf G. et al., *Toroidal embeddings*, I. Lecture Notes in Mathematics, **339**, Springer-Verlag, Berlin - New York, 1973, 209 pp.
- [28] Танкеев С. Г., “О стандартной гипотезе для комплексных абелевых схем над гладкими проективными кривыми”, *Изв. РАН. Сер. матем.*, **67**:3 (2003), 183–224; English transl.: Tankeev S. G., “On the standard conjecture for complex Abelian schemes over smooth projective curves”, *Izv. Math.*, **67**:3 (2003), 597–635.
- [29] Делинь П., “Теория Ходжа. II”, *Математика. Сборник переводов иностранных статей*, **17**:5 (1973), 3–56; French transl.: Deligne P., “Théorie de Hodge. II”, *Inst. Hautes Études Sci. Publ. Math.*, **40** (1971), 5–57.
- [30] Zucker S., “Hodge theory with degenerating coefficients: L_2 cohomology in the Poincaré metric”, *Ann. Math. (2)*, **109**:3 (1979), 415–476.
- [31] Танкеев С. Г., “О стандартной гипотезе типа Лефшеца для комплексных проективных трехмерных многообразий”, *Изв. РАН. Сер. матем.*, **74**:1 (2010), 175–196; English transl.: Tankeev S. G., “On the standard conjecture of Lefschetz type for complex projective threefolds”, *Izv. Math.*, **74**:1 (2010), 167–187.
- [32] Deligne P., “Théorie de Hodge. III”, *Inst. Hautes Études Sci. Publ. Math.*, **44** (1974), 5–77.
- [33] Okamoto M., “On a certain decomposition of 2-dimensional cycles on a product of two algebraic surfaces”, *Proceeding of Japan Academy, Series A*, **57**:6 (1981), 321–325.
- [34] Shimura G., “Reduction of algebraic varieties with respect to a discrete valuation of the basic field”, *Amer. J. Math.*, **77**:1 (1955), 134–176.
- [35] Kleiman S. L., “Algebraic cycles and the Weil conjectures”, *Dix exposés sur la cohomologie des schémas*, North-Holland, Amsterdam; Masson, Paris, 1968, 359–386.

Nikol'skaya O. V., "On Algebraic Cycles on Fibre Products of Non-isotrivial Families of Regular Surfaces with Geometric Genus 1", *Modeling and Analysis of Information Systems*, **23**:4 (2016), 440–465.

DOI: 10.18255/1818-1015-2016-4-440-465

Abstract. Let $\pi_k : X_k \rightarrow C$ ($k = 1, 2$) be a projective family of surfaces (possibly with degenerations) over a smooth projective curve C . Assume that the discriminant loci $\Delta_k = \{\delta \in C \mid \text{Sing}(X_{k\delta}) \neq \emptyset\}$ ($k = 1, 2$) are disjoint, $h^{2,0}(X_{ks}) = 1$, $h^{1,0}(X_{ks}) = 0$ for any smooth fibre X_{ks} and the period map associated with the variation of Hodge structures $R^2\pi'_{k*}\mathbb{Q}$ (where $\pi'_k : X'_k \rightarrow C \setminus \Delta_k$ is a smooth part of the morphism π_k), is non-constant. If for generic geometric fibres X_{1s} and X_{2s} the following conditions hold:

- (i) $b_2(X_{1s}) - \text{rank NS}(X_{1s})$ is an odd integer;
- (ii) $b_2(X_{1s}) - \text{rank NS}(X_{1s}) \neq b_2(X_{2s}) - \text{rank NS}(X_{2s})$,

then for any smooth projective model X of the fibre product $X_1 \times_C X_2$ the Hodge conjecture on algebraic cycles is true.

If, besides, the morphisms π_k are smooth, $p_k = b_2(X_{ks}) - \text{rank NS}(X_{ks})$ ($k = 1, 2$) are odd prime numbers and $p_1 \neq p_2$, then for $X_1 \times_C X_2$ and for the fibre square $X_1 \times_C X_1$ the Hodge conjecture and the Grothendieck standard conjecture on algebraicity of operators $*$ and Λ of Hodge theory hold as well.

This result yields new examples of smooth projective 5-dimensional varieties satisfying the Hodge and the Grothendieck conjectures, because one can take as smooth fibres of the morphism $\pi_k : X_k \rightarrow C$

some $K3$ surfaces, minimal regular surfaces of general type (of Kodaira dimension $\kappa = 2$) with geometric genus 1 belonging to one of the following types: (a) surfaces with $K^2 \leq 2$; (b) surfaces with $3 \leq K^2 \leq 8$, whose moduli are in the same component of the space of moduli as Todorov surface; (c) surfaces with $K^2 = 3$ with torsion of the Picard group $\mathbb{Z}/3\mathbb{Z}$.

Keywords: Hodge conjecture, standard conjecture, fibre product, Hodge group, Poincaré cycle

On the authors:

Nikol'skaya Olga Vladimirovna, orcid.org/0000-0002-6742-8453, PhD,
A.G. and N.G. Stoletov Vladimir State University,
Gorky str., 87, Vladimir, 600000, Russia, e-mail: papichonok@yandex.ru.

Acknowledgments:

This work was supported by the Russian Foundation for basic research under the Grant No 16-31-00266.

©Смирнов А. В., 2016

DOI: 10.18255/1818-1015-2016-4-466-478

УДК 519.179.2, 519.854.3

Сетевая модель для задачи целочисленного сбалансирования четырехмерной матрицы

Смирнов А. В.

получена 18 апреля 2016

Аннотация.

Рассматривается задача целочисленного сбалансирования четырехмерной матрицы. В исходной вещественной матрице элементы внутренней части (все четыре индекса больше нуля) просуммированы по каждому направлению и каждому плоскому и трехмерному сечению матрицы, а также найдена общая сумма. Данные суммы размещаются в элементах матрицы, у которых один или несколько индексов равны нулю (в соответствии с направлениями суммирования). Ищется целочисленная матрица той же структуры, получаемая из исходной заменой элементов на округления до целого сверху или целого снизу. При этом элемент с четырьмя нулевыми индексами получается по обычным правилам округления.

В статье рассматривается также задача о наибольшем кратном потоке в сети произвольной натуральной кратности k . Определяется три типа дуг в сети: обычная дуга, кратная дуга, мультидуга. Каждая кратная и мультидуга представляет собой объединение k связанных дуг, согласованных между собой. Задаются правила построения сети. Вводится понятие делимой сети и ряд связанных определений.

Определяются общие принципы сведения задачи целочисленного сбалансирования l -мерной матрицы ($l \geq 3$) к задаче о максимальном потоке в делимой кратной сети кратности k .

Задаются правила сведения четырехмерной задачи сбалансирования к задаче о наибольшем потоке в сети кратности 5. Для этой сети формулируется алгоритм нахождения максимального потока, удовлетворяющего условиям разрешимости задачи сбалансирования.

Ключевые слова: целочисленное сбалансирование, кратные сети, кратные потоки, делимые сети, NP -полнота, обобщенный алгоритм пометок, четырехмерные матрицы

Для цитирования: Смирнов А. В., "Сетевая модель для задачи целочисленного сбалансирования четырехмерной матрицы", *Моделирование и анализ информационных систем*, **23:4** (2016), 466–478.

Об авторах:

Смирнов Александр Валерьевич, orcid.org/0000-0002-0980-2507, канд. физ.-мат. наук, доцент кафедры теоретической информатики,

Ярославский государственный университет им. П.Г. Демидова,

ул. Советская, 14, г. Ярославль, 150000 Россия, e-mail: alexander_sm@mail.ru

Введение

Различные задачи целочисленного сбалансирования возникают в сфере управления, экономики, финансов. В частности, подобная задача ставится при планировании железнодорожных грузоперевозок. Имеется матричный план по отправке вагонов, который группируется по некоторым показателям (например, направление, тип вагона, владелец вагона и т.п.). Данный план составляется на месяц и естественно является целочисленным. Однако вагоны необходимо отправлять ежедневно. При делении на количество дней в месяце план перестает быть целочисленным. Поэтому возникает проблема такого округления основных параметров, чтобы суммирующие показатели не выходили за определенные рамки. Данный план может быть представлен в виде l -мерной матрицы, где l – это число показателей, по которым ведется суммирование. Соответствующая проблема округления основных параметров называется *задачей целочисленного сбалансирования l -мерной матрицы*.

Очевидно, что задача целочисленного сбалансирования l -мерной матрицы является частным случаем многоиндексных задач линейного программирования (см. [1] – [2]). Часто такие задачи являются NP -полными (например, рассмотренные в работе [3] аксиальная и планарная трехиндексная задача), однако существуют и полиномиально разрешимые многоиндексные задачи (например, трехиндексные задачи линейного программирования с вложенной структурой, см. [4]).

В работах [5] – [6] предложено сведение двухиндексной задачи целочисленного сбалансирования матрицы ($l = 2$) к классической задаче о наибольшем потоке в транспортной сети (см. [7]). Соответственно, задача целочисленного сбалансирования двумерной матрицы является полиномиально разрешимой. В случае же трех и более индексов такое сведение невозможно. Более того, задача целочисленного сбалансирования трехмерной матрицы является NP -полной (см. [8]). При доказательстве соответствующего утверждения строится полиномиальное сведение классической NP -полной задачи о трехмерном сочетании (см. [9] – [10]) к трехиндексной задаче сбалансирования.

Цель данной работы – обобщение для случая четырехмерной задачи сбалансирования результатов, полученных ранее для трехмерной задачи (см. [8], [11] – [13]), а также получение модели кратной сети для решения четырехмерной задачи.

1. Постановка задачи

Имеется четырехмерная вещественная матрица A с элементами $a_{ijpq} \geq 0$ ($i \in \overline{0, n}$, $j \in \overline{0, m}$, $p \in \overline{0, t}$, $q \in \overline{0, s}$), для которых выполнены условия баланса:

$$a_{0000} = \sum_{i=1}^n \sum_{j=1}^m \sum_{p=1}^t \sum_{q=1}^s a_{ijpq};$$

$$a_{i000} = \sum_{j=1}^m \sum_{p=1}^t \sum_{q=1}^s a_{ijpq} \quad (i \in \overline{1, n}); \quad a_{j000} = \sum_{i=1}^n \sum_{p=1}^t \sum_{q=1}^s a_{ijpq} \quad (j \in \overline{1, m});$$

$$a_{00p0} = \sum_{i=1}^n \sum_{j=1}^m \sum_{q=1}^s a_{ijpq} \quad (p \in \overline{1, t}); \quad a_{000q} = \sum_{i=1}^n \sum_{j=1}^m \sum_{p=1}^t a_{ijpq} \quad (q \in \overline{1, s});$$

$$\begin{aligned}
a_{ij00} &= \sum_{p=1}^t \sum_{q=1}^s a_{ijpq} \quad (i \in \overline{1, n}, j \in \overline{1, m}); & a_{i0p0} &= \sum_{j=1}^m \sum_{q=1}^s a_{ijpq} \quad (i \in \overline{1, n}, p \in \overline{1, t}); \\
a_{i00q} &= \sum_{j=1}^m \sum_{p=1}^t a_{ijpq} \quad (i \in \overline{1, n}, q \in \overline{1, s}); & a_{0jp0} &= \sum_{i=1}^n \sum_{q=1}^s a_{ijpq} \quad (j \in \overline{1, m}, p \in \overline{1, t}); \\
a_{0j0q} &= \sum_{i=1}^n \sum_{p=1}^t a_{ijpq} \quad (j \in \overline{1, m}, q \in \overline{1, s}); & a_{00pq} &= \sum_{i=1}^n \sum_{j=1}^m a_{ijpq} \quad (p \in \overline{1, t}, q \in \overline{1, s}); \\
a_{ijp0} &= \sum_{q=1}^s a_{ijpq} \quad (i \in \overline{1, n}, j \in \overline{1, m}, p \in \overline{1, t}); & a_{ij0q} &= \sum_{p=1}^t a_{ijpq} \quad (i \in \overline{1, n}, j \in \overline{1, m}, q \in \overline{1, s}); \\
a_{i0pq} &= \sum_{j=1}^m a_{ijpq} \quad (i \in \overline{1, n}, p \in \overline{1, t}, q \in \overline{1, s}); & a_{0jpq} &= \sum_{i=1}^n a_{ijpq} \quad (j \in \overline{1, m}, p \in \overline{1, t}, q \in \overline{1, s}).
\end{aligned}$$

Задача 1. Требуется найти целочисленную матрицу D той же размерности, для которой сохраняются условия баланса и выполняются условия:

$$d_{ijpq} \in \{\lfloor a_{ijpq} \rfloor, \lceil a_{ijpq} \rceil\} \quad (i \in \overline{0, n}, j \in \overline{0, m}, p \in \overline{0, t}, q \in \overline{0, s}); \quad d_{0000} = \lfloor a_{0000} + 0.5 \rfloor.$$

Как и в трехмерном случае (см. [11]), будем считать, что элементы a_{ijpq} ($i > 0, j > 0, p > 0, q > 0$) исходной матрицы A находятся в интервале $[0, 1)$.

Отметим, что ранее мы рассматривали трехмерные задачи сбалансирования с ограничениями первого (см. [11]) и второго (см. [12] – [13]) рода. В первом случае суммирующие элементы итоговой целочисленной матрицы могут отклоняться от исходных менее чем на 1, а во втором – менее чем на 2 (кроме элемента со всеми нулевыми индексами). Таким образом, задача 1 является задачей целочисленного сбалансирования четырехмерной матрицы с ограничениями первого рода.

Теорема 1. Если исходная матрица A содержит только рациональные элементы, то задача 1 является NP -полной.

При $s = 1$ задача 1 становится трехмерной и утверждение теоремы 1 следует из NP -полноты трехмерной задачи целочисленного сбалансирования (см. [8]).

2. Кратные сети и потоки. Необходимые определения

Для решения трехмерной задачи в работах [11] – [13] предложено их сведение к задаче поиска максимального потока в кратной сети целочисленного сбалансирования трехмерной матрицы (частный случай делимой кратной сети, см. [14] – [15]). Поскольку мы будем строить модель кратной сети и для задачи 1, необходимо напомним несколько определений, касающихся кратных сетей и потоков.

Определение 1. Кратной (транспортной) сетью натуральной кратности $k > 1$ назовем ориентированный мультиграф $G(X, U)$, между вершинами которого могут быть дуги одного из 3 видов:

1) обычная дуга u^o с пропускной способностью $c(u^o)$, поток по которой не связан с потоком по другим дугам; множество обычных дуг обозначим через U^o ;

2) кратная дуга u^k между двумя вершинами, которая состоит из k дуг одной ориентации с одинаковой пропускной способностью $c(u^k)$ и одинаковым потоком по каждой из них; множество кратных дуг обозначим через U^k ;

3) связанная дуга u между двумя вершинами, имеющая один общий конец с другими $k - 1$ связанными дугами; множество связанных дуг, выходящих из одной вершины или входящих в одну вершину, будем называть мультидугой u^m ; пропускная способность всех связанных дуг одной мультидуги одинакова; поток по каждой связанной дуге одной мультидуги одинаков; множество мультидуг обозначим через U^m .

Множество выходящих из вершины дуг может быть либо только кратными дугами, либо только одной мультидугой (k связанных дуг), либо только обычными дугами. Из источника x_0 сети выходят только кратные дуги, а в сток z сети входит только одна мультидуга. Если из вершины выходят связанные дуги мультидуги, то в нее обязательно входит кратная дуга. Если в вершину входит мультидуга, то из нее может выходить только кратная дуга. Все пропускные способности дуг целочисленны.

Определение 2. Кратным потоком по сети называется целочисленная функция, определенная на множестве дуг $U = U^o \cup U^k \cup U^m$, для которой выполнены условия неотрицательности, ограниченности (пропускными способностями дуг) и неразрывности потока (в каждой вершине).

Определение 3. Величиной кратного потока называется сумма φ_z входящего потока для стока z , равная сумме выходящего из источника потока. В силу того, что поток по каждой обычной дуге и по каждой связанной дуге каждой кратной и мультидуги должен быть целочислен, величина φ_z должна быть кратна k .

Обозначим через $c(u)$ пропускную способность дуги u , а через $f(u)$ – поток на ней.

Определение 4. Пусть имеется кратная сеть произвольной кратности k . Пусть при удалении всех мультидуг сеть распадается на $k + 2$ слабо связанных компонента, при этом одна компонента состоит только из вершины z , компонента, содержащая вершину x_0 , содержит только кратные дуги, а остальные k компонент содержат только обычные дуги. Если при этом каждая мультидуга имеет ровно один конец в каждой из k компонент, содержащих обычные дуги, то такую сеть мы будем называть делимой.

Обозначим через P_0 компоненту делимой сети, содержащую кратные дуги, через $P_1, \dots, P_k$ обозначим компоненты, содержащие обычные дуги.

Определение 5. Частью G_i делимой сети ($i \in \overline{1, k}$) назовем объединение соответствующей компоненты P_i с инцидентными ей связанными дугами всех мультидуг, кроме мультидуги с концом в z , а также с i -й связанной дугой каждой кратной дуги компоненты P_0 .

Обозначим начальные вершины мультидуги с концом в z через $z_1, \dots, z_k$. Вершины, являющиеся началом остальных мультидуг, будем обозначать через y_j .

Определение 6. Объединение мультидуги u_z^m , идущей из вершин $z_1, \dots, z_k$ в z и k путей μ_r ($r \in \overline{1, k}$), каждый из которых является ориентированным путем в соответствующей части G_r из вершины x_0 в вершину z_r , назовем обобщенным путем, если каждый из путей μ_r проходит через одну и ту же вершину y_j .

Определение 7. Кратный поток назовем полным, если любой обобщенный путь из x_0 в z содержит дугу u (обычную, кратную или мультидугу), поток по которой равен ее пропускной способности $f(u) = c(u)$.

Определение 8. Проекция C_i ($i \in \overline{1, k}$) подграфа C на часть сети G_i – это часть подграфа C , образованная его вершинами и дугами, принадлежащими G_i .

Определение 9. Так как части G_i ($i \in \overline{1, k}$) сети G представляют собой обычные транспортные сети с источником x_0 и стоком z_i , то будем называть некоторый путь из x_0 в z_i путем прорыва в части G_i , если $f(u) < c(u)$ на прямых дугах и $f(u) > 0$ на обратных дугах этого пути.

Пусть в делимой сети определен некоторый поток φ величины $\varphi_z \geq 0$.

Определение 10. Кратным циклом в делимой сети $G(X, U)$, где X – множество вершин, U – множество дуг (обычных, кратных или мультидуг), назовем такой подграф $C(X', U')$, $X' \subseteq X$, $U' \subseteq U$, для которого:

1) проекции $C_1, \dots, C_k$ на части $G_1, \dots, G_k$ соответственно есть объединение некоторых циклов, причем дуги, поток по которым ненулевой, могут проходиться в обратном направлении;

2) проекции $C_1, \dots, C_k$ согласованы (одинаковы) на общей части подграфов $G_1, \dots, G_k$;

3) C_1 можно представить в виде $C_1 = \cup \{C_1^j\}$, где C_1^j – некоторые циклы и $C_1^j \not\subseteq C_1^k \forall j \neq k$; при этом для любой дуги u из G_1 выполняется неравенство

$$0 \leq f(u) + a^+(u) - a^-(u) \leq c(u),$$

где $a^+(u)$ – это число циклов C_1^j , в которых дуга u проходится в прямом направлении, а $a^-(u)$ – это число циклов C_1^j , в которых дуга u проходится в обратном направлении. Такое же условие должно выполняться и для $C_2, \dots, C_k$.

Определение 11. Обобщенным путем прорыва в делимой сети $G(X, U)$ для некоторого кратного потока φ назовем такой подграф $S(X', U')$, $X' \subseteq X$, $U' \subseteq U$, для которого:

1) каждая из проекций $S_1, \dots, S_k$ на части $G_1, \dots, G_k$ соответственно есть объединение ровно одного пути прорыва из x_0 в z_i и некоторых циклов, причем дуги, поток по которым ненулевой, могут проходиться в обратном направлении;

2) проекции $S_1, \dots, S_k$ согласованы (одинаковы) на общей части подграфов $G_1, \dots, G_k$;

3) S_1 можно представить в виде $S_1 = \mu_1 \cup \{C_1^j\}$, где μ_1 – путь прорыва, C_1^j – некоторые циклы и $C_1^j \not\subseteq C_1^k \forall j \neq k$; при этом для любой дуги u из G_1 выполняется неравенство

$$0 \leq f(u) + a^+(u) - a^-(u) \leq c(u),$$

где $a^+(u)$ – это число элементов множества $\mu_1 \cup \{C_1^j\}$, в которых дуга u проходит в прямом направлении, а $a^-(u)$ – это число элементов множества $\mu_1 \cup \{C_1^j\}$, в которых дуга u проходит в обратном направлении. Такое же условие должно выполняться и для $S_2, \dots, S_k$;

4) Дуга $(\{z_1, \dots, z_k\}, z) \in U'$ и $f(\{z_1, \dots, z_k\}, z) < c(\{z_1, \dots, z_k\}, z)$.

5) S не содержит кратного цикла.

Обозначим $G_\varphi = G(X, U_\varphi)$; $U_\varphi = \{u \mid f(u) \geq 0\}$.

Теорема 2 (см. [14]). Пусть $\varphi^1(U)$, $\varphi^2(U)$ – два полных потока в делимой сети $G(X, U)$, причем $\varphi_z^1 \leq \varphi_z^2$. Пусть $\varphi^0(U) = \varphi^2(U) - \varphi^1(U)$. Тогда граф $G_{\varphi^0} = \{S_i\} \cup \{C_j\}$, где $\{S_i\}$ – множество всех обобщенных путей прорыва из x_0 в z , а $\{C_j\}$ – множество всех кратных циклов. В случае, когда $\varphi_z^1 = \varphi_z^2 = \varphi_z^{\max}$, $\{S_i\} = \emptyset$.

3. Общие принципы сведения задачи сбалансирования к задаче о наибольшем кратном потоке

Прежде чем перейти к описанию сетевой модели для задачи 1, определим общие принципы сведения задачи целочисленного сбалансирования к задаче о наибольшем кратном потоке.

В трехмерном случае делимая кратная сеть целочисленного сбалансирования строилась так, чтобы любой обобщенный путь из x_{000} в z , согласованный на общей части сети, фактически соответствовал обходу элементов a_{000} , a_{i00} , a_{0j0} , a_{00p} , a_{ij0} , a_{i0p} , a_{0jp} , a_{ijp} исходной трехмерной матрицы для некоторых $i > 0$, $j > 0$, $p > 0$ (см. [11] – [12]). Этот обход начинается с элемента a_{000} , завершается в нем же, а все остальные элементы посещаются ровно один раз. Данные элементы формируют куб в исходной матрице A , и увеличение потока на таком пути приводит к увеличению значения a_{ijp} и согласованному увеличению всех связанных с ним суммирующих элементов. Такой обход осуществляется в последовательности

$$a_{000} \rightarrow a_{i00} \rightarrow a_{ij0} \rightarrow a_{ijp} \rightarrow \left\{ \begin{array}{l} a_{0jp} \rightarrow a_{0j0} \\ a_{i0p} \rightarrow a_{00p} \end{array} \right\} \rightarrow a_{000}.$$

Используя мультидуги, мы можем осуществить указанные в последовательности два ветвления. Важным моментом при этом является то, что для каждого конкретного a_{ijp} пара элементов $\{a_{0jp}, a_{i0p}\}$ определяется единственным образом (в кратной сети вершина может быть концом только одной мультидуги).

Для произвольного $l > 3$ логично потребовать, чтобы сведение l -мерной задачи сбалансирования к поиску максимального потока в кратной сети осуществлялось похожим образом. В этом случае мы сможем обобщить многие результаты для трехмерной задачи на случай $l > 3$. Формально можно поставить следующую задачу.

Задача 2. Для задачи целочисленного сбалансирования l -мерной матрицы A ($l \geq 3$) требуется осуществить такое ее сведение к задаче нахождения максимального кратного потока в сети $G(X, U)$ кратности k , чтобы выполнялись условия:

1) каждому элементу матрицы A соответствует хотя бы одна вершина сети $G(X, U)$; каждая вершина сети $G(X, U)$ соответствует ровно одному элементу матрицы A ;

2) источник и сток сети соответствуют корневому элементу матрицы (все индексы равны 0);

3) любой обобщенный путь из источника в сток проходит ровно через одну вершину, соответствующую элементу x из внутренней части матрицы A (все индексы положительны). Этот путь реализует обход гиперкуба, состоящего из элемента x и всех связанных с ним суммирующих элементов, причем каждый суммирующий элемент (кроме корневого) учитывается только один раз;

4) сеть $G(X, U)$ является делимой;

5) лучшим считается такое сведение, для которого параметр k меньше;

6) сведение полиномиально.

Обсудим, каким образом задача 2 может быть решена в случае $l = 4$.

В четырехмерном случае кратная сеть должна строиться таким образом, чтобы любой обобщенный путь из x_{0000} в z , согласованный на общей части сети, соответствовал обходу 16 вершин гиперкуба. При попытке свести задачу 1 к сетевой задаче кратности 3 с использованием подхода, аналогичного трехмерному случаю, мы получим следующий вариант обхода (или подобный ему):

$$a_{0000} \rightarrow a_{i000} \rightarrow a_{ij00} \rightarrow a_{ijp0} \rightarrow a_{ijpq} \rightarrow \left\{ \begin{array}{l} a_{0j pq} \rightarrow a_{0j p0} \rightarrow a_{0j 00} \\ a_{i0 pq} \rightarrow a_{00 pq} \rightarrow a_{00 p0} \\ a_{ij 0q} \rightarrow a_{0j 0q} \rightarrow a_{000q} \end{array} \right\} \rightarrow a_{0000}.$$

Нетрудно убедиться, что при подобном подходе всегда остаются неучтенными два суммирующих элемента (в данном случае это элементы a_{i00q} и a_{i0p0}), то есть задача 2 не может быть решена таким образом.

Исправить ситуацию можно двумя способами.

Первый способ состоит в том, чтобы использовать 4 ветвления вместо 2:

$$a_{0000} \rightarrow a_{i000} \rightarrow \left\{ \begin{array}{l} a_{ij00} \rightarrow a_{ijp0} \\ a_{i0p0} \rightarrow a_{i0pq} \\ a_{i00q} \rightarrow a_{ij0q} \end{array} \right\} \rightarrow a_{ijpq} \rightarrow a_{0j pq} \rightarrow \left\{ \begin{array}{l} a_{0j p0} \rightarrow a_{0j 00} \\ a_{00 pq} \rightarrow a_{00 p0} \\ a_{0j 0q} \rightarrow a_{000q} \end{array} \right\} \rightarrow a_{0000}.$$

Данный способ является не самым удачным, поскольку для a_{i000} существует несколько вариантов ветвлений, соответствующих разным значениям j, p, q , следовательно, в кратной сети требуется допустить возможность связывания с одной вершиной более одной мультидуги, что существенно усложняет модель в целом. Естественно, такая сеть не будет делимой, следовательно, такой способ также не подходит для задачи 2.

Второй способ состоит в сведении задачи 1 к поиску максимального потока в

сети кратности 5. Для этого необходимо обойти вершины в последовательности

$$a_{0000} \rightarrow a_{i000} \rightarrow a_{ij00} \rightarrow a_{ijp0} \rightarrow a_{ijpq} \rightarrow \left\{ \begin{array}{l} a_{0j pq} \rightarrow a_{0j p0} \rightarrow a_{0j00} \\ a_{i0 pq} \rightarrow a_{00 pq} \rightarrow a_{00 p0} \\ a_{ij0q} \rightarrow a_{0j0q} \rightarrow a_{000q} \\ a_{i00q} \\ a_{i0p0} \end{array} \right\} \rightarrow a_{0000}.$$

Данный способ удовлетворяет условиям задачи 2, однако этого нам удастся добиться только за счет увеличения значения параметра k до 5. Тем не менее, этот способ сведения выглядит наиболее предпочтительным, поэтому в следующем разделе мы рассмотрим именно его.

4. Сетевая модель для четырехмерной задачи сбалансирования

Формально сведение задачи 1 к задаче о наибольшем потоке в сети кратности 5 выполняется по следующим правилам.

Каждому элементу a_{ijpq} матрицы A соответствует вершина сети x_{ijpq} . Каждой вершине x_{ijpq} , где более чем 1 индекс имеет нулевые значения, соответствует дополнительная вершина x'_{ijpq} ; вершинам z_r ($r \in \overline{1, 5}$) соответствуют дополнительные вершины z'_r . Источником сети является вершина x_{0000} , стоком – вершина z .

Пропускные способности дуг определяются по следующим правилам ($i \in \overline{1, n}$, $j \in \overline{1, m}$, $p \in \overline{1, t}$, $q \in \overline{1, s}$).

Пропускные способности кратных дуг:

$$\begin{aligned} c(x_{0000}, x_{i000}) &= 5 \lfloor a_{i000} \rfloor; \quad c(x_{0000}, x'_{0000}) = 5 \left(\lfloor a_{0000} + 0.5 \rfloor - \sum_{i=1}^n \lfloor a_{i000} \rfloor \right); \\ c(x'_{0000}, x_{i000}) &= 5(\lceil a_{i000} \rceil - \lfloor a_{i000} \rfloor); \quad c(x_{i000}, x_{ij00}) = 5 \lfloor a_{ij00} \rfloor; \\ c(x_{i000}, x'_{i000}) &= 5 \left(\lceil a_{i000} \rceil - \sum_{j=1}^m \lfloor a_{ij00} \rfloor \right); \quad c(x'_{i000}, x_{ij00}) = 5(\lceil a_{ij00} \rceil - \lfloor a_{ij00} \rfloor); \\ c(x_{ij00}, x_{ijp0}) &= 5 \lfloor a_{ijp0} \rfloor; \quad c(x_{ij00}, x'_{ij00}) = 5 \left(\lceil a_{ij00} \rceil - \sum_{p=1}^t \lfloor a_{ijp0} \rfloor \right); \\ c(x'_{ij00}, x_{ijp0}) &= 5(\lceil a_{ijp0} \rceil - \lfloor a_{ijp0} \rfloor); \quad c(x_{ijp0}, x_{ijpq}) = 5 \lfloor a_{ijpq} \rfloor. \end{aligned}$$

Пропускные способности мультидуг:

$$c(x_{ijpq}, \{x_{0j pq}, x_{i0 pq}, x_{ij0q}, x_{i00q}, x_{i0p0}\}) = 5 \lfloor a_{ijpq} \rfloor; \quad c(\{z_1, z_2, z_3, z_4, z_5\}, z) = 5 \lfloor a_{0000} + 0.5 \rfloor.$$

Пропускные способности обычных дуг в части G_1 :

$$c(x_{0j pq}, x_{0j p0}) = \lfloor a_{0j pq} \rfloor; \quad c(x_{0j pq}, x'_{0j p0}) = \lceil a_{0j pq} \rceil - \lfloor a_{0j pq} \rfloor;$$

$$\begin{aligned}
c(x'_{0jp0}, x_{0jp0}) &= \lceil a_{0jp0} \rceil - \sum_{q=1}^s \lfloor a_{0jpq} \rfloor; & c(x_{0jp0}, x_{0j00}) &= \lfloor a_{0jp0} \rfloor; \\
c(x_{0jp0}, x'_{0j00}) &= \lceil a_{0jp0} \rceil - \lfloor a_{0jp0} \rfloor; & c(x'_{0j00}, x_{0j00}) &= \lceil a_{0j00} \rceil - \sum_{p=1}^t \lfloor a_{0jp0} \rfloor; \\
c(x_{0j00}, z_1) &= \lfloor a_{0j00} \rfloor; & c(x_{0j00}, z'_1) &= \lceil a_{0j00} \rceil - \lfloor a_{0j00} \rfloor; & c(z'_1, z_1) &= \lfloor a_{0000} + 0.5 \rfloor - \sum_{j=1}^m \lfloor a_{0j00} \rfloor.
\end{aligned}$$

Пропускные способности обычных дуг в части G_2 :

$$\begin{aligned}
c(x_{i0pq}, x_{00pq}) &= \lfloor a_{i0pq} \rfloor; & c(x_{i0pq}, x'_{00pq}) &= \lceil a_{i0pq} \rceil - \lfloor a_{i0pq} \rfloor; \\
c(x'_{00pq}, x_{00pq}) &= \lceil a_{00pq} \rceil - \sum_{i=1}^n \lfloor a_{i0pq} \rfloor; & c(x_{00pq}, x_{00p0}) &= \lfloor a_{00pq} \rfloor; \\
c(x_{00pq}, x'_{00p0}) &= \lceil a_{00pq} \rceil - \lfloor a_{00pq} \rfloor; & c(x'_{00p0}, x_{00p0}) &= \lceil a_{00p0} \rceil - \sum_{q=1}^s \lfloor a_{00pq} \rfloor; \\
c(x_{00p0}, z_2) &= \lfloor a_{00p0} \rfloor; & c(x_{00p0}, z'_2) &= \lceil a_{00p0} \rceil - \lfloor a_{00p0} \rfloor; & c(z'_2, z_2) &= \lfloor a_{0000} + 0.5 \rfloor - \sum_{p=1}^t \lfloor a_{00p0} \rfloor.
\end{aligned}$$

Пропускные способности обычных дуг в части G_3 :

$$\begin{aligned}
c(x_{ij0q}, x_{0j0q}) &= \lfloor a_{ij0q} \rfloor; & c(x_{ij0q}, x'_{0j0q}) &= \lceil a_{ij0q} \rceil - \lfloor a_{ij0q} \rfloor; \\
c(x'_{0j0q}, x_{0j0q}) &= \lceil a_{0j0q} \rceil - \sum_{i=1}^n \lfloor a_{ij0q} \rfloor; & c(x_{0j0q}, x_{000q}) &= \lfloor a_{0j0q} \rfloor; \\
c(x_{0j0q}, x'_{000q}) &= \lceil a_{0j0q} \rceil - \lfloor a_{0j0q} \rfloor; & c(x'_{000q}, x_{000q}) &= \lceil a_{000q} \rceil - \sum_{j=1}^m \lfloor a_{0j0q} \rfloor; \\
c(x_{000q}, z_3) &= \lfloor a_{000q} \rfloor; & c(x_{000q}, z'_3) &= \lceil a_{000q} \rceil - \lfloor a_{000q} \rfloor; & c(z'_3, z_3) &= \lfloor a_{0000} + 0.5 \rfloor - \sum_{q=1}^s \lfloor a_{000q} \rfloor.
\end{aligned}$$

Пропускные способности обычных дуг в части G_4 :

$$c(x_{i00q}, z_4) = \lfloor a_{i00q} \rfloor; \quad c(x_{i00q}, z'_4) = \lceil a_{i00q} \rceil - \lfloor a_{i00q} \rfloor; \quad c(z'_4, z_4) = \lfloor a_{0000} + 0.5 \rfloor - \sum_{i=1}^n \sum_{q=1}^s \lfloor a_{i00q} \rfloor.$$

Пропускные способности обычных дуг в части G_5 :

$$c(x_{i0p0}, z_5) = \lfloor a_{i0p0} \rfloor; \quad c(x_{i0p0}, z'_5) = \lceil a_{i0p0} \rceil - \lfloor a_{i0p0} \rfloor; \quad c(z'_5, z_5) = \lfloor a_{0000} + 0.5 \rfloor - \sum_{i=1}^n \sum_{p=1}^t \lfloor a_{i0p0} \rfloor.$$

Введем для полученной кратной сети целочисленного сбалансирования условия разрешимости ($i \in \overline{1, n}$, $j \in \overline{1, m}$, $p \in \overline{1, t}$, $q \in \overline{1, s}$):

$$f(x_{i000}, x_{ij00}) + f(x'_{i000}, x_{ij00}) \geq c(x_{i000}, x_{ij00});$$

$$\begin{aligned}
 f(x_{ij00}, x_{ijp0}) + f(x'_{ij00}, x_{ijp0}) &\geq c(x_{ij00}, x_{ijp0}); \\
 f(x_{0jpr}, x_{0jp0}) + f(x_{0jpr}, x'_{0jp0}) &\geq c(x_{0jpr}, x_{0jp0}); \\
 f(x_{0jp0}, x_{0j00}) + f(x_{0jp0}, x'_{0j00}) &\geq c(x_{0jp0}, x_{0j00}); \\
 f(x_{i0pr}, x_{00pr}) + f(x_{i0pr}, x'_{00pr}) &\geq c(x_{i0pr}, x_{00pr}); \\
 f(x_{00pr}, x_{00p0}) + f(x_{00pr}, x'_{00p0}) &\geq c(x_{00pr}, x_{00p0}); \\
 f(x_{ij0q}, x_{0j0q}) + f(x_{ij0q}, x'_{0j0q}) &\geq c(x_{ij0q}, x_{0j0q}); \\
 f(x_{0j0q}, x_{000q}) + f(x_{0j0q}, x'_{000q}) &\geq c(x_{0j0q}, x_{000q}).
 \end{aligned}$$

Теорема 3. Для того, чтобы задача 1 имела решение, необходимо и достаточно, чтобы в соответствующей ей кратной сети существовал максимальный кратный поток φ величины $\varphi_z = 5[a_{0000} + 0.5]$, для которого выполняются условия разрешимости.

Справедливость теоремы 3 следует непосредственно из правил построения сети и условия неразрывности потока в каждой вершине. Отметим также, что решение задачи о максимальном кратном потоке указанного вида будет индуцировать решение задачи 1. При этом d_{ijpq} полагается равным одной пятой величины потока, проходящего через вершину x_{ijpq} , если она является концом кратной дуги; величине потока, проходящего через вершину x_{ijpq} , если она является концом обычной дуги.

Примечательно, что условия разрешимости необходимо формулировать только для кратных дуг, а также для обычных дуг частей G_1 , G_2 и G_3 . Если задача 1 разрешима, то соблюдение ограничений, связанных с минимально допустимым потоком на дугах в частях G_4 , G_5 , обеспечивается автоматически за счет того, что сумма пропускных способностей дуг, входящих в z_4 и z_5 , равна $[a_{0000} + 0.5]$.

Алгоритм решения задачи 1 получается в результате простого обобщения алгоритма решения трехмерной задачи.

Алгоритм 1. Алгоритм выполняется в три этапа.

1. *Этап получения полного потока.* На данном этапе увеличения потока можно добиваться, находя все возможные обобщенные пути прорыва из x_{0000} в z через вершины x_{ijpq} ($i \in \overline{1, n}$, $j \in \overline{1, m}$, $p \in \overline{1, t}$, $q \in \overline{1, s}$) без обратных дуг и увеличивая поток по ним. В итоге будет получен полный поток.

2. *Этап увеличения потока.* Если полученный полный поток не является максимальным, то увеличиваем поток при помощи обобщенного алгоритма пометок (см. [15]) до тех пор, пока поток не станет максимальным. Напомним, что идея обобщенного алгоритма пометок состоит в следующем: в проекциях $G_1, \dots, G_5$ поочередно строятся пути прорыва $\mu_1, \dots, \mu_5$ (возможно, в объединении с некоторыми циклами) до тех пор, пока пути во всех пяти проекциях не станут согласованными, либо же не останется вариантов для продолжения построения пути. В первом случае объединение $\mu_1, \dots, \mu_5$ с мультидугой с концом в z даст обобщенный путь прорыва, во втором случае производится откат до «точки ветвления», после чего выполнение алгоритма возобновляется. Если точкой ветвления оказывается x_{0000} и при этом x_{0000} не помечена, то задача целочисленного сбалансирования не имеет решения.

3. *Этап коррекции потока.* Если максимальный поток имеет величину $5[a_{0000} + 0.5]$, но не удовлетворяет условиям разрешимости, то выполняем коррекцию потока

при помощи модификации обобщенного алгоритма пометок, пока не будут выполнены все условия разрешимости или не будет установлена невозможность такой коррекции.

На шаге 3 модификация обобщенного алгоритма пометок для сети кратности 5 получается таким же образом, как и для сети кратности 2 (см. [11]), с точностью до количества частей G_r .

Теорема 4. *Задача 1 может быть решена с помощью алгоритма 1.*

Доказательство. Из теоремы 4 из статьи [15] следует, что если в сети $G(X, U)$ существует максимальный поток величины $5[a_{0000} + 0.5]$, то он может быть найден с помощью шагов 1–2 алгоритма 1.

Для доказательства того, что шаг 3 выполнит требуемую коррекцию потока всегда, когда это возможно, достаточно повторить рассуждения из части II доказательства теоремы 5 из статьи [11]. При этом все выводы, сделанные в статье [11] для частей G_1 и G_2 кратной сети целочисленного сбалансирования трехмерной матрицы, дословно сохраняются для частей G_r ($r \in \overline{1, 5}$) рассматриваемой сети $G(X, U)$.

Теорема 4 доказана.

Отметим, что по аналогии с трехмерной задачей целочисленного сбалансирования с ограничениями второго рода (см. [12]) можно поставить также задачу целочисленного сбалансирования четырехмерной матрицы с ограничениями второго рода. Сетевая модель при этом будет отличаться только пропускными способностями дуг, а теоремы 3 и 4 будут по-прежнему справедливы.

Заключение

Данная работа является продолжением исследований, начатых в [8], [11] – [15].

Основным результатом статьи является постановка задачи целочисленного сбалансирования с ограничениями первого рода в четырехмерном случае и сведение данной задачи к поиску максимального кратного потока в сети кратности 5. Сведение осуществляется таким образом, что результаты, полученные ранее для трехмерной задачи сбалансирования, легко обобщаются для четырехмерного случая.

Список литературы / References

- [1] Корбут А. А., Финкельштейн Ю. Ю., *Дискретное программирование*, Наука, Москва, 1969; [Korbut A. A., Finkelstein J. J., *Diskretnoe programmirovaniye*, Nauka, Moskva, 1969, (in Russian).]
- [2] Раскин Л. Г., Кириченко И. О., *Многоиндексные задачи линейного программирования*, Радио и связь, Москва, 1982; [Raskin L. G., Kirichenko I. O., *Mnogoindeksnyye zadachi lineynogo programmirovaniya*, Radio i svyaz, Moskva, 1982, (in Russian).]
- [3] Spieksma F. C. R., “Multi index assignment problems: complexity, approximation, applications”, *Nonlinear Assignment Problems. Algorithms and Applications*, eds. P. M. Pardalos, L. S. Pitsoulis, Kluwer Academic Publishers, 2000, 1–11.
- [4] Афраимович Л. Г., “Трехиндексные задачи линейного программирования с вложенной структурой”, *Автоматика и телемеханика*, 2011, № 8, 109–120; English transl.: Afraimovich L. G., “Three-index linear programs with nested structure”, *Automation and Remote Control*, **72**:8 (2011), 1679–1689.
- [5] Кондаков А. С., Рублев В. С., “Задача сбалансирования матрицы плана”, *Доклады Одесского семинара по дискретной математике*, Астропринт, Одесса, 2005, 24–26; [Kondakov A. S., Roublev V. S., “Zadacha sbalansirovaniya matritsy plana”, *Doklady Odesskogo seminar po diskretnoy matematike*, Astroprint, Odessa, 2005, 24–26, (in Russian).]
- [6] Коршунова Н. М., Рублев В. С., “Задача целочисленного сбалансирования матрицы”, *Современные проблемы математики и информатики*, ЯрГУ, Ярославль, 2000, 145–150; [Korshunova N. M., Roublev V. S., “Zadacha tselochislenno sbalansirovaniya matritsy”, *Sovremennyye problemy matematiki i informatiki*, Yaroslavl State University, Yaroslavl, 2000, 145–150, (in Russian).]
- [7] Ford L. R., Fulkerson D. R., *Flows in Networks*, Princeton University Press, 1962.
- [8] Рублев В. С., Смирнов А. В., “NP-полнота задачи целочисленного сбалансирования трехмерной матрицы”, *Доклады Академии Наук*, **435**:3 (2010), 314–316; English transl.: Roublev V. S., Smirnov A. V., “NP-Completeness of the Integer Balancing Problem for a Three-Dimensional Matrix”, *Doklady Mathematics*, **82**:3 (2010), 912–914.
- [9] Garey M. R., Johnson D. S., *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.
- [10] Karp R., “Reducibility among combinatorial problems”, *R. E. Miller, J. W. Thatcher (Eds.). Complexity of Computer Computations*, Plenum, 1972, 85–103.
- [11] Рублев В. С., Смирнов А. В., “Задача целочисленного сбалансирования трехмерной матрицы и алгоритмы ее решения”, *Моделирование и анализ информационных систем*, **17**:2 (2010), 72–98; [Roublev V. S., Smirnov A. V., “The Problem of Integer-Valued Balancing of a Three-Dimensional Matrix and Algorithms of Its Solution”, *Modeling and Analysis of Information Systems*, **17**:2 (2010), 72–98, (in Russian).]
- [12] Smirnov A. V., “Some Solvability Classes for the Problem of Integer Balancing of a Three-Dimensional Matrix with Constraints of the Second Type”, *Automatic Control and Computer Sciences*, **48**:7 (2014), 543–553.
- [13] Smirnov A. V., “Heuristic Algorithms for the Problem of Integer Balancing of a Three-Dimensional Matrix with Constraints of the Second Type”, *Automatic Control and Computer Sciences*, **49**:7 (2015), 473–483.
- [14] Смирнов А. В., “Задача о наибольшем кратном потоке в делимой сети и ее частные случаи”, *Моделирование и анализ информационных систем*, **22**:4 (2015), 533–545; [Smirnov A. V., “The Problem of Finding the Maximal Multiple Flow in the Divisible Network and its Special Cases”, *Modeling and Analysis of Information Systems*, **22**:4 (2015), 533–545, (in Russian).]
- [15] Рублев В. С., Смирнов А. В., “Потоки в кратных сетях”, *Ярославский педагогический вестник*, **3**:2 (2011), 60–68; [Rublev V. S., Smirnov A. V., “Flows in Multiple Networks”, *Yaroslavsky Pedagogichesky Vestnik*, **3**:2 (2011), 60–68, (in Russian).]

Smirnov A. V., "Network Model for The Problem of Integer Balancing of a Four-dimensional Matrix", *Modeling and Analysis of Information Systems*, **23**:4 (2016), 466–478.

DOI: 10.18255/1818-1015-2016-4-466-478

Abstract. The problem of integer balancing of a four-dimensional matrix is studied. The elements of the inner part (all four indices are greater than zero) of the given real matrix are summed in each direction and each two- and three-dimensional section of the matrix; the total sum is also found. These sums are placed into the elements where one or more indices are equal to zero (according to the summing directions). The problem is to find an integer matrix of the same structure, which can be produced from the initial one by replacing the elements with the largest previous or the smallest following integer. At the same time, the element with four zero indices should be produced with standard rules of rounding-off. In the article the problem of finding the maximum multiple flow in the network of any natural multiplicity k is also studied. There are arcs of three types: ordinary arcs, multiple arcs and multi-arcs. Each multiple and multi-arc is a union of k linked arcs, which are adjusted with each other. The network constructing rules are described. The definitions of a divisible network and some associated subjects are stated. There are defined the basic principles for reducing the integer balancing problem of an l -dimensional matrix ($l \geq 3$) to the problem of finding the maximum flow in a divisible multiple network of multiplicity k . There are stated the rules for reducing the four-dimensional balancing problem to the maximum flow problem in the network of multiplicity 5. The algorithm of finding the maximum flow, which meets the solvability conditions for the integer balancing problem, is formulated for such a network.

Keywords: integer balancing, multiple networks, multiple flows, divisible networks, NP -completeness, generalized labeling algorithm, four-dimensional matrices

On the authors:

Smirnov Alexander Valeryevich, orcid.org/0000-0002-0980-2507, PhD, Associate Professor
P.G. Demidov Yaroslavl State University,
Sovetskaya str., 14, Yaroslavl, 150000, Russia, e-mail: alexander_sm@mail.ru

©Таранин С. М., 2015

DOI: 10.18255/1818-1015-2016-4-479-491

УДК 004.056.3

Резервное копирование с хранением в базе данных

Таранин С. М.

получена 9 сентября 2015

Аннотация.

В данной работе представлен обзор некоторых технологий, которые используются в современных системах резервного копирования, кратко описаны их преимущества и недостатки. Далее рассматривается подход к реализации системы резервного копирования с сохранением файлов в базе данных. Предлагается разбивать копируемые файлы на блоки фиксированной длины. Каждый блок представляет собой последовательность байт. Длина блока может быть адаптивной, т.е. меняться в зависимости от типа или размера файла. В таком виде содержимое файлов предлагается хранить в одной таблице, а информацию о них: имена, атрибуты и связи между ними – хранить в другой таблице. Сведения о сохраненных файлах и папках предлагается хранить не только в базе данных на сервере, но и на стороне клиента в некоторой иерархической структуре. Она содержит набор записей и представляет собой модель копируемой директории. Наличие такой модели позволяет отслеживать изменения в файловой системе клиента без выполнения дополнительных запросов к базе данных. В случае если файл изменен, в базу копируются только его изменившиеся блоки. При этом в модели на стороне клиента также обновляется информация, например дата изменения отредактированного документа. Удаляются записи об удаленных файлах и папках. Таким образом, уменьшается нагрузка на канал передачи данных. В статье описаны алгоритмы сохранения и восстановления данных, а также рассмотрены факторы, влияющие на скорость их работы. Наглядно показана зависимость скорости сохранения и восстановления данных от мелкости разбиения файлов, а также от структуры копируемой директории.

Ключевые слова: файл, данные, резервное копирование, база данных, блок, модель

Для цитирования: Таранин С. М., "Резервное копирование с хранением в базе данных", *Моделирование и анализ информационных систем*, **23:4** (2016), 479–491.

Об авторах:

Таранин Сергей Максимович, orcid.org/0000-0001-8117-7358, аспирант,
Ярославский государственный университет им. П.Г. Демидова,
ул. Советская, 14, г. Ярославль, 150003 Россия, e-mail: staranin0208@yandex.ru

Введение

Несмотря на широкое применение цифровых устройств для обработки документов, еще рано говорить о полном отказе от бумажных носителей. И если защите информации, представляющей собой персональные данные или коммерческую тайну, уделяется много внимания, то защита «иных» форм данных – это личное дело их

владельца. Например, для студентов это могут быть лекции, рефераты, незавершенные курсовые работы. Сюда же можно отнести информацию, содержащуюся в записных книжках и ежедневниках, личные архивы и многое другое. Вся эта информация представляет определенную ценность для ее владельца.

Существует проблема обеспечения целостности такой информации на автоматизированном рабочем месте (АРМ) с надежностью, близкой к надежности бумажных носителей. Разумеется, надежность бумажного носителя – это, в данном случае, идеал, к которому нужно стремиться. Автоматизированное рабочее место – это программно-технический комплекс, предназначенный для автоматизации деятельности определенного вида. Он представляет собой профессионально-ориентированную малую вычислительную систему, расположенную непосредственно на рабочем месте специалиста и предназначенную для автоматизации его работы. Обычно в качестве АРМ выступает персональный компьютер (ПК) с предустановленным программным обеспечением (ПО). Это может быть система проектирования, система электронного документооборота, справочная система и др.

Обеспечение целостности данных на АРМ в основном достигается путем резервного копирования. Это процесс создания копии данных на носителе, предназначенном для восстановления данных в оригинальном или новом месте их расположения в случае их повреждения или разрушения. Копии записываются либо на магнитную ленту, либо на жесткий диск. Скорость записи на последний выше, однако его надежность ниже, поэтому еще рано говорить о полном отказе от ленточных носителей. Для того, чтобы объединить в одной системе резервного копирования скорость сохранения информации на диски с надежностью ленточных накопителей, была придумана схема D2D2T [1], при которой данные сначала копируются на жесткий диск, затем на ленту. Во время резервного копирования основные приложения приостанавливают свою работу. Применение технологии D2D2T позволяет свести к минимуму время простоя системы, поскольку приложения останавливаются на время, необходимое для копирования данных на жесткий диск. Потом приложения возобновляют работу, а данные копируются с диска на ленту. Однако применение данной технологии не решает следующие проблемы традиционного резервного копирования:

- избыточность копируемой информации;
- непредсказуемое время восстановления данных;
- риск потери данных из-за невозможности восстановления.

Последнее время активно развивается технология непрерывной защиты данных (CDP) [3, 6, 7]. В отличие от традиционного резервного копирования, здесь не требуется определять время создания копий. Специальный сервис постоянно отслеживает изменения в файловой системе и копирует только изменившиеся данные. Данная статья посвящена вопросам разработки подобной системы.

1. Современные технологии в системах резервного копирования

1.1. Непрерывная защита данных

Система резервного копирования, реализующая технологию CDP, должна иметь возможность восстановить любую версию каждого хранимого файла на любой момент времени. Существуют системы, которые в отличие от традиционных производят операции копирования очень часто и позиционируются разработчиками как CDP-решения. В связи с этим системы стали делить на «настоящие» CDP (англ. True CDP), и на «почти непрерывные» CDP решения (англ. near-Continuous CDP). Технология True CDP отличается от традиционного резервного копирования и near-Continuous CDP прежде всего тем, что операции копирования проходят сразу при изменении данных, а не по заданному расписанию.

Существуют споры вокруг преимуществ True CDP. Данная технология имеет ограниченное применение, и можно привести примеры, когда ее использование является нежелательным. Современные операционные системы для ускорения файловых операций используют кэширование. Измененные блоки файла, предназначенные для записи на диск, некоторое время хранятся в кэше, и лишь в свободное от других операций время сохраняются физически. Такой подход называют методом отложенной записи (англ. lazy write) [5]. По этой причине система резервного копирования, реализующая технологию True CDP, не может видеть все изменения в файлах, пока операционная система не сбросит их на диск из кэша. Одним из недостатков также является негативное влияние на производительность сети при обработке файлов большого размера, например видеозаписей или файлов систем проектирования.

1.2. Исключение дублирования данных

Для исключения повторяющихся данных применяется механизм дедубликации (англ. de-duplication) [8, 9]. Он представляет собой метод сжатия информации, когда поиск копий производится по всему массиву данных, а не в пределах одного файла. Главным преимуществом использования данной технологии является существенная экономия дискового пространства. В процессе дедубликации данные разбиваются на блоки. Каждый блок идентифицируется уникальным числом (хэш, номер корректировки). Каждый новый хэш сравнивается с вычисленными ранее. Если он уникален, то добавляется в список, иначе блок заменяется ссылкой на уже существующий хэш. Сложность реализации заключается в том, чтобы не снизить производительность системы резервного копирования. Так, использование дедубликации может существенно ухудшить показатели скорости восстановления данных. Дедубликация данных может быть выполнена до того, как данные будут записаны в целевое хранилище (англ. in-line de-duplication), или уже после записи (англ. post-process de-duplication). Системы, в которых процесс дедубликации выполняется после записи, требуют хранилища большего объема, так как вначале данные сохраняются полностью.

1.3. Сервисы удаленного резервного копирования

Удаленное резервное копирование (англ. remote backup или online backup) представляет собой сервис, предлагающий пользователям систему хранения и резервного копирования данных [1, 2]. Пользователю предоставляется клиентская часть системы, которая собирает, сжимает, шифрует и отправляет файлы на сервер поставщика услуг резервного копирования. Доступ к файлам можно получить в любом месте, где есть подключение к сети Интернет. Основные преимущества сервисов удаленного резервного копирования в том, что копии хранятся отдельно от оригиналов, а участие пользователя практически не требуется. Использование таких сервисов избавляет пользователя от необходимости устанавливать и поддерживать собственные системы хранения. Однако скорость создания резервных копий и восстановления значительно ограничена и зависит от скорости доступа к сети Интернет.

2. Резервное копирование с хранением информации в базе данных

Рассмотрим подход к реализации системы резервного копирования на клиент-серверной архитектуре [10], где клиентская часть генерирует запросы на сохранение или восстановление данных, а серверная часть обрабатывает запросы от клиентов и упорядочивает сохраненные данные.

Для хранения пользовательских данных на сервере и управления ими можно ввести две таблицы Model и Data. Первая таблица представляет собой модель копируемой директории. Каждая запись в модели хранит информацию об объекте файловой системы. Далее под таким объектом будем понимать файл или папку копируемой директории. Связь между таблицами Model и Data – «один ко многим» соответственно.

Таблица Model содержит следующие поля:

1. ID – первичный ключ записи
2. NAME – имя объекта файловой системы. Имя файла должно быть указано с расширением.
3. PARENT – идентификатор «родителя» – записи, соответствующей некоторой папке.
4. ISFOLDER – запись соответствует папке.
5. HASH – хэш-код файла для проверки целостности данных после восстановления из резервной копии. Для ускорения обработки файлов вместо вычисления хэш-кода можно хранить номер корректировки или дату последнего изменения.

Поля NAME и PARENT представляют собой естественный ключ записи таблицы, поскольку в одной папке не может быть 2-х файлов с одинаковым именем и расширением.

Таблица Data хранит данные файлов в виде блоков и содержит следующие поля:

1. ID – первичный ключ записи
2. FILEID – идентификатор записи таблицы Model.
3. BLOCK – последовательность байт некоторого файла копируемой директории в виде строки символов.
4. HASH – хэш-код. Для проверки целостности данных после восстановления из резервной копии.

Модель копируемой директории предлагается хранить не только на сервере, но и на клиентской части в некоторой иерархической структуре. Это позволяет отслеживать изменения без выполнения дополнительных запросов к БД [11, 12, 13]. В случае если файл изменен, то копируются только его изменившиеся блоки [14, 15]. Размер блоков может быть фиксированным, а может выбираться в зависимости от типа или размера файла (адаптивная дедубликация) [4]. Таким образом, уменьшается нагрузка на канал передачи данных.

2.1. Сохранение данных

Предлагается следующий рекурсивный алгоритм резервного копирования:

- 1 Для каждого файла из текущей директории:
 - 1.1 Создаем запись для таблицы Model и сохраняем ее insert-запросом к БД.
 - 1.2 Получаем идентификатор сохраненной записи.
 - 1.3 Читаем файл с диска блоками, формируя массив записей для сохранения в таблицу Data.
 - 1.4 Сохраняем записи в таблице Data insert-запросом к БД.
- 2 Для каждой папки из текущей директории:
 - 2.1 Формируем запись для таблицы Model и сохраняем ее insert-запросом к БД.
 - 2.2 Запускаем данный алгоритм для файлов в этой директории.
- 3 Считаем хэш-код для текущей директории и сохраняем его в поле HASH соответствующей записи таблицы Model update-запросом.

2.1.1. Оценка трудоемкости алгоритма

Глубина дерева рекурсии равна глубине копируемой директории. При обработке очередного узла сканируемой директории наиболее трудоемкой задачей является обработка файла, т.к. от его размера зависит количество обращений к БД. Поэтому необходимо свести к минимуму количество запросов при сохранении приемлемого уровня обеспечения целостности данных за счет разбиения файлов на блоки. Оценим количество запросов к БД при сканировании очередного узла дерева рекурсии.

Пусть n – количество файлов в текущей директории, m – количество папок в текущей директории, sgm – длина блока. Тогда количество запросов можно выразить следующей формулой:

$$F(m, n) = \sum_{i=1}^n \left(\frac{l_i}{sgm} + 2 \right) + m + 1,$$

где $sgm < l_i, l_i$ – длина i -го файла в текущей директории.

Обозначим $s_i = \frac{l_i}{sgm} + 1$ – число блоков i -го файла. Тогда формула примет более компактный вид:

$$F(m, n) = \sum_{i=1}^n (s_i + 1) + m + 1.$$

Для каждого файла выполняется s_i запросов для записи блоков в таблицу Data, один запрос для записи информации о файле в таблицу Model, m запросов для записи информации о папках в таблицу Model и один запрос для обновления записи в таблице Model (сохранение хэша текущей директории). Каждой директории ставится в соответствие уникальный идентификатор – первичный ключ записи в таблице Model. Общее количество запросов к БД при обработке директории можно представить следующей формулой:

$$F_k = F(m_k, n_k) = \sum_{i=1}^{n_k} \left(\frac{l_{k_i}}{sgm} + 2 \right) + \sum_{j=1}^{m_k} F(m_{k_j}, n_{k_j}) + m + 1 =$$

$$\sum_{i=1}^{n_k} (s_{k_i}) + \sum_{j=1}^{m_k} F(m_{k_j}, n_{k_j}) + m + 1,$$

где k – идентификатор текущей директории. Здесь мы учитываем результаты обработки вложенных папок. Переменную l_{k_i} мы понимаем здесь как «длина i -го файла директории с идентификатором k ». Таким образом, скорость работы алгоритма зависит от глубины вложенности папок и от отношения $\frac{l_{k_i}}{sgm}$. Длина блока может быть константой, а может меняться в зависимости от типа или объема файла.

Посмотрим, как меняется скорость обработки данных в зависимости от длины блока. Тестовый стенд представляет собой два компьютера, объединенные в сеть. Один из них играет роль сервера с установленной СУБД Linter 6.0.18.9 (процессор: Intel Celeron 2.8 GHz, память: 3 GB) [11, 16]. Второй ПК представляет собой клиентскую часть системы, которая содержит пользовательские данные (процессор: Intel Core i5-2500K, 3.3 GHz, память 8 GB). На клиентском ПК запускается приложение, которое обрабатывает данные пользователя и отправляет их на сервер. В качестве тестовых данных возьмем папку с файлами, близкими по размеру (360 фото 1600x1200, 0,8–1,5 МБ каждая) без вложенных директорий. Результат обработки данных по описанному алгоритму с разными размерами блока представлен на рисунке 1:

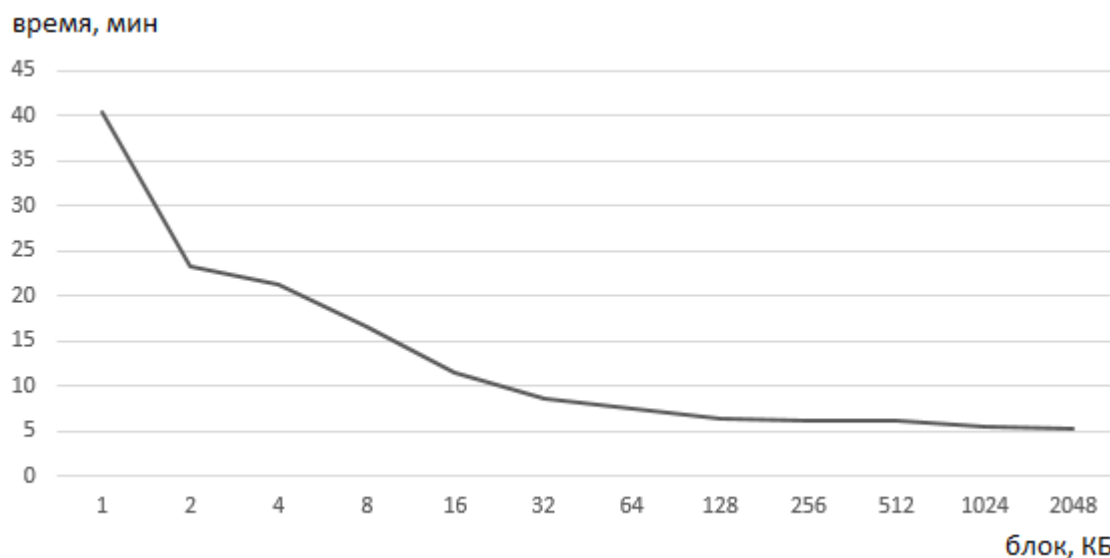
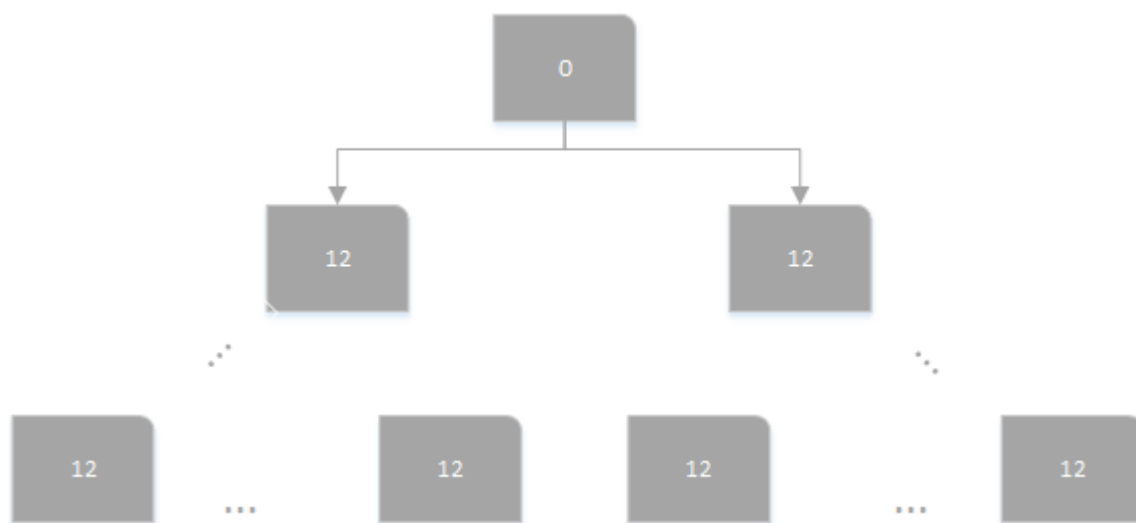


Рис. 1. Зависимость времени сохранения от размера блока
(все файлы в одной папке)
Fig. 1. The dependence of the saving time of the block size
(all the files in the one folder)

Наибольший рост производительности достигается после увеличения размера блока с 1 до 2 КБ (42%). При последующем увеличении его размера производительность растет медленнее. Поэтому оптимальный размер блока для данного случая составляет 64 – 128 КБ, т.к. дальнейшее его увеличение незначительно влияет на скорость работы. При размере блока 64 КБ каждый файл тестового стенда был разбит в среднем на 22 блока, а при 128 КБ – на 11 блоков.

Посмотрим теперь, как изменится скорость работы, если распределить файлы по папкам. Пусть каждая папка имеет 2 вложенные папки. В каждой папке, кроме корневой, по 12 файлов. Тогда директория будет иметь структуру как на рисунке 2.



Результат обработки данных представлен на рисунке 3:

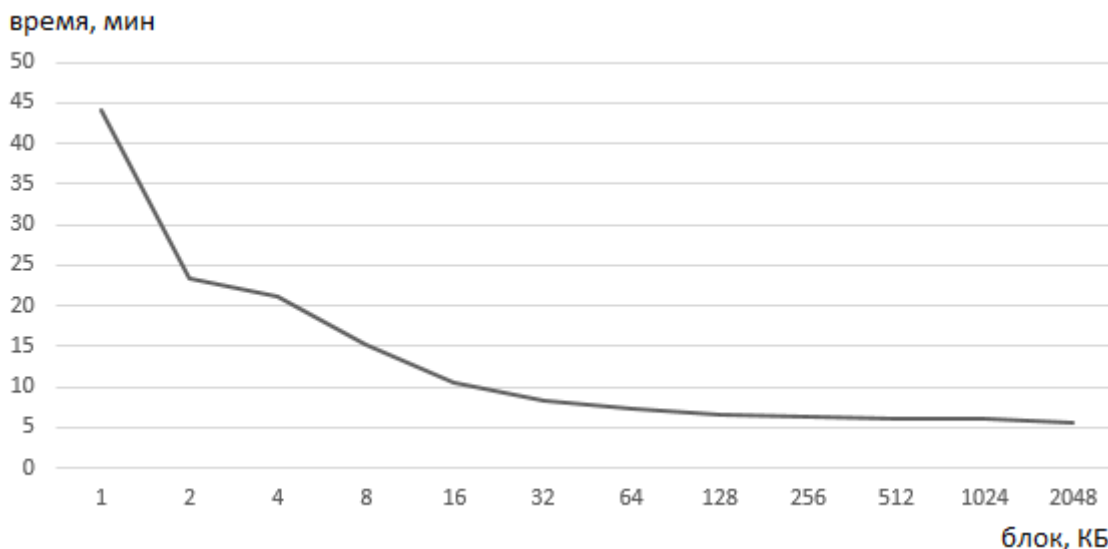


Рис. 3. Зависимость времени сохранения от размера блока
(все файлы равномерно распределены по вложенным папкам)

Fig. 3. The dependence of the saving time of the block size
(all the files are evenly distributed in the subfolders)

Исходя из опытных данных можно сделать предположение о том, что оптимальный размер блока 64 – 128 КБ. Ветвление структуры файловой системы практически не повлияло на скорость работы.

2.2. Восстановление данных

Для восстановления файлов из БД можно использовать рекурсивный алгоритм, входными параметрами которого являются путь до корневой директории – source, а также идентификатор соответствующей записи таблицы Model – parentID.

- 1 По parentID получим набор записей из таблицы Model с информацией о файлах и папках корневой директории.
- 2 Для каждой записи из этого набора:
 - 2.1 Формируем путь к объекту файловой системы (source плюс значение поля NAME)
 - 2.2 Если значение поля ISFOLDER равно «истина», то
 - 2.2.1 Создаем директорию, если такой не существует
 - 2.2.2 Восстанавливаем содержимое созданной директории, передав на вход данного алгоритма путь до нее и идентификатор записи.
 - 2.3 Иначе:
 - 2.3.1 Создаем файл.
 - 2.3.2 По ключу записи получаем содержимое этого файла из таблицы Data в виде набора записей.

2.3.3 Записываем последовательно содержимое полей BLOCK в созданный файл.

Наиболее трудоемкой процедурой является последовательная запись данных в файл. Она зависит от размера блоков, на которые файл был разбит перед сохранением. Чем меньше размер блока, тем больше последовательных операций записи в файл будет выполнено. Зависимость времени восстановления данных от размера блока представлена на рисунке 4.

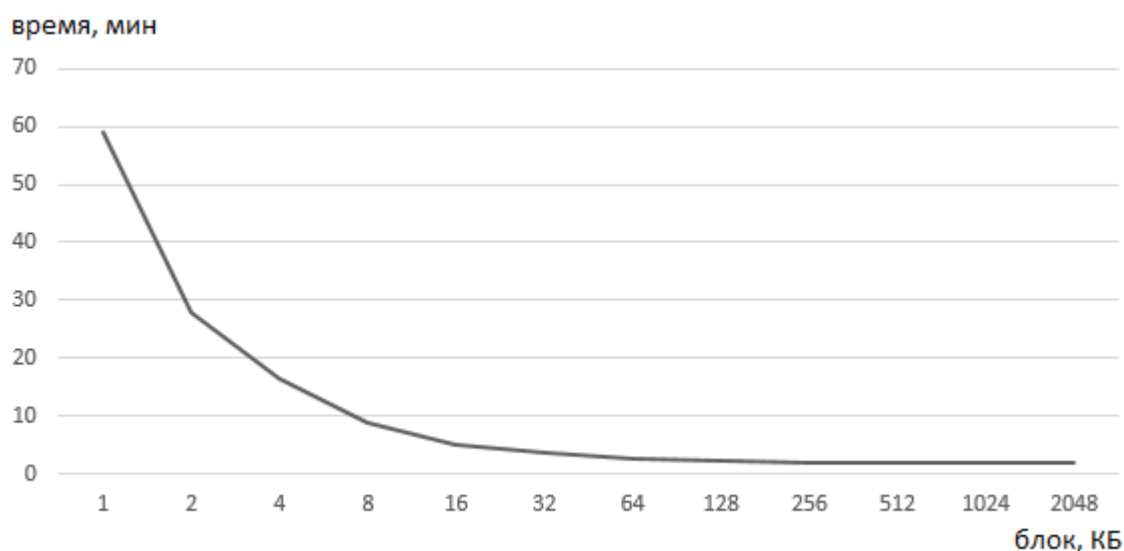


Рис. 4. Зависимость времени восстановления от размера блока

Fig. 4. The dependence of the recovery time of the block size

Как и в случае сохранения, оптимальный размер блока составляет 64 – 128 КБ, поскольку дальнейшее его увеличение незначительно влияет на скорость работы алгоритма.

Мелкость разбиения файлов влияет не только на скорость сохранения и восстановления. В каждой записи таблицы Data хранится хэш-код и идентификатор записи таблицы Model. Чем меньше размер блока, тем больше дополнительной информации, кроме пользовательских данных, приходится хранить в БД. Зависимость размера резервной копии от размера блока для десяти файлов размером 0.8 МБ представлена на рисунке 5.

Таким образом, оптимальное разбиение для сохранения и восстановления данных составляет около 11–22 блоков на файл. Такое разбиение незначительно влияет на размер БД, а также позволяет сохранить приемлемую скорость обработки данных.

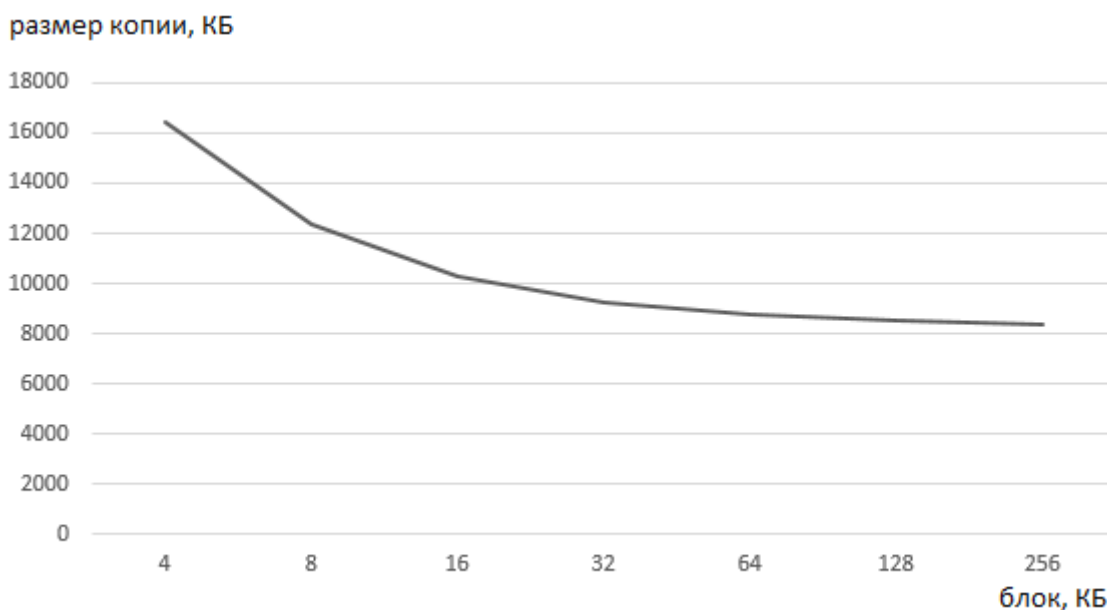


Рис. 5. Зависимость размера резервной копии от размера блока

Fig. 5. The dependence of the backup size of the block size

2.3. Поиск изменившихся данных

После полного резервного копирования данных запускается непрерывный процесс сканирования файловой системы на наличие изменений. Как только процесс обнаружит, что файловая система изменена, он должен сравнить новые данные с данными в модели, обновить модель и отправить изменения на сервер.

Ниже предлагается рекурсивный алгоритм сканирования файловой системы, состоящий из двух этапов. Сначала производится поиск удаленных файлов и папок. Далее происходит поиск изменений в существующих файлах, и сохранение новых данных.

На первом этапе выполняется следующая последовательность действий:

- 1 По parentID получим набор записей из таблицы Model с информацией о файлах и папках корневой директории.
- 2 Для каждой записи из этого набора:
 - 2.1 Формируем путь к объекту файловой системы (source плюс значение поля NAME)
 - 2.2 Если значение поля ISFOLDER равно «истина», то
 - 2.2.1 Если такой директории не существует, удаляем данные вложенных файлов из таблицы Data и соответствующие записи из таблицы Model.
 - 2.2.2 Иначе, выполняем аналогичные действия для данной директории.
 - 2.3 Иначе (ISFOLDER равно «ложь»):

2.3.1 Если файл не существует, удаляем соответствующие данные из таблиц Data и Model.

2.3.2 Иначе, переходим к следующей записи выборки.

На втором этапе выполняется следующая последовательность действий:

1 Для каждого файла из текущей директории:

2 Если в модели на стороне клиента нет записи о файле, то:

2.1 Создаем запись для таблицы Model и сохраняем ее insert-запросом к БД.

2.2 Получаем идентификатор сохраненной записи.

2.3 Читаем файл с диска блоками, формируя массив записей для сохранения в таблицу Data.

2.4 Сохраняем записи в таблице Data insert-запросом к БД.

3 Иначе (в модели на стороне клиента есть запись о файле):

3.1 Вычисляем его хэш-код и сравниваем с хэш-кодом клиентской модели.

3.2 Если коды совпадают, переходим к следующему файлу.

3.3 Иначе:

3.3.1 Читаем файл с диска блоками.

3.3.2 Синхронизация данных файла. От каждого блока считаем хэш-код и сравниваем с вычисленным ранее. Если хэш-коды очередного блока не совпадают, сохраняем этот блок в БД. Лишние блоки удаляем, недостающие – сохраняем.

4 Для каждой папки из текущей директории:

4.1 Если в таблице Model нет соответствующей записи, то

4.1.1 Сохраняем информацию о данной папке в БД.

4.1.2 Обработываем данные в новой директории по рассмотренному выше алгоритму сохранения.

4.2 Иначе, выполняем аналогичные действия для данной директории.

Чтобы сократить количество запросов к БД при сканировании, можно хранить список хэш-кодов блоков данных на стороне клиента и поддерживать в актуальном состоянии, так же как модель данных сканируемой директории.

Заключение

В статье были рассмотрены современные технологии резервного копирования данных. Далее был предложен подход к реализации системы резервного копирования с хранением файлов в базе данных. Одним из преимуществ такого подхода является его универсальность. Решение на его основе можно адаптировать под разные

схемы сети. Предлагаемый подход может служить основой для разработки сервиса удаленного резервного копирования, решения для локальной сети небольшой организации или распределенной сети крупной фирмы с филиалами. В статье были предложены варианты алгоритмов сохранения и восстановления резервной копии из БД, дана оценка их производительности и приведены результаты соответствующих экспериментов. Далее был предложен алгоритм поиска изменившихся данных. Также была показана зависимость скорости сохранения и восстановления данных от величины разбиения файлов.

Направлениями дальнейшего исследования являются доработка предложенных алгоритмов с применением технологии адаптивной дедубликации данных, экспериментальное исследование влияния данной технологии на скорость сохранения и восстановления, оценка границ применимости технологии адаптивной дедубликации.

Список литературы / References

- [1] Казаков В. Г., Федосин С. А., “Технологии и алгоритмы резервного копирования”, *Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы»*, 2008, 1–49; [Kazakov V. G., Fedosin S. A., “Technologii i algoritmi rezervnogo kopirovaniya”, *Vserossiyskiy konkursniy otbor obzorno-analiticheskikh statey po prioritetnomu napravleniu "Informacionno-telekommunikacionnie sistemi"*, 2008, 1–49, (in Russian).]
- [2] Алиев А. А., Самедов Р. Б., “Алгоритм создания полного резервного копирования в облачных вычислениях”, *Вестник Бакинского Университета*, 2013, № 4, 120–127; English transl.: Aliev A. A., Samadov R. B., “An algorithm of a full backup in cloud computing”, *Baku State University News*, 2013, № 4, 120–127.
- [3] Mugoh L., Ateya I. L., Shibwabo B. K., “Continuous Data Protection Architecture as a Strategy for Reduced Data Recovery Time”, *Journal of Systems Integration*, **2**:4 (2011), 54–69.
- [4] Казаков В. Г., Федосин С. А., Плотникова Н. П., “Способ адаптивной дедубликации с применением многоуровневого индекса размещения копируемых блоков данных”, *Фундаментальные исследования*, 2013, № 8 (часть 6), 1322–1325; Kazakov V. G., Fedosin S. A., Plotnikova N. P., “Method of adaptive dedublication with multilevel block indexing”, *Fundamental research*, 2013, № 8 (part 6), 1322–1325.
- [5] Kathuria V., Dhamankar R., Kodavalla H., “Transaction Isolation and Lazy Commit”, *IEEE 23rd International Conference on Data Engineering*, 2007, 1204–1211.
- [6] Curtis Preston Mugoh W., “Data Protection Strategies In Today’s Data Center”, *Oracle White Paper*, 2012, 1–8.
- [7] Zhu N., Chiueh T., “Portable and Efficient Continuous Data Protection for Network File Servers”, *Stony Brook University*, 2007, 1–17.
- [8] Meyer D. T., Bolosky W. J., “A Study of Practical Deduplication”, *ACM Transactions on Storage*, **7**:4 (2012), 1–13.
- [9] Storer M. W., Greenan K., Long D. D. E., Miller E. L., “Secure Data Deduplication”, *Proceedings of the 4th ACM international workshop on Storage security and survivability*, 2008, 1–10.
- [10] Renzel K., Keller W., “Client/Server Architectures for Business Information Systems”, *A Pattern Language*, 1997, 1–25.
- [11] Дейт К. Дж., *Введение в системы баз данных*, **8**, Издательский дом "Вильямс", 2005; Date C. J., *An Introduction to Database Systems*, **8**, Pearson Education, Inc., 2004.
- [12] Грофф Д., Вайнберг П., Оппель Э., *SQL: полное руководство*, **3**, Издательский дом "Вильямс", 2015; Groff J., Weinberg P., Oppel A., *SQL The Complete Reference*, **3**, The McGraw-Hill Companies, 2010.

- [13] Дейт К. Дж., *SQL и реляционная теория. Как грамотно писать код на SQL*, Символ-Плюс, 2010; Date C. J., *SQL and Relational Theory. How to Write Accurate SQL Code*, O'Reilly Media Inc., 2009.
- [14] Sebastian J., Aelterman S., *The Art of SQL Server FILESTREAM*, Simple Talk Publishing, 2012.
- [15] Sears R., Catharine van Ingen, Gray J., "To BLOB or Not To BLOB: Large Object Storage in a Database or a Filesystem?", *Technical Report MSR-TR-2006-45*, 2006, 1–11.
- [16] Максимов В., Козленко Л. А., Маркин С. П., Бойченко И. А., "Защищенная реляционная СУБД Линтер", *Открытые системы. СУБД*, 1999, № 11–12; English transl.: Maksimov V., Kozlenko L. A., Markin S. P., Bojchenko I. A., "Zashchishchennaya relyacionnaya SUBD Linter", *Otkrytye sistemy. SUBD*, 1999, № 11–12, (in Russian).]

Taranin S. M., "Backup with Storage in a Database", *Modeling and Analysis of Information Systems*, **23**:4 (2016), 479–491.

DOI: 10.18255/1818-1015-2016-4-479-491

Abstract. This paper presents an overview of some technologies that are used in modern backup systems. We consider their advantages and disadvantages. Next, we consider an example of the realisation of the backup system with files store in the database. We propose to divide the copied files into blocks of fixed length. Each block is a sequence of bytes. The block length may be adaptive, i.e. it can vary depending on the type or file size. We can store the file content in one table, and information of them such as names, attributes, and relationships between them, store in another table. The information of retained files and folders can be stored also on the client side in a hierarchical structure. It is a set of records and a model of the copied directory. The presence of such a model allows to find changes of the copied directory without additional queries to the database. If a file is modified, it is copied only the changed blocks. The model is also updated on the client side. Thus, the load on the data channel reduces. This paper presents the algorithms of saving and restoring data, and describes the factors that affect to the speed of their work. It demonstrates the dependence of the rate of saving and recovery of the fineness of the partition files, as well as the structure of the copied directory.

Keywords: file, data, backup, database, block, model

On the authors:

Taranin Sergey Maksimovich, orcid.org/0000-0001-8117-7358, graduate student,
P.G. Demidov Yaroslavl State University,
Sovetskaya str., 14, Yaroslavl, 150003, Russia, e-mail: staranin0208@yandex.ru

©Цирлин А. М., Ахременков Ал. А., 2016

DOI: 10.18255/1818-1015-2016-4-492-507

УДК 519.8

Вероятностный анализ систем организации турниров

Цирлин А. М., Ахременков Ал. А.

получена 4 февраля 2016

Аннотация.

В работе предложен критерий для сравнения структуры организации турниров в спортивных соревнованиях по вероятности победы в турнире объективно сильнейшего участника. Проанализированы две вероятностные модели результатов парной игры. Получены расчетные формулы для оценки такой вероятности и для плотности распределения вероятности числа очков, набранных в турнире тем или иным игроком. С использованием метода стохастического моделирования проанализированы некоторые реально использующиеся структуры турниров. Полученные результаты определяют и порядок предъявления экспертам объектов при организации экспертизы посредством серии парных сравнений. Получена аналитическая оценка вероятности результатов турнира или серии парных сравнений, позволяющая во многих случаях избежать трудоемкую процедуру перебора допустимых вариантов.

Ключевые слова: организация турниров, вероятность, парные сравнения

Для цитирования: Цирлин А. М., Ахременков Ал. А., "Вероятностный анализ систем организации турниров", *Моделирование и анализ информационных систем*, **23:4** (2016), 492–507.

Об авторах:

Цирлин Анатолий Михайлович, orcid.org/0000-0002-3637-6160, д-р техн. наук., профессор,
Институт программных систем им. А.К. Айламазяна РАН

ул. Петра Первого, 4а, с. Веськово, Переславский р-он, Ярославская обл. 152020 Россия, e-mail: tsirlin@sarc.botik.ru

Ахременков Александр Александрович, orcid.org/0000-0002-5090-6041, канд. физ.-мат. наук,

Институт программных систем им. А.К. Айламазяна РАН

ул. Петра Первого, 4а, с. Веськово, Переславский р-он, Ярославская обл. 152020 Россия, e-mail: alex@eco.botik.ru

Введение

Возникновение и развитие теории вероятностей во многом связано с потребностью анализа игр (см. [1]). Методы теории вероятностей позволяют при некоторых условиях, наложенных на случайный элемент игры, предсказать средние величины, характеризующие ее результаты, при многократном повторении.

Вероятностным методам оценки решения комбинаторных задач посвящена обширная литература (см. [2] – [13]). Ниже рассматривается постановка и анализ решения подобной задачи как экстремальной, критерий которой определен на множестве возможных систем обработки результатов парных сравнений.

Для многих видов игр целью является выявление относительной силы игроков, при этом организация турниров может быть разной для разных соревнований, различных видов спорта и пр. В одних случаях турнир проводят по круговой системе, в других — по кубковой или по «швейцарской». В некоторых случаях игроков предварительно разбивают на группы с последующим розыгрышем первенства между победителями групп и пр. При этом в основе вывода о победителе при любой организации турнира лежат результаты парных сравнений [14] – [19].

В связи с этим возникают вопросы:

–Как оценить организацию турнира?

–Каково должно быть минимальное число игр, чтобы при заданном числе игроков выявить сильнейшего с заданной вероятностью?

–Какова вероятность того, что в результате турнира порядок занятых игроками мест совпадает с их фактической «силой»?

–Как сказывается на этой вероятности правило начисления очков по результатам игры?

Заданными будем считать число игроков и суммарное число игр. Так как целью турнира является выявление сильнейшего игрока, интуитивно ясно, что турнир должен быть организован так, чтобы число встреч между близкими по силе игроками было больше, а число встреч, в которых результат может быть предсказан с вероятностью близкой к единице, меньше.

Для анализа схемы турнира используем вероятностную модель результата встречи двух игроков, причем вероятность того или иного результата встречи должна зависеть от их „силы“. Этот показатель может быть введен по-разному: как вероятность нахождения игрока в том или ином из его возможных состояний, как среднее значение некоторого случайного показателя, сравнение которого для двух участников определяет результат встречи, и пр.

Ниже сделана попытка провести аналитический и численный анализ различных систем розыгрыша. При этом мы первоначально используем наиболее простые вероятностную модель игрока и правило начисления очков. Затем обобщим рассмотрение на модель, более близкую к реальности.

Структура турнира тем лучше, чем больше при заданном общем числе игр N вероятность того, что игрок с наибольшей „силой“ окажется победителем (наберет по результатам турнира не меньшее число очков, чем любой другой его участник).

1. Дискретное распределение состояния игроков и плотность распределения вероятности числа набранных очков

Пусть состояние ξ_j j -го игрока — случайная дискретная величина, принимающая значение 1 с вероятностью s_j и ноль с вероятностью $1 - s_j$. Величину s_j будем называть силой j -го игрока. Состояния игроков независимы друг от друга. Общее число игроков обозначим через M .

Турнир представляет собой последовательность парных сравнений (игр), в которой сравниваются состояния игроков. Если у j -го игрока оно оказалось больше,

чем у k -го, он получает два очка, а его соперник ноль, если состояния одинаковы, то каждый получает по очку. Игроки упорядочены так, что $s_1 > s_2 > \dots s_M$. Результат игры, а значит, набранное в ней каждым участником число очков r – случайная величина, характеризующаяся плотностью распределения вероятностей $P(r)$, определенной на множестве значений $r = 0, 1, 2$ очка.

Вероятностный характер результатов каждой игры приводит к появлению случайных ошибок. Турнир можно рассматривать как фильтр, отделяющий полезный сигнал (априорное распределение сил) от помехи. Этот фильтр тем лучше, чем ближе финальная расстановка игроков к априорной расстановке их сил.

Упорядочим игроков по величине s_j так, что $s_1 > s_2 > s_3 > \dots$. Результат розыгрыша назовем *идеальным*, если занятые игроками места (набранные числа очков) $r_1 \geq r_2 \geq r_3 \dots$, т.е. если порядок мест, определяющийся набранным числом очков и системой розыгрыша, соответствует распределению сил. Результат розыгрыша назовем *правильным*, если игрок, имеющий силу s_1 , оказался на первом месте или поделил его с другими игроками, т.е. $r_1 \geq r_j, j = 2, \dots, M$.

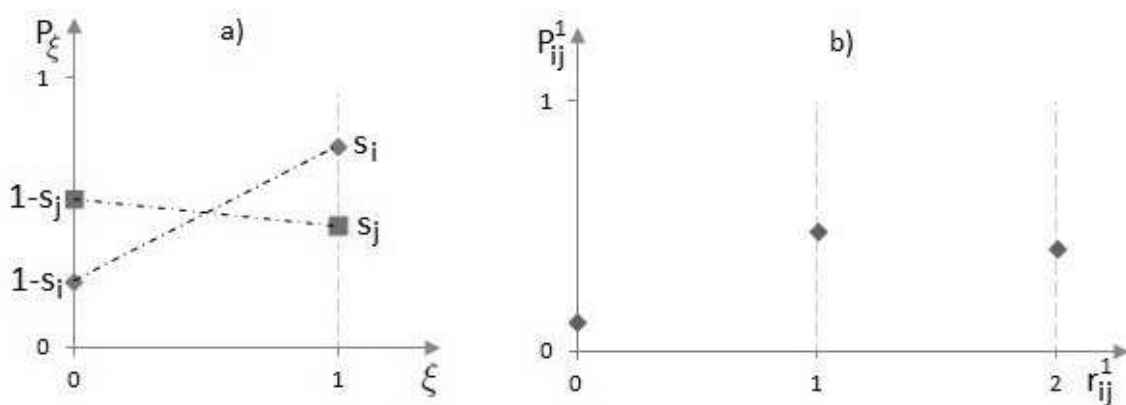


Рис. 1. Плотность распределения состояний i -го и j -го игроков (а) и числа очков, набранных i -м игроком по итогам их встречи (б) для $i < j$

Т.к. число очков, набранное игроком по итогам турнира, случайно, то идеальный или правильный результаты можно ожидать с некоторой вероятностью $P_I(M, N)$ и $P_p(M, N)$, где N — общее количество игр, проведенных в турнире. Далее сосредоточимся на определении вероятности правильности турнира.

Будем полагать, что система розыгрыша A является лучшей, чем система B , если

$$P_p^A(M, N) > P_p^B(M, N). \quad (1)$$

В данной работе проведено сравнение систем розыгрыша: круговой, круговой с предварительным разбиением на группы и выбыванием, кубковой с выбыванием после каждой игры, и пр.

1.1. Плотность распределения числа очков, набранных по итогам турнира

Единичная игра. Состояние j -го игрока — случайная величина, ее математическое ожидание

$$E_{\xi_j} = s_j, \quad (2)$$

а дисперсия

$$D_{\xi_j} = s_j(1 - s_j). \quad (3)$$

Математическое ожидание числа очков, набранных j -м игроком в игре с k -м игроком:

$$E_{r_{jk}} = 2P_{jk}(2) + 1P_{jk}(1) + 0P_{jk}(0),$$

где $P_{jk}(2)$ — вероятность победы j -го, $P_{jk}(0)$ — вероятность поражения j -го, $P_{jk}(1)$ — вероятность ничьей. Число очков k -го игрока $r_{jk}(1)$ в одной встрече j -го и k -го игроков принимает значения 0, 1, 2 с вероятностями

$$\begin{aligned} P_{jk}^1(0) &= P_{kj}^1(2) = s_k(1 - s_j) \\ P_{jk}^1(1) &= P_{kj}^1(1) = 1 - (s_j + s_k - 2s_js_k) = 1 - P_{jk}^1(0) - P_{jk}^1(2) \\ P_{jk}^1(2) &= P_{kj}^1(0) = s_j(1 - s_k). \end{aligned} \quad (4)$$

Так что среднее число очков, набранных j -м игроком во встрече с k -м, и дисперсия r_{jk} равны:

$$E_{r_{jk}} = 1 + s_j - s_k, \quad D_{r_{jk}} = s_j(1 - s_j) + s_k(1 - s_k). \quad (5)$$

Для числа очков $r_j < 0$ и $r_j > 2$ $P^1(r_j) = 0$. Сумма числа очков, набранных обоими игроками в каждой игре, равна 2.

При подсчете числа очков, набранных в турнире j -м игроком, нужно учесть, что он не встречается с самим собой, а значит, и не набирает ни одного очка. Проще всего положить, что

$$P_{jj}^1(0) = 1, \quad P_{jj}^1(1) = P_{jj}^1(2) = 0. \quad (6)$$

В дальнейшем будем предполагать, что результаты игр не зависят друг от друга. Тогда число очков в двух встречах равно $\nu = r_{jk}(2)$, принимает значение от 0, до 4, плотность распределения этой величины равна свертке

$$P_{jk}^2(\nu) = \sum_{\mu=0}^2 P_{jk}^1(\mu)P_{jk}^1(\nu - \mu), \quad \nu = 0, 1, 2, 3, 4. \quad (7)$$

Так, для $\nu = 0$

$$P_{jk}^2(0) = (P_{jk}^1(0))^2,$$

для $\nu = 1$

$$P_{jk}^2(1) = P_{jk}^1(0)P_{jk}^1(1) + P_{jk}^1(1)P_{jk}^1(0) = 2P_{jk}^1(0)P_{jk}^1(1),$$

и т.д.

С учетом (6) для n встреч справедливо рекуррентное соотношение:

$$P_{jk}^n(\nu) = \sum_{\mu=0}^2 P_{jk}^{n-1}(\nu - \mu)P_{jk}^1(\mu), \quad \nu = 0, \dots, 2n. \quad (8)$$

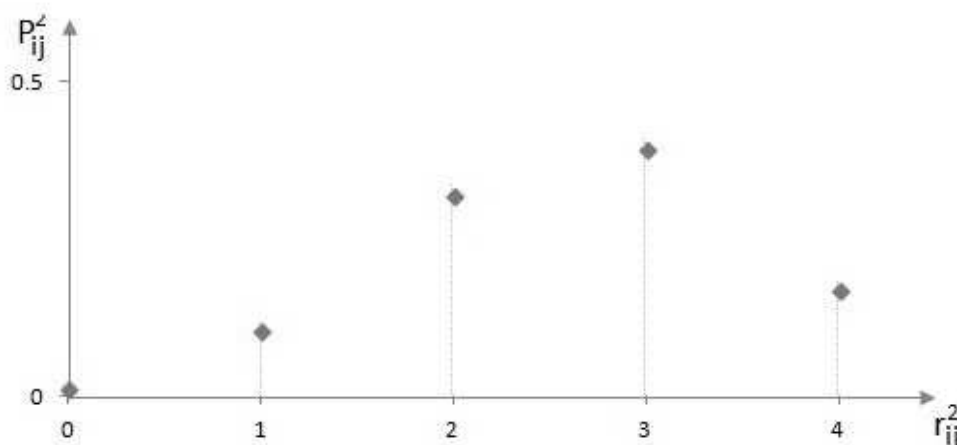


Рис. 2. Плотность распределения числа очков, набранных i -м игроком в двух встречах с j -м игроком при $i < j$

Чем больше n , тем ближе распределение суммарного числа очков к нормальному дискретному закону распределения, определенному на множестве натуральных чисел.

Так как ниже операция свертки

$$z(\nu) = x(\nu) * y(\nu) = \sum_t x(t)y(\nu - t) = \sum_t y(t)x(\nu - t) \quad (9)$$

будет использована неоднократно, напомним некоторые ее свойства [20], [21]:

1. Свертка перестановочна, т.е. результат не зависит от того, в каком порядке под знаком суммы стоят сворачиваемые функции (см. (9)).
2. Область определения z представляет собой объединение областей определения x и y . Так, если одна из функций определена для $\nu = 0, 1, 2$, а другая для $\nu = 0, 1, 2, 3$, то функция z_ν имеет область определения $\nu = 0, 1, \dots, 5$.
3. Так как каждая из сворачиваемых плотностей распределения неотрицательна и сумма ее значений для всех точек области определения равна единице, то тем же условиям отвечает и результат свертки.
4. Математическое ожидание и дисперсия суммы независимых случайных величин равны сумме математических ожиданий и дисперсий каждой из них, что позволяет найти математическое ожидание и дисперсию распределения, являющегося результатом свертки.

1.2. Круговой турнир

Пусть M – число игроков. Общее число игр $N = 0,5M(M - 1)$, а общее число очков, набранных всеми игроками, $R = M(M - 1)$. Число игр, проведенных каждым игроком, $n = M - 1$. Так как порядок игр не влияет на число набранных очков, то будем предполагать, что каждый j -й игрок встречается последовательно с первым, вторым и т.д. до M .

Обозначим число очков, полученных j -м игроком во всех $(M - 1)$ играх турнира, как r_j . Число очков в каждой встрече — дискретная случайная величина, имеющая плотность распределения (4). Плотность распределения числа очков после m встреч связана с плотностью распределения числа очков после $m - 1$ игры соотношением, аналогичным (8)

$$P_j^m(r_j) = \sum_{\mu=0}^2 P_j^{m-1}(r_j - \mu) P_{jm}^1(\mu) = P_j^{m-1} * P_{jm}^1, \quad r_j = 0, \dots, 2m, \quad (10)$$

где $*$ — знак операции свертки. Здесь учтено, что в последней игре j -й игрок встречается с m -м.

В соответствии с (4)

$$\begin{aligned} P_{j1}^1(0) &= s_1(1 - s_j) \\ P_{j1}^1(1) &= 1 - (s_j + s_1 - 2s_j s_1) \\ P_{j1}^1(2) &= s_j(1 - s_1). \end{aligned} \quad (11)$$

Случайная величина r_j принимает целочисленные значения в диапазоне от нуля до $2(M - 1)$ и ее плотность распределения находят рекуррентно с начальными условиями (11) по формуле (10):

$$P_j(r_j) = \sum_{\mu=r_j-2}^{r_j} P_j^{M-2}(\mu) P_{jM}^1(r_j - \mu) = \sum_{\mu=0}^2 P_j^{M-2}(r_j - \mu) P_{jM}^1(\mu). \quad (12)$$

Пределы суммирования определены тем, что аргумент в функции P_{jM}^1 принимает значения от нуля до двух.

Среднее число очков $\bar{r}_j$, набранных в однокруговом турнире j -м игроком, равно сумме по $k (k \neq j) \bar{r}_{jk}$, а дисперсия — такой же сумме $D_{r_{jk}}$. Так что

$$\begin{aligned} \bar{r}_j &= (M - 1)(1 + s_j) - \sum_{k=1, k \neq j}^M s_k, \\ D_{r_{jk}} &= (M - 1)s_j(1 - s_j) + \sum_{k=1, k \neq j}^M s_k(1 - s_k). \end{aligned} \quad (13)$$

Расчет вероятности правильности турнира. Зная $P_i(r_i)$ для всех игроков, нужно найти вероятность того, что один игрок (для определенности первый) набрал не меньше очков, чем j -й. Область определения плотностей распределения вероятностей удовлетворяет условиям

$$0 \leq r_i \leq 2(M - 1), i = 1, \dots, M, \quad (14)$$

$$\sum_{i=1}^M r_i = R = M(M - 1).$$

Пусть первый игрок набрал максимально возможное число очков $2(M - 1)$, тогда с вероятностью единица он набрал больше любого j -го. Турнир заведомо правильный, если по итогам турнира $r_1 \geq 2(M - 1) - 1 = 2M - 3$. В противном случае при вычислении вероятности того, что первый игрок наберет r_1 очков, и при этом j -й наберет $r_j \geq r_1$, нужно учесть отдельно результаты личной встречи, так как эти события не являются независимыми.

Обозначим через $\check{P}_{1j-}(r)$ плотность распределения вероятности того, что во всех встречах 1-го игрока, кроме его встречи с j -м игроком, общее количество набранных очков равно r . Эта плотность вероятности для $0 \leq r \leq 2(M-2)$ находится следующим образом:

$$\check{P}_{1j-}(r) = P_{12}(r) * \dots * P_{1(j-1)}(r) * P_{1(j+1)}(r) * \dots * P_{1M}(r), \quad (15)$$

где $P_{ik}(r)$ – плотность распределения вероятности получения i -м игроком r очков в игре с k -м игроком. Символ $*$ используется для операции свертки:

$$p_1(r) * p_2(r) = \sum_{r_0=0}^2 p_1(r_0)p_2(r-r_0). \quad (16)$$

Аналогично находится $\check{P}_{j1-}(r)$ – плотность распределения очков, набранных во всех играх j -м игроком, кроме игры этого игрока с первым:

$$\check{P}_{j1-}(r) = P_{j3}(r) * P_{j4}(r) * P_{j5}(r) * \dots * P_{jM}(r). \quad (17)$$

Рассмотрим 3 возможных результата встречи между 1-м и j -м игроками.

А) Игрок 1 проиграл игроку j . В этом случае вероятность $P_1(r)$ того, что 1-й игрок за весь турнир набрал r очков равна

$$P_1(r) = \check{P}_{1j-}(r).$$

В) Игрок 1 сыграл вничью с игроком j :

$$P_1(r) = \check{P}_{1j-}(r-1).$$

С) Игрок 1 выиграл у игрока j :

$$P_1(r) = \check{P}_{1j-}(r-2).$$

Найдем вероятность $P(r_1 > r_j)$ того, что количество очков r_1 , набранное 1-м игроком, окажется не меньше, чем количество очков r_j j -го игрока. Выпишем выражения для этой вероятности в зависимости от результатов личной встречи 1-го и j -го игроков:

А)

$$P_a(r_1 \geq r_j) = \sum_{r_1=2}^{2(M-2)} \left(\check{P}_{1j-}(r_1) \sum_{r_j=2}^{r_1} \check{P}_{j1-}(r_j-2) \right)$$

В)

$$P_b(r_1 \geq r_j) = \sum_{r_1=1}^{2(M-2)+1} \left(\check{P}_{1j-}(r_1-1) \sum_{r_j=1}^{r_1} \check{P}_{j1-}(r_j-1) \right)$$

С)

$$P_c(r_1 \geq r_j) = \sum_{r_1=0}^{2(M-2)+2} \left(\check{P}_{1j-}(r_1-2) \sum_{r_j=0}^{r_1} \check{P}_{j1-}(r_j) \right)$$

Вероятность $P(r_1 \geq r_j)$ того, что первый игрок в итоговой таблице окажется не ниже j -го, находят как средневзвешенную из записанных величин с учетом вероятностей результатов личной встречи следующим образом:

$$P(r_1 \geq r_j) = P_{1j}(0)P_a(r_1 \geq r_j) + P_{1j}(1)P_b(r_1 \geq r_j) + P_{1j}(2)P_c(r_1 \geq r_j), \quad (18)$$

где $P_{1j}(r)$ – плотность распределения вероятности того, что 1-й игрок наберет r очков в игре с j -м игроком.

Однако для подсчета вероятности правильности турнира нельзя использовать произведение вероятностей $P(r_1 \geq r_j)$ того, что первый игрок наберет больше очков, чем j -й, для $j = 2, 3, \dots, M$, так как каждая из этих вероятностей зависит от того, набрал ли первый игрок больше очков, чем k -ый. Справедливо утверждение: *Турнир правильный, если для любого возможного количества r_1 очков, набранных первым игроком, никто из остальных игроков не набрал большего числа очков.* Минимальное число очков, которое должен набрать первый игрок, чтобы с вероятностью, большей нуля, разделить первое место, равна $M - 1$. В этом случае все игры закончатся вничью и он разделит первое место с остальными участниками. Если, как указывалось выше, он наберет $2M - 3$ очка, то он заведомо окажется первым или разделит первое место с одним из остальных участников турнира. Так что вероятность правильного турнира в этом случае равна единице.

Нужно учесть, что общее число очков, набранных всеми участниками по итогам турнира, фиксировано и равно $M(M - 1)$. Значит, если первый игрок набрал r_1 очков, то остальные игроки в сумме наберут $M(M - 1) - r_1$ очков. Вероятность того, что все игроки в сумме наберут некоторое число очков r_σ , обозначим через P_σ . Она равна свертке плотностей распределения числа очков, набранных по итогам турнира вторым, третьим и т.д. M -м игроками:

$$P_\sigma(r_\sigma) = P_2(r_2) * P_3(r_3) * \dots * P_M(r_M).$$

Вероятность правильности турнира в общем случае может быть выражена формулой

$$P_p = \sum_{r_1=M-1}^{2(M-1)} P_1(r_1) \sum_{r_j \in V, j=2 \dots M} P_2(r_2)P_3(r_3) \dots P_M(r_M). \quad (19)$$

При этом множество V допустимых значений аргументов $r_j, j = 2, 3, \dots, M$ определено ограничениями:

$$\sum_{j=2}^M r_j = M(M - 1) - r_1, \quad r_j \leq r_1 \forall j.$$

Практически использование формулы (19) требует громоздкого перебора и при $M > 10$ становится слишком затратным по времени вычисления. Ниже приведены формулы, позволяющие получить оценку вероятности правильности турнира с гораздо меньшей трудоемкостью.

Таблица 1. Вероятность правильности турнира, ее оценка и относительная погрешность для различных вариантов распределения сил

P_p	$P_p(M, N)$	$\Delta\%$
0.4021	0.4206	4.60
0.4580	0.4844	5.76
0.5071	0.5254	3.61
0.5513	0.5739	4.10
0.5990	0.6157	2.79
0.6533	0.6807	4.19
0.6952	0.7211	3.73
0.7506	0.7808	4.02
0.8095	0.8111	0.20

Так как $r_\sigma = M(M - 1) - r_1$, то вероятность P_σ определена при фиксированном числе игроков для каждого значения r_1 . Поэтому, имея в виду такую замену, будем писать $P_\sigma(r_1)$. Для оценки вероятности правильности турнира получим выражение

$$P_p(M, N) = \sum_{r_1=M-1}^{2M-3} P_1(r_1) \prod_{i=2}^M \left[1 - \sum_{r_i=r_1+1}^{2M-3} P_i(r_i) \right] P_\sigma(r_1). \quad (20)$$

Здесь случайные события „ни один игрок, кроме первого, не наберет больше очков, чем r_1 “, и „все игроки, кроме первого, наберут в сумме r_σ очков“ предполагаются независимыми. Поэтому найденная таким образом оценка правильности турнира выше истинного значения, найденного по формуле (19), однако расчет не требует перебора.

Для оценки точности последней формулы были проведены численные эксперименты для различных вариантов первоначального распределения сил $s_1, s_2, \dots, s_M$. В таблице 1 приведены результаты этих расчетов для различного априорного распределения сил. В ней показаны вероятность P_p , полученная по формуле (19), оценка вероятности $P_p(M, N)$, найденная по формуле (20), и погрешность $\Delta\%$, полученная по формуле

$$\Delta\% = \frac{P_p(M, N) - P_p}{P_p} 100\%.$$

. Эти расчеты показали, что оценка (20) имеет погрешность порядка 4%, что позволяет ее использовать для сравнения различных схем турнира.

2. Розыгрыш с предварительным разбиением на группы и выбыванием

Будем оценивать структуру турнира значением вероятности его правильности P_p .

Пусть число участников M делится на 4 и команды разбиты на подгруппы по правилу: в первую подгруппу – команды с нечетными, а во вторую – с четными

номерами. Розыгрыш проводится в подгруппах по круговой системе. Отсеивается половина команд, занявших последние места в подгруппах, а оставшиеся разыгрывают первенство по круговой системе. Общее число игр

$$N_{ng} = \frac{3}{4}M \left(\frac{M}{2} - 1 \right). \quad (21)$$

Число игр N_{ng} меньше, чем при круговой системе. Сравним величину P_p для двух систем розыгрыша при $M = 4$, когда система с отбором становится кубковой, а число игр при круговой системе $N = 6$ вдвое превышает число игр при розыгрыше по кубковой.

Для круговой системы P_p может быть найдено по формулам, полученным в предыдущем разделе. Для системы с отбором подгруппы включают команды с s_1, s_3 и s_2, s_4 . Чтобы победить в розыгрыше, игрок с «силой» s_1 должен победить в подгруппе, а затем в финале. Так как эти события независимые, то вероятность победы равна

$$P_p = P_{13}P_{1f} = s_1(1 - s_3)P_{1f}. \quad (22)$$

В свою очередь, вероятность победы в финальной встрече равна

$$P_{1f} = P_{12} \cdot P_{24} + P_{14} \cdot P_{42} = s_1(1 - s_2)s_2(1 - s_4) + s_1(1 - s_4)s_4(1 - s_2) = s_1(1 - s_2)(1 - s_4)(s_2 + s_4). \quad (23)$$

После подстановки этого выражения в (22) получим

$$P_p = s_1^2(1 - s_3)(1 - s_2)(1 - s_4)(s_2 + s_4). \quad (24)$$

Так как число игр при круговой системе вдвое больше, то можно при кубковой системе проводить не по одной, а по две встречи между игроками, считая победителем того, кто набрал больше очков в двух встречах. Тогда вероятность победы первого игрока в своей подгруппе (вероятность набрать более двух очков) по формуле (7)

$$\bar{P}_{13} = s_1(1 - s_3)[1 + 2(1 - s_1 - s_3 + 2s_1s_3)]. \quad (25)$$

Вероятность выхода в финал второго и четвертого игрока равны соответственно

$$\begin{aligned} \bar{P}_{24} &= s_2(1 - s_4)[1 + 2(1 - s_2 - s_4 + 2s_1s_4)], \\ \bar{P}_{42} &= 1 - \bar{P}_{24}. \end{aligned} \quad (26)$$

Вероятность победы первого игрока в розыгрыше

$$\begin{aligned} \bar{P}_p &= \bar{P}_{13}(\bar{P}_{24} \cdot \bar{P}_{12} + \bar{P}_{42} \cdot \bar{P}_{14}) = \\ &= \bar{P}_{13} \left\{ \bar{P}_{24} \cdot s_1(1 - s_2)[1 + 2(1 - s_1 - s_2 + 2s_1s_2)] + \right. \\ &\quad \left. + \bar{P}_{42} \cdot s_1(1 - s_4)[1 + 2(1 - s_1 - s_4 + 2s_1s_4)] \right\}. \end{aligned} \quad (27)$$

Здесь черта соответствует вероятности победы в розыгрыше из двух игр.

Численные расчеты, проведенные по формуле (27), при равномерном распределении сил игроков (а именно когда $s_1 = 0,8$; $s_2 = 0,6$; $s_3 = 0,4$; $s_4 = 0,2$) показали, что при такой организации вероятность правильности турнира равна $\bar{P}_p = 0,645$, при этом оценка вероятности правильного турнира при круговом розыгрыше, найденная по формуле (20), $P_p = 0,670$.

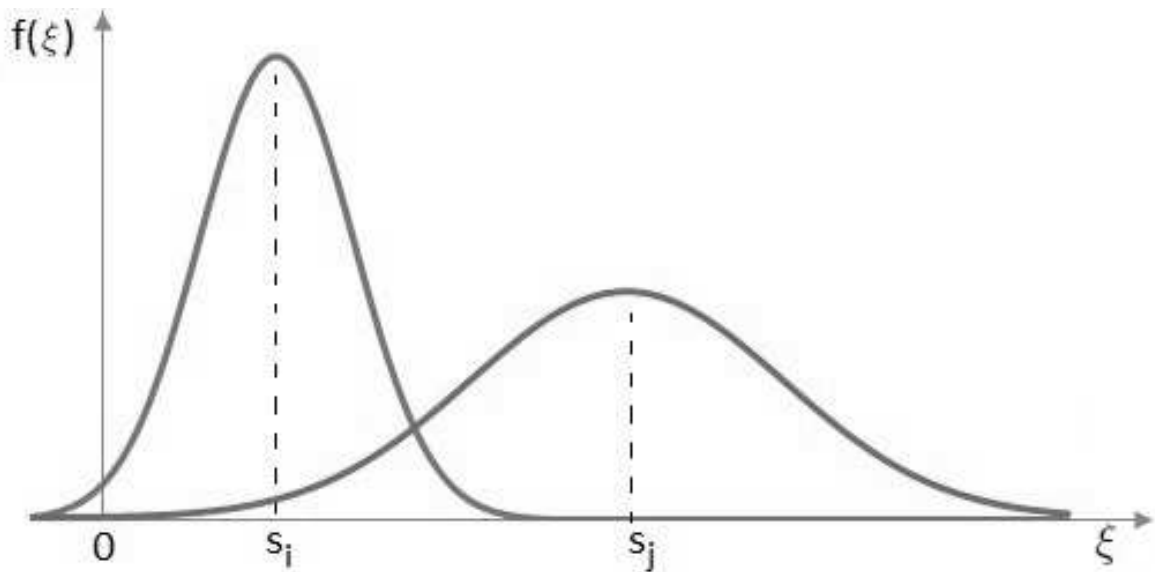


Рис. 3. Плотности распределения состояния i -го и j -го игроков для нормального закона распределения при $s_i < s_j$ и различных D_i и D_j

3. Состояние игрока — случайная величина с непрерывным распределением

Выше рассмотрен случай, когда сила каждого j -го игрока s_j представляет собой вероятность того, что его состояние равно единице. Это обстоятельство облегчало расчет результатов парных сравнений. Рассмотрим, что изменится, если плотность распределения состояния ξ $f_j(\xi_j)$ — непрерывна с математическим ожиданием $\bar{s}_j$ и дисперсией $D_j = \sigma_j^2$. Игрок j побеждает игрока k и получает два очка, когда при парном сравнении $\xi_j > \xi_k$. Так как плотности распределения непрерывны, то $\xi_j = \xi_k$ со сколь угодно малой вероятностью, и таким образом ничьи исключаются. Формулы для вычисления плотности распределения суммы набранных в турнире очков остаются справедливыми, так как они используют только результаты парных сравнений.

Вероятность победы j -го игрока в одиночной встрече с k -м (см. (4)) равна

$$P_{jk}(2) = \int_0^{\infty} f_j(\xi_j) F_k(\xi_j) d\xi_j, \quad P_{jk}(0) = 1 - P_{jk}(2) = P_{kj}(2), \quad (28)$$

где

$$F_k(\xi_j) = \int_0^{\xi_j} f_k(\xi_k) d\xi_k, \quad (29)$$

$$P_{kj} = 1 - P_{jk}. \quad (30)$$

Эти выражения позволяют анализировать исход турниров и находить вероятность их «правильности» при фиксированном числе игр.

Если $f_j(\xi_j)$ — нормальный закон распределения

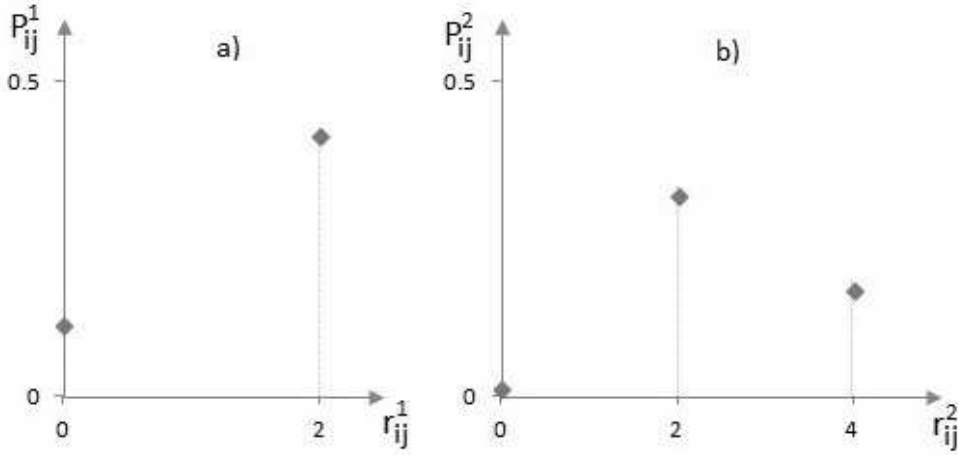


Рис. 4. Плотность распределения результатов встречи i -го и j -го игроков после одной (а) и двух (б) встреч

$$f_j(\xi_j) = \frac{1}{\sqrt{2\pi D_j}} e^{-\frac{(\xi_j - \bar{s}_j)^2}{2D_j}}, \quad (31)$$

как и $f_k(\xi_k)$, то вероятность выигрыша j -го игрока у k -го (проигрыша k -го) равны:

$$P_{jk}(2) = \frac{1}{2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi D_j}} e^{-\frac{(\xi_j - \bar{s}_j)^2}{2D_j}} \left[\Phi\left(\frac{\xi_j - \bar{s}_k}{\sigma_k \sqrt{2}}\right) - \Phi\left(\frac{-\bar{s}_k}{\sigma_k \sqrt{2}}\right) \right] d\xi_j, \quad (32)$$

где

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^x \exp^{-t^2} dt$$

– функция Лапласа, значения которой заданы в таблицах [21].

Для равномерного распределения, равного для каждого игрока нулю за пределами отрезка $\bar{s}_k - \delta_k, \bar{s}_k + \delta_k$ и равного $\frac{1}{2\delta_k}$ внутри этого отрезка

$$\begin{aligned} P_{jk}(2) &= 0 \quad \text{при} \quad (\bar{s}_j - \bar{s}_k) < -(\delta_j + \delta_k), \\ P_{jk}(2) &= \frac{1}{4} \left(\frac{\bar{s}_j - \bar{s}_k}{\delta_j + \delta_k} + 1 \right)^2 \quad \text{при} \quad |\bar{s}_j - \bar{s}_k| < (\delta_j + \delta_k), \\ P_{jk}(2) &= 1 \quad \text{при} \quad (\bar{s}_j - \bar{s}_k) > (\delta_j + \delta_k). \end{aligned} \quad (33)$$

4. Сравнение различных типов турниров методом стохастического моделирования

Проведем сравнение нескольких систем розыгрыша, в которых общее количество проведенных игр одинаково.

Кроме круговой системы, были рассмотрены еще 3 системы розыгрыша.

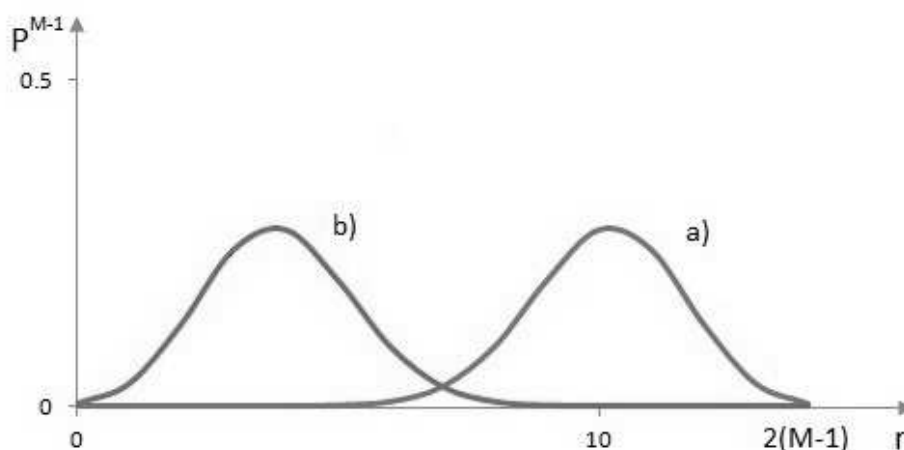


Рис. 5. Плотность распределения числа очков r , набранных в турнире сильнейшим (a) и слабейшим (b) игроками

Система 1.

Всех игроков разбивают на 2 одинаковые по средней силе подгруппы. В каждой подгруппе проводится круговой турнир. Затем половина игроков каждой подгруппы, занявших последние места, образуют новую подгруппу, в которой проводится еще один круговой турнир и выявляется его победитель. Этот победитель и оставшиеся лучшие игроки из каждой подгруппы проводят четвертый круговой турнир. По общей сумме набранных в этом турнире очков выявляются 2 игрока, занявшие 1 и 2 места. Между ними проводятся несколько встреч из расчета, чтобы общее количество проведенных игр совпадало с количеством игр при круговой системе, т.е. было равно $M(M-1)$. Победителем объявляется игрок, набравший максимальное количество очков в этих встречах.

Система 2.

Всех игроков разбивают на 2 одинаковые подгруппы, в каждой подгруппе проводится круговой турнир. По одному игроку, занявшему последнее место в подгруппах, выбывают из турнира. Оставшиеся игроки проводят круговой турнир, после чего два лучших игрока проводят серию финальных встреч, количество которых определяется из условия, чтобы общее число проведенных игр совпадало с числом игр при круговой системе.

Система 3.

Всех игроков разбивают на 2 одинаковые подгруппы, в каждой из которых проводится круговой турнир. Половина игроков каждой подгруппы, занявших последние места, выбывают из турнира. Оставшиеся игроки проводят два круговых турнира, затем два лучших игрока проводят серию финальных встреч, число которых таково, чтобы общее число проведенных игр было равно числу игр при круговой системе.

Кроме этого, были рассмотрены два варианта первоначального разбиения всех игроков на подгруппы. В первом случае (способ А) разбиение проводилось с учетом силы игроков s_i : все игроки выстраивались по уменьшению их силы, первый игрок попадал в подгруппу 1, второй – в подгруппу 2, третий – в 1, четвертый – в

2, и т.д. Второй способ (Б) имитирует широко используемое в спортивных соревнованиях разбиение на подгруппы с помощью жеребьевки, при этом способе игроки разбиваются случайным образом.

При распределении сил игроков были использованы два подхода: равномерное распределение сил, когда различие между s_i одинаковы, и случайное распределение сил. При случайном распределении сил были выбраны 2 характерных случая: “Случайное 1” – когда силы сильнейших игроков близки друг к другу и “Случайное 2” – когда сила сильнейшего игрока существенно больше, чем сила второго игрока.

Приведенные ниже результаты были получены методом стохастического моделирования. При заданных силах игроков $s_1, s_2, \dots, s_M$ на каждой итерации находилась случайная реализация турнира и определялась случайная величина ζ_i , принимающая значение 1, если искомое событие (в данном случае сильнейший игрок занял первое место) произошло, либо 0 в противном случае. Оценка вероятности P^* правильности турнира определялась как оценка математического ожидания $E\zeta$ случайной величины ζ .

Количество итераций K находилось из условия, чтобы с заданной вероятностью Q отклонение оценки P^* от истинной вероятности правильности турнира было меньше заданной точности ε

$$P(|P^* - p| < \varepsilon) = Q.$$

Для этих условий K определяется следующим образом [22]:

$$K = \left(\frac{\sigma_\zeta}{\varepsilon}\right)^2 \left[\Phi^{-1}\left(\frac{Q}{2}\right)\right]^2,$$

где σ_ζ – оценка среднего квадратичного отклонения случайной величины ζ , Φ^{-1} – функция, обратная функции Лапласа.

Чтобы найти первую оценку среднего квадратичного отклонения σ_ζ , сначала находилось некоторое количество k_0 значений ζ_i , после чего вычислялись оценка для σ_ζ и значение K . Далее эти величины уточнялись после каждой итерации.

Результаты сравнительных экспериментов приводятся в таблице 2, в ней показаны вероятности того, что сильнейший игрок займет первое место. В данных экспериментах:

- количество игроков $M = 8$;
- общее число игр в турнире $N = 56$;
- уровень доверия $Q = 0,999$;
- точность $\varepsilon = 0,001$;
- начальное число итераций $k_0 = 1\,000\,000$.

Анализируя результаты проведенных экспериментов, можно сделать следующие выводы:

1. Выбор системы розыгрыша может оказать существенное влияние на эффективность турнира.

2. Третья система организации турнира для всех вариантов априорного распределения сил превосходит остальные.

Таблица 2. Вероятность правильности турнира для различных схем его организации

Распределение сил	Равномерное		Случайное 1		Случайное 2	
Способ разбиения	А	Б	А	Б	А	Б
Круговой	0,41		0,26		0,60	
Система 1	0,39	0,42	0,20	0,27	0,58	0,61
Система 2	0,49	0,49	0,30	0,32	0,69	0,69
Система 3	0,55	0,56	0,32	0,37	0,76	0,76

5. Заключение

Предложен критерий для оценки способа организации турнира с заданным числом участников и ограниченным количеством игр. Получены выражения для аналитической оценки качества организации турнира и проведено статистическое моделирование для расчета критерия правильности той или иной структуры розыгрыша.

Все полученные рекомендации в иных терминах относятся к задаче определения лучшего изделия посредством экспертизы, в которой эксперт выбирает лучшее путем серии парных сравнений.

Список литературы / References

- [1] Гнеденко Б. В., *Курс теории вероятностей*, Наука, М., 1988; [Gnedenko B. V., *Kurs teorii veroyatnostey*, Nauka, M., 1988, (in Russian).]
- [2] Эрдеш П., Спенсер Дж., *Вероятностные методы в комбинаторике*, Мир, М., 1976; [Erdesh P., Spenser Dzh., *Veroyatnostnye metody v kombinatorike*, Mir, M., 1976, (in Russian).]
- [3] Эрдеш П., Мун Дж. В., “О множестве согласованных дуг в турнире”, *Теория графов*, Мир, М., 1976, 160–162; [Erdesh P., Mun Dzh. V., “O mnozhestve soglasovannykh dug v turnire”, *Teoriya grafov*, Mir, M., 1976, 160–162, (in Russian).]
- [4] Alspach B., Mason D. W., Pullman N. J., “Path numbers of tournaments”, *J. of Comb Theory*, **20**:3 (1976), 222–228.
- [5] Daniels H. E., “Round-robin touranmet scores”, *Biomentrika*, **56**:2 (1969), 295–299.
- [6] Ford L. R., Jahnsen S. M., “A tournament problem”, *Am. Math. Monthly*, **66** (1959), 387–389.
- [7] Freund J. E., “Round Robin Mathematics”, *Am. Math. Monthly*, **63** (1956), 112–114.
- [8] Hartigan J. A., “Probabilistic competition of knockout tournament”, *Ann. Math. Statist.*, **37** (1966), 495–503.
- [9] Moon J. W. Topics on tournaments in graph theory. Dover Publications, Inc., Meniola, NY, 2015.
- [10] Narayana T. V., Bent B. H., “Computation of the number of score sequence in round-robin tournamets”, *Canad. Math. Bull.*, **7** (1964), 133–136.
- [11] Narayana T. V., Zidek J., “Statistical inference in random tournaments”, *Rev. Roum. Math. Pures et Appl.*, **10** (1969), 1563–1576.
- [12] Searls D. T., “On the probability of winning with different tournament procedures”, *J. Amer. Statist. Assoc.*, **58**:304 (1963), 1064–1081.

- [13] Thompson G.L., *Lectures on game theory, Markov chains and related topics*, Sandia Corporation Monograph, 1958.
- [14] David H. A., *The Method of paired comparisons*, London, 1976.
- [15] David H. A., "Ranking the Players in a Round Robin Tournament", *Rev. Int. Statist. Inst.*, **39**, 1971, 137–147.
- [16] David H. A., "Tournaments and paired comparisons", **46** (1959), 139–149.
- [17] Glason J. R., Haplerin S. A., "A paired comparisons model for Round-Robin experiments", *Psichomentika*, **40**:4 (1975), 433–454.
- [18] Glenn W. A., "A comparison for the effectiveness of tournaments", *Biometrika*, **47** (1960), 253–262.
- [19] Брук Б. Н., Бурков В. Н., "методы экспертных оценок в задачах упорядочивания объектов", *Техн. кибернетика*, 1972, 29–39; [Bruck B. N., Burkov V. N., "Metody ekspertnykh otsenok v zadachakh uporyadochivaniya obektov", *Tekhn. kibernetika*, 1972, 29–39, (in Russian).]
- [20] Хиршман И. И., Уиддер Д. В., *Преобразования типа свертки*, Мир, М., 1958; [Khirshman I. I., Uidder D. V., *Preobrazovaniya tipa svertki*, Mir, M., 1958, (in Russian).]
- [21] Диткин В. А., Кузнецов П. И., *Справочник по операционному исчислению: Основы теории и таблицы формул*, Государственное издательство технико-теоретической литературы, М., 1951; [Ditkin V. A., Kuznetsov P. I., *Spravochnik po operatsionnomu ischisleniyu: Osnovy teorii i tablitsy formul*, Gosudarstvennoe izdatelstvo tekhniko-teoreticheskoy literatury, M., 1951, (in Russian).]
- [22] Вентцель Е. С., *Исследование операций*, Советское радио, 1972; [Venttsel E. S., *Issledovanie operatsiy*, Sovetskoe radio, M., 1972, (in Russian).]

Tsirlin A. M., Akhremenkov Al. A., "Probabilistic Analysis of Tournament Organization Systems", *Modeling and Analysis of Information Systems*, **23**:4 (2016), 492–507.

DOI: 10.18255/1818-1015-2016-4-492-507

Abstract. In this paper a criteria of comparison of different tournament organization systems in sporting contests is offered; the criteria uses the probability of winning the fairly strongest player. Two probabilistic models have been analyzed. Calculating formulas for estimating the probability and probability density of score points gained by one or another player were obtained. Some really used tournament systems were analyzed with the stochastic modeling method. The available results also provide an order of objects presenting to experts while organizing the examination by paired comparison. An analytical estimation of probability of tournament results (or pared comparison) was obtained. In many cases it allows to avoid a time-consuming procedure of sorting out possible variants.

Keywords: tournament organization, probability, paired comparisons

On the authors:

Tsirlin Anatoly Michailovich, orcid.org/0000-0002-3637-6160, Prof,
Program Systems Institute of RAS
Petra 1 str., 4a, Veskovo Jaroslavskoy, 152020, Russia, e-mail: tsirlin@sarc.botik.ru

Akhremenkov Alexander Alexandrovich, orcid.org/0009-0203-2514-7755, PhD,
Program Systems Institute of RAS
Petra 1 str., 4a, Veskovo Jaroslavskoy, 152020, Russia, e-mail: alex@eco.botik.ru