

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Advanced science

Научный журнал

№ 3 (3)

Киров

2013

Главный редактор:

Пугач В. Н.,

кандидат экономических наук,
ректор ФГБОУ ВПО «ВятГУ»

Заместитель главного редактора:

Литвинец С. Г.,

кандидат сельскохозяйственных наук, доцент,
проректор по НиИ ФГБОУ ВПО «ВятГУ»

Редакционная коллегия:

Частиков А. В.,

доктор технических наук, профессор,
декан ФПМТ ФГБОУ ВПО «ВятГУ»

Князьков В. С.,

доктор технических наук, профессор кафедры ЭВМ ФГБОУ ВПО «ВятГУ»

Шатров А. В.,

доктор физико-математических наук, профессор,
заведующий кафедрой ММЭ ФГБОУ ВПО «ВятГУ»

Черепанов В. В.,

доктор технических наук, профессор кафедры ЭПС ФГБОУ ВПО «ВятГУ»

Эфрос Е. И.,

доктор технических наук, профессор кафедры ТиГ ФГБОУ ВПО «ВятГУ»

Прозоров Д. Е.,

доктор технических наук, профессор кафедры РЭС ФГБОУ ВПО «ВятГУ»

Петров Е. П.,

доктор технических наук, профессор кафедры РЭС ФГБОУ ВПО «ВятГУ»

Решетников С. М.,

доктор технических наук, профессор, заведующий кафедрой физики ФГБОУ ВПО
«ВятГУ»

Тюкалов Ю. Я.,

доктор технических наук, профессор кафедры СК ФГБОУ ВПО «ВятГУ»

Красных А. А.,

доктор технических наук, профессор, заведующий кафедрой ЭиЭ ФГБОУ ВПО
«ВятГУ»

Ковалевский А. В.,

доктор технических наук, профессор, заведующий кафедрой ОХ ФГБОУ ВПО
«ВятГУ»

Флегентов И. В.,

доктор технических наук, профессор, заведующий кафедрой ПЭБ ФГБОУ ВПО
«ВятГУ»

Бывалов А. А.,

доктор биологических наук, профессор кафедры БТ ФГБОУ ВПО «ВятГУ»

Дармов И. В.,

доктор медицинских наук, профессор, заведующий кафедрой
микробиологии ФГБОУ ВПО «ВятГУ»

Палкина М. В.,

доктор экономических наук, профессор, заведующая кафедрой РЭУ ФГБОУ ВПО
«ВятГУ»

Печенкин А. А.,

доктор исторических наук, профессор, декан ГФ ФГБОУ ВПО «ВятГУ»

Ненашев М. И.,

доктор философских наук, профессор кафедры ФСП ФГБОУ ВПО «ВятГУ»

Стяжкина С. А.,

начальник редакционно-издательского отдела ФГБОУ ВПО «ВятГУ»

Редакторы: Куликова А. В., Мухачева М. С.

Верстка: Мухачева М. С.

«Вятский государственный университет» ПРИП ФГБОУ ВПО «ВятГУ» 610000,
Киров, ул. Московская, 36 Тел.: (8332) 64-23-56, <http://vyatsu.ru>

Свидетельство о регистрации средства массовой информации
(Министерство по делам печати, телерадиовещания и средств массовых
коммуникаций)

Эл. № ФС77-50605

Содержание

В третий номер журнала «Advanced science» вошли статьи и доклады с VII Всероссийской научной конференции «Математическое моделирование развивающейся экономики, экологии и биотехнологий» (ЭКОМОД 2012)

Андреев М. Ю.

Дезагрегирование задачи благосостояния в задачу межвременного равновесия с однозначным определением равновесных цен

Арсеньев-Образцов С. С.

Численное решение трехмерной динамической задачи теории вязкоупругости

Белотелов Н. В., Коноваленко И. А.

Модель коллективного поведения животных

Белотелов Н. В.

Моделирование миграции с учетом образовательного процесса

Березкин В. Е., Лотов А. В.

Использование функций включения для сравнения точечных аппроксимаций многомерной границы Парето

Брацун Д. А., Захаров А. П.

Численное исследование динамики распределенной системы с запаздыванием в модели циркадианных ритмов *Neurospora crassa*

Волченская Т. В., Князьков В. С.

Спецпроцессоры для решения задач на графах

Вржещ В. П., Масютин А. А.

Модельное дезагрегирование макроэкономической статистики на примере России, Украины и Финляндии

Гасникова Е. В.

О новых достаточных условиях существования равновесия макросистемы

Гасников А. В., Холодов Я. А.

Обобщенная СТМ-модель транспортных потоков и исследование устойчивости её стационарных режимов

Горбачев В. А., Оленёв Н. Н.

Идентификация модели добывающего сектора экономики Монголии

Князьков В. С., Исупов К. С.

Программный пакет высокоточных модулярно-позиционных вычислений с плавающей точкой

Князьков В. С., Осинин И. П.

Однородный арифметический процессор с реконфигурируемой микроархитектурой для массовых арифметических операций

Кошкин Ю. Л., Сошникова Е. М.

Две модели прогнозирования валового регионального продукта Кировской области

Кошкин Ю. Л.

Имитация структурно-временных данных системной регрессии

Ростовцев В. С., Смехова В. М.

Моделирование процесса логического вывода с использованием ДНК-цепочек

**ДЕЗАГРЕГИРОВАНИЕ ЗАДАЧИ БЛАГОСОСТОЯНИЯ В ЗАДАЧУ
МЕЖВРЕМЕННОГО РАВНОВЕСИЯ С ОДНОЗНАЧНЫМ
ОПРЕДЕЛЕНИЕМ РАВНОВЕСНЫХ ЦЕН***

**WELFARE PROBLEM DISAGGREGATION INTO
INTERTEMPORAL EQUILIBRIUM PROBLEM WITH UNIQUE
EQUILIBRIUM PRICES**

Рассматривается модель межвременного равновесия с конечным числом однотипных агентов-потребителей. Потребители обмениваются друг с другом единственным продуктом, выбирая оптимальное потребление, будучи стесненными бюджетным ограничением и ограничением ликвидности. Показано, что в модели существует эффективное равновесие, а цены определяются однозначно. Таким образом, добавление ограничения ликвидности решает проблему неоднозначного определения цен в моделях типа Эрроу-Дебре и их производных.

Intertemporal equilibrium model with the finite number of agents-consumers is considered. Consumers exchange the only commodity with each other. They choose optimal consumption plan subject to budget constraint and liquidity constraint. Existence of efficient equilibrium is proven. The established fact of the uniqueness of equilibrium prices gives hope that equilibrium constraint could remove equilibrium indeterminacy problem of the stochastic Arrow-Debreu and derivative models.

* Работа выполнена при поддержке РФФИ (проекты № 11-01-00644, 11-01-12136-офи-м-2011), ПФИ ОМН РАН №3 (проект 3.14), ПФИ Президиум РАН №14 (проект 109), а также частично выполнена в научно-учебной лаборатории макроструктурного моделирования экономики России факультета экономики ГУ-ВШЭ.

Введение

В модели Эрроу-Дебре и во многих других моделях общего равновесия равновесные цены определяются с точностью до нормировки [1]: можно умножить все цены и все денежные величины на одно и то же число и полученные цены и денежные величины также будут равновесными. Этот факт не вызывает неудобств, пока модели детерминированные: цена в начальный момент времени обычно выбирается совпадающей со статистикой, а совпадение цен на всей траектории уже зависит от качества модели.

Трудности возникают, когда речь заходит о стохастических моделях типа Эрроу-Дебре. В стохастических моделях тоже можно одновременно перенормировать все цены и получить то же самое равновесие. Но появляется дополнительная степень свободы: цены можно перенормировать еще и для каждого варианта развития будущего, т. е. для каждой реализации случайного процесса. В результате, если в детерминированных моделях неоднозначность определения цен (возможность перенормировки) приводит лишь к номинальной неопределённости равновесия (*nominal indeterminacy*), т. е. реальные величины экономики остаются прежними, то в стохастических моделях это приводит к реальной неопределенности (*real indeterminacy* [2, 3]), т. е. различным векторам (равновесных) цен будут соответствовать различные вектора (равновесных) реальных величин экономики.

Равновесия могут доминировать друг друга. Как показано в [4], иногда доминирующими равновесиями являются равновесия, у которых одна из цен бесконечна. Такие равновесия трактуются в [4] как равновесия с дефолтами: при определенных событиях кредиторы прощают долги должникам. Однако с прикладной точки зрения возможность существования таких равновесий кажется недостатком теоретических конструкций, связанных с отсутствием механизма определения цен.

В теории вопрос реальной неопределённости зачастую обходится благодаря обеспечению свойства полноты рынков [5]. Есть несколько определений понятия «полных рынков», но автор склоняется к следующему: в

модели равновесия рынки полны, если все равновесия реализуют одно и то же решение задачи благосостояния (задачи идеального планирования). В силу совпадения равновесий вопрос неоднозначности несущественен. Полнота же обычно обеспечивается, грубо говоря, наличием в экономике финансовых инструментов в количестве не меньшем, чем число случайных исходов в следующий момент времени. Однако преобладает мнение, что реальная экономика соответствует случаю неполных рынков, поскольку инструменты конечны, а сомнения, или неопределённости, бесконечны.

Было решено отбросить стохастическую постановку задачи и рассмотреть детерминированную модель. В модель чистого обмена было включено ограничение ликвидности, которое и определяет равновесные цены однозначно. Однозначное определение равновесных цен позволяет надеяться, что в будущем и в стохастическом аналоге модели цены определятся однозначно.

В рассматриваемую ниже модель была включена возможность обмениваться кредитами, что приводит к эффективности равновесия.

1. Модель чистого обмена с ограничением ликвидности

Задача потребителя. Будем считать, что экономика состоит из конечного числа потребителей $n \in \mathcal{N}$, которые живут бесконечно долго. В последовательные периоды времени $t \in \mathcal{T} = \mathbb{N} = \{1, 2, \dots\}$, начиная с первого, каждый потребитель $n \in \mathcal{N}$ получает запас продукта $w_t^n \in \mathbb{R}_{++}$. Продукт один, рассмотрение нескольких продуктов не привносит в модель ничего существенного.

Будем считать, что все потребители обладают полным предвидением [5]. Это означает, что все будущие запасы продукта $w_t^n, t \in \mathcal{T}$ известны потребителю уже в первый момент времени. Запас продукта w_t^n потребитель продаёт на рынке в момент времени t по сложившейся цене p_t и приобретает тот же продукт для потребления в количестве $c_t^n \in \mathbb{R}_+$. Потребитель может взять кредит у других потребителей или, наоборот, одолжить. Будем обозначать кредиты

через L_t^n : если $L_t^n > 0$, то потребитель взял кредит в момент времени t , если $L_t^n < 0$ – выдал. В момент $t+1$ кредит необходимо вернуть с процентом R_t (R_t – брутто-процент, $R_t = 1 + r_t, r_t \geq 0$). Количество денег, имеющееся у потребителя к моменту времени t , будем обозначать через Φ_t^n .

В первый момент времени потребитель планирует все свои управления – c_t^n, Φ_t^n, L_t^n – на текущий $t=1$ и все будущие периоды $t \in \mathcal{T}$. При выборе плана $\langle c_t^n, \Phi_t^n, L_t^n |_{t \in \mathcal{T}} \rangle$ потребитель стеснён бюджетными ограничениями

$$\Phi_t^n = \Phi_{t-1}^n + p_t (w_t^n - c_t^n) + L_t^n - R_{t-1} L_{t-1}^n, \quad t \in \mathcal{T}, \quad (1)$$

и ограничениями ликвидности

$$\Phi_{t-1}^n \geq p_t c_t^n + L_t^n - R_{t-1} L_{t-1}^n, \quad t \in \mathcal{T}. \quad (2)$$

В соответствии с используемым нами принципом полного предвидения, каждый потребитель при выборе плана $\langle c_t^n, \Phi_t^n, L_t^n |_{t \in \mathcal{T}} \rangle$ наперёд знает траектории всех информационных переменных: цен p_t , процентов R_t , и, как уже говорилось, запаса продукта w_t^n .

Смысл ограничения ликвидности (2) состоит в том, что потребитель должен сначала накопить некоторое количество наличности Φ_{t-1}^n , чтобы купить товар c_t^n . Он не может сказать продавцу: я возьму твой товар, а деньги верну тебе вечером, когда заплатят за мой товар w_t^n . Но, если ему не хватает денег для покупки, он может взять кредит L_t^n . При этом он обязан имеющимися деньгами Φ_{t-1}^n и кредитом L_t^n оплатить вчерашние займы $R_{t-1} L_{t-1}^n$.

Заметим, что в ограничении ликвидности (2) стоит полная стоимость потребляемого товара: $p_t c_t^n$. Более естественно было бы написать, что некая денежная масса, Φ_{t-1}^n , должна быть не меньше товарооборота $p_t [c_t^n - w_t^n]_+$, где

$[a]_+ = \max\{a, 0\}$, поскольку в действительности потребитель будет покупать только ту часть потребления c_t^n , которая превосходит его запас w_t^n . Однако задача с подобным ограничением не кажется автору решаемой.

Стоит также отметить, что рассматривалось несколько вариантов ограничения (2) относительно наличия в нём членов с кредитами. Представленный вариант оказался наиболее корректным и интерпретируемым.

Будем считать, что план $\langle c_t^n, \Phi_t^n, L_t^n |_{t \in T} \rangle$ потребитель выбирает так, чтобы максимизировать полезность потребления:

$$\sum_{t=1}^{\infty} U^n(c_t^n) e^{-\delta t} \longrightarrow \max \quad (3)$$

Функция U^n выпукла, и для простоты дважды дифференцируема.

Чтобы потребитель не занимал слишком много и не потреблял больше своих возможностей, необходимо ввести некоторое условие отсутствия пирамиды. В данном случае будем считать выбираемый план допустимым, если

$$\lim_{T \rightarrow \infty} (\Phi_T^n - R_T L_T^n) / \prod_{\tau=1}^T R_{\tau} \geq 0 \quad (4)$$

Это условие требует, чтобы приведенный к настоящему времени предельный долг был не больше приведённых предельных денежных средств. К этому условию мы еще вернемся далее.

Задачу максимизации (3) при условиях (1), (2), (4) будем называть *задачей потребителя*.

Условия равновесия. Выбранные потребителями планы кредитов, денег, потреблений должны быть согласованы в равновесии: спрос не должен превышать предложения. Поэтому в модели присутствуют следующие условия равновесия рынков:

$$\sum_{n \in \mathcal{N}} w_t^n \geq \sum_{n \in \mathcal{N}} c_t^n, \quad t \in \mathcal{T}, \quad (5)$$

$$\sum_{n \in \mathcal{N}} L_t^n = 0, \quad t \in \mathcal{T}, \quad (6)$$

$$\sum_{n \in \mathcal{N}} \Phi_t^n = \sum_{n \in \mathcal{N}} \Phi_0^n, \quad t \in \mathcal{T}. \quad (7)$$

Условие равновесия на рынке продукта (5) неявно определяет цену p_t , условие (6) неявно определяет процент R_t , а условие (7) ничего не определяет, так как неявно является следствием других условий равновесия и ограничений (1), (2).

Будем считать, что нулевой период времени (начальные условия) ничем не отличается от других, поэтому $\sum_{n \in \mathcal{N}} L_0^n = 0$.

Итак, моделью чистого обмена с ограничением ликвидности будем называть задачу (1)-(7).

2. Дезагрегирование задачи благосостояния

Задачей благосостояния или идеального планирования называется задача управления ресурсами экономики одним лицом. Как следствие, в задаче отсутствует какое-либо взаимодействие между агентами экономики – потребителями. Управляющее лицо максимизирует линейную свёртку функционалов (3)

$$\sum_{n \in \mathcal{N}} a^n \sum_{t=1}^{\infty} U^n(c_t^n) e^{-\delta t} \longrightarrow \max \quad (8)$$

при ограничениях материального баланса (5) по потреблением $\langle c_t^n |_{t \in \mathcal{T}} \rangle$.

Отметим, что решений *задачи благосостояния* (8), (5) много, они зависят от выбранных взвешивающих коэффициентов $\langle a_n |_{n \in \mathcal{N}} \rangle$. Но при конкретном наборе взвешивающих коэффициентов решение, как правило, одно.

Задаче (8), (5) соответствует функционал Лагранжа:

$$L = \sum_{n \in \mathcal{N}} a^n \sum_{t=1}^{\infty} U^n(c_t^n) e^{-\delta t} + \sum_{t=1}^{\infty} \tilde{p}_t \left(\sum_{n \in \mathcal{N}} w_t^n - \sum_{n \in \mathcal{N}} c_t^n \right), \quad (9)$$

где $\tilde{p}_t \left(\sum_{n \in \mathcal{N}} w_t^n - \sum_{n \in \mathcal{N}} c_t^n \right) = 0$, $\tilde{p}_t \geq 0$. Решение задачи благосостояния (8), (5)

вместе с множителями Лагранжа $\tilde{p}_t$ будем обозначать как $\langle \tilde{c}_t^n, \tilde{p}_t |_{t \in \mathcal{T}} \rangle$.

Дезагрегировать задачу благосостояния (8), (5) в задачу равновесия (1)-(7) означает представить решение задачи равновесия $\langle c_t^n, \Phi_t^n, L_t^n, p_t, L_t |_{t \in \mathcal{T}} \rangle$ через решение $\langle \tilde{c}_t^n, \tilde{p}_t |_{t \in \mathcal{T}} \rangle$ при

$$c_t^n = \tilde{c}_t^n. \quad (10)$$

Тогда решение задачи равновесия называют *эффективным*.

В экономической теории возможность дезагрегирования интерпретируется как свидетельство о том, что рыночная система даёт не худшее распределение благ, чем идеальное планирование (Первая и Вторая теоремы теории благосостояния). На практике дезагрегирование это часто удобный способ получить решение задачи равновесия.

Итак, проведём дезагрегирование. Для начала составим функционал Лагранжа для задачи (1)-(3):

$$\begin{aligned} L^n = & \sum_{t=1}^{\infty} U^n(c_t^n) e^{-\delta t} + \sum_{t=1}^{\infty} \lambda_t^n \left(\Phi_{t-1}^n - \Phi_t^n + p_t (w_t^n - c_t^n) + L_t^n - R_{t-1} L_{t-1}^n \right) + \\ & + \sum_{t=1}^{\infty} \mu_t^n \left(\Phi_{t-1}^n - p_t c_t^n + L_t^n - R_{t-1} L_{t-1}^n \right) \end{aligned}$$

Здесь λ_t^n, μ_t^n – множители Лагранжа, причём в силу знака неравенства в (2) считаем

$$\mu_t^n \geq 0, \mu_t^n (\Phi_{t-1}^n - p_t c_t^n + L_t^n - R_{t-1} L_{t-1}^n) = 0. \quad (11)$$

Условия первого порядка следующие:

$$\frac{\partial U^n(c_t^n)}{\partial c_t^n} e^{-\delta t} = p_t (\lambda_t^n + \mu_t^n), \quad (12)$$

$$\frac{\partial U^n}{\partial \Phi_t^n} e^{-\delta t} = -\lambda_t^n + \lambda_{t+1}^n + \mu_{t+1}^n = 0, \quad (13)$$

$$\frac{\partial U^n}{\partial L_t^n} e^{-\delta t} = \lambda_t^n + \mu_t^n - R_t (\lambda_{t+1}^n + \mu_{t+1}^n) = 0. \quad (14)$$

Будем предполагать, что дезагрегирование возможно. Тогда $c_t^n = \tilde{c}_t^n$, а из условий (13), (12) и дифференцирования (9) по $\tilde{c}_t^n$ следует $\lambda_t^n = \sum_{\tau=t+1}^{\infty} \mu_{\tau}^n$ и

$\tilde{p}_t/a^n = p_t \sum_{\tau=t+1}^{\infty} \mu_{\tau}^n$. Тогда условие $\mu_t^n \geq 0$ (см. (11)) можно переписать как $\mu_t^n = \sum_{\tau=t}^{\infty} \mu_{\tau}^n - \sum_{\tau=t+1}^{\infty} \mu_{\tau}^n = \frac{1}{a^n} (\tilde{p}_t/p_t - \tilde{p}_{t+1}/p_{t+1}) \geq 0$ при $a^n \neq 0$, а условие $\mu_t^n (\Phi_{t-1}^n - p_t c_t^n + L_t^n - R_{t-1} L_{t-1}^n) = 0$ как $(\tilde{p}_t/p_t - \tilde{p}_{t+1}/p_{t+1}) (\Phi_{t-1}^n - p_t c_t^n + L_t^n - R_{t-1} L_{t-1}^n) = 0$. После суммирования по n с учетом условий равновесия это даст $(\tilde{p}_t/p_t - \tilde{p}_{t+1}/p_{t+1}) \left(\sum_{n \in \mathcal{N}} \Phi_0^n - p_t \sum_{n \in \mathcal{N}} w_t^n \right) = 0$, причем в силу (2) и (11) каждое из слагаемых неотрицательно. Из этого равенства и неравенств следует, что

$$p_t = \min \left\{ \sum_{n \in \mathcal{N}} \Phi_0^n / \sum_{n \in \mathcal{N}} w_t^n, p_{t+1} \tilde{p}_t / \tilde{p}_{t+1} \right\}. \quad (15)$$

Из этой формулы следует, что для определения p_t ключевой является последовательность

$$b_t = \sum_{n \in \mathcal{N}} \Phi_0^n / \left(\sum_{n \in \mathcal{N}} w_t^n \tilde{p}_t \right). \quad (16)$$

Если мы хотим, чтобы цены p_t существовали, нужно потребовать, чтобы последовательность (16) была возрастающей (быть может, не монотонно). Точнее, чтобы для любого $t \in \mathcal{T}$ существовал $\tau^* = \underset{\tau \geq t}{\text{Arg min}} \{b_\tau\}$ и $\forall k > \tau^* : b_k > b_{\tau^*}$. Обозначим через $\mathcal{T}^*$ множество всех таких периодов времени τ^* :

$$\mathcal{T}^* = \left\{ \tau^* \mid \tau^* \in \mathcal{T} \cap \exists t \in \mathcal{T} : \tau^* = \underset{\tau \geq t}{\text{Arg min}} \{b_\tau\} \cap \forall k > \tau^* : b_k > b_{\tau^*} \right\}. \quad (17)$$

На самом деле мы будем предполагать более сильное свойство, чем возрастание (16). Будем предполагать, что сумма $\sum_t \sum_{n \in \mathcal{N}} w_t^n \tilde{p}_t$ имеет конечный предел:

$$\exists \lim_{T \rightarrow \infty} \sum_t^T \sum_{n \in \mathcal{N}} w_t^n \tilde{p}_t < +\infty \quad (18)$$

Тогда *возрастание* (16) и бесконечность множества (17) автоматически выполнено.

Выражение (15) можно теперь переписать в виде:

$$p_t = \tilde{p}_t / \tilde{p}_\tau \sum_{n \in \mathcal{N}} \Phi_0^n / \sum_{n \in \mathcal{N}} w_\tau^n, \quad \tau = \min_{k \geq t, k \in \mathcal{T}^*} k. \quad (19)$$

Из (14) определяются проценты:

$$R_t = (\tilde{p}_t / p_t) / (\tilde{p}_{t+1} / p_{t+1}). \quad (20)$$

Можно отметить, что при $t \in \mathcal{T} \setminus \mathcal{T}^*$ в силу (15) верно $R_t = 1$ – потребители коллективно не испытывают потребности в деньгах, поскольку в сумме ограничения (2) выполнены строго. При $t \in \mathcal{T}^*$ $R_t > 1$.

Кредиты L_t^n и деньги Φ_t^n определяются из ограничений (1), (2) в зависимости от принадлежности момента времени t множеству $\mathcal{T}^*$:

$$L_t^n = \begin{cases} \Phi_{t-1}^n - p_t \tilde{c}_t^n + R_{t-1} L_{t-1}^n, & t \in \mathcal{T}^*, \\ \text{любое, удовлетворяющее } L_t^n \geq -\Phi_{t-1}^n + p_t \tilde{c}_t^n + R_{t-1} L_{t-1}^n, & t \in \mathcal{T} \setminus \mathcal{T}^*. \end{cases} \quad (21)$$

$$\Phi_t^n = \begin{cases} p_t w_t^n, & t \in \mathcal{T}^*, \\ \Phi_{t-1}^n + p_t (w_t^n - \tilde{c}_t^n) + L_t^n - R_{t-1} L_{t-1}^n, & t \in \mathcal{T} \setminus \mathcal{T}^*. \end{cases} \quad (22)$$

В формуле (21) для случая $t \in \mathcal{T} \setminus \mathcal{T}^*$ кредиты L_t^n , удовлетворяющие указанному условию $L_t^n \geq -\Phi_{t-1}^n + p_t \tilde{c}_t^n + R_{t-1} L_{t-1}^n$ и балансу (6), всегда найдутся, поскольку в сумме условия дают $0 \geq -\sum_{n \in \mathcal{N}} \Phi_{t-1}^n + p_t \sum_{n \in \mathcal{N}} \tilde{c}_t^n = -\sum_{n \in \mathcal{N}} \Phi_0^n + p_t \sum_{n \in \mathcal{N}} w_t^n$, что верно по построению множества $\mathcal{T}^*$.

Формулы (21), (22) рекуррентные, они определяют значения кредитов и денег в момент времени t через значения этих переменных в момент $t+1$.

Формулы (19)-(22) вместе с правилом (17) определения множества $\mathcal{T}^*$ дают решение задачи равновесия через решение задачи благосостояния.

3. Достаточные условия оптимальности для задачи потребителя и теорема теории благосостояния

Остаются неразрешенными следующие два вопроса. Во-первых, почему планы $\langle c_t^n, \Phi_t^n, L_t^n |_{t \in T} \rangle$, найденные по формулам (10), (21), (22), будут оптимальными для задачи потребителя (1)-(4) при процентах и ценах, определённых формулами (19), (20). И, наконец, какое именно решение $\langle \tilde{c}_t^n, \tilde{p}_t |_{t \in T} \rangle$ задачи благосостояния (т.е. какие коэффициенты a^n в (9)) нужно взять в качестве объекта дезагрегирования. Искомые коэффициенты a_n , $n \in \mathcal{N}$ должны быть некоторым образом связаны с начальными значениями Φ_0^n , L_0^n , $n \in \mathcal{N}$ и предопределёнными рядами w_t^n , $n \in \mathcal{N}$, $t \in T$.

Первый вопрос вызывает трудность, поскольку задача поставлена на бесконечном интервале времени. Стандартных необходимых или достаточных условий существования решения для таких задач не существует. Заменить формулировку задачи на бесконечности задачей на конечном интервале нельзя, иначе потеряется суть ограничения ликвидности (2): для существования решения будет достаточно, чтобы это ограничение выполнялось лишь в конечный период времени, и в итоге это ограничение не будет определять уровень цен.

Итак, просуммируем выражения (1), умноженные на λ_t^n , и выражения (11) по $t = 1, \dots, T$. Так как согласно построению стратегии потребителя (10), (21), (22) формулы (1), (11) верны, то указанная сумма равна нулю:

$$\sum_{t=1}^T \lambda_t^n \left(\Phi_{t-1}^n - \Phi_t^n + p_t (w_t^n - c_t^n) + L_t^n - R_{t-1} L_{t-1}^n \right) + \sum_{t=1}^T \mu_t^n \left(\Phi_{t-1}^n - p_t c_t^n + L_t^n - R_{t-1} L_{t-1}^n \right) = 0.$$

Воспользуемся условиями первого порядка (13), (14), которые выполнены по построению, и полученным выражением $\tilde{p}_t / a^n = p_t \sum_{\tau=t+1}^{\infty} \mu_{\tau}^n$. Тогда последнее выражение преобразуется в конечном итоге в следующее:

$$\tilde{p}_1/p_1(\Phi_0^n - R_0 L_0^n) + \sum_{t=1}^T \tilde{p}_{t+1} w_t^n - \sum_{t=1}^T \tilde{p}_t \tilde{c}_t^n - (\Phi_T^n - R_T L_T^n) / \prod_{\tau=1}^T R_\tau = 0. \quad (23)$$

Будем предполагать, что коэффициенты a_n , $n \in \mathcal{N}$ определяются из условий

$$\tilde{p}_1/p_1(\Phi_0^n - R_0 L_0^n) + \sum_{t=1}^{\infty} \tilde{p}_{t+1} w_t^n - \sum_{t=1}^{\infty} \tilde{p}_t \tilde{c}_t^n = 0, \quad n \in \mathcal{N}, \quad (24)$$

где $\langle \tilde{c}_t^n, \tilde{p}_t |_{t \in T} \rangle$ как решения задачи благосостояния (9), (5) неявно зависят от a_n , $n \in \mathcal{N}$, а цена p_1 определяется по (19). Суммы в (24) конечны в силу (18).

Покажем тогда, что стратегия (10), (21), (22) оптимальна для задачи потребителя.

Действительно, пусть существует другая допустимая стратегия $\langle \hat{c}_t^n, \hat{\Phi}_t^n, \hat{L}_t^n |_{t \in T} \rangle$, при которой значение функционала (3) больше.

Выпуклость функции полезности означает, что:

$$\frac{\partial U^n(\tilde{c}_t^n)}{\partial c_t^n}(\hat{c}_t^n - \tilde{c}_t^n) + \frac{\partial U^n(\tilde{c}_t^n)}{\partial L_t^n}(\hat{L}_t^n - L_t^n) + \frac{\partial U^n(\tilde{c}_t^n)}{\partial \Phi_t^n}(\hat{\Phi}_t^n - \Phi_t^n) \geq U^n(\hat{c}_t^n) - U^n(\tilde{c}_t^n)$$

Тогда

$$\begin{aligned} 0 < \delta &= \lim_{T \rightarrow \infty} \sum_{t=1}^T (U^n(\hat{c}_t^n) - U^n(\tilde{c}_t^n)) e^{-\delta t} \leq \\ &\leq \lim_{T \rightarrow \infty} \sum_{t=1}^T \frac{\partial U^n(\tilde{c}_t^n)}{\partial c_t^n} e^{-\delta t} (\hat{c}_t^n - \tilde{c}_t^n) + \frac{\partial U^n(\tilde{c}_t^n)}{\partial L_t^n} e^{-\delta t} (\hat{L}_t^n - L_t^n) + \frac{\partial U^n(\tilde{c}_t^n)}{\partial \Phi_t^n} e^{-\delta t} (\hat{\Phi}_t^n - \Phi_t^n) \stackrel{(a)}{=} \\ &\stackrel{(a)}{=} \lim_{T \rightarrow \infty} \sum_{t=1}^T \left[(\hat{c}_t^n - \tilde{c}_t^n) p_t (\lambda_t^n + \mu_t^n) + (\hat{\Phi}_t^n - \Phi_t^n) (\lambda_t^n - \lambda_{t+1}^n - \mu_{t+1}^n) + \right. \\ &\quad \left. + (\hat{L}_t^n - L_t^n) (-\lambda_t^n - \mu_t^n + R_t (\lambda_{t+1}^n + \mu_{t+1}^n)) \right] \stackrel{(b)}{=} \end{aligned}$$

$$\begin{aligned}
& \stackrel{(b)}{=} \lim_{T \rightarrow \infty} \left\{ \sum_{t=1}^T \left[\lambda_t^n (L_t^n - R_t L_{t-1}^n - p_t (w_t^n - \tilde{c}_t^n) - \Phi_t^n + \Phi_{t-1}^n + \alpha^n Z_t) + \mu_t^n (L_t^n - R_t L_{t-1}^n + \Phi_{t-1}^n - p_t \tilde{c}_t^n) - \right. \right. \\
& \quad \left. \left. - \lambda_t^n (\hat{L}_t^n - R_t \hat{L}_{t-1}^n - p_t (w_t^n - \hat{c}_t^n) - \hat{\Phi}_t^n + \hat{\Phi}_{t-1}^n + \alpha^n Z_t) - \mu_t^n (\hat{L}_t^n - R_t \hat{L}_{t-1}^n + \hat{\Phi}_{t-1}^n - p_t \hat{c}_t^n) \right] \right\} \stackrel{(c)}{\leq} \\
& \quad - (\lambda_1^n + \mu_1^n) (\Phi_0^n - \hat{\Phi}_0^n) + (\lambda_1^n + \mu_1^n) R_0 (L_0^n - \hat{L}_0^n) + \\
& \quad + (\lambda_{T+1}^n + \mu_{T+1}^n) (\Phi_T^n - \hat{\Phi}_T^n) - (\lambda_{T+1}^n + \mu_{T+1}^n) R_T (L_T^n - \hat{L}_T^n) \\
& \stackrel{(c)}{\leq} \lim_{T \rightarrow \infty} \sum_{t=1}^T \left[- (L_T^n - \hat{L}_T^n) R_T (\lambda_{T+1}^n + \mu_{T+1}^n) + (\Phi_T^n - \hat{\Phi}_T^n) (\lambda_{T+1}^n + \mu_{T+1}^n) \right] = \\
& = \lim_{T \rightarrow \infty} \left[(\Phi_T^n - R_T L_T^n) / \prod_{\tau=1}^T R_\tau - (\hat{\Phi}_T^n - R_T \hat{L}_T^n) / \prod_{\tau=1}^T R_\tau \right] \stackrel{(d)}{\leq} 0
\end{aligned}$$

Равенство (а) в предыдущей выкладке обусловлено использованием условий первого порядка (13)-(14). Равенство (b) это тождественное преобразование, полученное группировкой членов, связанных со стратегиями $\langle \hat{c}_t^n, \hat{\Phi}_t^n, \hat{L}_t^n |_{t \in T} \rangle$, $\langle \tilde{c}_t^n, \Phi_t^n, L_t^n |_{t \in T} \rangle$, при λ_t^n, μ_t^n . При этом искусственно добавлены и вычтены некоторые члены, чтобы получить при λ_t^n, μ_t^n выражения (1), (2). Неравенство (с) обусловлено несколькими причинами. Первую строчку в фигурных скобках можно отбросить, поскольку в ней все члены нулевые в силу условия первого порядка (13)-(14). Во второй стоят не положительные члены, поскольку $\langle \hat{c}_t^n, \hat{\Phi}_t^n, \hat{L}_t^n |_{t \in T} \rangle$ допустимая стратегия. Третья строчка обнуляется, поскольку для всех стратегий начальные условия одинаковые: $\Phi_0^n = \hat{\Phi}_0^n$, $L_0^n = \hat{L}_0^n$. Неравенство (d) следует из неравенства (4) для допустимой стратегии $\langle \hat{c}_t^n, \hat{\Phi}_t^n, \hat{L}_t^n |_{t \in T} \rangle$ и равенства $\lim_{T \rightarrow \infty} (\Phi_T^n - R_T L_T^n) / \prod_{\tau=1}^T R_\tau = 0$, следующего из (23) и (24).

В результате получено противоречие. Таким образом, доказано следующее утверждение.

Утверждение. Пусть для некоторых начальных значений Φ_0^n , L_0^n , $n \in \mathcal{N}$ и запасов продукта w_t^n , $n \in \mathcal{N}$, $t \in \mathcal{T}$ существует набор взвешивающих коэффициентов a_n , $n \in \mathcal{N}$ для задачи благосостояния (9), (5) такой, что соответствующее решение $\langle \tilde{c}_t^n, \tilde{p}_t |_{t \in T} \rangle$ задачи благосостояния удовлетворяет

условиям (24), (19) и (18). Тогда переменные $\langle c_t^n, \Phi_t^n, L_t^n, p_t, L_t |_{t \in T} \rangle$, определённые по формулам (10), (19)-(22), являются равновесием модели чистого обмена с ограничением ликвидности (1)-(7).

Это утверждение является по сути обеими теоремами теории благосостояния: если под заданные начальные значения задачи равновесия, согласно системе (24), удаётся подобрать решение задачи благосостояния, то это Первая теорема теории благосостояния, а если наоборот, то Вторая.

Заключение

Мы рассмотрели модель чистого обмена, в которой потребители стеснены ограничением ликвидности. Благодаря вводу в модель ограничения ликвидности, удалось однозначно определить равновесные цены. Таким образом, равновесие модели, в отличие от классических моделей типа Эрроу-Дебре [1], перестало быть инвариантно относительно растяжения траектории цен, т.е. замены $\langle p_t |_{t \in T} \rangle$ на $\langle C p_t |_{t \in T} \rangle$, $C > 0$. Равновесные цены определяются постоянной денежной массой экономики $\sum_{n \in \mathcal{N}} \Phi_0^n > 0$ и объемами покупаемого на эти деньги продукта $\sum_{n \in \mathcal{N}} w_t^n$.

Однозначное определение цен позволяет рассчитывать, что и в стохастических моделях с помощью механизмов типа ограничений ликвидности можно избежать проблемы множественности равновесий, приводящей к существованию таких интересных и странных равновесий, как равновесия с дефолтами [4].

Однако не следует забывать, что в рассмотренной модели результат получен благодаря свойству эффективности равновесий. Если же для подобного результата и в стохастической модели потребуется эффективность равновесий, то может просто оказаться, что мы будем рассматривать по-прежнему множество равновесий, но реализующих одну точку – решение задачи благосостояния, тогда как в стохастических моделях зачастую

интересны именно неэффективные равновесия, соответствующие случаю неполных рынков.

Наконец, полученные условия для существования равновесия в форме возрастания последовательности (16) или существования предела (18) не является чем-то необычным. Если рассмотреть случай функции полезности из класса CRRA, т.е. $U^n(c_t^n) = (c_t^n)^{1-\eta} / (1-\eta)$, $\eta \in (0,1)$, а запасы взять растущими по экспоненте, $w_t^n = w_0^n e^{gt}$, то оба условия сведутся к условию на параметры $g(1-\eta) < \delta$, достаточно типичному для оптимизационных задач математической экономики.

Литература

1. Поспелов И. Г. Модели экономической динамики, основанные на равновесии прогнозов экономических агентов. – М.: ВЦ РАН, 2003. – 200 с.
2. Cass D. Sunspots and incomplete financial markets: the leading example. CARESS Working Paper №84-06R (1984).
3. Geanakoplos J., Mas-Colell A. Real Indeterminacy with Financial Assets // Journal of Economic Theory. 1989. Vol. 47. Pp. 22-38.
4. Андреев М. Ю., Поспелов И. Г. Стохастическая задача чистого обмена и актуально бесконечно малые цены // Экономика и математические методы. – 2008. – Т.44, №2. – С.68-82.
5. Radner R. Equilibrium under uncertainty, in: K.J. Arrow and M.D. Intriligator (Eds.), Handbook of Mathematical Economics. North-Holland. 1982. Vol. II. Pp. 923-1006.

Информация об авторах

Data on authors

Андреев Михаил Юрьевич, к.ф.-м.н., старший научный сотрудник ВЦ РАН, доцент НИУ ВШЭ г. Москва

Тел.(499) 783-33-28

Электронная почта: m.andreyev@inbox.ru

Andreev Mikhail, docent, senior researcher at CCAS, Associate Prof. on Math
Modeling in Economics NSU HSE, Moscow

Tel. (499) 783-33-28

E-mail: m.andreyev@inbox.ru

ЧИСЛЕННОЕ РЕШЕНИЕ ТРЕХМЕРНОЙ ДИНАМИЧЕСКОЙ ЗАДАЧИ ТЕОРИИ ВЯЗКОУПРУГОСТИ

NUMERICAL SIMULATION OF 3D VISCOUS-ELASTODYNAMICS PROBLEM

В работе представлены математическая модель и результаты тестирования параллельного программного обеспечения для вычислительных кластеров, предназначенного для моделирования распространения сейсмических волн в вязкоупругих средах. Двустадийный конечно-разностный метод дискретизации второго порядка точности по пространству и времени дает устойчивую процедуру вычисления синтетических сейсмограмм для поглощающих сред с сильно меняющимися параметрами. Предложенное программное обеспечение позволяет задавать граничные условия свободной поверхности для произвольной гладкой топографии. С целью уменьшения влияния отраженных от фиктивных границ волн были использованы поглощающие энергию граничные условия. В программе предусмотрена возможность локального измельчения вычислительной сетки в областях, где волны высокой частоты представляют особенный интерес. Для обоснования адекватности вычислительной модели был проведен счет на задачах, имеющих аналитическое решение.

This paper presents mathematical model and evaluation of parallel cluster software for the numerical simulation of seismic wave propagation in the viscous-elastic continuum. Two stages, second order accurate in space and time finite-difference discretization provide stable numerical procedure for generating synthetic seismograms for wave energy absorbing, rapidly changing rock. Proposed software

allows the free surface boundary condition to be imposed on a curved topography. To diminish the influence of artificial boundaries energy absorbing boundary conditions were implemented. Local mesh refinement can be used to make the computational mesh finer in the area of interest, where more resolution is needed for short length wave propagation. Some numerical experiments with the problems, having analytical solution, were conducted to show the validity of this computational model.

Ключевые слова: распространение сейсмических волн в вязкоупругих средах, параллельное программное обеспечение, численное решение сейсмических волн

Key words: seismic wave propagation in the viscous-elastic continuum, parallel cluster software, numerical simulation of seismic wave

Применение высокопроизводительных вычислений в задачах поиска и разведки, разработки, транспорта и переработки углеводородного сырья становится определяющим фактором для принятия обоснованных решений. Это, с одной стороны, обусловлено высоким уровнем неопределенности входных данных, а с другой большим объемом оперативной памяти, которая необходима для проведения численных экспериментов с достаточным уровнем достоверности. В первом случае для определения влияния неопределенности, например, геологической модели месторождения, обычно получаемой методами геостатистики, на результаты гидродинамического моделирования требуется выполнить очень большой объем вычислений для различных возможных реализаций модели. Во втором, как показывает практика, для расчета одной гидродинамической или геофизической модели требуется объем памяти, исчисляемый сотнями гигабайт. Некоторые исследователи, которых становится все больше, например, из компании Aramco указывают, что для получения результатов моделирования, приемлемых для принятия обоснованных решений в области стратегии разработки оперативная память должна составлять десятки терабайт. Это связано с тем, что число

вычислительных ячеек, необходимых для нужного уровня разрешающей способности компьютерной модели, исчисляется тераблоками.

В данной работе рассматривается численная реализация одной из большого числа задач, требующих применения суперкомпьютеров, а именно, создание параллельных алгоритмов для решения динамической задачи распространения волн в линейно-вязкоупругой среде. Естественно, что в сейсморазведке приоритет отдается обратным задачам. Однако, без хорошей вычислительной модели для прямой задачи, математическая модель которой является более общей по сравнению с положенными в основу методов решения обратной задачи, корректный сравнительный анализ альтернативных алгоритмов невозможен.

Прямая задача распространения сейсмических волн в поглощающей сплошной среде в настоящее время приобретает большое значение, особенно если ее вычислительная реализация допускает моделирование распространения волн в жидких и твердых фазах с произвольной топографией дневной поверхности. Это позволяет проводить оценку сейсмических рисков для нефтедобывающих платформ и других крупных морских сооружений, связанных с волновой нагрузкой из-за подводных землетрясений. Также можно моделировать экологические проблемы, порожденные как естественными, так и рукотворными землетрясениями: разрушение инфраструктуры, плотин, дамб, сооружений, трубопроводов и т. д.

С практической точки зрения в связи с тем, что 3D сейсморазведка стала исключительно популярным средством поиска, разведки и мониторинга разработки нефтегазовых месторождений, решение прямой задачи позволяет с учетом предварительных сведений о геологической модели коллектора определить оптимальное расположение сейсмоприемников и выбрать эффективную процедуру обработки сейсмических трасс. Это можно отнести и к 4D сейсмике – методу мониторинга разработки, который становится все более популярным. Это особенно важно для нашей страны, поскольку потенциальные запасы углеводородов находятся в шельфе Северного Ледовитого океана,

акватория которого в зоне, принадлежащей Российской Федерации, недостаточно исследована.

Предлагаемая модель включает механизмы диссипации энергии, что позволяет разрабатывать и оценивать алгоритмы оконтуривания месторождений углеводородного сырья, так как известно, что насыщенная пористая среда поглощает сейсмическую энергию в большей мере по сравнению с вмещающей породой.

Задача построения вычислительного алгоритма в данном случае распадается на серию подзадач: создание математической модели и выбор численного метода, задание граничных условий на фиктивных границах области решения, реализация граничных условий на дневной поверхности со сложной топографией, выбор источника или распределения источников для моделирования сейсмических возмущений, вызванных, например, смещением горных пород вдоль сбросов, параллельная реализация вычислительного алгоритма. В программе реализована процедура эффективного использования дисковой памяти, построенная на основе применения рейд-массивов, информация на которых хранится в базе данных в виде октодеревьев Эйлера. Как показала практика, такой подход обеспечивает скорость работы с дисковой памятью практически сравнимую с оперативной.

Общая постановка задачи

В последнее время наиболее популярной формой представления задач вычислительной эластодинамики является совместная система уравнений, объединяющая поля смещений и тензора напряжений. В последнем случае уравнения получаются дифференцированием по времени определяющих соотношений. Хотя, представление уравнений в компонентах вектора смещений и тензора деформаций кажется предпочтительным для моделей динамической вязкоупругости, в данной работе был использован классический подход, основанный на определении решения только для поля смещений. Это позволяет достаточно сильно уменьшить объем необходимой оперативной памяти.

Сплошная среда является линейно-упругой, если между напряжением и деформацией имеется простая линейная зависимость, например, изотропно-упругая среда описывается соотношением $\sigma_{ij} = \lambda \delta_{ij} \varepsilon_{kk} + 2\mu \varepsilon_{ij}$. Если в определяющих линейных соотношениях допустить наличие производных по времени, что можно интерпретировать как зависимость скорости деформации от интенсивности или истории нагружения, то в результате получим линейную задачу вязкоупругости, определяющая модель которой в общем случае имеет вид:

$$\sum_{i=0}^N a_i \frac{\partial^i}{\partial t^i} \varepsilon = \sum_{i=0}^M a_i \frac{\partial^i}{\partial t^i} \sigma. \quad (1)$$

На практике наиболее часто используются следующие модели:

$$\frac{\partial \varepsilon_{ij}}{\partial t} = \frac{1}{2\mu} \frac{\partial \sigma_{ij}}{\partial t} + \frac{\sigma_{ij}}{2\eta} \text{ тело Максвелла,} \quad (2)$$

$$\sigma_{ij} = 2\eta \frac{\partial \varepsilon_{ij}}{\partial t} + 2\mu \varepsilon_{ij} \text{ тело Кельвина – Фогта,} \quad (3)$$

и

$$\eta \frac{\partial \sigma_{ij}}{\partial t} + (E_1 + E_2) \sigma_{ij} = E_2 \left(\frac{\partial \varepsilon_{ij}}{\partial t} + E_1 \varepsilon_{ij} \right) \text{ стандартное линейное твердое тело.} \quad (4)$$

Это обусловлено тем, что использование большого числа параметров сильно усложняет решение обратной задачи, а также делает практически невозможной физическую интерпретацию, а, следовательно, и разработку методов их экспериментального определения.

Подставляя определяющие соотношения, связывающие напряжение, деформацию и вектор смещений, в уравнение сохранения импульса получим математическую модель распространения волн в вязкоупругой среде:

$$\rho v_{tt} + b v_t = \partial_x (\mu(v_x + u_y)) + \partial_y ((2\mu + \lambda)v_y + \lambda(u_x + w_x)) + \partial_z (\mu(v_z + w_y)) + \rho f_y \quad (5)$$

$$\rho u_{tt} + b u_t = \partial_x ((2\mu + \lambda)u_x + \lambda(v_y + w_x)) + \partial_y (\mu(v_x + u_y)) + \partial_z (\mu(u_z + w_x)) + \rho f_x \quad (6)$$

$$\rho w_{tt} + b w_t = \partial_x (\mu(u_z + w_x)) + \partial_y (\mu(v_z + w_y)) + \partial_z ((2\mu + \lambda)w_y + \lambda(u_x + v_y)) + \rho f_z, \quad (7)$$

где (u, v, w) – вектор смещений в точке $x = (x, y, z)$, $\rho(x)$ плотность породы, а $b(x)$ коэффициент релаксации тела Максвелла. В случае моделирования линейно упругой среды или тела Максвелла λ, μ являются параметрами Ламе, а, в случае тела Кельвина-Фохта это уже дифференциальные операторы вида: $\lambda = \lambda_1 + \lambda_2 \partial_t$ и $\mu = \mu_1 + \mu_2 \partial_t$. Как и раньше λ_1, μ_1 можно рассматривать как параметры Ламе, а λ_2, μ_2 определяют вязкость среды. Приведенная система уравнений может быть использована и для моделирования стандартного линейного твердого тела, являющегося объединением моделей Кельвина-Фохта и Максвелла. Связь между E_1, E_2, μ, η и обобщенной системой параметров λ, μ, b очевидна. Здесь использованы две системы обозначений, так как в определяющих соотношениях необходимо отразить физический смысл параметров. Например, E_1, E_2 это дополнительные модули упругости, а η задает вязкость среды. В то время как λ, μ сохраняют связь с сейсмикой. Поскольку скорости продольной и поперечной упругих волн соответственно равны $c_p = \sqrt{(2\mu_1 + \lambda_1)/\rho}$ и $c_s = \sqrt{\mu_1/\rho}$. Это позволяет достаточно просто интерпретировать результаты сравнением с упругой моделью, а также провести ее верификацию на основе задач, для которых известны аналитические решения, например, задача Лэмба.

В соотношениях (5)-(7) $f(x, t)$ представляет источник, который выбирается в зависимости от решаемой задачи. Для моделирования землетрясений она одна, а для решения прямых задач 3D сейсморазведки и 4D сеймики – другая. Система (5)-(7) должна быть определена или на геоиде или в полупространстве $\Omega' = \{\bar{x} \mid -\infty \leq x \leq \infty, -\infty \leq y \leq \infty, c_1(x, y) \leq z \leq \infty\}$, где $c_1(x, y)$ – функция, задающая топографию дневной поверхности, на которой должны быть заданы условия свободной границы. Если планируется применить систему для моделирования распространения волн, вызванных землетрясением по поверхности и в толще Земли, то необходимо дополнить программу модулем, преобразующим географические координаты долготу и широту (λ, φ)

в декартовы (x, y) , используя различные картографические проекции. Задача определена для полупространства, однако, в силу конечности объема оперативной памяти необходимо ограничить до разумных размеров область вычислений $\Omega = \{(x, y, z) | 0 \leq x \leq a, 0 \leq y \leq b, c_1(x, y) \leq z \leq c\}$. Тогда на фиктивных границах необходимо задать условия, исходя из того, что отраженные от них волны не сильно испортят волновое поле. Существует несколько альтернативных подходов к их построению. Некоторые из них будут рассмотрены ниже. В данной работе были использованы поглощающие граничные условия, полученные исходя из принципа минимальности отраженной энергии. Например, для границы $x=0$ они близки по форме к представленным в работе [2], хотя и полученным, исходя из другого принципа:

$$u_t = c_p u_x + \lambda_1 \rho (v_y + w_z) / c_p, \quad v_t = c_s v_x + c_s u_y, \quad w_t = c_s w_x + c_s u_z, \quad (8)$$

На дневной поверхности заданы условия свободной границы, учитывающие ее топографию:

$$\mu u_z + \mu w_x = 0, \quad \mu v_z + \mu w_y = 0, \quad (2\mu + \lambda)w_z + \lambda u_x + \lambda y_x$$

$$\text{для } z = c_1(x, y), \quad t > 0. \quad (9)$$

Наконец, необходимо определить начальные условия:

$$u(x, 0) = u_0(x), \quad u_t(x, 0) = u_1(x), \dots, \quad w(x, 0) = w_0(x), \quad w_t(x, 0) = w_1(x). \quad (10)$$

Таким образом, распространение волн в вязкоупругих средах, определяющие соотношения для которых содержат производные не выше первого порядка, задается математической моделью (5)-(10).

Поглощающие граничные условия

Одной из первых работ, посвященных построению поглощающих граничных условий, была статья Клейтона и Энkvиста [2], в которой приводилась целая иерархия граничных условий, определяемая степенью аппроксимации псевдо-дифференциального оператора. Порядок аппроксимации задает интервал угла падения волн на фиктивную границу, в котором будет работать механизм поглощения. Все граничные условия приведенной иерархии являются неотражающими для нормально падающих волн. Хигдон в работе [3] использовал другой подход, при котором граничные условия получаются применением скалярных неотражающих условий для каждой компоненты поля смещений. Это также порождает определенную иерархию поглощающих граничных условий. Для скалярных волновых полей оба подхода дают один и тот же результат. Данные граничные условия интенсивно использовались в вычислительных экспериментах, которые показали их неустойчивость, начиная с аппроксимации третьего порядка. В последнее время большую популярность приобрел метод, впервые примененный для решения уравнений Максвелла и называемый «идеально согласованный слой» [4]. Хотя поглощающие условия, полученные с использованием этого подхода, работают лучше граничных условий Энkvиста и Хигдона низкого порядка, но они существенно сложнее в программной реализации, а также требуют точной настройки мощности поглощающего слоя и степени поглощения в нем. Этот подход напоминает аналоговые модели, которые применяли в докомпьютерную эру.

Наш подход в построении поглощающих граничных условий состоит в задании диссипации внутренней энергии на фиктивных границах. В работе [1] для построения поглощающих граничных условий для вязкопластичных сред были использованы основные идеи [2], однако, это потребовало достаточно больших усилий. Если в основу метода построения положить условие невозрастания внутренней энергии, то в силу диссипативных свойств вязкой составляющей можно показать, что в энергетической норме решение упругой

задачи мажорирует вязкоупругую. Это дает возможность непосредственно использовать поглощающие условия для упругой задачи в вычислительной схеме, что существенно сокращает формальную процедуру их вывода. В общем случае уравнения эластодинамики имеют вид:

$$\rho u_{tt} = \operatorname{div}(\sigma(u)) + f(x, t), \quad x \in \Omega, \quad t \geq 0, \quad (11)$$

где $u = (u, v, w)$, $x = (x, y, z)$. Тензор напряжения в удобной для программирования матричной нотации можно представить в виде:

$$\sigma(u) = \lambda \operatorname{div}(u)I + 2\mu D(u). \quad (12)$$

Здесь I единичная матрица, а оператор D имеет вид:

$$D(u) = \begin{pmatrix} 2u_x & 2u_y + v_x & u_z + w_x \\ 2u_y + v_x & 2v_y & v_z + w_y \\ u_z + w_x & v_z + w_y & 2w_z \end{pmatrix} \quad (13)$$

Если положить $f(x, t) = 0$, что соответствует случаю, когда источник полностью отработал в момент достижения волнами фиктивной границы, то изменение общей упругой энергии среды в области Ω будет определяться только ее диссипацией через фиктивную границу, поскольку для свободной границы и внутренней области выполняется закон сохранения энергии. В результате получим, что для (11) должно выполняться следующее условие сохранения энергии:

$$\frac{1}{2} \frac{d}{dt} (\|\sqrt{\rho} u_t\|^2 + \int_{\Omega} [\lambda (\operatorname{div}(u))^2 + 2\mu (D(u) : D(u))] dx) = \int_{\partial\Omega} u_t^T \sigma(u) n dS, \quad (14)$$

где n – единичный вектор внешней нормали, dS – элемент поверхности границы $\partial\Omega$, а символ $:$ означает операцию сжатия тензора –

$$(A : B) = \sum_{i=1}^3 \sum_{j=1}^3 a_{ij} b_{ij}.$$

Из (14) следует, что любое граничное условие такое, что

$$u_t^T \sigma(u) n \leq 0, \quad x \in \partial\Omega, \quad t \geq 0 \quad (15)$$

приводит к решению, для которого внутренняя энергия не возрастает, что в

частности обеспечивает корректность постановки задачи. Поскольку задача (11), (15) энергетически мажорирует вязкоупругую (5)-(10), то условие (15) можно использовать для доказательства ее корректности. К сохраняющим энергию граничными условиям относятся: условие Дирихле $u = 0$ и условие свободной границы $\sigma(u)n = 0$.

Из неравенства (15), в частности, следует, что любое условие вида

$$u_t = -A\sigma(u)n, \quad x \in \partial\tilde{\Omega}, \quad t \geq 0, \quad (16)$$

где A – положительно определенная матрица, дает диссипативные граничные условия на $\partial\tilde{\Omega}$. Например, условия Клейтона-Энквиста получаются, если положить:

$$A = \frac{1}{\sqrt{\rho}} I \left(\frac{1}{\sqrt{\lambda + 2\mu}}, \frac{1}{\sqrt{\mu}}, \frac{1}{\sqrt{\mu}} \right)^T.$$

Именно таким образом были получены поглощающие условия (8), которые показали свою вычислительную эффективность на ряде примеров

Вычислительная модель

В качестве основы для построения вычислительной модели можно использовать различные подходы, основанные на методах конечных элементов, конечных объемов, конечных разностей и некоторые другие. В данной работе использован метод конечных разностей, который был выбран из-за простоты его реализации на вычислительных кластерах. Проиллюстрируем построение разностной схемы на простом примере, в котором дневная поверхность полагается плоскостью.

Зададим равномерную по пространству вычислительную сетку: $x_i = (i-1)h$, $y_j = (j-1)h$, $z_k = (k-1)h$, где $h > 0$ – фиксированный шаг по пространству, $x_{N_x} = a$, $y_{N_y} = b$, $z_{N_z} = c$, а N_x, N_y, N_z размеры сеточной области. Аналогично задав шаг по времени $\Delta t > 0$ можно определить сетку $t_n = n\Delta t$. Введем сеточную функцию $u_{i,j,k}^n = u(x_i, y_j, z_k, t_n)$ и определим пространственные

разностные операторы, которые будут использованы для аппроксимации производных:

$$D_{\beta}^{\alpha}, \text{ где } \alpha \in \{x, y, z\}, \beta \in \{l, c, r\}, \quad (17)$$

где β означает тип разности: левая, центральная или правая. Например, для

$$\alpha = x \quad D_r^x u_{i,j,k} = \frac{u_{i+1,j,k} - u_{i,j,k}}{h},$$

$$D_l^x u_{i,j,k} = D_r^x u_{i-1,j,k} \quad \text{и} \quad D_c^x u_{i,j,k} = \frac{(D_r^x u_{i,j,k} + D_l^x u_{i,j,k})}{2}.$$

Кроме того, для единообразия представления разностной схемы в окрестности границы введем оператор

$$\hat{D}_0^{\alpha} = \text{sign}(\alpha_m - 1) D_c^{\alpha} + \overline{\text{sign}(\alpha_m - 1)} D_l^{\alpha} \quad (18)$$

и оператор осреднения (сдвига), который необходим в разностной схеме для уравнений с переменными коэффициентами:

$$E^x(\gamma_{i,j,k}) = \frac{\gamma_{i+1,j,k} + \gamma_{i,j,k}}{2}, \quad (19)$$

где $\gamma_{i,j,k} = \gamma(x_i, y_j, z_k)$. Можно использовать и другие операторы пространственного осреднения, но приведенный выше наиболее простой и вероятно самый распространенный. Осуществим частичную дискретизацию (5)-(7) и в результате получим систему дифференциально-алгебраических уравнений

$$\begin{aligned} \rho \frac{d^2 u}{dt^2} + b \frac{du}{dt} &= D_l^x (E^x(2\mu + \lambda) D_r^x u) + D_l^y (E^y(\mu) D_r^y u) + \\ &D_0^x (\lambda D_0^y v + \lambda \hat{D}_0^z w) + D_0^y (\mu D_0^x v) + \hat{D}_0^z (\mu D_0^x v) \triangleq L^u(u, v, w), \\ \rho \frac{d^2 v}{dt^2} + b \frac{dv}{dt} &= L^v(u, v, w), \quad \rho \frac{d^2 w}{dt^2} + b \frac{dw}{dt} = L^w(u, v, w). \end{aligned} \quad (20)$$

Здесь $L^{\alpha}, \alpha \in \{u, v, w\}$, в некотором смысле, представляет собой конечно-разностную аппроксимацию пространственной части системы дифференциальных уравнений. Это верно для динамической теории упругости и тела Максвелла. Однако уже для модели Кельвина-Фогта правая часть

содержит производные по времени. В этом случае параметры λ и μ являются дифференциальными операторами, содержащими производные по времени. Данную проблему можно решить двумя способами: перенести производные по времени с учетом связанных с ними разностных операторов в правую часть, что изменит b или заморозить производные по времени в правой части, а при дискретизации по времени в трехслойной схеме аппроксимировать их на двух предыдущих слоях. В первом случае аппроксимация по времени сохранит второй порядок, а во втором она будет иметь первый порядок. Однако это не играет существенной роли, поскольку, как показал Рихтмайер [5], качество разностной схемы в основном определяется порядком аппроксимации старшей производной.

Аналогичным образом строится аппроксимация граничных условий, например, условия на дневной поверхности имеют вид:

$$\begin{aligned}
& \frac{1}{2}(E^z(\mu_{i,j,1})D_r^z u_{i,j,1} + E^z(\mu_{i,j,0})D_r^z u_{i,j,0}) + \mu_{i,j,1}D_c^z w_{i,j,1} = 0 \\
& \frac{1}{2}(E^z(\mu_{i,j,1})D_r^z v_{i,j,1} + E^z(\mu_{i,j,0})D_r^z v_{i,j,0}) + \mu_{i,j,1}D_c^z w_{i,j,1} = 0 \\
& \frac{1}{2}(E^z(2\mu_{i,j,1} + \lambda_{i,j,1})D_r^z w_{i,j,1} + E^z(2\mu_{i,j,0} + \lambda_{i,j,0})D_r^z w_{i,j,0}) + \\
& \lambda_{i,j,1}(D_c^x u_{i,j,1} + D_c^x v_{i,j,1}) = 0
\end{aligned} \tag{21}$$

Можно доказать справедливость следующей теоремы:

Теорема 1 Полудискретная схема (20) с граничными условиями (21) и однородными условиями Дирихле на фиктивных границах аппроксимирует задачу (5)-(7) со вторым порядком точности.

Если провести дискретизацию (20) по времени, то получим полную систему конечно-разностных уравнений, которая и была реализована:

$$\begin{aligned}
& \rho\left(\frac{u^{n+1} - 2u^n + u^{n-1}}{(\Delta t)^2}\right) + b\left(\frac{u^{n+1} - u^{n-1}}{2\Delta t}\right) = L^{(u)}(u^n, v^n, w^n) \\
& \rho\left(\frac{v^{n+1} - 2v^n + v^{n-1}}{(\Delta t)^2}\right) + b\left(\frac{v^{n+1} - v^{n-1}}{2\Delta t}\right) = L^{(v)}(u^n, v^n, w^n)
\end{aligned} \tag{22}$$

$$\rho \left(\frac{w^{n+1} - 2w^n + w^{n-1}}{(\Delta t)^2} \right) + b \left(\frac{w^{n+1} - w^{n-1}}{2\Delta t} \right) = L^{(w)}(u^n, v^n, w^n).$$

Можно показать, что для этой явной разностной схемы в первом приближении должно выполняться условие устойчивости справедливое для упругого случая:

$$\Delta t < \frac{h}{\sqrt{(3\mu + \lambda) / \rho}} = \frac{h}{\sqrt{c_p^2 + c_s^2}} \quad (23)$$

Последним элементом, который необходимо определить для завершения принципиальной вычислительной схемы является задание функции источника. Она выбирается в зависимости от физической постановки задачи. Если необходимо моделировать распространение волн в задачах 3D сейсмологии, то используют импульсные точечные источники, если же необходимо моделировать распространение волн, вызванных землетрясением, порожденным смещением горных пород вдоль разлома (сброса), обычно применяют распределенный по некоторой поверхности источник с монотонно возрастающей амплитудой. Точечный источник можно представить в виде:

$$f(x, t) = g(t, t_0, \omega) f_0 (f_x, f_y, f_z)^T \delta(x - x_0),$$

где $g_c(t, t_0, \omega) = \int_{-\infty}^t g(\tau, t_0, \omega) d\tau$ зависимость амплитуды от времени для источника, расположенного в точке $x_0 = (x_0, y_0, z_0)$, $(f_x, f_y, f_z)^T$ – вектор, состоящий из компонент силы, действующей в этой точке, а f_0 – ее амплитуда. Такое представление дает возможность использовать не только источники типа центра расширения, но и момента сил. Все временные функции должны быть такими, что $\lim_{t \rightarrow -\infty} g(t, t_0, \omega) = 0$ и $\lim_{t \rightarrow \infty} g(t, t_0, \omega) = g_\infty$. Значение $g_\infty \neq 0$ соответствует остаточным смещениям, порожденным моделируемым сейсмическим событием.

Наиболее часто применяют источники с временными функциями: Гаусса, Рикерта, Рэмпа, Лю, треугольной и многими другими. Кроме того, в набор входят и их интегралы, т. е. $g_{new}(t, t_0, \omega) = \alpha(\omega) \int_{-\infty}^t g(\tau, t_0, \omega) d\tau$. Например, для

источника Гаусса $g(t, t_0, \omega) = \omega \sqrt{2} e^{-1/2 \omega^2 (t-t_0)^2} / (2\sqrt{\pi})$, причем интеграл от этой функции используется для моделирования землетрясений, дающих остаточные смещения пород. Для импульса Рикерта эта функция имеет вид: $g(t, t_0, \omega) = (2\pi^2 \omega^2 (t-t_0)^2 - 1) e^{-\pi^2 \omega^2 (t-t_0)^2}$.

Параллельная реализация вычислительной модели

Приведенная выше вычислительная модель была реализована на вычислительном кластере с использованием библиотеки MPI. В программную реализацию дополнительно включена возможность численного моделирования распространения волн в среде, имеющей произвольную топографию дневной поверхности, что позволяет оценить ее влияние на результаты сейсморазведки или сейсмоопасность определенного региона. Это сделано с помощью специальной адаптации вычислительной сетки и разностной схемы для учета искривлений дневной поверхности. Кроме того, программа допускает в процессе счета использование больших объемов дисковой памяти как альтернативу оперативной, практически не оказывая влияния на скорость вычислений. Это сделано на основе применения Эйлеровых октодеревьев, которые используются для хранения параметров, распределенных на больших регулярных вычислительных сетках. Их можно интерпретировать как базы данных специального вида, выборка из которых осуществляется очень быстро, особенно, если они работают на хорошо настроенных рейд-массивах. И, наконец, в программе имеется возможность применения регулярных адаптивных сеток, что согласуется с хранением данных на октодеревьях.

Программа была протестирована на достаточном числе примеров. В качестве иллюстраций приведены результаты решения двух задач. Первая, это задача Лэмба – распространение волн в полупространстве, вызванных точечным источником, расположенным на дневной поверхности. Размер области вычислений $x=y=10000$ м., $z=5000$ м., шаг по пространству $h=5$ м. Это гарантирует точное представление волн длиной больше 50 м. Объем

необходимой оперативной памяти приблизительно составляет 140 Гб. Время счета 30 мин. на 64 ядрах (16 узлах) кластера. Рисунок 1 показывает волновые поля на дневной поверхности в указанные моменты времени. Относительная ошибка решения не превышает одного процента до начала отражения от фиктивной границы. На рисунке 1 видно, что амплитуда паразитных волн, по крайней мере, на порядок меньше падающих. Структура данного волнового поля лучше прослеживается на рисунке 2, который отображает его в градациях серого, причем точка наблюдения расположена над дневной поверхностью.

Вторая задача – это моделирование распространения волнового поля в твердой породе ($x=y=45000$ м, $z=14000$ м), покрытой сверху километровой толщей воды, шаг по пространству – $h=15$. Возмущения порождаются источником, моделирующим смещение пород вдоль сброса, падающего под углом 90° в центре области. Разлом простирается по глубине до конца области вычислений. Для решения задачи необходима оперативная память порядка 300 Гб. Время счета составляет 90 мин. на 128 ядрах. На рисунке 3 показано волновое поле на дне, а на рисунке 4 волны на поверхности воды. Необходимо отметить, что на рисунке 3 возникают сильные паразитные возмущения у фиктивных границ. Это может быть следствием или некоторой вычислительной неустойчивости в области резкого скачка поля скоростей на разделе фаз, или недостаточным уровнем поглощения волновой энергии на этих границах, что в любом случае требует переопределения матрицы A в (16).

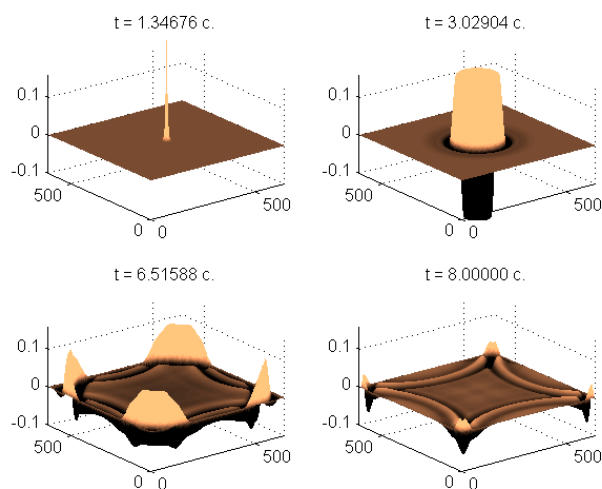


Рис. 1. Задача Лэмба

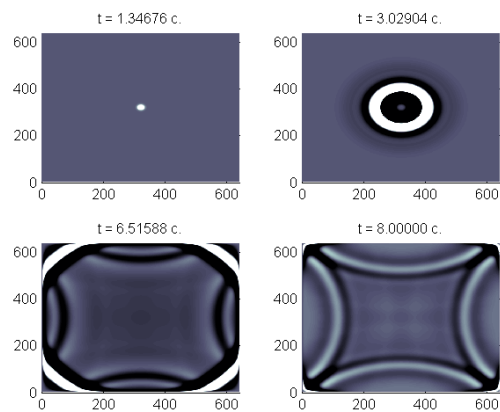


Рис. 2. Задача Лэмба, вид сверху

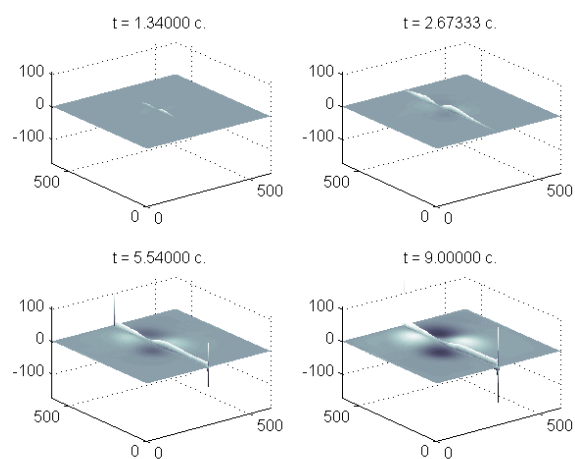


Рис. 3. Волны на поверхности раздела воды и твердой фазы

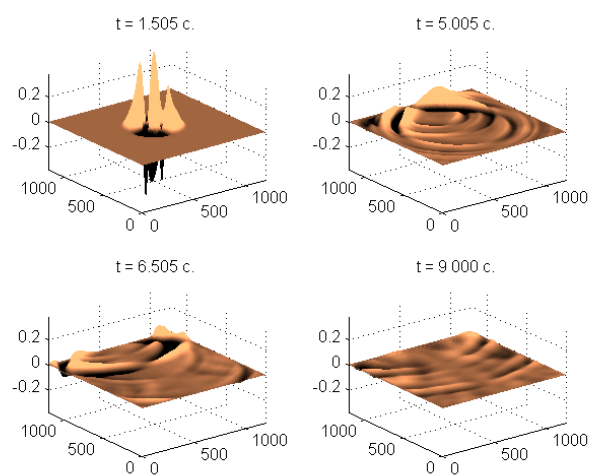


Рис. 4. Волны на поверхности воды

Заключение

Таким образом, разработанное программное обеспечение для моделирования процесса распространения сейсмических волн в вязкоупругих средах доказало свою работоспособность и дает возможность решать широкий класс задач: от анализа и настройки алгоритмов 3D-сейсморазведки и 4D-сейсмомониторинга разработки месторождений углеводородного сырья до оценки потенциальных разрушений, вызванных подземными или подводными землетрясениями. В качестве целей дальнейшего развития программного комплекса можно указать на применение метода конечных элементов или метода конечных объемов при использовании обобщенной процедуры метода прямых. Это позволит моделировать более широкие классы волновых процессов, а также использовать методы кодогенерации для ускорения разработки программного обеспечения. Однако в этом случае вместо октодеревьев необходимо будет использовать β -деревья, которые ориентированы на хранение и быструю выборку данных, заданных на нерегулярных сетках.

Литература

1. Арсеньев-Образцов С. С., Жукова Т. М. Поглощающие граничные условия для численного решения задач теории вязкоупругости. ЖВМ и МФ, 1987. – Том 27. – № 2. – Стр. 301–306.
2. R. Clayton, B. Engquist. Absorbing boundary conditions for acoustic and elastic wave equations. Bull. Seismo. Soc. Amer., 67 (1977) pp. 1529-1540.
3. R. L. Higdon, Absorbing boundary conditions for acoustic and elastic waves in stratified media. J. Comput. Phys., 101 (1992), 386-418.
4. J. P. Berenger, A perfectly matched layer for the absorption of electromagnetic waves, J. Comput. Phys., 114 (1994), 185-200.
5. Рихтмайер Р., Мортон К. Разностные методы решения краевых задач. М.: Мир, 1972.

6. Н. Lamb. On the propagation of tremors over the surface of an elastic solid. Phil. Trans. Roy. Soc. London, Ser. A, 1904.

Информация об авторах

Data on authors

Арсеньев-Образцов Сергей Сергеевич, к.ф.-м.н., доцент НИУ НиГ им.
И.М. Губкина, г. Москва

Тел. (499) 135-71-36

Электронная почта: arseniev@gubkin.ru

Arseniev-Obraztsov Sergey, PhD of Physics and Mathematics, Associate
Prof. in NSU of Oil & Gas Gubkin, Moscow

Tel. (499) 135-71-36

E-mail: arseniev@gubkin.ru

МОДЕЛЬ КОЛЛЕКТИВНОГО ПОВЕДЕНИЯ ЖИВОТНЫХ*

MODEL OF THE COLLECTIVE BEHAVIOR ANIMALS

В работе рассматривается пространственно распределенная эколого-физиологическая модель популяции, взаимодействующей с возобновимым ресурсом. Описываются энергетический баланс особи, процессы рождения и смертности. Перемещение особей по пространству определяется наличием ресурса, а также нелокальным взаимодействием между особями. Приводятся некоторые численные результаты.

The spatial distributed ecological – physiological model of the population consuming renew recourse is under consideration. Individual energy balance, the birth and dead processes are described. Spatial motion is depend on available recourses and non local interaction between individuals, some results is demonstrated.

Ключевые слова: модель популяции, нелокальное взаимодействие между особями, численное решение для пространственно распределенной эколого-физиологической модели.

Key words: model of the population, non local interaction between individuals, numerical solution of ecological-physiological model.

Одной из важнейших задач популяционной экологии является получение (обоснование) популяционных уравнений на основании характеристик

* Работа выполнена в рамках гранта РФФИ № 09-0700-398а.

отдельных особей. Ряд авторов (Охонин Н. Н., Горбань А. Н.) [1] пытались подойти к ее решению с позиции методов статистической физики, используя формализм цепочек Боголюбова. Существует целый ряд работ, так называемые индивидуально ориентированные модели, в которых авторы численно моделируют динамику популяции. Например, гар-модели лесной растительности [2]. В первой части данной работе мы, исходя из некоторых упрощенных эколого-физиологических представлений об энергетическом балансе особи с учетом возрастных особенностей, с помощью имитационного моделирования исследуем некоторых популяционных макро характеристики. Во-второй части работы мы вводим нелокальное взаимодействие между особями популяции, зависящее от возраста особей, с целью описать эффект образования стада.

Описание модели

Рассматривается целочисленная решетка $(i=1,...,L, j=1,...,F)$, на который равномерно произрастает ресурс (трава). Уравнение роста ресурса имеют вид:

$$m^{i,j}_{t+1} = m^{i,j}_t + K - \theta_t, \text{ если } m^{i,j}_t < H \text{ и } m^{i,j}_t = H, \text{ если } m^{i,j}_t \geq H,$$

где $m^{i,j}_t$ – текущее количество ресурса в точке (i, j) , K – скорость роста травы, H – максимально возможное количество ресурса, а θ_t – доля изъятия ресурсом особью, если она есть в точке (i, j) .

На этой решетке находится некоторое множество подвижных особей, которые потребляют ресурс, размножаются и могут перемещаться из узла в узел. Считается, что особь имеет возраст (τ) . Гибель особи определяется условием недостатка «энергии» (ресурса) $(n_\tau \leq 0)$, который необходим для поддержания затрат на основной метаболизм. Считается, что с возрастом, потребление ресурса уменьшается. $\theta = n_\tau * C_1(1 - \frac{\tau}{T})(1 - \frac{n_\tau}{N})$, где C_1, T, N – некоторые постоянные. По достижении определенного размера особь случайным образом делится (процесс рождения), при этом теряется часть «энергии». Перемещение особи определяется «длиной прыжка» (S) за один временной такт и при этом

теряется часть «энергии», имеющейся у особи. Особь характеризуется способностью «видеть» ресурс на расстоянии радиуса обзора (R). Направление движения особи определяется расположением ближайшего узла, в котором есть ресурс.

Уравнение баланса «энергии» особи имеет вид:

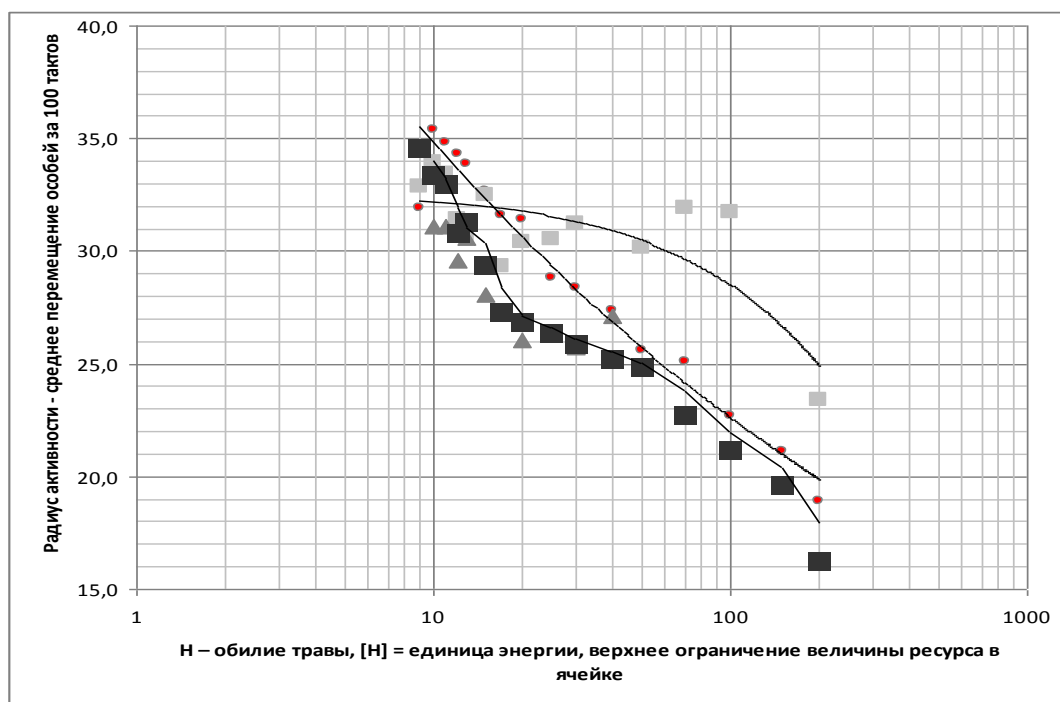
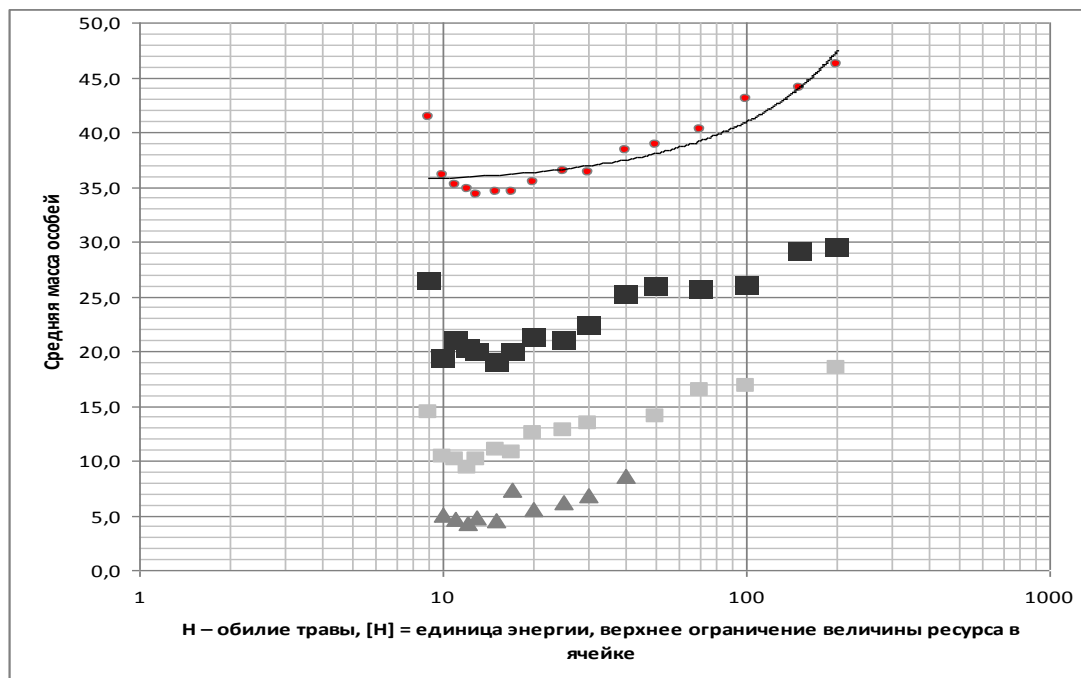
$$n_{t+1} = n_t * C_1(1 - \frac{\tau}{T})(1 - \frac{n_t}{N}) - C_2 n_t S - \alpha n_t^\beta - \lambda \frac{n_t}{2}$$

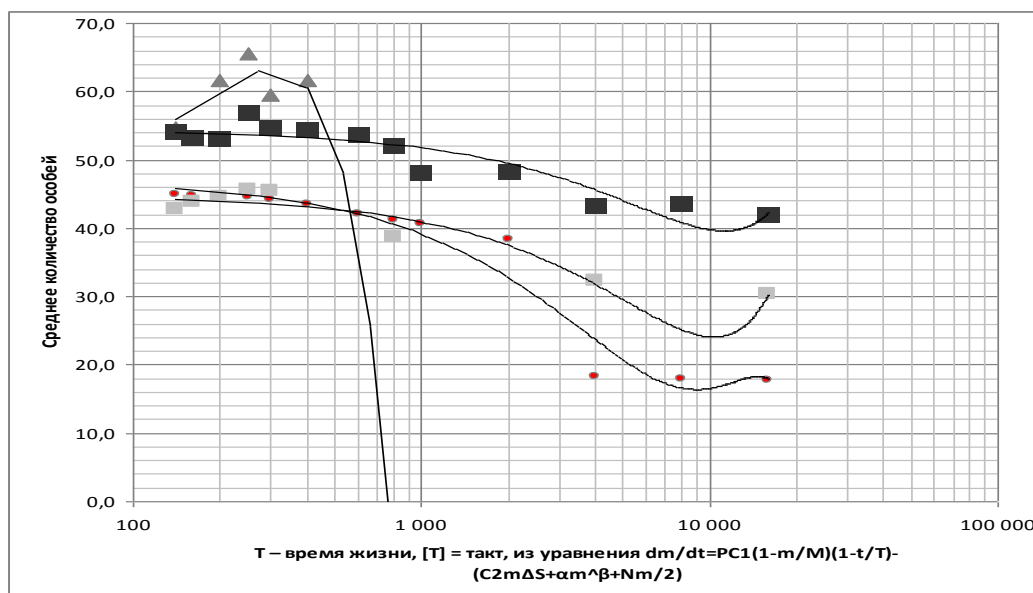
где $C_2 n_t S$ – затраты на перемещение; αn_t^β – основной метаболизм; $\lambda \frac{n_t}{2}$ – затраты на рождение ($\lambda = 1$ – происходит рождение в момент времени t , в противном случае $\lambda = 0$). В модели считается, что рождение происходит в среднем один раз за некоторый период времени.

Численные эксперименты с моделью проводились с целью описать как «взаимосвязаны» параметры модели при стационарных режимах. Для этого брался один из параметров модели, бралась некоторая начальная конфигурация особей на решетке, и проводился численный эксперимент. После переходного периода, устанавливался некоторый колебательный режим. И для этого режима определялись значение другого параметра, например, первым параметром могла выступать K – скорость роста травы, а другим – средний радиус индивидуальной активности.

Некоторые численные результаты

Ниже представлены некоторые результаты обработки имитационных экспериментов с моделью.





На графиках кружки соответствуют значениям $\alpha = 1.0$; $\beta = 0.00$, квадратики – соответствуют значениям $\alpha = 0.2$; $\beta = 0.79$ треугольники $\alpha = 0.2$; $\beta = 0.90$ – крестики $\alpha = 0.2$; $\beta = 0.50$ – где α, β параметры потерь энергии на основной метаболизм αn_t^β . По горизонтальной оси откладываются, соответствующий изменяющийся параметр модели, а по вертикальной среднее значение изучаемого параметра

Введение нелокального взаимодействия между особями

Если в вышеописанной модели перемещение определяется лишь наличием ресурса, то в дополненной модели вводится предположение о том, что между особями имеет место нелокальное взаимодействие (рис. 1). В рассматриваемом варианте модели взаимодействие между особями определяется выражением

$$f_{ij} = \begin{cases} A, r_i - r_j < r \\ B, r \leq r_i - r_j < R \\ 0, r_i - r_j \geq R \end{cases}$$

где A – значение «отталкивание» особей друг от друга на расстояниях меньше r , а B – «притягивание» на расстояниях между R и r .

Перемещение Δ i -ой особи определяется векторной суммой

$$\Delta = \vec{l}_i + \sum_j f_{ij} \frac{(r_i - r_j)}{|r_i - r_j|}$$

где первый член суммы $\vec{l}_i$ определяет «ресурсное» перемещение, а второй – популяционное взаимодействие между особями.



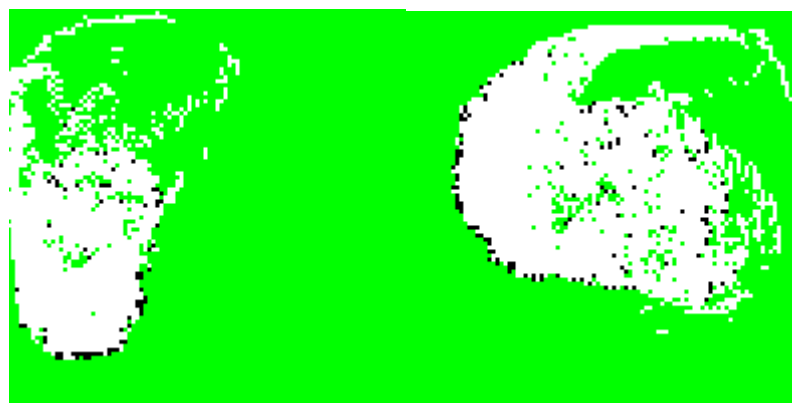
Рис. 1. Зависимость перемещения f_{ij} за такт от расстояния между особями

Такая модель позволяет получить при определенных параметрах коллективное поведение особей. Некоторые результаты моделирования приведены на рис. 2.



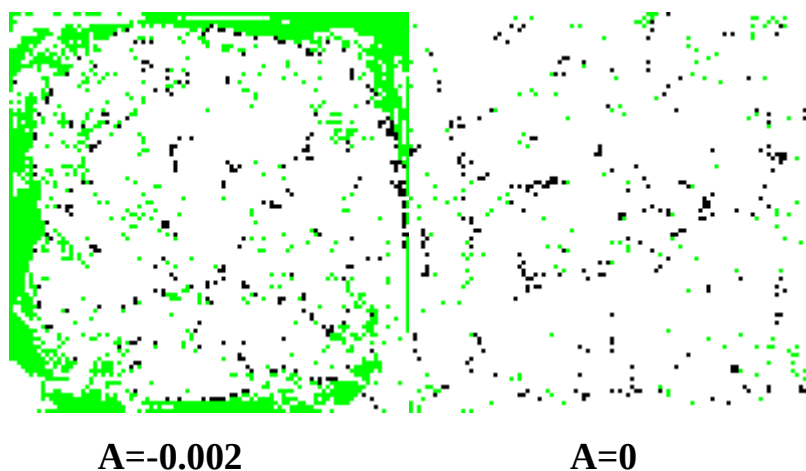
A=-0.2

A=-0.02



A=-0.01

A=-0.005



На рис. 2 представлены результаты расчета при разных значениях параметра A . Черные точки изображают особей, белая область – места, где нет ресурса.

Заключение

Промежуточные результаты, приведенные выше, показывают, с одной стороны, что некоторые зависимости вполне интуитивно ожидаемы (зависимость массы особи от обилия ресурса), с другой – «контринтуитивное поведение» (среднее число особей от обилия ресурса). Более детальный анализ работы модели требует участия биолога, который бы мог дать конкретные значения параметров. Введение в модель нелокального взаимодействия между особями требует тщательного анализа взаимоотношения между характерными временами демографических и миграционных процессов.

Литература

1. Горбань А. Н. Обход равновесия. Новосибирск. Наука, 1984.
2. A Shugart H. Y. A Theory of the Forest Dynamics. Springer, New York, 1984.
3. South. Extrapolating from individual movement behavior to population spacing patterns in a ranging mammal. Ecological Modelling, 117, 1999, p. 343-360.

4. R. Mac Nally. Modelling confinement experiments in community ecology: differential mobility among competitors. *Ecological Modelling*, 129, 2000, p. 65-85.
5. Шмидт-Ниельсон К. Размеры животных: Почему они так важны? Москва. – Мир, 1987.
6. Angela Gallegosa, Barbara Mazzagb, Alex Mogilner, Two Continuum Models for the Spreading of Myxobacteria Swarms // *Bulletin of Mathematical Biology* (2006), p. 837–861.
7. C. T. Lee, M. F. Hoopse, J. Diehl, W. Gilliland, G. Huxel, E. V. Leaver, K. Mccann, J. Umbanhowar, A. Moglner. Non-local Concepts and Models in Biology // *J. theor. Biol.* (2001), p. 210-219.

Информация об авторах

Data on authors

Белотелов Николай Вадимович, к.ф.-м.н., старший научный сотрудник
ВЦ РАН, доцент НИУ МФТИ, г. Москва

Тел. (499)-135-24-89

Электронная почта: belotel@mail.ru

Belotelov Nicolas, PhD of Physics & Mathematics, senior researcher at CCAS,
Associate Prof. on System Research in NSU MPTI, Moscow

Tel. (499)-135-24-89

E-mail: belotel@mail.ru

Коноваленко Иван Александрович, студент НИУ МФТИ, г. Долгопрудный

Тел. (495)-408-80-88

Электронная почта: alatkon@yandex.ru

Konovalenko Ivan, student of NSU MPTI, Dolgoprunny

Tel. (495)-408-80-88

E-mail: alatkon@yandex.ru

МОДЕЛИРОВАНИЕ МИГРАЦИИ С УЧЕТОМ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА*

MODELING MIGRATION TAKING INTO ACCOUNT EDUCATIONAL PROCESSE

В работе рассматривается модель миграции между несколькими странами. Страны описываются однотипными эколого-демографо-экономическими моделями. Обсуждается влияние образовательных процессов на процессы миграции и ассимиляции.

Model of the population migration between several countries is under consideration. Countries are described by the ecological-demographical-economical models. The influence of the educational process on migration and assimilation process is investigated.

Ключевые слова: модель миграции между несколькими странами, влияние образовательных процессов на процессы миграции и ассимиляции.

Key words: Model of the population migration between several countries, the influence of the educational process on migration and assimilation process.

Рассматриваемая в работе модель является блоком модели эколого-демографо-экономической модели (ЭДЭМ), разработанной ранее в ВЦ РАН коллективом авторов [1, 2]. Отличие от базовой модели заключается в дополнении демографического блока модели моделью миграции населения

* Работа выполнена при поддержке РФФИ, грант 10-07-00420-а.

между странами и детализации подмодели образовательного процесса. В модели рассматривается конечное множество стран. Каждая страна описывается своей эколого-демографо-экономической моделью. При этом считается, что схема модели в каждой стране одна и та же, и отличия заключаются в различных начальных данных и значениях параметров, входящих в функции, описывающие демографические, экономические и экологические процессы [2, 3].

В модели считается, что каждой стране соответствует своя культура и своя структура потребления. Носителями этой культуры являются жители страны и при переезде они могут либо сохранить свой культурный статус и, соответственно, структуру потребления, либо ассимилироваться, то есть принять культуру страны пребывания.

Схема модели в общих чертах следующая. В каждой стране существует некоторое количество людей, неудовлетворенных текущей социально-экономической или экологической ситуацией в стране. Мы считаем, что этот уровень определяется следующими факторами: уровнем безработицы в стране, среднедушевым доходом, качеством образования, информационным воздействием. Причем важны не только текущие значения этих факторов, но и предыстория. А именно, если ситуация в прошлом была плохой, но наблюдаются улучшение ситуации, то доля недовольных сокращается.

В каждом году часть из недовольных жителей страны мигрируют в другие страны. При этом в зависимости от возраста и уровня образованности выбирается одна из стран, где уровень жизни максимизирует определенный критерий, соответствующий представлениям мигрирующей группы.

Следующим этапом является процесс ассимиляции мигрантов. В модели считается, что на это влияют следующие факторы: величина доли мигрантов из страны в общей численности страны, интенсивности миграционных потоков, а также образовательная политика страны. Считается, что существуют некоторые критические значения вышеперечисленных факторов, при достижении которых

интенсивность ассимиляционных процессов замедляется и начинается процесс формирования диаспор.

Особо отметим, что в данной версии модели мы не рассматриваем процессы перемешивания (взаимопроникновения) культур за счет образования смешанных браков, а также за счет процесса образования. В описываемой модели существует культуры, которые не перемешиваются.

В данной работе мы описываем уравнения блока миграции, включающей в себя процесс формирования недовольных жителей, формирования потока мигрантов, процессов ассимиляции и образования диаспор. Центральным вопросом, которому посвящена статья, является вопрос влияния факторов информационных воздействий на скорости протекания указанных выше процессов. К информационным воздействиям мы относим образовательный процесс, а именно, вводим два типа образования – технологическое, которое влияет на спрос рабочей силы, имеющий определенный уровень этого образования, в производственном процессе, и культурный, уровень которого влияет на скорость ассимиляции: чем он выше, тем медленнее идут процессы ассимиляции. Так же мы обсуждаем возможность стран информационно влиять на свою привлекательность.

Описание уравнений модели

Обозначим через $x_{e,\tau,j}^{i,t}(\zeta)$ численность в i -ой стране в момент времени t людей уровня образованности e , возраста τ , культуры j . Величина ζ равна количеству лет нахождения данной группы вне своей культуры. $i, j \in M; e \in E$, где E – рассматриваемые уровни образованности (в ЭДЭМ $e=1$ соответствует неполное среднее образование, $e=2$ – среднее и т. д.), $\tau = (0,100)$ – возраст людей, $\zeta = (0,100)$ время пребывания вне страны рождения.

В каждый момент времени t демографический состав страны описывается вектором $\overrightarrow{x^{i,t}} = (x_{1,0,1(0)}^{i,t}, \dots, x_{E,100,M(100)}^{i,t})$

Запишем уравнения, описывающие изменение подвектора, характеризующего динамику изменения численности коренного населения в i -ой стране

$$x_{e,\tau+1,i}^{i,t+1} = x_{e,\tau,i}^{i,t} (1 - \beta_{e,\tau,i}^{i,t}) + \varphi_{e-1,\tau,j(\zeta)}^{i,t} x_{e-1,\tau,i}^{i,t} + \sum_{j \neq i}^{M-1} \theta_{e,\tau,j(\zeta)}^{i,t} x_{e,\tau,j(\zeta)}^{i,t} + \\ + \sum_{k \neq i} (\tilde{\phi}_{e,\tau,i}^{i,k,t} \tilde{x}_{e,\tau,i(\zeta)}^{k,t} - \tilde{\phi}_{e,\tau,i}^{k,i,t} \tilde{x}_{e,\tau,i}^{i,t}) \quad (1)$$

Первое слагаемое описывает смертность, второе слагаемое – образовательный процесс, третье – процесс ассимиляции, четвертое – процесс миграции людей культуры j из остальных стран ($k \neq i$), последний член описывает отток из страны людей культуры i . Уравнение, описывающее изменение численности мигрантов, имеет вид:

$$x_{e,\tau+1,j(\zeta+1)}^{i,t+1} = x_{e,\tau,j(\zeta)}^{i,t} (1 - \beta_{e,\tau,j(\zeta)}^{i,t}) + \varphi_{e-1,\tau,j(\zeta)}^{i,t} x_{e-1,\tau,j(\zeta)}^{i,t} + \\ + \sum_{k \neq i} (\tilde{\phi}_{e,\tau,j(\zeta)}^{i,k,t} \tilde{x}_{e,\tau,j(\zeta)}^{k,t} - \tilde{\phi}_{e,\tau,j(\zeta)}^{k,i,t} \tilde{x}_{e,\tau,j(\zeta)}^{i,t}) \quad (2)$$

К этим уравнениям надо добавить уравнение, описывающее рождаемость в i -ой стране людей культуры j в k -ую страну

$$x_{0,0,j(0)}^{i,t+1} = \sum_{\tau=15}^{\tau=50} \sum_e \sum_{\zeta} x_{e,\tau,j(\zeta)}^{i,t} \gamma_{e,\tau,j(\zeta)}^{i,t} \quad (3)$$

Здесь $\beta_{e,\tau,j(\zeta)}^{i,t}, \gamma_{e,\tau,j(\zeta)}^{i,t}$ – коэффициенты смертности и рождаемости соответственно. Коэффициент смертности вычисляется по формуле:

$$\beta_{e,\tau,j(\zeta)}^{i,t} = deat_{e,\tau,j(\zeta)}^{i,t} \cdot deatq_{e,\tau,j(\zeta)}^{i,t} \cdot deatql_{e,\tau,j(\zeta)}^{i,t} (ql_t^i) \cdot deatwh_{e,\tau,j(\zeta)}^{i,t} (wh_t^i)$$

$deat_{e,\tau,j(\zeta)}^{i,t}$ – номинальный коэффициент смертности;

$deatq_{e,\tau,j(\zeta)}^{i,t}$ – мультипликативная поправка, учитывающая зависимость коэффициента смертности от уровня образованности;

$deatql_{e,\tau,j(\zeta)}^{i,t}(ql_t^i)$ – мультипликативная поправка, учитывающая зависимость коэффициента смертности от уровня качества ql_t^i окружающей среды в стране i ;

$deatwh_{e,\tau,j(\zeta)}^{i,t}(wh_t^i)$ – мультипликативная поправка, учитывающая зависимость коэффициента смертности от расходов wh_t^i на здравоохранение в стране i .

Коэффициент рождаемости вычисляется по формуле

$$\gamma_{e,\tau,j(\zeta)}^{i,t} = birth_{e,\tau,j(\zeta)}^{i,t} \cdot birq_{e,\tau,j(\zeta)}^{i,t} \cdot birql_{e,\tau,j(\zeta)}^{i,t}(ql_t^i) \cdot birwh_{e,\tau,j(\zeta)}^{i,t}(wh_t^i)$$

Здесь $birth_{e,\tau,j(\zeta)}^{i,t}$ – номинальный коэффициент рождаемости;

$birq_{e,\tau,j(\zeta)}^{i,t}$ – мультипликативная поправка, учитывающая зависимость коэффициента рождаемости от уровня образованности;

$birql_{e,\tau,j(\zeta)}^{i,t}(ql_t^i)$ – мультипликативная поправка, учитывающая зависимость коэффициента рождаемости от уровня качества ql_t^i окружающей среды в стране i ;

$birwh_{e,\tau,j(\zeta)}^{i,t}(wh_t^i)$ – мультипликативная поправка, учитывающая зависимость коэффициента смертности от расходов wh_t^i на здравоохранение в стране i ;

$\tilde{x}_{e,\tau,j(\zeta)}^{k,t}$ – количество людей, недовольных уровнем своей жизни в стране k в момент времени t , имеющих уровень образованности e , уехавших из страны j ζ тактов в прошлом, возраста τ .

$$\tilde{X}_{e,\tau,j(\zeta)}^{k,t} = x_{e,\tau,j(\zeta)}^{k,t} * \left[\frac{q_{e,\tau,j(\zeta)} * ql^{k,t}}{\sum_{l=0}^T ql^{k,t-l}} + \frac{b_{e,\tau,j(\zeta)} * B^{k,t}}{\sum_{l=0}^T B^{k,t-l}} \right] * \frac{\sum_{j \neq k} Inf_j^{k,t}}{M-1}$$

где $ql^{k,t}$ – качество окружающей среды в стране k в момент времени t ;

$B^{k,t}$ – государственные доходы в стране k в момент времени t ;

T – время «социальной памяти»;

$\frac{\sum_{j \neq k} Inf_j^{k,t}}{M-1}$ – информационное воздействие со стороны других стран на

население страны k ;

$\tilde{\phi}_{e,\tau,j(\zeta)}^{i,k,t}$ – коэффициент миграции из стран $k \neq i$ в страну i , который равен доле людей, проживающих в стране k в момент времени t , имеющих уровень образованности e , уехавших из страны j ζ тактов в прошлом, возраста τ , недовольных уровнем своей жизни.

Он определяется из условия

$$\tilde{\phi}_{e,\tau,j(\zeta)}^{i,k,t} = \lambda_{e,\tau,j(\zeta)}^{i,k,t} * \max \left(0, \frac{A_{e,\tau,j(\zeta)}^{i,k,t} + (QL^{i,k,t})}{\sum_{k \neq i} |A_{e,\tau,j(\zeta)}^{i,k,t} + (QL^{i,k,t})|} \right)$$

где $A_{e,\tau,j(\zeta)}^{i,k,t} = (w_{e,\tau,j(\zeta)}^{i,t} - w_{e,\tau,j(\zeta)}^{k,t})$ – разница в зарплатах в разных странах;

$QL^{i,k,t} = (ql_t^i - ql_t^k)$ – различие в качестве окружающей среды.

$$\lambda_{e,\tau,j(\zeta)}^{i,t} = \frac{\sum_{j \neq k} Inf_j^{k,t} - Inf_j^{j,t}}{M-1}$$

$\phi_{e-1,\tau,j(\zeta)}^{i,t}$ – коэффициент, описывающий переход соответствующей группы людей на другой уровень образованности. Считается, что понизить уровень образованности нельзя. Коэффициент состоит из двух частей: первый определяется долей доходов населения, затрачиваемых на образование, а второй – затратами государства на образование. Он равен

$$\phi_{e-1,\tau,j(\zeta)}^{i,t} = prte_e * w_{e,\tau,j(\zeta)}^{i,t} + \frac{artE^{i,t}}{\sum_{e,\tau,j(\zeta)} x_{e,\tau,j(\zeta)}^{i,t}}$$

где $w_{e,\tau,j(\zeta)}^{i,t}$ – зарплата в стране i в момент времени t людей определенной культуры, имеющих соответствующий уровень образования и возраст;

$artE^{i,t}$ – затраты на образование;

$\sum_{e,\tau,j(\zeta)} x_{e,\tau,j(\zeta)}^{i,t}$ – общая численность населения в стране i в момент времени t ;

$\theta_{e-1,\tau,j(\zeta)}^{i,t}$ – коэффициент ассимиляции равен доли людей культуры j , принимающих культуру страны i в момент времени t . Он до определенного предела зависит от относительной численности населения с культурой i к общей численности жителей культуры j . Но, если доля коренного населения в общей численности достигает уровня 0,5 и ниже, то коэффициент ассимиляции в стране становится равным нулю и начинают расти диаспоры.

$$\theta_{e-1,\tau,j(\zeta)}^{i,t} = \begin{cases} 1, & \text{если } \zeta \geq \zeta^* \\ \alpha \left(\frac{x_{e,\tau,j(\zeta)}^{i,t}}{\sum_{e,\tau,j(\zeta)} x_{e,\tau,j(\zeta)}^{i,t}} \right) & \text{если } \zeta < \zeta^* \text{ и } \frac{x_{e,\tau,j(\zeta)}^{i,t}}{\sum_{e,\tau,j(\zeta)} x_{e,\tau,j(\zeta)}^{i,t}} < 0,5 \\ 0 & \text{если } \frac{x_{e,\tau,j(\zeta)}^{i,t}}{\sum_{e,\tau,j(\zeta)} x_{e,\tau,j(\zeta)}^{i,t}} \geq 0,5 \end{cases}$$

где значения α и ζ^* зависят от многих факторов, в частности от $Inf_j^{k,t}$ – информационного влияния культуры k на культуру j . В настоящее время считается, что это некоторые постоянные.

Вновь введенный фактор – информационное влияние культур друг на друга является внешней переменной модели и задается сценарно при распределении бюджета стран (смотри [1]). Описанная выше модель позволяет учесть миграционные взаимодействия между странами, но пока не учитывает

экономические обмены. Для учета последнего необходимо изменение экономического блока модели.

Литература

1. Павловский Ю. Н., Белотелов Н. В., Бродский Ю. И., Оленев Н. Н. Опыт имитационного моделирования при анализе социально-экономических явлений. – Москва: МЗ Пресс, 2005. – 136 с.
2. Белотелов Н. В. Эколого-демографо-экономическая модель с учетом миграционных процессов между странами // Труды V Всероссийской научной конференции «Математическое моделирование развивающейся экономики и экологии», ЭКОМОД – 2010. – Киров, 5-11 июля 2010. – Киров: ГОУ ВПО ВятГУ, 2010. – С. 19-27.
3. Белотелов Н. В. Модель миграционных процессов между несколькими странами. // Моделирование, декомпозиция и оптимизация сложных динамических процессов / Ю. Н. Павловский. – М.: ВЦ РАН, 2010. – С. 25-40.
4. Вайдлих В. Социодинамика: системный подход к математическому моделированию в социальных науках. – М.: Едиториал УРСС, 2005. – 480 с.
5. Чернавский Д. С. Синергетика и информация (динамическая теория информации). – М.: Едиториал УРСС, 2004. – 288 с.

Информация об авторах

Data on authors

Белотелов Николай Вадимович, к.ф.-м.н., старший научный сотрудник
ВЦ РАН, доцент НИУ МФТИ, г. Москва

Тел. (499)-135-24-89

Электронная почта: belotel@mail.ru

Belotelov Nicolas, PhD of Physics & Mathematics, senior researcher at CCAS,
Associate Prof. on System Research in NSU MPTI, Moscow

Tel. (499)-135-24-89

E-mail: belotel@mail.ru

Березкин В. Е., Лотов А. В.

Beryozkin V. E., Lotov A. V.

**ИСПОЛЬЗОВАНИЕ ФУНКЦИЙ ВКЛЮЧЕНИЯ ДЛЯ
СРАВНЕНИЯ ТОЧЕЧНЫХ АППРОКСИМАЦИЙ МНОГОМЕРНОЙ
ГРАНИЦЫ ПАРЕТО¹**

**INCLUSION FUNCTION APPLICATION FOR COMPARING
POINT-WISE APPROXIMATIONS OF MULTI-DIMENSIONAL
PARETO FRONTIER**

Предлагается новый способ сравнения двух точечных аппроксимаций границы Парето, метод функций включения, который является обобщением хорошо известного показателя отношения ошибок (Error Ratio).

A new method for comparing of two point-wise approximations of the Pareto frontier, the method of inclusion functions, is proposed. The method is a generalization of the well-known Error Ratio indicator.

Ключевые слова: двух точечные аппроксимации границы Парето, метод функций включения, показатель отношения ошибок.

Key words: the method for comparing of two point-wise approximations of the Pareto frontier, the method of inclusion functions, Error Ratio.

¹ Работа выполнена при частичной финансовой поддержке РФФИ (гранты № 10-01-00199 и 11-01-12136-офи-м), ПФИ Президиума РАН П-14, ПФИ Президиума РАН П-17, ПФИ ОМН РАН № 3.

В методах решения задач многокритериальной оптимизации, основанных на аппроксимации границы Парето [1], одной из важнейших проблем является оценка качества аппроксимации границы Парето и методов их построения (см. обзор [2]). В частности, для оценки качества аппроксимации границы Парето и метода, который использовался для построения аппроксимации, можно использовать ее сравнение с другой аппроксимацией границы Парето, построенной с помощью альтернативного метода для той же задачи многокритериальной оптимизации. Для сравнения двух точечных аппроксимаций границы Парето в данной работе предлагается новый способ, метод функций включения (ФВ), который является обобщением хорошо известного показателя отношения ошибок (Error Ratio), рассмотренного в [3].

Рассмотрим следующую задачу многокритериальной оптимизации. Пусть допустимые решения x являются элементами множества X , а векторы критериев y , связанные с допустимыми решениями соотношением $y=f(x)$, являются элементами линейного пространства R^m . Будем считать, что желательно уменьшение значения каждого из критериев в отдельности (при равных значениях остальных критериев). Такое предпочтение в рамках задачи многокритериальной минимизации формализуется в виде бинарного отношения Парето, при использовании которого критериальная точка y' более предпочтительна, чем точка y , тогда и только тогда, когда $y' \leq y$ и $y' \neq y$. Теоретическим решением этой задачи является граница Парето (недоминируемая граница) $P(Y)$ множества достижимых критериальных векторов $Y=f(X)$, определяемая как

$$P(Y) = \{ y \in Y : \{ y' \in Y : y' \leq y, y' \neq y \} = \emptyset \}.$$

В дальнейшем нам понадобится понятие оболочки Эджворта-Парето множества $Y=f(X)$ (или множества $P(Y)$, что одно и то же), которая определяется как

$$Y^* = \{ y \in R^m : \exists y' \in Y : y \geq y' \}.$$

Отметим заранее, что используемые нами методы визуализации границы Парето [1] основываются на аппроксимации оболочки Эджворта-Парето и на дальнейшей визуализации границы Парето с использованием интерактивной визуализации наборов двухкритериальных сечений построенной аппроксимации оболочки Эджворта-Парето.

Пусть имеются две аппроксимации A и B паретовой границы $P(Y)$, состоящие из конечного числа точек (т.е. $\text{card } A$ и $\text{card } B$ – конечные числа). В дальнейшем через A^* и B^* будем обозначать оболочку Эджворта-Парето множеств A и B , соответственно. Таким образом, для $z \in R^m$ утверждение $z \in A^*$ эквивалентно утверждению о том, что найдется такой $y \in A$, что $z \geq y$. Отметим, что поскольку множества A и B являются аппроксимациями $P(Y)$, то множества A^* и B^* являются аппроксимациями множества Y^* . Будем для определенности предполагать, что множества A и B не содержат доминируемых точек.

Для сравнения точечных аппроксимаций паретовой границы было предложено несколько показателей [3]. Среди них наиболее практичен показатель, называемый Error Ratio (отношение ошибок) и удобно записываемый с использованием понятия оболочки Эджворта-Парето

$$ER(A, B^*) = \frac{1}{\text{card } A} \sum_{i=1}^{\text{card } A} \delta(a^i), \text{ где } \delta(a^i) = \begin{cases} 1, & a^i \notin B^* \\ 0, & a^i \in B^* \end{cases}.$$

Ясно, что $0 \leq ER(A, B^*) \leq 1$. Величина $ER(A, B^*)$ показывает, в какой степени ошибочно утверждение о том, что аппроксимация A доминируется аппроксимацией B , т.е. $A \subset B^*$. При этом равенство $ER(A, B^*) = 1$ означает, что ни одна точка из A не принадлежит B^* , а равенство $ER(A, B^*) = 0$ означает, что $A \subset B^*$. В свою очередь, равенство $ER(B, A^*) = 1$ означает, что ни одна точка из B не принадлежит A^* , а равенство $ER(B, A^*) = 0$ означает, что $B \subset A^*$.

Для сравнения качества двух аппроксимаций предлагается сравнить величины $ER(A, B^*)$ и $ER(B, A^*)$. Отметим, что равенство $ER(A, B^*) = 1$ не

означает, что аппроксимация A лучше аппроксимации B . Это подтверждается примером, приведенном на рис. 1, где имеет место случай, когда и $ER(A, B^*)=1$, и $ER(B, A^*)=1$.

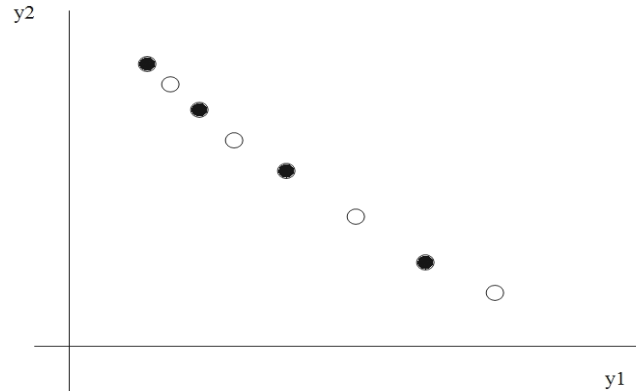


Рис. 1. Здесь две аппроксимации A (черные точки) и B (белые точки) не отличаются по качеству, причем $ER(A, B^*) = ER(B, A^*)=1$

Если $ER(A, B^*) < ER(B, A^*)$, то обычно утверждают, что аппроксимация A хуже аппроксимации B , и, наоборот, если $ER(B, A^*) < ER(A, B^*)$, то аппроксимация B считается хуже аппроксимации A . То, что эти утверждения неточны, показывает пример, приведенный на рис. 2, где аппроксимация A (черные точки) лучше, чем аппроксимация B (белые точки), но, как легко заметить, $ER(A, B^*) = 1/3$, $ER(B, A^*)=1$.

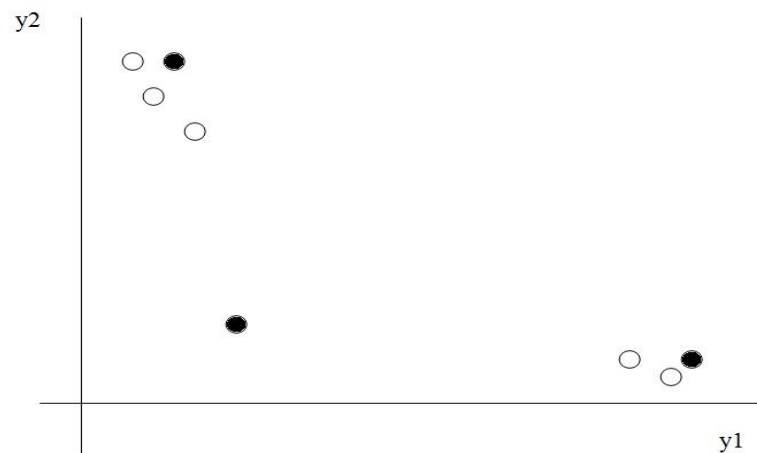


Рис. 2. Здесь аппроксимация A (черные точки) точнее, чем B (белые точки), но $ER(A, B^*) < ER(B, A^*)$

Отметим, что причиной такой «неточности» показателя Error Ratio является то, что он не учитывает количественных аспектов доминирования-отклонения точек аппроксимации A от границы множества B^* (см. рис. 3).

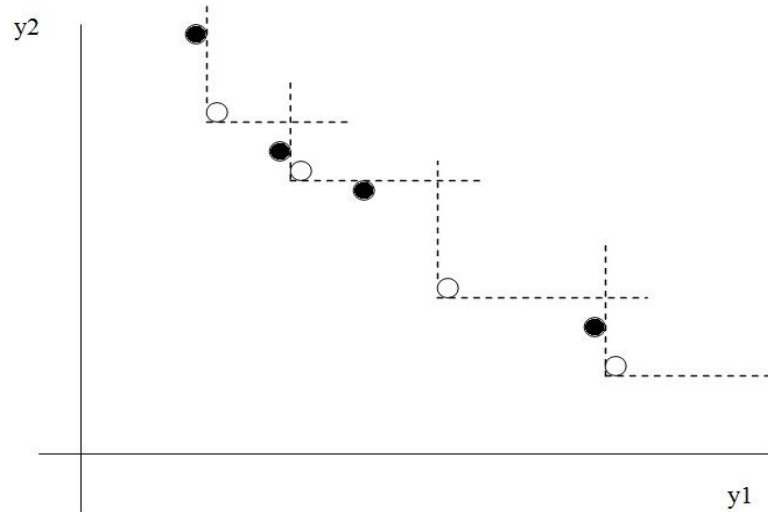


Рис. 3. Хотя $ER(A, B^*) = ER(B, A^*)=1$, точки аппроксимации A (черные точки) близки к B^* (конусы с вершинами в белых точках) и не улучшают аппроксимацию ОЭП, задаваемую множеством B . В то же время, устранение белых точек может резко ухудшить аппроксимацию

Для того чтобы избавиться от указанного недостатка, нами на основе показателя «Error Ratio» была предложена новая характеристика, позволяющая сравнить качество двух аппроксимаций более детально при помощи анализа принадлежности точек множества A некоторой окрестности множества B^* в некоторой метрике. Точнее говоря, рассматривается ε окрестность множества B^* , которая обозначается через $(B^*)_\varepsilon$, и используется характеристика, которая является функцией ε . Это функция включения $\nu_A(\varepsilon, B^*)$, определяемая следующим образом:

$$\nu_A(\varepsilon, B^*) = \frac{1}{\text{card } A} \sum_{i=1}^{\text{card } A} \delta(\varepsilon, a^i), \text{ где } \delta(\varepsilon, a^i) = \begin{cases} 1, & a^i \in (B^*)_\varepsilon, \\ 0, & a^i \notin (B^*)_\varepsilon \end{cases}, a^i \in A.$$

Функция включения показывает, какая доля точек базы A содержится в ε -окрестности множества B^* . Аналогичным образом определяется функция включения $\nu_B(\varepsilon, A^*)$. Как легко заметить,

$$ER(A, B^*) = 1 - \nu_A(0, B^*).$$

На рис. 4 и рис. 5 приведены функции включения $\nu_A(0, B^*)$ и $\nu_B(\varepsilon, A^*)$ для двух аппроксимаций некоторой восьмимерной паретовой границы, которые содержали порядка нескольких сотен точек каждая.

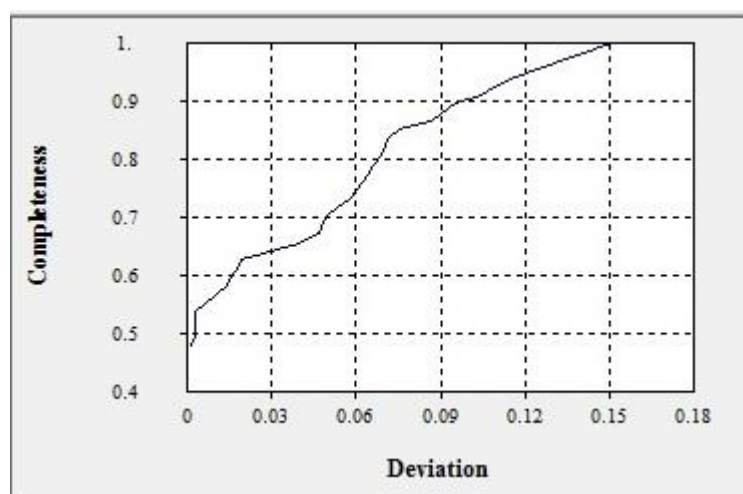


Рис. 4. График функции включения $\nu_A(\varepsilon, B^*)$

По горизонтальной оси – величина ε (в долях от размера множества $P(Y)$), по вертикальной оси – значение функции. Эти аппроксимации были построены с использованием гибридных методов аппроксимации границы Парето, описанных в [4], при применении к задаче построения диспетчерских графиков управления водохранилищем ГЭС с учетом как экономических, так и экологических факторов [5].

Как видим на рис. 4, функция включения $\nu_A(\varepsilon, B^*)$ равна приблизительно 0.5 при $\varepsilon=0$ и достигает единицы при $\varepsilon=0.15$. Таким образом, $ER(A, B^*)=0.5$, т.е. аппроксимация B^* включает примерно половину точек A . В отличие от $\nu_A(\varepsilon, B^*)$, функция включения $\nu_B(\varepsilon, A^*)$ на

рис. 5 при $\varepsilon=0$ равна нулю, т.е. аппроксимация A^* вообще не включает точек B , но достигает единицы уже при $\varepsilon=0.083$.

Таким образом, на основе показателя ER можно сделать вывод о том, что аппроксимация границы Парето, задаваемая множеством точек B , значительно лучше аппроксимации, задаваемая множеством A . Использование функций включения показывает, что дело обстоит не так просто.

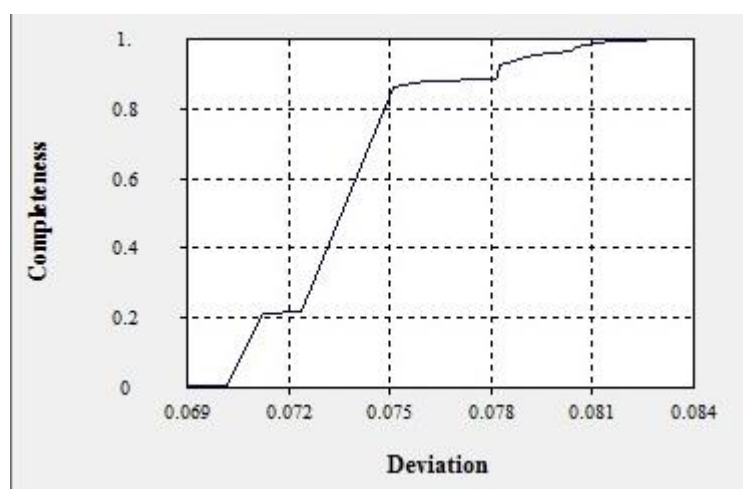


Рис. 5. График функции включения $\nu_B(\varepsilon, A^*)$

Действительно, множество $(A^*)_\varepsilon$ содержит все точки множества B уже при $\varepsilon=0.083$, в то время как множество $(B^*)_\varepsilon$ содержит все точки множества A только при $\varepsilon=0.15$. Эта дополнительная информация особенно наглядно видна на рис. 6, где графики функций включения $\nu_A(\varepsilon, B^*)$ и $\nu_B(\varepsilon, A^*)$ изображены на одном рисунке.

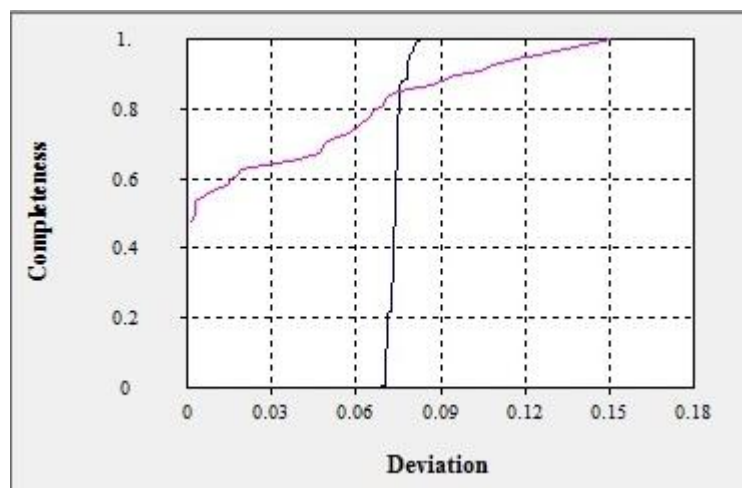


Рис. 6. Графики функций $\nu_A(\varepsilon, B^*)$ и $\nu_B(\varepsilon, A^*)$

Из этого рисунка следует, что все точки множества B расположены на расстоянии от аппроксимации A^* , принадлежащем диапазону в 7-8% от размеров множества Парето. В то же время, множество A содержит значительное число точек (около 10%), находящихся на расстоянии от аппроксимации A^* , превышающем 10% от размеров множества Парето. Таким образом, вопрос о сравнении качества аппроксимаций A^* и B^* в данном случае сложен и должен решаться пользователем в зависимости от его потребностей. Информация для такого решения предоставляется методом, описанным в данной работе.

Литература

1. Lotov A. V., Bushenkov V. A., Kamenev G. K. Interactive Decision Maps. Boston: Kluwer Academic Publishers, 2004, 310 p.
2. Zitzler E., Knowles J., and Thiele L. Quality assessment of Pareto set approximations // J.Branke, K.Deb, K.Miettinen, R.Slowinski (eds.) Multiobjective Optimization. Interactive and Evolutionary Approaches, Lecture Notes in Computer Science. V.5252. Springer, Berlin-Heidelberg, 2008. P. 373-404.

3. Zitzler, E., L. Thiele, M. Laumanns, C. M. Foneseca, and V. Grunert da Fonseca. Performance Assessment of Multiobjective Optimizers: An Analysis and Review // IEEE Transactions on Evolutionary Computation. 2003. V. 7 (2). P. 117–132.

4. Березкин В. Е., Каменев Г. К., Лотов А. В. Гибридные адаптивные методы аппроксимации невыпуклой многомерной паретовой границы // ЖВМиМФ. 2006. Т. 46(11). С. 2009-2023.

5. Каменев Г. К., Лотов А. В., Рябиков А. И. Использование параллельных вычислений при аппроксимации многомерной границы Парето в задачах многокритериальной оптимизации // Труды V Международной конференции «Параллельные вычисления и задачи управления» РАСО'2010 (Москва, 26-28 октября 2010 г.). М.: ИПУ РАН, 2010. – С.241-263.

Информация об авторах

Data on authors

Берёзкин Вадим Евгеньевич, к.ф.-м.н., старший научный сотрудник
ВЦ РАН, г. Москва

Тел.: (499)-135-12-09

Электронная почта: berezkin@ccas.ru

Beryozkin Vadim, Cand. of Physics & Mathematics, senior researcher at
CCAS, Moscow

Tel.: (499)-135-12-09

E-mail: berezkin@ccas.ru

Лотов Александр Владимирович, д.ф.-м.н., профессор ВЦ РАН, г.
Москва

Тел.: (499)-135-12-09

Электронная почта: avlotov@ccas.ru

Lotov Alexander, Doctor of Physics & Mathematics, Prof. at CCAS,
Moscow

Tel.: (499)-135-12-09

E-mail: avlotov@ccas.ru

**ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ДИНАМИКИ
РАСПРЕДЕЛЕННОЙ СИСТЕМЫ С ЗАПАЗДЫВАНИЕМ В МОДЕЛИ
ЦИРКАДИАННЫХ РИТМОВ *NEUROSPORA CRASSA***

**NUMERICAL STUDY OF THE DYNAMICS OF A DISTRIBUTED
SYSTEM WITH TIME-DELAY IN A MODEL OF CIRCADIAN RHYTHMS
OF *NEUROSPORA CRASSA***

В работе предложена новая модель циркадианных колебаний нейроспоры, которая описывает пространственно-временную динамику белков, ответственных за механизм биоритмов. Модель основывается на нелинейном взаимодействии белков *FRQ* и *WCC*, кодируемых генами *frequency* и *white collar*, и включает в себя как положительную, так и отрицательную петлю обратной связи. Главным элементом механизма колебаний является эффект запаздывания транскрипции генов в биохимических реакциях. Показано, что модель воспроизводит такие свойства циркадианных колебаний нейроспоры, как захват частоты под действием внешнего периодического освещения, сброс фазы биоритмов при воздействии импульса света, устойчивость механизма колебаний по отношению к случайным флуктуациям и т. д. Исследованы волновые структуры, возникающие в ходе пространственной эволюции системы. Показано, что волны синхронизации биоритмов среды возникают под воздействием базального транскрипционного фактора.

We derive a new model of circadian oscillations in *Neurospora crassa*, which is suitable to analyze both temporal and spatial dynamics of proteins responsible for mechanism of rhythms. The model is based on the non-linear interplay between proteins *FRQ* and *WCC* which are products of transcription of *frequency* and *white collar* genes

forming a feedback loop comprised both positive and negative elements. The main component of oscillations mechanism is supposed to be time-delay in biochemical reactions of transcription. We show that the model accounts for various features observed in *Neurospora*'s experiments such as entrainment by light cycles, phase shift under light pulse, robustness to action of fluctuations and so on. Wave patterns excited during spatial development of the system are studied. It is shown that the wave of synchronization of biorythms arises under basal transcription factors.

Ключевые слова: запаздывание, стохастические системы, циркадианные ритмы, генная регуляция

Key words: time-delay, stochastic systems, circadian rythms, gene regulation

Введение

Циркадианные ритмы, или их еще называют биоритмы, присущи всем живым организмам [6]. Также достоверно известно, что существуют гены, определяющие функционирование этих ритмов. Для модельного организма генетических исследований *Neurospora crassa* такими генами являются *frequency* и *white collar* [8]. Данные гены кодируют белки FRQ и WCC, соответственно, которые вступают в нелинейное взаимодействие, включающее в себя положительную и отрицательную обратную связь. Но главным элементом механизма колебаний является эффект запаздывания в биохимических реакциях транскрипции генов. В отличие от работ других исследователей [4,5], впервые рассмотрена пространственно-временная эволюция, генерируемая циркадианными колебаниями.

Математическая модель

Молекулярные процессы, которые являются основой построения нашей модели, представлены в таблице 1.

Список реакций

Процесс димеризации:	$F + F \xrightarrow{k_1^F} F_2$, $W + W \xrightarrow{k_1^W} W_2$
Процесс дедимеризации:	$F_2 \xrightarrow{k_2^F} F + F$, $W_2 \xrightarrow{k_2^W} W + W$
Динамика оператор-сайтов:	$D_0^F + W_2 \xrightarrow{k_3^F} D_1^F$, $D_0^W + F_2 \xrightarrow{k_3^W} D_1^W$ $D_1^F \xrightarrow{k_4^F} D_0^F + W_2$, $D_1^W \xrightarrow{k_4^W} D_0^W + F_2$
Процесс синтеза белков:	$D_1^F(t) \xrightarrow{k_F} D_1^F + F^{t+\tau_F}$, $D_1^W(t) \xrightarrow{k_W} D_1^W + W^{t+\tau_W}$
Процесс распада белков:	$F \xrightarrow{\gamma_F} \emptyset$, $W \xrightarrow{\gamma_W} \emptyset$
Процесс образования гетеродимера:	$F + W \xrightarrow{k} \emptyset$
Базальная транскрипция:	$\emptyset \xrightarrow{b_F} F$, $\emptyset \xrightarrow{b_W} W$

Таким образом, пространственно-временная эволюция *Neurospora crassa* может быть выражена следующим образом:

$$\frac{\partial F}{\partial t} = \frac{1}{(1+4K_1^F F)} \left(b_F + k_F \frac{K_1^W K_2^F W^2 (t-\tau_F)}{1+K_1^W K_2^F W^2 (t-\tau_F)} - \gamma_F F - kFW \right) + D\Delta_{xy} F, \quad (1)$$

$$\frac{\partial W}{\partial t} = \frac{1}{(1+4K_1^W W)} \left(b_W + k_W \frac{K_1^F K_2^W F^2 (t-\tau_W)}{1+K_1^F K_2^W F^2 (t-\tau_W)} - \gamma_W W - kFW \right) + D\Delta_{xy} W, \quad (2)$$

где $F(x, y, t)$ и $W(x, y, t)$ – функции колебаний концентрации белков *FRQ* *WCC* соответственно, D – коэффициент диффузии, Δ_{xy} – оператор Лапласа, заданный на плоскости координат x и y . Будем предполагать, что коэффициенты диффузии белков *FRQ* и *WCC* равны.

Расчётная область $\Sigma(x, y)$ может интерпретироваться как область плоского физического пространства, занятая мицелием нейроспоры. Нейроспора является многоклеточным организмом. Однако транскрипция и кодирование белков происходят внутри отдельных клеток. Наш анализ упрощает важная особенность *N. crassa*: мицелий организма состоит из многоядерных разветвленных гиф, обладающих апикальным полярным ростом. Септальная перегородка с центральной порой до 0.5 мкм в диаметре разделяет гифы на клетки или компартменты (100x20 мкм). Пора проницаема для цитоплазмы,

ядер и митохондрий. Таким образом, белки, вырабатываемые в отдельных клетках нейроспоры, по-видимому, могут преодолевать межклеточные стенки. По крайней мере, в ряде работ было показано наличие функционально связанных групп клеток в гифах мицелия. Оказалось, что такие сообщества клеток на верхушке растущей грибной гифы эффективно функционируют, успешно решая задачу энергетической кооперации и продвижения организма в пространстве. Верхушка растущей грибной гифы – не просто высокоорганизованное образование клеток: в ходе роста все структуры находятся в непрерывной динамике, меняя место, интенсивность работы, связь с соседями. На кончике гифа непрерывно выделяется клеточная стенка и ферменты, переваривающие окружающий субстрат. В нашем случае это означает, что можно говорить об общем белковом облаке внутри организма. Динамика этого облака и определяет биоритмы всего мицелия нейроспоры в целом.

Опишем кратко численный метод. Система уравнений (1-2) интегрировалась в области $\Sigma: (0 < x < 200, 0 < y < 200)$, на границах которой задавались следующие условия:

$$\Delta F = 0, \quad \Delta W = 0. \quad (3)$$

Для решения использовалась явная схема и, чтобы обеспечить её устойчивость, шаг по времени вычислялся по формуле:

$$\Delta t = \frac{\Delta x^2}{2(2 + \max(|F|, |W|))}, \quad (4)$$

где Δx – шаг сетки по пространству.

Главная трудность моделирования системы заключается в вычислении запаздывающих значений. Прямолинейный подход к данной проблеме подразумевает хранение в машинной памяти данных для всех временных слоёв в пределах диапазона запаздывания. В нашем случае из-за большой величины запаздывания, протяжённой области интегрирования и малого шага по времени, общее количество хранимых данных может превысить возможности компьютера. Чтобы преодолеть эту трудность,

хранение данных осуществлялось не на каждом временном шаге, а только в определённые моменты времени, которые мы называем «опорными». Значения переменных F и W для промежуточных слоев времени аппроксимировались по хранимым в памяти опорным слоям. В процессе расчета с целью повышения точности вычислений число таких слоев в пределах диапазона запаздывания могло меняться. Тестовые расчеты показали, что при времени запаздывания $\tau = 6$ для получения корректных результатов достаточно введения не более 20 опорных слоёв.

Результаты

Было рассмотрено два принципиально разных случая начальных условий. В первом случае в качестве начального состояния эволюции системы рассматривалось случайное распределение концентрации белков FRQ и WCC по всей области интегрирования Σ . Данная постановка начальных условий моделирует ситуацию, когда в каждой точке мицелия задана своя начальная фаза циркадианных колебаний. В начальный момент времени все эти биоритмы никак не связаны друг с другом, но в процессе эволюции вступают в нелинейное взаимодействие, результатом которого являются разные виды пространственно-временных структур. С точки зрения функционирования реального организма такая постановка задачи является несколько искусственной, так как внутренняя синхронизация биоритмов в клетках осуществляется на стадии эмбрионального развития организма и структурно поддерживается в ходе его физического роста. Таким образом, ситуация, при которой бы во взрослом организме неожиданно обнулялись все фазы локальных биоритмов в клетках, и эволюция стартует с самого начала, представляется противоестественной. Тем не менее, исследование системы с такими начальными условиями позволяет глубже понять потенциал системы по отношению к пространственно-временной самоорганизации и оценить её вероятные формы структурообразования.

Второй случай начальных условий, представляющий собой малое симметричное возмущение концентрации протеинов FRQ или WCC на фоне нулевых значений полей в остальных точках области Σ , является более естественным. Экспериментально установлено, что при выращивании нейроспоры из небольшого

зародыша в однородной питательной среде на плоскости (в чашке Петри) организм развивается равномерно в обоих пространственных направлениях [2]. В этом случае фронт продвижения границы мицелия математически представляет собой концентрическую ударную волну, распространяющуюся с постоянной скоростью по плоскости. Такое распространение мицелия феноменологически может быть описано соответствующим уравнением, однако мы не вводим его, ограничиваясь только уравнениями (1-2), которые напрямую вытекают из генетико-транскрипционного анализа. Стоит подчеркнуть, что локальная фаза колебаний, вызываемых распространяющейся волной, определяется фазой начального возмущения в зародыше, а вся пространственно-временная структура в целом формируется в ходе морфогенеза организма.

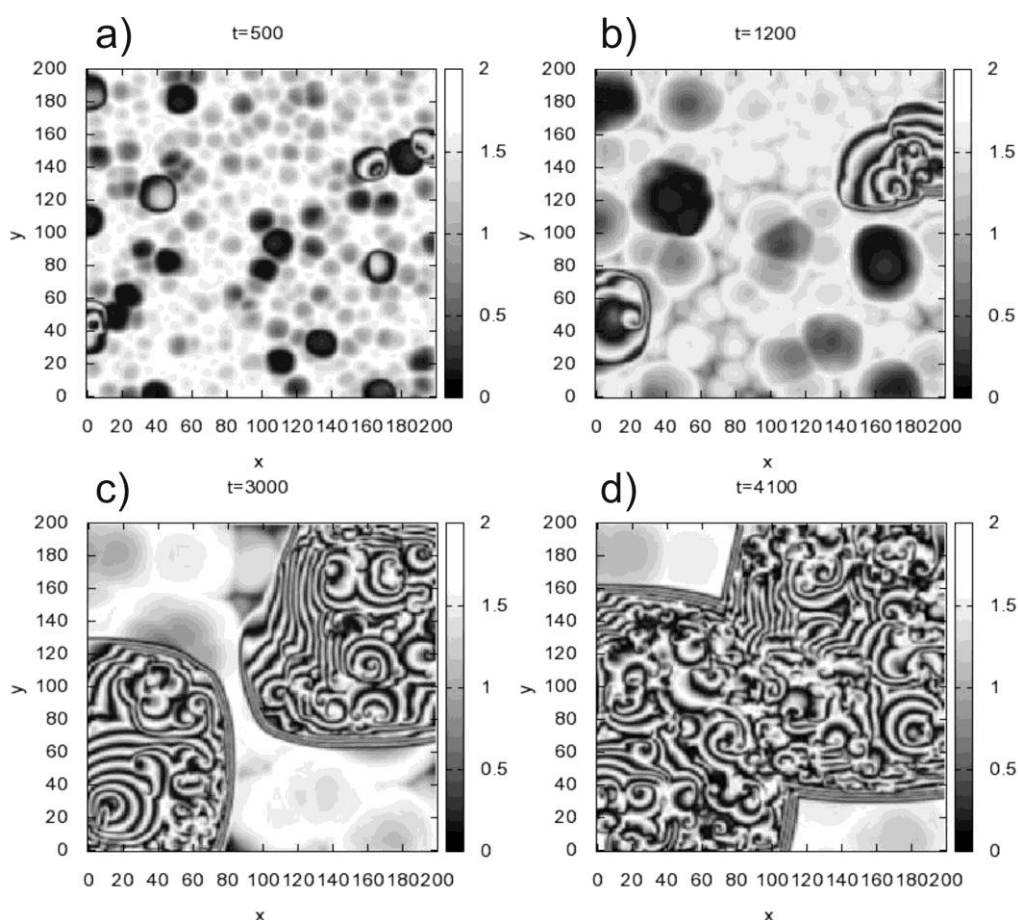


Рис. 1. Поля концентрации белка *FRQ* для четырех последовательных моментов времени t : (a) – 500, (b) – 1200, (c) – 3000, (d) – 4100. Внешние потоки белков в систему отсутствуют: $b_F = 0$, $b_W = 0$.

Рассмотрим сначала первый случай начальных условий. На рисунке 1 представлены пространственные структуры, образованные полем концентрации мономеров FRQ , для четырех последовательных моментов времени. Расчёты выполнены для случая, когда факторы базальной транскрипции отсутствуют: $b_F = 0$, $b_W = 0$. Как видно из рисунка, пространственная динамика системы складывается из двух колебательных режимов. Один из них представляет собой осцилляции концентрации FRQ в квазистационарных ячейках, чей характерный размер медленно растёт со временем (рис.1a, b, c). Одновременно на фоне ячеек распространяются волны (рис.1b, c, d), возникающие из отдельных центров возмущения (рис.1a). Эти волны продвигаются во всех направлениях пространства, пока не захватывают всю область интегрирования. Отметим, что, если сам фронт волны выглядит более-менее упорядоченным, то внутри области неустойчивости возникают многочисленные центры возбуждения вторичных спиральных волн, нелинейное взаимодействие которых приводит к хаотическому поведению среды за гребнем первичной волны. Расчёт взаимокорреляционной функции между парами точек среды, принадлежащих областям до и после прохождения волны, показывает, что корреляции между точками в волновой зоне практически отсутствуют. Таким образом, с точки зрения динамики биоритмов эволюция системы проходит три стадии: сначала из случайных начальных условий возникает медленно меняющаяся во времени ячеистая структура, которая устанавливает определенную синхронизацию между колебаниями различных точек среды. Затем очень быстро в системе возбуждаются волны рассинхронизации, которые сбрасывают все локальные установки фаз часов в среде. Тем не менее, поведение системы нельзя назвать пассивным – скорее она находится в «ждущем» режиме: поле концентрации белка FRQ хаотически осциллирует как по пространству, так и по времени, но при любой команде извне система готова упорядочить фазу своих колебаний нужным образом. Внешнее управление этой активной средой может осуществляться, например, с помощью фактора базальной транскрипции. Но внешний поток белков FRQ и WCC действует на систему по-разному. Особенно чувствительной к возмущениям система становится в случае действия фактора базальной транскрипции WCC .

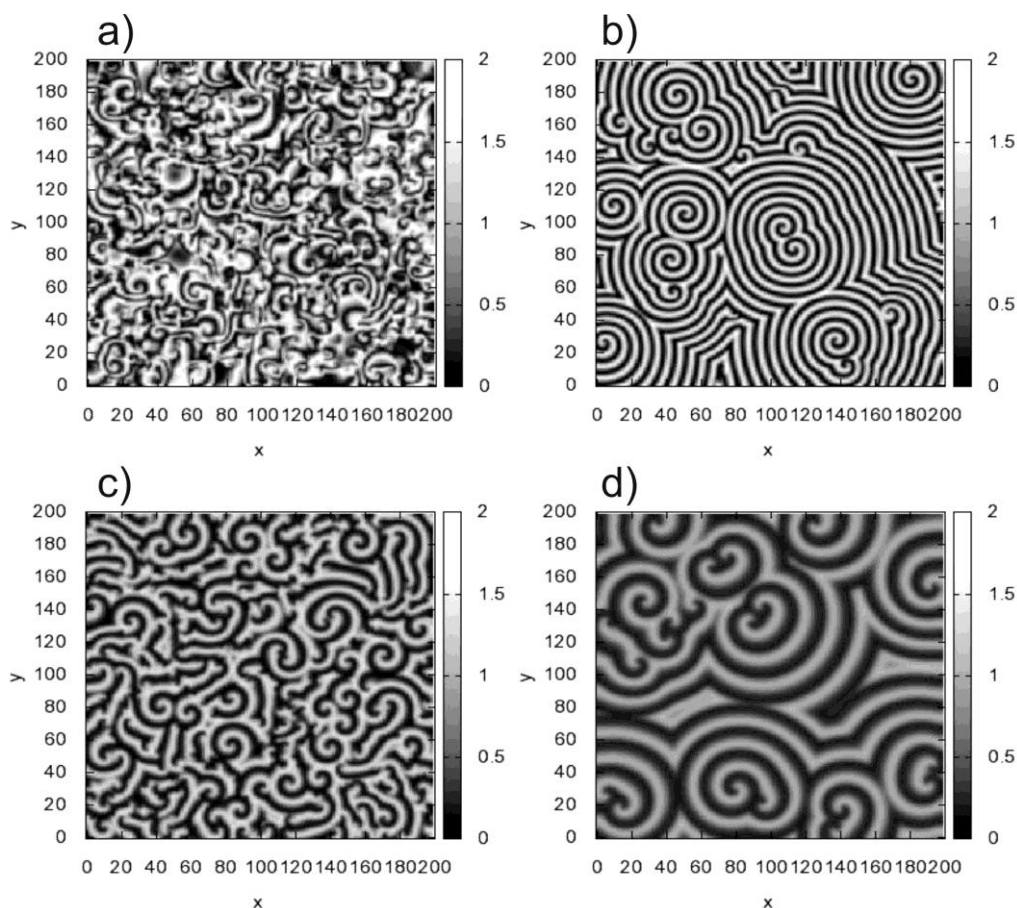


Рис. 2. Структуры, образуемые белком *FRQ* в зависимости от разных значений параметра b_w : (a) – 0, (b) – 1, (c) – 2, (d) – 2.8. Каждая структура является результатом 5000-часовой эволюции системы (1-2) в полной темноте из случайных начальных условий.

На рисунке 2 представлены поля концентрации мономеров *FRQ*, которые являются результатом 5000-часовой эволюции системы, для четырёх значений параметра b_w . Как видно из рисунка, с увеличением внешнего потока белка *WCC* количество организующих центров волновой неустойчивости резко возрастает, а сама пространственно-временная структура становится более упорядоченной. Если в изолированной системе структура имела ярко выраженный нестационарный характер (рис. 2 a), то в случае $b_w \neq 0$ образования являются квазистационарными: пространственной динамики здесь уже нет – структуры просто «вморожены» в пространство, а каждая точка среды осциллирует периодически в соответствии со

своей фазой колебаний, которую ей навязывает паттерн (рис. 2 *b, c, d*). При этом период колебаний несколько увеличивается с ростом b_w . Таким образом, фактор базальной транскрипции играет здесь синхронизирующую роль.

Численные эксперименты для случая внешнего потока белка FRQ в систему показали, что волновые структуры здесь не возникают. Динамика системы представляет собой медленно эволюционирующие ячеистые возмущения, которые походят на фоновые структуры на рис.1*a*. Это поведение можно связать с резким уменьшением относительной величины флуктуаций поля концентрации FRQ . В каком-то смысле влияние этого фактора аналогично действию внешнего освещения нейроспоры, которое также вызывает резкий рост синтеза FRQ и прекращение колебаний.

Рассмотрим теперь второй случай начальных условий: пусть эволюция системы начинается из локального возмущения x^*, y^* в центре области интегрирования. Расчеты показали, что первоначальное возмущение $F(t, x^*, y^*) - W(t, x^*, y^*)$ определяет фазу начальных колебаний в системе. Так как весь организм вырастает из этой точки, колебания в x^*, y^* можно определить как «глобальные часы» нейроспоры. Именно эти часы управляют процессом роста организма.

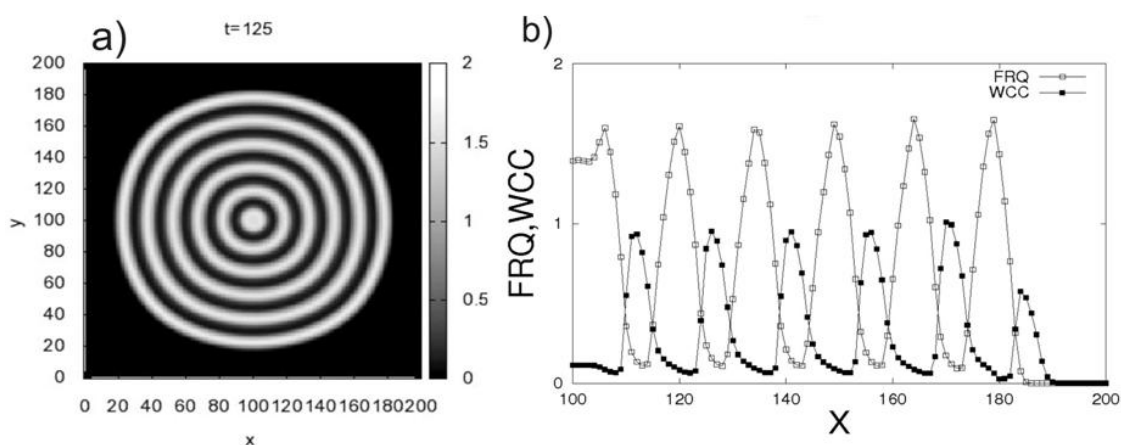


Рис. 3. (а) Структура, образованная белком FRQ после 125-часовой эволюции в полной темноте из одиночного возмущения в центре расчетной области в точке $x^* = 100$, $y^* = 100$; (б) распространение концентрационных волн белков FRQ и WCC вдоль координаты x (из-за симметрии показана только половина волны).

На рисунке 3 приведены: поле концентрации *FRQ* после 125 часов эволюции (a) и одномерный срез поля $y=100$ (b). Хорошо видно, что возникающая белковая структура представляет собой систему концентрических окружностей с центром в точке первоначального возмущения $x^*=100, y^*=100$. Предложенная модель отражает все основные особенности пространственной эволюции *N. crassa*: во-первых, фронт организма расширяется радиально во всех направлениях с постоянной скоростью, начиная с места, куда зародыш был помещен изначально (рис. 3 a). Во-вторых, в течение «дня», т.е. когда уровень белка *FRQ* доминирует над *WCC*, начинается процесс конидиирования. Условие того, будет происходить рост конидий или нет, определяется сразу же за фронтом распространения мицелия (рис.3b). Положительное решение принимается, если уровень *FRQ* высокий («день»), в противном случае конидиирование не происходит («ночь»). После того, как дифференцирование клеток произошло (определились полосы конидиирования), изменения во внешнем облике организма уже не происходят, несмотря на продолжающиеся локальные периодические колебания протеина *FRQ*. На рисунке 3 видно, что по истечении 125 часов эволюции у нейроспоры сформировалось шесть полос конидий. В-третьих, в пространственной модели максимумы (минимумы) значений концентрации белка *FRQ* соответствуют минимумам (максимумам) уровня *WCC*. Период колебаний уровня белков в темноте составляет приблизительно 22.5 часа.

В рамках исследования было рассмотрено влияние внешнего освещения на динамику циркадианных ритмов. Как известно, свет влияет на динамику циркадианных колебаний нейроспоры через белок *WC-1*. Являясь хорошим фоторецептором, *WC-1* получает прямой световой сигнал и образует гетеродимер с белком *WC-2*. Вместе они запускают процесс транскрипции гена *frq*. Таким образом, свет напрямую влияет на механизм генетических часов, который лежит в основе циркадианных осцилляций нейроспоры. Как выражение этого факта, экспериментально наблюдаются такие явления как синхронизация биоритмов нейроспоры под действием внешнего периодического освещения или

сброс текущей фазы колебаний при импульсном освещении [8]. В рамках нашей модели описанный механизм влияния света может быть учтён в (1).

Введём в рассмотрение функцию $\phi(t) = 1 + R(1 + \cos(\Omega t))$, которая описывает изменение интенсивности внешнего освещения во времени (R , Ω – амплитуда и частота колебаний). В этом случае уравнение (1) принимает следующий вид:

$$\frac{\partial F}{\partial t} = \frac{1}{(1 + 4K_1^F F)} \left(b_F + k_F \frac{K_1^W K_2^F W^2(t - \tau_F)}{1 + K_1^W K_2^F \phi(t) W^2(t - \tau_F)} - \gamma_F F - kFW \right) + D\Delta_{xy} F \quad (5)$$

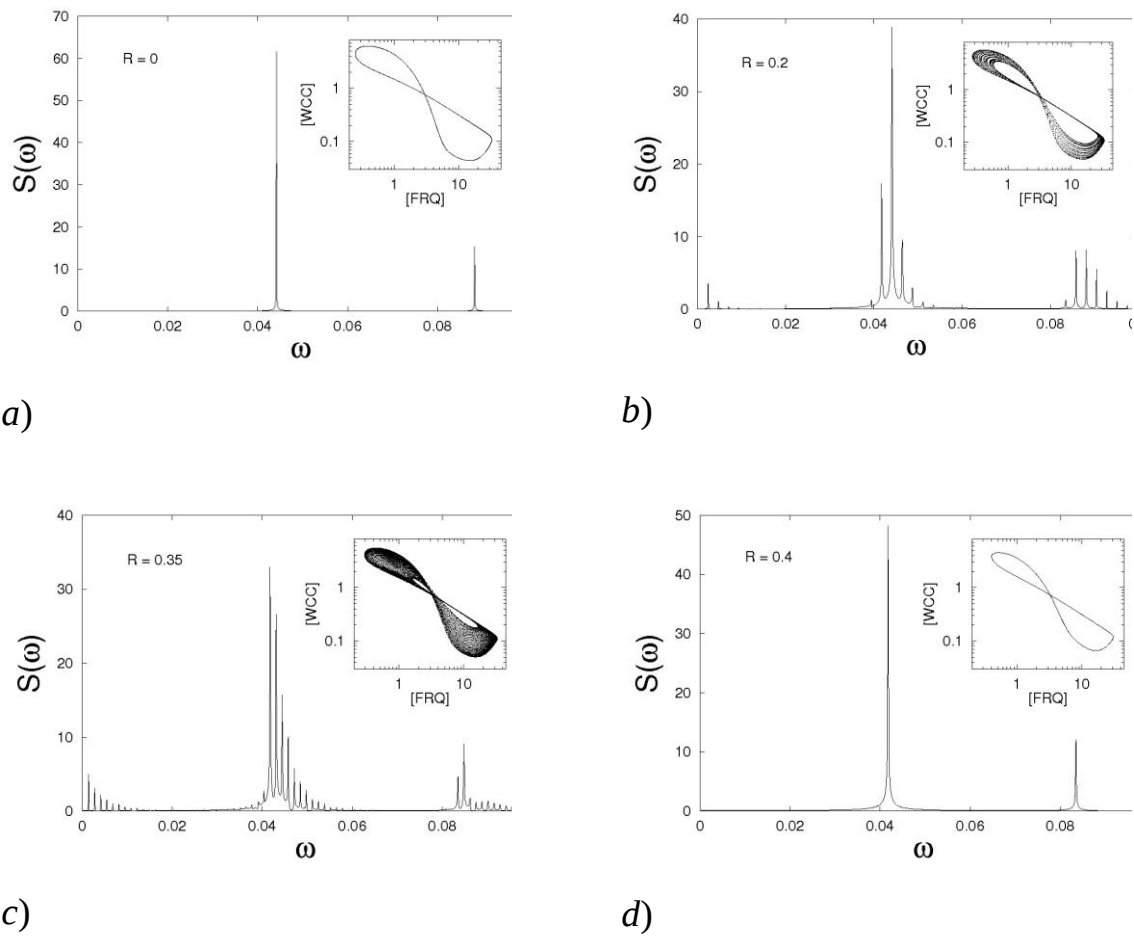


Рис. 4. Спектры Фурье и фазовые портреты системы (во врезке), полученные при численном интегрировании уравнений (2, 5) для четырёх значений амплитуды модуляции внешнего освещения R : (a) – 0, (b) – 0.2, (c) – 0.35, (d) – 0.4.

Также экспериментально установлено, что неожиданный импульс света в темноте производит сброс текущей фазы колебаний у нейроспоры и устанавливает стрелки биочасов на момент, соответствующий максимальному уровню белка *FRQ* в системе [3]. Предложенная нами модель легко описывает данный эффект.

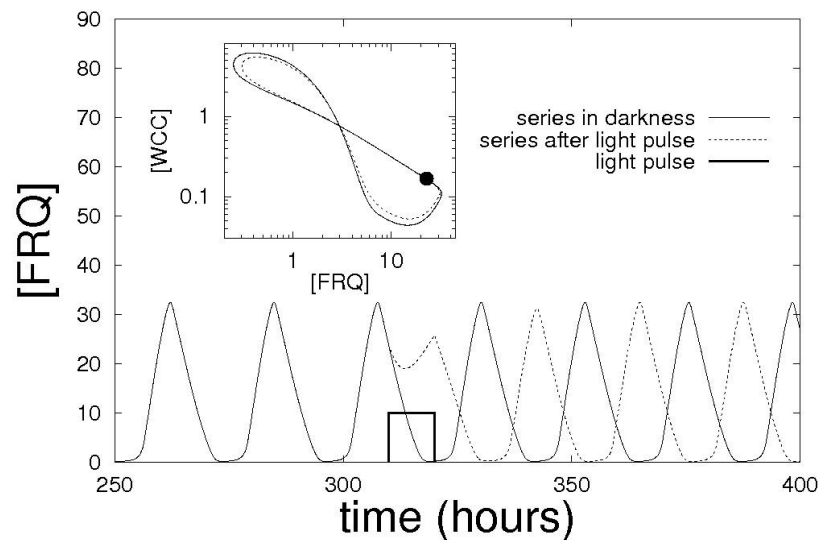


Рис. 5. Эффект сброса фазы колебаний под действием внезапного импульса света в промежутке времени $310 < t < 320$. Во врезке представлен фазовый портрет системы, демонстрирующий что происходит с системой в этот момент.

Заключение

Предложенная в работе модель биоритмов *Neurospora crassa* на основе запаздывания является дальнейшим развитием существующих в литературе моделей. Однако данная модель построена не феноменологически, а напрямую выведена из цепочки биохимических реакций, происходящих при транскрипции генов.

Впервые было проведено исследование пространственно-временной динамики системы запаздывающих циркадных колебаний, которое продемонстрировало разнообразие и сложность возникающих колебательных

режимов. Выявлена зависимость относительной величины флуктуаций от факторов базальной транскрипции генов.

Кроме всего прочего, в работе предложен метод расчета пространственно-распределённых систем с запаздыванием по времени, который основан на хранении данных не всей предшествующей эволюции системы, а только нескольких опорных слоёв. Значения полей для промежуточных моментов времени восстанавливаются путем интерполяции данных опорных слоёв. Описываемая схема может быть адаптирована для решения широкого круга задач с запаздыванием.

Литература

1. Bratsun D., Volfson D., Hasty J., Tsimring L. Delay-induced stochastic oscillations in gene regulation // PNAS, 2005, V.102, N.41, pp. 14593-14598.
2. Lakin-Thomas P.L., Brody S. Circadian rhythms in microorganisms: New complexities // Annu. Rev. Microbiol., 2004, V. 58, pp. 489–519.
3. Liu Y., Loros J., Dunlap J. C.. Phosphorylation of the *Neurospora* clock protein FREQUENCY determines its degradation rate and strongly influences the period length of the circadian clock // PNAS, 2000, V. 97, N. 1, pp. 234-239.
4. Smolen P., Baxter D. A., Byrne J. H. Reduced models of the circadian oscillators in *Neurospora crassa* and *Drosophila melanogaster* illustrate mechanistic similarities // OMICS, 2003, V. 7, N.4, pp. 337-54.
5. Sriram K., Gopinathan M.S. A two variable delay model for the circadian rhythm of *Neurospora crassa* // J. Theor. Biol., 2004, Vol. 231, pp.23-38.
6. Sweeney, B. M. Circadian rhythms, definition and general characterization. // The Molecular Basis of Circadian Rhythms. / edited J. W. Hastings and H.-G. Schweiger. Berlin, Dahlem Konferenzen, 1976, pp. 77-83.
7. Романовский Ю. М. Математическое моделирование в биофизике. – Москва-Ижевск: Институт компьютерных исследований, 2003. – 402 с.

8. Соколовский В. Ю. Действие стрессоров на дифференциальную экспрессию генов в ходе развития *Neurospora crassa*.// Успехи биологической химии, Т. 40, 2000. – С. 85-152.

Информация об авторах

Data on authors

Брацун Дмитрий Анатольевич, д.ф.-м.н., профессор, зав. кафедрой физики ПГПУ, г. Пермь

Тел.: (342) 238-63-25

Электронная почта: dmitribratsun@rambler.ru

Bratsun Dmitri, Doctor of Physics & Mathematics, Prof., Head of Department of Physics at PSPU, Perm

Tel.: (342) 238-63-25

E-mail: dmitribratsun@rambler.ru

Захаров Андрей Павлович, аспирант ПГПУ, г. Пермь

Тел.: (342) 238-63-25

Электронная почта: andrrei1@rambler.ru

Zakharov Andrei, aspirant at PSPU, Perm

Tel.: (342) 238-63-25

E-mail: andrrei1@rambler.ru

СПЕЦПРОЦЕССОРЫ ДЛЯ РЕШЕНИЯ ЗАДАЧ НА ГРАФАХ

SPECIAL PROCESSOR FOR SOLVING THE TASK ON GRAPHS

Рассматриваются вопросы проектирования специализированного аппаратного обеспечения вычислительных систем, ориентированных на решение задач на графах и сетях. Предлагаются аппаратные решения функциональных расширителей для выполнения операций выделения максимальных внутренне устойчивых подграфов и выбора вершин, входящих в путь между вершинами графа.

The problems of designing specialized hardware computing, orientirovannyh to solve problems on graphs and networks. Offers hardware solution of functional extenders to perform allocation of maximum internally stable subgraphs and select vertices contained in the path between the vertices of the graph.

Ключевые слова: спецпроцессоры, графы, решение задач на графах.

Keywords: special processor, graph, solving the task on graphs.

Одной из основных тенденций развития современных вычислительных систем является расширение их инструментальной базы за счет включения средств аппаратной поддержки системного и прикладного математического обеспечения. В частности, весьма интенсивно ведутся разработки специализированного аппаратного обеспечения решения задач, в которых графы и сети выступают в качестве математических моделей для описания достаточно широкого класса прикладных задач. По оценкам специалистов без такого аппаратного обеспечения в ближайшем будущем системы

математического моделирования окажутся практически неработоспособными ввиду несоответствия вычислительной мощности аппаратных платформ требованиям прикладных задач. В связи с этим производители следующих поколений вычислительных систем придают большое значение включению в их состав ускоряющих процессоров, ориентированных на решение задач высокой вычислительной сложности с использованием специфической техники обработки данных. Так, в серии систем схемного моделирования [2] за счет этого подхода удалось повысить скорость моделирования схем в 20-100 раз по сравнению с программной реализацией на VAX-11/780. Однако реализация концепции аппаратной поддержки компонент специализированного математического обеспечения хотя и позволяет достичь более скоростной обработки данных, но ограничивает возможность использования в дальнейшем более прогрессивных методов и алгоритмов из-за сложности модификации вычислительных платформ. В связи с этим более целесообразно включение в состав вычислительных систем наборов функциональных расширителей, обеспечивающих высокоэффективную реализацию ряда операций, которые требуют значительных временных затрат при их программной реализации и часто используются при решении прикладных задач в достаточно широком ряде областей.

Такой подход позволяет повысить скорость решения задач за счет аппаратной реализации базового набора операций с высокой вычислительной сложностью и обеспечивает возможность реализации различных алгоритмических решений. При алгоритмизации многих прикладных задач используется математический аппарат теории графов. Причем ряд операций, например, такие операции как выделение максимально внутренне устойчивых подграфов и выбора вершин, входящих в кратчайшие пути в графе, широко применяются при решении задач из различных прикладных областей. Ниже рассматриваются вопросы реализации специализированных вычислителей для выполнения перечисленных операций.

Максимальным внутренне устойчивым называется такой подграф,

который не является собственным подграфом никакого другого внутренне устойчивого подграфа. Процедура выделения максимальных внутренне устойчивых подграфов (МВУПГ) можно выполнить следующим образом.

1. Первоначально для некоторой вершины X_i графа $G(X, A)$ находится его прямое отображение первого порядка $\Gamma^{+1}(X_i)$, т. е. определяется подмножество вершин, с которыми имеется связь вершины X_i ;

2. Затем определяется подмножество вершин, которые непосредственно не имеют связи с вершиной X_i , т. е. $Y = G \setminus \Gamma^{+1}(X_i)$;

3. Далее вычисляется подмножество вершин графа, которые связаны с элементами подмножества Y , т. е. $Z = \Gamma^{+1}(Y)$, $Z = \Gamma^{+1}[G \setminus \Gamma^{+1}(X_i)]$;

4. Находится объединение подмножеств $\Gamma^{+1}(X_i) \cap Z$ или $\Psi = \Gamma^{+1}(X_i) \cap \Gamma^{+1}[G \setminus \Gamma^{+1}(X_i)]$, которое отражает подмножество связанных элементов графа;

5. В итоге выполняется формирование максимального подмножества несвязанных между собой элементов:

$$W = G \setminus \Psi \text{ или } W = G \setminus \{\Gamma^{+1}(X_i) \cup \Gamma^{+1}[G \setminus \Gamma^{+1}(X_i)]\}.$$

Организация специализированного вычислителя, обеспечивающего реализацию рассмотренной операции, приведена на рис. 1. Основными узлами устройства являются следующие.

- Блок матричной памяти (МП), реализованный на $(p \times p)$ триггерах с выходными вентильными схемами, используемый для хранения исходной матрицы смежности обрабатываемого графа;
- Вычислительный блок – решающее поле (РП), состоящее из p специализированных элементарных процессоров (СЭП), каждый из которых обеспечивает выполнение операции логического сложения и хранения результатов вычислений;
- Коммутационное поле (КП), в котором выполняется анализ подмножеств элементов графа на связность и на его основе выполняется формирование на выходах КП сигналов управления

для доступа к данным матрицы памяти;

- Блок буферной памяти (БП) для хранения результатов преобразований;
- Блок управления – местное устройство управления (МУУ);
- Процедура выделения максимально внутренне устойчивых подграфов (МВУпГ) выполняется в рассматриваемом вычислителе следующим образом;
- На первом этапе КП формирует сигнал считывания на первую горизонтальную шину матрицы памяти. В результате в СЭП РП считываются состояния элементов памяти, то есть значения элементов первой строки матрицы смежности обрабатываемого графа, и выполняется их логическое сложение. Результат этой операции заносится в память СЭП. Таким образом, на данном шаге производится вычисление значения $\Gamma^{+1}(X_i)$;
- На втором этапе на вход КП подается двоичный вектор, нулевые элементы которого соответствуют вершинам подмножества $G \setminus \Gamma^{+1}(X_i)$. Данный вектор является результатом выполненной операции в РП. На выходах КП формируются сигналы для считывания данных в тех позициях, которые соответствуют элементам подмножества Y . Таким образом, на входы СЭП РП поступает результат выполненной операции $Z = \Gamma^{+1}(Y)$ и формируется значение $\Psi = \Gamma^{+1}(X_i) \cap \Gamma^{+1}[G \setminus \Gamma^{+1}(X_i)]$, соответствующее третьему и четвертому шагам приведенного выше алгоритма. Далее описанный процесс продолжается до полного исключения из формируемого максимального внутренне устойчивого подмножества всех вершин являющихся инцидентными по отношению к первой вершин. По завершению этого процесса МУУ вырабатывает сигнал записи для занесения в БП из памяти СЭП выделенного подграфа;
- На третьем этапе коммутируется следующая горизонтальная шина

управления считыванием данных, соответствующая строке, содержащей элементы следующего подграфа для вершины X_2 . На последующих этапах процесс анализа графа повторяется аналогично для всех N элементов графа $G(X, A)$;

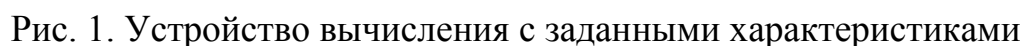
- После завершения работы устройства в БП будет записана информация о выделенных максимальных внутренне устойчивых подмножествах вершин, содержащих все вершины графа.

На базе рассмотренного устройства за счет расширения его функциональных возможностей может быть реализована также и операция выбора вершин графа, входящих в кратчайший путь.

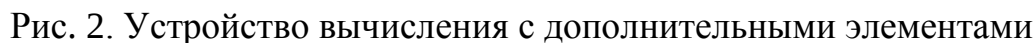
Задача нахождения кратчайшего пути между начальной и конечной вершинами X_n и X_k в графе $G=(X, A)$ состоит в выделении такого подграфа $G'=(X', A')$, что $X'=\{X_j\}$, $j=1,2,3 \dots, N'$, $X_n \in X'$, N' – минимально, а элементы этого подграфа $P=\{X_i\}$, $i=1, 2, 3, \dots, N'$, представляют собой множество вершин кратчайшего пути между $X_i = X_n$ и $X_{N'} = X_k$.

На основе операции транзитивного замыкания эта задача решается следующим образом.

На первом шаге для вершины X_n находится прямое транзитивное замыкание $\Gamma^+(X_n) = X_n \cup \Gamma^1(X_n) \cup \Gamma^2(X_n) \dots \cup \Gamma^{+1}(X_n)$ до тех пор, пока не будет в него включена искомая вершина X_k . Если пересечение $\Gamma^+(X_n) \cap X_k \neq \emptyset$, то это означает, что требуемый подграф G' дает множество вершин, входящих в кратчайший путь невзвешенного графа.



При решении задачи нахождения кратчайших путей в графе в память БП, реализованную на сдвиговых регистрах, записывается информация о вершине X_n . Информация о конечной вершине заносится в соответствующий разряд П.



На выходе матрицы памяти формируются единичные сигналы, соответствующие вершинам, имеющим ребра с вершиной X_n , т. е. определяется $\Gamma^{+1}(X_n)$. Далее выполняется сдвиг информации в БП и запись в первые разряды ее регистра результата из РП, т. е. $\Gamma^{+1}(X_n)$.

На втором шаге по матрице смежности исследуемого графа, хранящейся в матрице памяти, находится значение $\Gamma^{+2}(X_H) = \Gamma^+(\Gamma^{+1}(X_H))$. Одновременно с этим на каждом шаге выделяемое множество $\Gamma^{+i}(X_H)$ анализируется на достижение вершины X_K .

Если $X_k \notin T^{+i}(X_n)$, то процесс выделения следующего отображения $\Gamma^+(X_n)$ продолжается. Если вершина X_k достигнута, то выполняется переключение работы устройства: информация о вершине X_k записывается в некоторый $(N+1)$ -й разряд соответствующего регистра БП.

В КП коммутируются потоки данных, поступающих с единичных выходов $(N+1)$ -х разрядов регистров БП. Таким образом, через l тактов будут определены все вершины кратчайшего пути из X_n в X_k графа G . После окончания вычислений результат формируется в регистрах памяти БП. Количество вершин графа, составляющих кратчайший путь из вершины X_n в вершину X_k , соответственно также будет равно l . Реализация рассмотренных устройств на функциональном уровне приведена в [4].

Литература

1. Быстродействующая система схемного моделирования с высокими точностными показателями // Электроника. – 1985. – № 13. – С. 90-91.
2. Мануэль Т. Расширение возможностей АРМ за счет параллельно работающих процессоров // Электроника. – 1985. – № 13. – С. 103-104.
3. Нейгл Т. Процессор-акселератор с быстродействием 10 млн команд для САПР/САИТ // Электроника. – 1986. – № 14. – С. 73-74.
4. А. с. 1363237 СССР. Устройство для исследования графов / Т. В. Волченская, В. С. Князьков и др. – Опул. БИ, 1987. – № 48.

Информация об авторах

Data on authors

Волченская Тамара Викторовна, к.т.н., доцент, доцент кафедры «Информатики и вычислительная техника» НОУ ВПО «Вятский социально-экономический институт»

Тел. (8332) 67-02-35

Электронная почта: volchenskaya@list.ru

Volchenskaya Tamara V., Ph.D., Associate Professor of the Department
«Computer science and computer engineering» LEU Institution «Vyatka economic
and social institution»

Tel. (8332) 67-02-35

Email: volchenskaya@list.ru

Князьков Владимир Сергеевич, доктор технических наук, профессор,
руководитель Научно-образовательного центра «Супервычислительные
технологии и системы», профессор кафедры «Электронные вычислительные
машины», ФГБОУ ВПО «ВятГУ».

Тел.: 8(912)7343173

Электронная почта: kniazkov@list.ru

Knyazkov Vladimir, Doctor of Technical Science, Prof., Head of Science &
Education Centre «High Power Computer Technology & Systems», Prof. of
Department «Computer Engineering», FGEI HPE «VyatSU».

Tel.: 8(912)7343173

E-mail: kniazkov@list.ru

Вржещ В. П., Масютин А. А.

Vrzheshch V. P., Masyutin A. A.

МОДЕЛЬНОЕ ДЕЗАГРЕГИРОВАНИЕ МАКРОЭКОНОМИЧЕСКОЙ СТАТИСТИКИ НА ПРИМЕРЕ РОССИИ, УКРАИНЫ И ФИНЛЯНДИИ*

NONLINEAR NATIONAL ACCOUNTS DATA DESAGREGATION: RUSSIA, UKRAINE AND FINLAND

Рассматривается трехпродуктовое разложение макроэкономической статистики. За основу взят макроэкономический баланс по использованию, который разворачивается в балансы трех модельных продуктов: экспортного, импортного и внутреннего. Предлагается упрощенная модель общего равновесия, в которой потребление и валовое накопление представляются как CES-функции полезности от потоков импортного и внутреннего продуктов, а ВВП – как CES-функция замещения от объемов производства внутреннего и экспортного продуктов. Показано успешное применение этого подхода для статистики России, Украины и Финляндии.

In this paper three-product desagregation is applied to Russia, Ukraine and Finland quarter national accounts data. The object of examination is national accounts in division by five consumption categories. These accounts are desagregated into three products: export, import and interior product. Constant elasticity of substitution production functions is applied. Unobserved (calculated) series of national accounts are revealed.

* Работа выполнена при поддержке РФФИ проект № 11-01-00644, РГНФ проект № 11-02-00241а, ПФИ ОМН РАН №3, проект 3.14, ПФИ Президиум РАН №14, проект 109.

Ключевые слова: трехпродуктовое разложение макроэкономической статистики, макроэкономический баланс, модель общего равновесия, CES-функция полезности, CES-функция замещения.

Key words: three-product desegregation of national accounts data, balance of macro-economics, model of CGE, CES-function.

1. Потребность в трехпродуктовом описании экономики

Цель работы – предложить модель, которая за счет введения некоторых внутримодельных **ненаблюдаемых** величин объясняет наблюдаемую динамику основных макроэкономических показателей, как в текущих ценах, так и в реальном выражении [2]. Будем исходить из основного макроэкономического баланса по использованию (ОМБ) **в конечных ценах**, который имеет вид

$$ВВП = \text{Валовое накопление} + \text{Конечное потребление} + \text{Экспорт} - \text{Импорт}. \quad (0.1)$$

В выражении (0.1) левая и правая части собираются статистическими службами независимо. ВВП собирается как суммарная добавленная стоимость произведенных товаров и услуг в рыночных ценах всех резидентов страны за определенный период и равен конечному выпуску по конечным ценам за вычетом промежуточных затрат. В правой части (0.1) учитываются товары и услуги также по конечным ценам отдельно по каждой компоненте: **конечное потребление, валовое накопление, экспорт и импорт**.

Статистические службы собирают первичную информацию по ОМБ в деньгах (в текущих ценах), а затем ОМБ приводится к «реальным величинам» (неизменным ценам). ОМБ в неизменных ценах мы будем записывать как

$$Y(t) = J(t) + C(t) + E(t) - I(t), \quad (0.2)$$

где $Y(t)$ – реальный ВВП, $J(t)$ – реальное валовое накопление, $C(t)$ – реальное конечное потребление, $E(t)$ – реальный экспорт, $I(t)$ – реальный импорт.

Тогда исходный баланс (0.1) в текущих ценах можно записать как

$$p_Y(t)Y(t) = p_J(t)J(t) + p_C(t)C(t) + p_E(t)E(t) - p_I(t)I(t), \quad (0.3)$$

где $p_Y(t)$, $p_J(t)$, $p_C(t)$, $p_E(t)$, $p_I(t)$ – базовые **дефляторы** ВВП, валового накопления, конечного потребления, экспорта и импорта соответственно. Все дефляторы измеряются в ценах базового периода (года). Их несглаженная квартальная статистика с базовым 2003-им годом [6] показана на рис. 1.

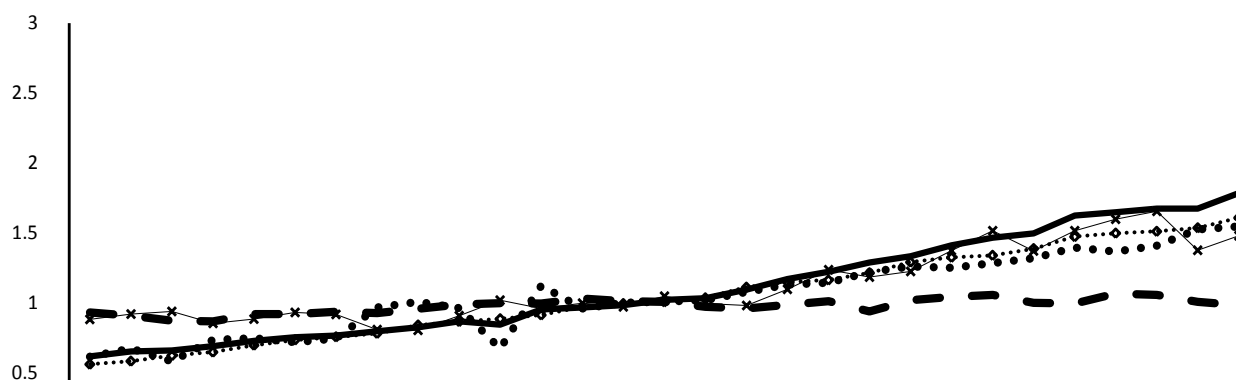


Рис. 1 .Дефлятор ВВП (P_Y), дефлятор конечного потребления (P_C), дефлятор валового накопления (P_J), дефлятор экспорта (P_E), дефлятор импорта (P_I).

В однопродуктовых моделях баланс (0.2) рассматривается как условие равновесия на рынке однородного агрегированного продукта. В модели равновесия одному продукту (*аддитивному*) соответствует одна цена (дефлятор ВВП), поэтому однопродуктовая модель может одновременно воспроизвести реальные (0.2) и номинальные величины с точностью, не превосходящей различия в индексах цен составляющих ОМБ. В настоящее время такая точность становится уже явно недостаточной (см. рис. 1), что и послужило основным стимулом к проведению настоящего исследования.

Использовать обычный путь построения более подробной модели на основе данных межотраслевого баланса (МОБ) в данном случае не представляется возможным по причине отсутствия квартальных публикаций по МОБ. Модель экономики России, для которой преимущественно проводились эти исследования [1], строится в квартальном разрезе. Таким образом, мы решили пойти путем дезагрегирования ОМБ, используя, прежде всего,

зафиксированную статистикой разницу индексов цен его составляющих (см. рис. 1).

2. Нелинейное дезагрегирование макроэкономического баланса

Детальное пояснение логики наших рассуждений относительно агрегированных балансов всех товаров и услуг приводятся в [2]. Здесь мы просто опишем агрегаты, участвующие в исследовании.

Главная идея работы состоит в том, чтобы от модельного баланса одного продукта (0.2) перейти к естественным с экономической точки зрения модельным балансам трех продуктов экспортного E , импортного I и внутреннего X , хотя последний и оказывается ненаблюдаемым в статистике.

Будем считать, что использование в качестве сырья и комплектующих импортных продуктов (стройматериалов, оборудования, деталей, иностранной рабочей силы, лицензий и патентов на технологии, и т. п.) диктуется не столько технологическими ограничениями, сколько требованиями потребителей и заказчиков. Поэтому мы делим внутренний продукт X и импортный продукт I

$$X = J_x + C_x, \quad (0.4)$$

$$I = J_I + C_I, \quad (0.5)$$

где J_I – импорт, связанный с инвестициями, C_I – импорт, связанный с потреблением, J_x – внутреннее производство, связанное с инвестициями, и C_x – внутреннее производство, связанное с потреблением.

В рамках модели мы предполагаем, что величины C_I , C_x определяются рыночными механизмами с учетом предпочтений потребителя. Если предпочтения потребителя можно описать *линейно-однородной функцией полезности* $g(C_I, C_x)$, то эта функция определяется однозначно с точностью до множителя и может рассматриваться как агрегированный индекс объема потребления набора продуктов $\langle C_I, C_x \rangle$, [5, 7]. Именно так мы попытаемся смоделировать наблюдаемый объем потребления в (0.2)

$$C = g(C_I, C_X). \quad (0.6)$$

Аналогично *линейно-однородная функция полезности для инвестора* $h(J_I, J_X)$ определяет в модели валовое накопление в (0.2)

$$J = h(J_I, J_X). \quad (0.7)$$

Следует заметить, что ожидать простых равенств $C = C_I + C_X$ или $J = J_I + J_X$ не приходится, поскольку введенные нами модельные потоки J_I , C_I , J_X , C_X включают не только собственно потребление и инвестиции, но и какие-то потоки текущих затрат, сделанные по требованию потребителя или инвестора.

Второе наше предположение состоит в том, что неучтенные выше факторы производства: труд, основной капитал и природные ресурсы – определяют соотношением типа производственной функции именно величину реального ВВП, но реализована она может быть по-разному за счет выпуска внутреннего и экспортного продуктов. Возможности замещения при производстве описываются еще одной линейно-однородной функцией свертки

$$Y = f(X, E), \quad (0.8)$$

соотношение $Y = X + E$, опять-таки невозможно, поскольку X включает промежуточный продукт, а Y – нет.

Соотношения (0.2), (0.4)-(0.8), дополненные обсуждаемыми ниже соображениями рациональности поведения экономических агентов (см. разд. 3), задают предлагаемую нами схему дезагрегирования ОМБ в модели.

Следует заметить, что приведенные выше рассуждения про детальные балансы можно считать просто наводящими соображениями, а соотношения (0.4)-(0.8) рассматривать как чисто феноменологическую схему, практически однозначно определяемую требованием описать составляющие макроэкономического баланса как нелинейные свертки потоков минимального числа аддитивных благ, при выполнении условий рациональности (см. разд. 3).

Чтобы использовать для идентификации все ряды мы обратимся к анализу финансовых потоков. Соотношения (0.6)-(0.8) определяют индексы физического объема потребления, накопления и ВВП. Требование их **линейной однородности** обеспечивает сложение финансовых потоков при нелинейном свертывании реальных потоков. Пять соотношений (0.4)-(0.8) содержат как раз пять квартальных неизвестных временных рядов в ценах 2003 года: $X(t)$ – выпуск внутреннего продукта, $C_x(t)$ – использование внутреннего продукта в связи с потреблением, $J_x(t)$ – использование внутреннего продукта в связи с накоплением, $C_I(t)$ – использование импортного продукта в связи с потреблением, $J_I(t)$ – использование импортного продукта в связи с накоплением.

Поэтому соотношений (0.4)-(0.8) недостаточно, чтобы определить вид функций $f(\cdot, \cdot)$, $h(\cdot, \cdot)$, $g(\cdot, \cdot)$. Однако, у нас есть еще информация о финансовых потоках и дефляторах. Благодаря этой информации мы можем надеяться выявить элементы рационального поведения агентов, лежащие в основе имеющейся статистики. Чтобы перейти от натуральных величин к исследованию рационального поведения, мы будем рассматривать три дефлятора: $p_E(t)$, $p_I(t)$, $p_X(t)$ – первые два из которых наблюдаемы в статистике (0.3), а последний должен быть восстановлен в процессе дезагрегирования. Индекс цен $p_X(t)$ также должен быть нормирован в 2003 году.

3. Рациональность поведения экономических агентов

Чтобы использовать информацию о дефляторах сделаем некоторые предположения о рациональности поведения экономических агентов. Будем условно считать, что в экономике действуют 5 агентов (см. рис. 2):

три основных

- **Потребитель** покупает продукты C_I , C_x по ценам p_I и p_x соответственно;
- **Инвестор** покупает продукты J_I , J_x по ценам p_I и p_x соответственно;
- **Производитель** производит продукты X , E и продает их на внутреннем рынке и экспортеру соответственно по ценам p_x и p_E соответственно;

и два вспомогательных агента

- **Импортёр** продает на внутреннем рынке импортный продукт I по цене p_I ;
- **Экспортёр** покупает на внутреннем рынке продукт для экспорта E по цене p_E .

Балансы (0.4), (0.5) будут рассматриваться как условия равновесия на конкурентных рынках однородных аддитивных благ: внутреннего и импортного продукта, соответственно. На таких рынках все участники продают и покупают благо по одной и той же цене.

Для составляющих импортного продукта в балансе (0.5) в качестве единой цены будем использовать известный из статистики дефлятор p_I . Для составляющих чисто модельного внутреннего продукта введем модельную (ненаблюдаемую) цену-дефлятор p_X . Поскольку дефлятор экспорта p_E , согласно статистике, отличается от всех остальных (см. рис. 1), будем считать, что экспортный продукт продается на особом (внешнем) рынке по цене p_E .



Рис. 2. Схема взаимодействия экономических агентов
в трехпродуктовом дезагрегировании

Теперь опишем, как формируются спрос и предложение на перечисленных выше рынках. Начнем с Потребителя. Купив в период t продукты $C_I(t)$, $C_X(t)$, Потребитель понес денежные расходы $\Phi(t) = p_I(t)C_I(t) + p_X(t)C_X(t)$. Эта величина известна нам из статистики: $\Phi(t) = p_C(t)C(t)$ (0.3). В предположении (0.6) получаем, следовательно,

$$p_C(t)g(C_I(t), C_X(t)) = p_I(t)C_I(t) + p_X(t)C_X(t). \quad (0.9)$$

А почему был куплен набор продуктов $C_I(t)$, $C_X(t)$, а не какой-то другой? Простейший ответ состоит в том, что при другом наборе продуктов и данных ценах соотношение (0.6) несовместимо с бюджетным ограничением

$$p_C(t)g(C_I, C_X) \leq p_I(t)C_I + p_X(t)C_X, \quad \text{при} \quad \text{всех} \quad C_I, C_X \geq 0. \quad (0.10)$$

Из соотношений (0.9), (0.10) следует, что значения $C_I(t)$, $C_X(t)$ доставляют максимум выражению

$$p_C(t)g(C_I, C_X) - p_I(t)C_I - p_X(t)C_X \rightarrow \max_{C_I, C_X \geq 0} \quad (0.11)$$

Предполагая, что функция $g(\cdot, \cdot)$ – гладкая и выпуклая вверх, а также учитывая, что нас интересуют только внутренние точки максимума, получаем необходимые и достаточные условия выполнения² (0.11)

$$\partial_1 g(C_I(t), C_X(t)) = \frac{p_I(t)}{p_C(t)}, \quad \partial_2 g(C_I(t), C_X(t)) = \frac{p_X(t)}{p_C(t)}. \quad (0.12)$$

Однако, поскольку функция $g(\cdot, \cdot)$ – линейно-однородна, задача максимизации (0.11) вырождена. Она имеет нетривиальное решение не при всех соотношениях цен, а, когда решение есть, оно не единственно. При этом в точке максимума с необходимостью выполняется равенство (0.9). Алгебраически это проявляется в том, что частные производные $g(\cdot, \cdot)$ суть функции отношения аргументов. Поэтому уравнения (0.12) не определяют значения $C_I(t)$, $C_X(t)$ по отдельности. Из этих уравнений определяется отношение $C_I(t)/C_X(t)$ и цена $p_X(t)$. Из соображений симметрии удобнее использовать отношение равенств (0.12)

² Символом ∂_i здесь и далее обозначается частная производная по i -му аргументу.

$$\frac{\partial_1 g \ C_I(t), C_X(t)}{\partial_2 g \ C_I(t), C_X(t)} = \frac{p_I(t)}{p_X(t)} \quad (0.13)$$

и условие (0.9) в качестве строгого финансового ограничения.

Аналогичные рассуждения применимы и к агенту Инвестор. Соображения рациональности снова приводят к равенству отношения цен отношению предельных полезностей, аналогичному (0.13), которое для однородной функции $h(\cdot, \cdot)$ приводит к равенству аналогичному (0.9). В результате получаем соотношения

$$\frac{\partial_1 h \ J_I(t), J_X(t)}{\partial_2 h \ J_I(t), J_X(t)} = \frac{p_I(t)}{p_X(t)}, \quad p_J(t) h \ J_I(t), J_X(t) = p_I(t) J_I(t) + p_X(t) J_X(t). \quad (0.14)$$

Величина $\Pi(t) = p_X(t)X(t) + p_E(t)E(t)$ представляет собой выручку производителя. В статистике она выражается как $\Pi(t) = p_Y(t)Y(t)$. Реализующийся набор выпускаемых продуктов $X(t)$, $E(t)$ можно снова выделить среди всевозможных пар величин $X, E \geq 0$ тем, что остальные сочетания несовместимы с (0.8) при данных ценах. Для однородной функции $f(\cdot, \cdot)$ получаются соотношения

$$\frac{\partial_1 f \ X(t), E(t)}{\partial_2 f \ X(t), E(t)} = \frac{p_X(t)}{p_E(t)}, \quad p_Y(t) f \ X(t), E(t) = p_X(t)X(t) + p_E(t)E(t). \quad (0.15)$$

4. Система условий дезагрегирования

Итак, мы получаем следующую систему из 11 гипотетических соотношений, которые должны выполняться в каждый период наблюдений:

$$Y(t) = f \ X(t), E(t), \quad J(t) = h \ J_I(t), J_X(t), \quad C(t) = g \ C_I(t), C_X(t), \quad (0.16)$$

$$\frac{\partial_1 f \ X(t), E(t)}{\partial_2 f \ X(t), E(t)} = \frac{p_X(t)}{p_E(t)}, \quad \frac{\partial_1 h \ J_I(t), J_X(t)}{\partial_2 h \ J_I(t), J_X(t)} = \frac{p_I(t)}{p_X(t)}, \quad \frac{\partial_1 g \ C_I(t), C_X(t)}{\partial_2 g \ C_I(t), C_X(t)} = \frac{p_I(t)}{p_X(t)}, \quad (0.17)$$

$$p_Y(t)Y(t) = p_X(t)X(t) + p_E(t)E(t), \quad p_J(t)J(t) = p_I(t)J_I(t) + p_X(t)J_X(t), \quad (0.18)$$

$$p_C(t)C(t) = p_I(t)C_I(t) + p_X(t)C_X(t), \quad (0.19)$$

$$X(t) = J_X(t) + C_X(t), \quad I(t) = J_I(t) + C_I(t), \quad (0.20)$$

где функции $f(\cdot, \cdot)$, $g(\cdot, \cdot)$, $h(\cdot, \cdot)$ предполагаются гладкими и линейно-однородными, так что по теореме Эйлера

$$\begin{aligned} f(X, E) &= X \partial_1 f(X, E) + E \partial_2 f(X, E), \\ g(C_I, C_X) &= C_I \partial_1 g(C_I, C_X) + C_X \partial_2 g(C_I, C_X), \\ h(J_I, J_X) &= J_I \partial_1 h(J_I, J_X) + J_X \partial_2 h(J_I, J_X). \end{aligned} \quad (0.21)$$

Кроме того, исходные данные (после вычитания из ВВП статистических расхождений) удовлетворяют балансам (0.2), (0.3). Впрочем, баланс (0.3) оказывается следствием (0.18)–(0.20).

В (0.16)–(0.20) 10 рядов:

$$Y(t), J(t), C(t), E(t), I(t), p_Y(t), p_J(t), p_C(t), p_E(t), p_I(t) \quad (0.22)$$

известны из статистики, поэтому у нас остается 6 неизвестных временных рядов

$$X(t), C_X(t), J_X(t), C_I(t), J_I(t), p_X(t). \quad (0.23)$$

Таким образом, для проверки выполнения (0.16)–(0.20) на практике требуется подобрать не изменяющиеся со временем функции $f(\cdot, \cdot)$, $g(\cdot, \cdot)$, $h(\cdot, \cdot)$ так, чтобы за счет шести рядов (0.23) все 11 соотношений (0.16)–(0.20) выполнялись с приемлемой точностью во все периоды наблюдения.

5. Параметризация дезагрегирующих функций и постановка задачи их идентификации

Самый простой вид функций, для которого была проведена идентификация, это так называемые функций с постоянной эластичностью

замещения (CES – Constant Elasticity of Substitution) $u(X,Y) = A [\sigma X^e + (1-\sigma)Y^e]^{1/e}$,
где $A > 0$, $0 < \sigma < 1$, $e \neq 0$ – постоянные параметры. Итак, положим в (0.16)

$$\begin{aligned} f(X,E) &= A_f [\sigma_f X^{e_f} + (1-\sigma_f)E^{e_f}]^{1/e_f}, \quad h(J_I, J_X) = A_h [\sigma_h J_I^{e_h} + (1-\sigma_h)J_X^{e_h}]^{1/e_h}, \\ g(C_I, C_X) &= A_g [\sigma_g C_I^{e_g} + (1-\sigma_g)C_X^{e_g}]^{1/e_g}. \end{aligned} \quad (0.24)$$

Теперь получаем задачу идентификации, которую неформально можно сформулировать следующим образом: подобрать коэффициенты A_f , σ_f , e_f , A_h , σ_h , e_h , A_g , σ_g , e_g так, чтобы при некоторых рядах (0.23) по возможности точно выполнялась переопределенная система соотношений (0.16)–(0.20). Для корректной постановки этой задачи надо определить, что значит «по возможности точно».

Будем считать, что финансовые и материальные балансы (0.18)–(0.20) должны выполняться действительно точно, а в условиях агрегирования (0.16) и условиях рациональности поведения (0.17) допустим отклонения. Таким образом, мы заменяем равенства (0.16) на

$$\ln Y(t) = \ln A_f + \frac{1}{e_f} \ln [\sigma_f X^{e_f} + (1-\sigma_f)E^{e_f}] + \varepsilon_1(t), \quad (0.25)$$

$$\ln J(t) = \ln A_h + \frac{1}{e_h} \ln [\sigma_h J_I^{e_h} + (1-\sigma_h)J_X^{e_h}] + \varepsilon_2(t),$$

$$\ln C(t) = \ln A_g + \frac{1}{e_g} \ln [\sigma_g C_I^{e_g} + (1-\sigma_g)C_X^{e_g}] + \varepsilon_3(t), \quad (0.26)$$

а равенства (0.17) – на

$$\begin{aligned} \frac{\sigma_f}{1-\sigma_f} \left(\frac{X(t)}{E(t)} \right)^{e_f-1} - \frac{p_X(t)}{p_E(t)} &= \varepsilon_4(t), \quad \frac{\sigma_h}{1-\sigma_h} \left(\frac{J_I(t)}{J_X(t)} \right)^{e_h-1} - \frac{p_I(t)}{p_X(t)} = \varepsilon_5(t), \\ \frac{\sigma_g}{1-\sigma_g} \left(\frac{C_I(t)}{C_X(t)} \right)^{e_g-1} - \frac{p_I(t)}{p_X(t)} &= \varepsilon_6(t) \end{aligned} \quad (0.27)$$

и будем минимизировать среднеквадратичную погрешность

$$\sum_{t=1}^T \sum_{i=1}^6 \varepsilon_i(t)^2 \rightarrow \min, \quad \text{где } T \text{ — число наблюдений} \quad (0.28)$$

По $X(t)$, $C_X(t)$, $J_X(t)$, $C_I(t)$, $J_I(t)$, $p_X(t)$, a_f , b_f , a_h , b_h , a_g , $b_g > 0$, e_f , e_h , e_g , $t=1...T$ при ограничениях (0.18)–(0.20) и (0.25)–(0.27). Ограничения-равенства (0.18)–(0.20) в задаче (0.28) позволяют исключить часть переменных и свести задачу с $(9+6T)$ переменными к задаче с 9 неизвестными параметрами производственных функций (0.24).

6. Использованные данные и результаты идентификации

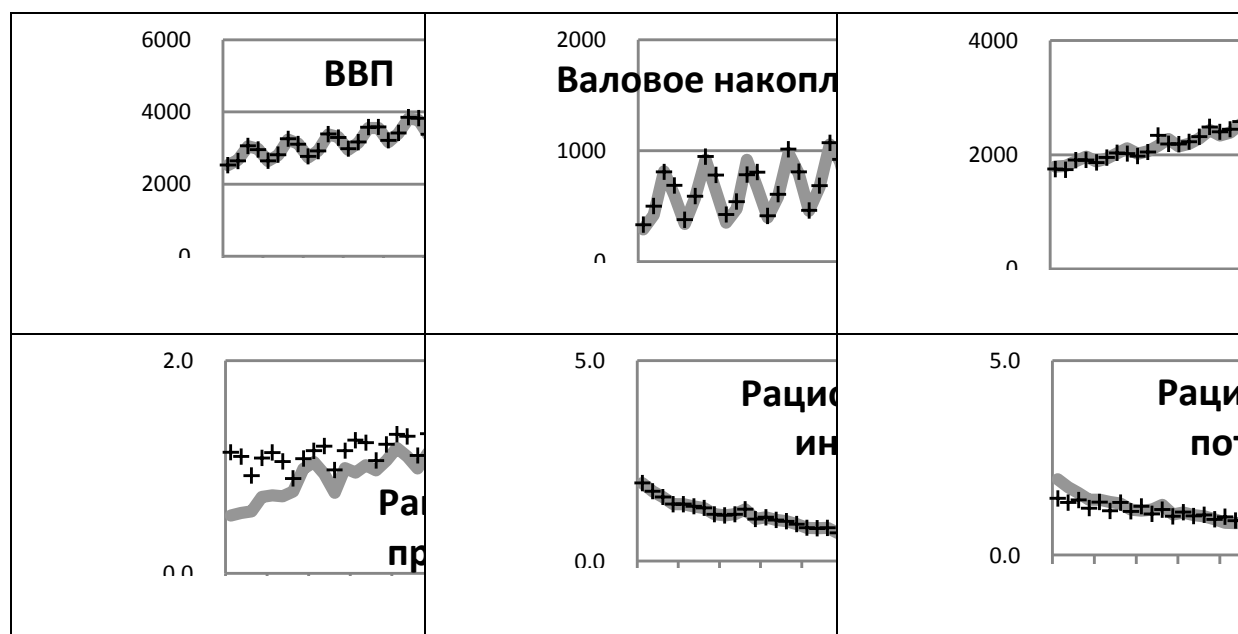
Для идентификации использовались несглаженные квартальные данные по основному макроэкономическому балансу в текущих рыночных ценах (номинальные показатели) (0.3) и в рыночных среднегодовых ценах 2003 г. (реальные показатели) (0.2) за период с I квартала 2000 г. по IV квартал 2010 г. [6] (всего 44 наблюдения). Эти данные позволили получить прямым вычислением ряды (0.23) и параметры производственных функций (0.24).

Вычисленные параметры функций (0.24) приведены в (0.29). Среднеквадратичное отклонение (среднее $\varepsilon_i^2(t)$) равно 0.023. Выполнение условий (0.16)–(0.17) при условии (0.18)–(0.20) показано в табл. **Ошибка! Источник ссылки не найден.** Всего условия (0.16)–(0.17) содержат 264 уравнения, которые неплохо выполняются за счет подбора девяти параметров.

$$\begin{aligned} f(X, E) &= 1.88 \cdot 68X^{.05} + .32E^{.05 \cdot 1/.05}, \quad h(J_I, J_C) = 1.75 \cdot 22J_I^{.005} + .78J_C^{.005 \cdot 1/.005}, \\ g(C_I, C_X) &= 1.85 \cdot 28C_I^{.003} + .72C_X^{.003 \cdot 1/.003}. \end{aligned} \quad (0.29)$$

Таблица 1.

Выполнение нелинейных сверток (0.16) (серая линия – статистика, черный маркер-крестик – расчет) и условий рациональности (0.17) (серая линия – отношения цен, черный маркер-крестик – отношение производных)



7. Введение пошлин и промежуточного продукта

Обратимся к вопросу влияния налогов и пошлин на результаты дезагрегирования. Дефляторы импорта и экспорта, представленные в статистике [6], содержат налоги и пошлины. В нашей постановке Производитель продает Экспортеру продукт до уплаты налогов. Таким образом, нам необходимо учесть влияние налогов и пошлин в экспорте. В отношении внутренней цены нам не приходится это делать, так как мы не берем в её основе статистику.

Итак, мы решаем задачу (0.28) при ограничениях (0.18)–(0.20) и (0.25)–(0.27), в которых первое балансовое уравнение в (0.18) и первое условие рациональности в (0.27) заменяются на:

$$p_Y(t)Y(t) = p_X(t)X(t) + \frac{1 - es(t)}{1 - nv} p_E(t)E(t), \quad (0.30)$$

$$\frac{\sigma_f}{1-\sigma_f} \left(\frac{X(t)}{E(t)} \right)^{e_f-1} - \frac{p_X(t)}{p_E(t)} \frac{1-nv}{1-es(t)} = \varepsilon_4(t), \quad (0.31)$$

где nv – ставка НДС (18%), $es(t)$ – агрегированная ставка вывозной пошлины, рассчитанная как сумма поступлений в консолидированный бюджет РФ за счет вывозных пошлин, деленная на объем экспорта. Результаты расчетов незначительно лучше результатов, показанных в табл. 1.

Для того чтобы исключить промежуточное потребление из нелинейных сверток потребления и накопления (см. разд. 3) необходимо ввести его в явном виде, а для удобства Производитель возьмет на себя функции Экспортера. В России порядка половины всего произведенного продукта приходится на промежуточный продукт. Для введения промежуточного потребления при производстве модель разд. 3 существенно меняется. Результаты расчетов также незначительно лучше результатов, показанных в табл. 1.

8. Применение для других стран

Аналогичное исследование без пошлин, налогов и промежуточного продукта (см. разд. 3) было проведено для статистики по ОМБ Финляндии за 1990-2010 гг. [4] (0.32) и Украины за 2002-2007 гг. [3] (0.33). В обоих случаях замечательно воспроизводятся натуральные величины ВВП (0.8), конечного потребления (0.6) и валового накопления (0.7), и несколько хуже выполняются условия рациональности (0.17). Аналогичные попытки в отношении статистических данных республики Казахстан не увенчались успехом.

$$\begin{aligned} f(X, E) &= 1.91 \cdot 0.59 X^{0.65} + 0.41 E^{0.65 \cdot 1/0.65}, \quad h(J_I, J_C) = 1.89 \cdot 0.39 J_I^{0.52} + 0.61 J_C^{0.52 \cdot 1/0.52}, \\ g(C_I, C_X) &= 2.0 \cdot 0.52 C_I^{1.08} + 0.48 C_X^{1.08 \cdot 1/1.08}. \end{aligned} \quad (0.32)$$

$$\begin{aligned} f(X, E) &= 2.0 \cdot 0.52 X^{0.56} + 0.48 E^{0.56 \cdot 1/0.56}, \quad h(J_I, J_C) = 2.14 \cdot 0.39 J_I^{1.05} + 0.61 J_C^{1.05 \cdot 1/1.05}, \\ g(C_I, C_X) &= 2.0 \cdot 0.52 C_I^{0.74} + 0.48 C_X^{0.74 \cdot 1/0.74}. \end{aligned} \quad (0.33)$$

Литература

1. Андреев М. Ю., Поспелов И. Г., Поспелова И. И., Хохлов М. А. Новая технология моделирования экономики и модель современной экономики России. М.: МИФИ, 2007.
2. Вржещ В. П., Поспелов И. Г., Хохлов М. А. Модельное дезагрегирование макроэкономической статистики: «Экономический журнал Высшей школы экономики», Том 14, №1, 2010. – С. 88-104.
3. Государственный статистический комитет Украины (<http://www.ukrstat.gov.ua/>)
4. Государственный статистический комитет Финляндии (<http://www.stat.fi/>)
5. Петров А. А., Поспелов И. Г., Шананин А. А. Опыт математического моделирования экономики. М.: Энергоатомиздат, 1996.
6. Федеральная служба государственной статистики. (<http://www.gks.ru>)
7. Шананин А. А. Проблема интегрируемости и обобщенный непараметрический метод анализа потребительского спроса // Труды МФТИ. Т. 4. 2009. – С. 84-96.

Информация об авторах

Data on authors

Вржещ Валентин Павлович, к.ф.-м.н., Московский государственный университет им. М.В. Ломоносова, факультет Вычислительной математики и кибернетики, кафедра исследования операций

Тел.: (495)939-24-91

Электронная почта: valentin.vrzheshch@gmail.com

Vrzheshch Valentin, Cand. of Physics & Mathematics, Moscow State University Lomonosov, Department of Research Operations

Tel.: (495)939-24-91

E-mail: valentin.vrzheshch@gmail.com

Масютин Алексей Александрович, студент, Национальный исследовательский университет – Высшая школа экономики, факультет Экономики, кафедра математической экономики и эконометрики

Тел - -

Электронная почта: alexey.masyutin@gmail.com

Masyutin Alexey, student, NRU HSE, Department of Econometrics

Tel - -

E-mail: alexey.masyutin@gmail.com

О НОВЫХ ДОСТАТОЧНЫХ УСЛОВИЯХ СУЩЕСТВОВАНИЯ РАВНОВЕСИЯ МАКРОСИСТЕМЫ*

NEW SUFFICIENT CONDITIONS FOR EXISTENCE OF EQUILIBRIUM FOR MACROSYSTEM

В работе описывается способ использования аппарата стохастической химической кинетики по отношению к изучению макросистем, возникающих в экономике, социологии и теории транспортных потоков. Ключевым понятием является равновесие макросистемы, которое определяется как наиболее вероятное макросостояние стационарной (инвариантной) меры, эргодической (марковской) динамики, порождающей рассматриваемую макросистему. В конце статьи приводятся, по-видимому, новые, довольно общие условия существования равновесия.

In this work we lead a general scheme (goes back to the works of M.A. Leontovich, 1935) describes a chemical kinetic stochastic approach to the macrosystems arise in economic, sociology and traffic flow theory. The main line is a definition of equilibrium of macrosystem as most probable macrostate of invariant measure of Markov dynamic (corresponds to the macrosystem). We demonstrate new

* Работа выполнена при поддержке грантов РФФИ 11-01-00494-а, 12-01-33007 мол-а-вед; гранта Президента РФ № МК-5285.2013.9.

results (under more general conditions) for well known models of migration of people. We propose a rather general sufficient condition for existence of equilibrium.

Ключевые слова: равновесие макросистемы, эргодический марковский процесс, энтропия, принцип детального равновесия, условие унитарности.

Keywords: macrosystem, equilibrium, ergodic theorem, invariant (stationary) measure, Markov chain Monte Carlo revolution, entropy, principle of detailed balance.

1. Предварительные сведения из теории марковских процессов

Напомним известные, необходимые нам в дальнейшем, факты из теории марковских процессов [1]. Пусть некоторая система может находиться в одном из n состояний. В начальный момент времени положение системы характеризуется вектором распределения вероятностей:

$$\vec{p}^T(0) = (p_1(0), \dots, p_k(0), \dots, p_n(0)), \quad \sum_{k=1}^n p_k(0) = 1,$$

где $p_k(0) \geq 0$ – вероятность того, что система находилась в момент времени $t = 0$ в состоянии с номером k .

Пусть система эволюционирует во времени согласно *матрице переходных вероятностей* $P = \|p_{ij}\|_{i,j=1,1}^{n,n}$. При этом $0 \leq p_{ij} \leq 1$ – вероятность того, что система перейдет за один шаг из состояния i в состояние j . Очевидно, что $\sum_{j=1}^n p_{ij} = 1$, $i = 1, \dots, n$. Такие матрицы также называют *стохастическими матрицами*. Таким образом, распределение вероятностей в момент времени $t = 1$ можно найти по формулам (*полной вероятности*):

$$p_k(1) = \sum_{i=1}^n p_i(0) p_{ik}, \quad k = 1, \dots, n.$$

Аналогично на любом шаге $(m+1) \in N$:

$$p_k(m+1) = \sum_{i=1}^n p_i(m) p_{ik}, \quad k=1, \dots, n.$$

Эти соотношения можно переписать более компактно:

$$\vec{p}^T(m+1) = \vec{p}^T(m) P \text{ (уравнение Колмогорова – Чэмпена).}$$

Осуществим предельный переход к непрерывному времени (скейлинг): шаг по времени $\Delta t \rightarrow 0+$ и $p_{ij} := \lambda_{ij} \Delta t$, $i \neq j$. Тогда уравнение Колмогорова – Чэмпена перейдет в уравнение:

$$\frac{d\vec{p}^T(t)}{dt} = \vec{p}^T(t) \Lambda \text{ (уравнение Колмогорова – Феллера),}$$

где $[\Lambda]_{ij} = \lambda_{ij}$ при $i \neq j$ (интенсивности переходов) и $[\Lambda]_{ii} = -\sum_{j: j \neq i} \lambda_{ij}$, $i=1, \dots, n$.

Матрицу Λ называют инфинитезимальной матрицей.

Таким образом, для того чтобы задать конечный однородный марковский процесс (цепь) нужно задать начальное распределение вероятностей и инфинитезимальную матрицу (или матрицу переходных вероятностей в случае дискретного времени). В следующем пункте будут рассмотрены конкретные примеры. А сейчас поставим основной вопрос: что будет происходить с марковским процессом на больших временах? Ответ подсказывают выписанные уравнения динамики: $\lim_{m \rightarrow \infty} \vec{p}(m) = \vec{p}^*$, где $\vec{p}^{*T} = \vec{p}^{*T} P$ (в дискретном времени) и $\lim_{t \rightarrow \infty} \vec{p}(t) = \vec{p}^*$, где $\vec{p}^{*T} \Lambda = \vec{0}^T$ (в непрерывном времени). Так определяемый вектор распределения вероятностей $\vec{p}^*$ называется *стационарным (инвариантным) распределением*. Предельное распределение также называют *финальным распределением*. Выше лишь сказано, что финальные распределения следует искать среди стационарных.

Но, во-первых, где гарантии, что предел существует? Возьмем марковскую цепь с матрицей переходных вероятностей: $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ и начальным состоянием $\vec{p}^T(0) = (1 \ 0)^T$. Очевидно, что $\vec{p}^T(m) = 0.5 \cdot (1 + (-1)^m, 1 + (-1)^{m+1})^T$ предела в обычном смысле не имеет. Легко при этом проверить, что стационарное распределение $\vec{p}^{*T} = (0.5, 0.5)^T$ (заметим, что стационарное распределение всегда существует) будет пределом $\vec{p}^T(m)$ по Чезаро (см. ниже).

Во-вторых, при каких условиях стационарное распределение единственно? Ведь если взять, например, в качестве матрицы переходных вероятностей единичную матрицу, то линейно независимых стационарных состояний будет столько, сколько состояний у цепи.

Чтобы ответить на поставленные вопросы напомним следующий простой факт: монотонное преобразование отрезка в себя имеет неподвижную точку. Причем это утверждение можно доказать с помощью *принципа сжимающих отображений*. Марковская динамика – это монотонная динамика в том смысле, что если $\vec{x} \geq \vec{y} \geq \vec{0}$ (покомпонентно, следовательно, « $\geq$ » задает лишь частичный порядок), то $\vec{x}^T P \geq \vec{y}^T P$ (это следует из того, что все элементы матрицы P неотрицательные). Поэтому возникает желание воспользоваться одним из основных инструментов прикладного функционального анализа «принципом сжимающих отображений», чтобы исследовать на устойчивость стационарные распределения (неподвижные точки марковской динамики). Далее осуществляется этот план.

Теорема (принцип сжимающих отображений, монотонные операторы, эргодическая теорема для конечных однородных марковских цепей [2]).

а) (принцип сжимающих отображений) Если оператор (вообще говоря, нелинейный) A действует в полном метрическом пространстве X и

$$\exists k \in \mathbb{N}: \forall x, y \in X \rightarrow \rho(A^k(x), A^k(y)) \leq \theta \rho(x, y), \theta \in (0, 1), \text{ то}$$

$$\exists! x^* \in X: A(x^*) = x^* \text{ и } \forall x \in X \rightarrow \rho(A^n(x), x^*) = O(\theta^{n/k}).$$

б) (монотонные операторы) Пусть $X = \mathbb{R}_{++}^n$ – множество лучей пространства $\mathbb{R}^n$, лежащих во внутренности неотрицательного ортанта, на котором введена метрика Биркгофа:

$$\rho(x, y) = \ln \min \left\{ \frac{\beta}{\alpha} : \alpha x \leq y \leq \beta x \right\} = \ln \min_{i,j=1,\dots,n} \frac{x_i/x_j}{y_i/y_j}.$$

Здесь под элементами x и y в левой части равенства понимаются лучи, а вот в правой части уже какие-то векторы, лежащие на соответствующих лучах. Какие именно векторы – не важно. Тогда X – полное метрическое пространство. И если линейный оператор $A: X \rightarrow X$ ($A = \|a_{ij}\|_{i,j=1,1}^{n,n}$ – матрица $n \times n$) положительный, т.е. $\forall i, j = 1, \dots, n \rightarrow a_{ij} > 0$, то

$$\exists \theta \in (0, 1): \forall x, y \in X \rightarrow \rho(Ax, Ay) \leq \theta \rho(x, y).$$

в) (стохастический вариант теоремы Фробениуса – Перрона, или эргодическая теорема для конечных однородных марковских цепей) Следующие условия для стохастической матрицы $n \times n$ P равносильны:

$$1) \exists m_0 \in \mathbb{N}: P^{m_0} = \|p_{ij}(m_0)\|_{i,j=1}^{n,n} > 0, \text{ т.е. } \forall i, j = 1, \dots, n \rightarrow p_{ij}(m_0) > 0;$$

2) $i)$ У системы

$$\vec{p}^{*T} = \vec{p}^{*T} P, \sum_{k=1}^n p_k^* = 1. \quad (S)$$

существует, притом единственно, решение $\vec{p}^*$.

ii) Причем $\vec{p}^* > \vec{0}$.

iii) $\forall \vec{p}(0) \geq \vec{0} \left(\sum_{k=1}^n p_k(0) = 1 \right) \rightarrow \lim_{m \rightarrow \infty} \vec{p}^T(m) = \lim_{m \rightarrow \infty} \vec{p}^T(0) P^m = \vec{p}^{*T}$, другими словами,

$$\lim_{m \rightarrow \infty} P^m = \underbrace{\begin{bmatrix} \vec{p}^* & \vec{p}^* & \dots & \vec{p}^* \end{bmatrix}}_n^T.$$

Заметим также, что условия 1), 2) пункта в) теоремы равносильны следующим требованиям: конечная однородная марковская цепь с матрицей переходных вероятностей P – неразложимая = неприводимая (т. е. из произвольного состояния «можно прийти» в любое, наперед заданное) и непериодическая Н.О.Д. $\left\{ k : \left[P^k \right]_{11} > 0 \right\} = 1$.

Например, для цепи с матрицей переходных вероятностей $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

период равняется 2. Если убрать условие непериодичности, то для неразложимой цепи с периодом d будет иметь место сходимость средних временных (*сходимость по Чезаро*):

$$\exists! \vec{p}^* > \vec{0} \left(\vec{p}^* \in S \right) : \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=1}^N P^m = \lim_{m \rightarrow \infty} \frac{1}{d} \sum_{k=1}^d P^{m+k-1} = \underbrace{\begin{bmatrix} \vec{p}^* & \vec{p}^* & \dots & \vec{p}^* \end{bmatrix}}_n^T.$$

Если перейти к непрерывному времени, осуществляя соответствующий скейлинг, то легко показать, что необходимость в условии непериодичности исчезает.

Если же цепь разложима, то система (S) уже будет разрешима не единственным образом (аналогично и в непрерывном времени). Финальное распределение существует, но уже может зависеть от того, с какого распределения мы стартуем. Кроме того, не обязательно будет верно условие $\vec{p}^* > \vec{0}$.

В заключение заметим, что эргодическая теорема переносится и на марковские процессы со счетным числом состояний с одним лишь нюансом: финальное распределение = стационарному распределению уже не обязательно будет распределением вероятностей. Это можно продемонстрировать на примере симметричного случайного блуждания на бесконечной прямой. Эта марковская цепь имеет единственное финальное распределение (хотя цепь и периодическая, зато неразложимая), которое совпадает с единственным стационарным распределением, в котором все вероятности равны нулю, т. е. не являющимся распределением вероятностей.

2. Примеры и общая схема поиска равновесия макросистем

Для наглядности сначала рассмотрим, пожалуй, один из самых простых примеров.

Пример (модель Эренфестов [1, 3]). Рядом стоят две собаки с номерами **1** и **2**. На собаках как-то расположились $M = 2n \gg 1$ блох. Скажем, в начальный момент все блохи собрались на собаке с номером **1**. На каждом шаге случайно и независимо от предыстории определяется блоха (с вероятностью $1/M$ будет выбрана любая из блох), которая перепрыгивает на другую собаку. Микросостояние системы – есть способ распределения M различных блох по двум различным собакам. Макросостояние системы – есть способ распределения M одинаковых блох между двумя различными собаками. Микросостояний будет 2^M , а макросостояний $M + 1$.

Обозначим через P матрицу (размера $2^M \times 2^M$) переходных вероятностей описанной выше микроскопической динамики. Для нас в дальнейшем будет важно лишь одно свойство этой стохастической матрицы: $P = P^T$, которое

следует из обратимости динамики во времени. Но поскольку P – стохастическая матрица, то

$$(1, \dots, 1) = (1, \dots, 1)P^T \Rightarrow (1, \dots, 1) = (1, \dots, 1)P.$$

Откуда с учетом нормировки распределения вероятностей на 1 имеем, что в стационарном распределении все микросостояния равновероятны, т. е. в стационарном распределении каждому микросостоянию приписана вероятность 2^{-M} . Но тогда вероятность макросостояния $(k, M-k)$ в стационарном распределении равна $C_M^k 2^{-M}$. Отсюда по теореме Муавра – Лапласа и по эргодической теореме для конечных однородных марковских процессов следует, что (в конце статьи будет продемонстрирован довольно общий способ получения подобного рода соотношений)

$$\lim_{m \rightarrow \infty} P \left(\frac{|n_1(m) - n_2(m)|}{M} \leq \frac{3}{\sqrt{M}} \right) \geq 0.99,$$

где $n_1(m)$ – число блох на первой собаке на шаге m ,

$n_2(m)$ – на второй (случайные величины). Т. е. относительная разность числа блох на собаках будет иметь порядок малости $O(1/\sqrt{M})$ на больших временах ($T \geq 2M$). Обратим внимание, что марковская цепь – периодическая (период 2). Однако, поскольку речь идет о вычислении вероятностей относительных величин, то в данной задаче это не играет роли.

Обозначим

$$\tau(k) = \inf \{m \in \mathbb{N} \cup \{0\} : n_1(m) = k\}, \quad \sigma(k) = \inf \{m \in \mathbb{N} : n_1(m) = k, n_1(0) = k\}.$$

Времена соответственно первого попадания и первого возвращения в состояние k . Тогда

а) $E\sigma(k) = 2^M \frac{k!(M-k)!}{M!}$, и, в частности, среднее время возвращения в

нулевое состояние $E\sigma(0) = 2^M$, где $E\sigma(k)$ – математическое ожидание времени первого возвращения в состояние k , если $n_1(0) = k$, $k = 0, \dots, n$;

б) $E_n\tau(0) = \frac{1}{M} 2^M (1 + o(M))$, где $E_n\tau(0)$ – математическое ожидание времени

первого попадания в состояние 0, если $n_1(0) = n$;

в) $E_0\tau(n) = n \ln n + n + O(1)$, где $E_0\tau(n)$ – математическое ожидание времени

первого попадания в состояние n , если $n_1(0) = 0$.

На примере этой модели можно говорить о том, что в макросистемах возврат к неравновесным макросостояниям вполне допустим, но происходить это может только через очень большое время (*циклы Пуанкаре*), так что нам может не хватить отведенного времени, чтобы это заметить (*парадокс Цермело*). Напомним, что описанный выше случайный процесс обратим во времени. Однако наблюдается необратимая динамика относительной разности числа блох на собаках (*парадокс Лошмидта*). Но в таком случае можно удивляться также и тому, что газ, собранный в начальный момент в одной половине сосуда, с течением времени равномерно распределится по сосуду.

Изложим теперь общую схему (ограничимся рассмотрением непрерывного времени, случай дискретного времени содержательно не добавляет ничего нового).

Предположим, что некоторая макросистема может находиться в различных состояниях, характеризующихся вектором $\vec{n}$ с неотрицательными целочисленными компонентами. Будем считать, что в системе происходят случайные превращения (химические реакции). Пусть $\vec{n} \rightarrow \vec{n} - \vec{\alpha} + \vec{\beta}$, $(\vec{\alpha}, \vec{\beta}) \in J$ – все возможные типы реакций. Введем, следуя М.А. Леонтовичу (1935),

интенсивность реакции (случай дискретного времени рассматривается аналогичным образом):

$$\lambda_{(\vec{\alpha}, \vec{\beta})}(\vec{n}) = \lambda_{(\vec{\alpha}, \vec{\beta})}(\vec{n} \rightarrow \vec{n} - \vec{\alpha} + \vec{\beta}) = M^{1 - \sum_i \alpha_i} K_{\vec{\beta}}^{\vec{\alpha}}(\vec{n}) \prod_{i: \alpha_i > 0} n_i \cdot \dots \cdot (n_i - \alpha_i + 1),$$

где $K_{\vec{\beta}}^{\vec{\alpha}} \geq 0$ – константы реакции (в химической кинетике – постоянные, а в социодинамике (В. Вайдлих [4]) – необязательно); при этом часто считают $\sum_i n_i(t) \equiv M$. Т.е. $\lambda_{(\vec{\alpha}, \vec{\beta})}(\vec{n})$ – вероятность осуществления в единицу времени перехода $\vec{n} \rightarrow \vec{n} - \vec{\alpha} + \vec{\beta}$. На макроуровне это соответствует принципам химической кинетики (закон действующих масс Гульдберга–Вааге, 1865). Таким образом, динамика макросистемы задается линейной полугруппой (однородный дискретный марковский случайный процесс), инфинитезимальный оператор которой определяется интенсивностями реакций $\lambda_{(\vec{\alpha}, \vec{\beta})}(\vec{n})$.

Следующее **утверждение** отображает известные результаты а) В. В. Веденяпина [5, 7], б) С. А. Пирогова и др. [6, 8] и в) обобщает результаты В. Вайдлиха и др. [4] на случай, когда рассматривается более общая схема, чем модель миграции населения:

$$\text{а) } \langle \vec{\mu}, \vec{n}(t) \rangle \equiv \langle \vec{\mu}, \vec{n}(0) \rangle \Leftrightarrow \vec{\mu} \perp \text{Lin} \{ \vec{\alpha} - \vec{\beta} \}_{(\vec{\alpha}, \vec{\beta}) \in J}. \quad (\text{inv})$$

б) Пусть выполняется условие унитарности, которое, следуя В. В. Веденяпину, будем называть условием Штюкельберга – Батищевой – Пирогова [5-8]

$$\exists \vec{\xi} > \vec{0}: \forall \vec{\alpha} \rightarrow \sum_{\vec{\beta}: (\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} = \sum_{\vec{\beta}: (\vec{\alpha}, \vec{\beta}) \in J} K_{\vec{\alpha}}^{\vec{\beta}} \prod_j \xi_j^{\beta_j} \quad (K_{\vec{\beta}}^{\vec{\alpha}}(\vec{n}) \equiv K_{\vec{\beta}}^{\vec{\alpha}}). \quad (\text{ШБП})$$

Тогда «пуассоновская» мера $\nu(\vec{n}) = \prod_i \lambda_i^{n_i} e^{-\lambda_i} / n_i!$ (точнее говоря, мера, индуцированная пуассоновской мерой на множестве (вообще говоря,

конечном!), задаваемом условиями (inv)), где $\lambda_i = \xi_i^* M$, а $\vec{\xi}^*$ – произвольное решение (ШБП), будет инвариантной относительно предложенной стохастической марковской динамики. Эта мера экспоненциально быстро концентрируется, с ростом M , в окрестности *наиболее вероятного состояния* (также удовлетворяющего условию (ШБП)), которое и принимается за *положение равновесия макросистемы*. Задача поиска наиболее вероятного макросостояния асимптотически эквивалентна задаче максимизации *энтропийного функционала* (воспользовались $n! = \sqrt{2\pi n} (n/e)^n (1 + o(1))$ – формулой Стирлинга):

$$E \approx - \sum_i n_i \cdot (\ln(n_i / \lambda_i) - 1)$$

на множестве, задаваемом условием (inv). Отметим, что условие (ШБП), называемое также *условием унитарности* [8], обобщает хорошо известное в физике и экономике *условие детального равновесия* [1, 4]:

$$\exists \vec{\xi} > \vec{0}: \forall (\vec{\alpha}, \vec{\beta}) \in J \rightarrow K_{\vec{\beta}}^{\vec{\alpha}} \prod_j \xi_j^{\alpha_j} = K_{\vec{\alpha}}^{\vec{\beta}} \prod_j \xi_j^{\beta_j}.$$

в) Пусть $\forall (\vec{\alpha}, \vec{\beta}) \in J \rightarrow \sum_i \alpha_i = \sum_i \beta_i; \alpha_i, \beta_i \in \{0, 1\}, K_{\vec{\beta}}^{\vec{\alpha}} = K_{\vec{\alpha}}^{\vec{\beta}},$

$$K_{\vec{\beta}}^{\vec{\alpha}}(\vec{n}) = K_{\vec{\beta}}^{\vec{\alpha}} \exp\left(\sum_{i: \beta_i=1} u_i(n_i+1) - \sum_{i: \alpha_i=1} u_i(n_i)\right), u_i'(n_i) \leq 0.$$

Тогда мера $\nu(\vec{n}) = \exp\left(\sum_i U_i(n_i)\right) \cdot \prod_i (n_i!)^{-1}$, где $U_i(n_i) = 2 \sum_{\nu=1}^{n_i} u_i(\nu)$ будет

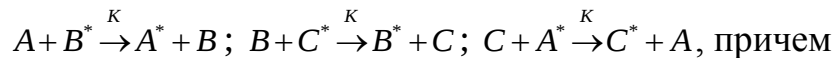
инвариантной относительно предложенной стохастической марковской динамики. Эта мера экспоненциально быстро концентрируется, с ростом $M \equiv \sum_i n_i(t)$, в окрестности наиболее вероятного состояния, которое и принимается за положение равновесия макросистемы. Задача поиска наиболее вероятного макросостояния асимптотически эквивалентна задаче максимизации энтропийного типа функционала:

$$E \approx \sum_i \{-n_i \ln(n_i) + U_i(n_i)\}$$

на множестве, задаваемом условием (inv).

В пунктах б) и в) предполагалось, что марковский процесс неразложим (неприводим) в классе (inv): из любого состояния можно со временем прийти в любое другое (по-прежнему оставаясь на множестве (inv)). Отсюда следует единственность инвариантной меры. Это условие не выполняется, например, для хорошо известной модели «хищник–жертва» (кролики – трава) [7], в которой имеется поглощающее состояние: без хищников. Нам еще будет встречаться ниже, и не один раз, эта модель.

Контрпример (С. А. Пирогов). Условие (ШБП) является только достаточным условием инвариантности «пуассоновской» меры. Действительно, рассмотрим систему уравнений химических реакций (константы реакций K одинаковы и постоянны):



$$n_A(t) + n_{A^*}(t) \equiv n_B(t) + n_{B^*}(t) \equiv n_C(t) + n_{C^*}(t) \equiv N.$$

Заметим, что есть и еще один независимый закон сохранения:

$$n_A(t) + n_B(t) + n_C(t) \equiv \text{const}.$$

Можно проверить, что «пуассоновская» мера (« $\sim$ » – знак пропорциональности)

$$\nu(\vec{n}) \sim C_N^{n_A} \cdot C_N^{n_B} \cdot C_N^{n_C} \sim (1^{n_A} e^{-1} / n_A!) \cdot \dots \cdot (1^{n_{C^*}} e^{-1} / n_{C^*}!)$$

при $\text{const} = 3N/2$ будет инвариантной, хотя условие (ШБП), очевидным образом, не выполняется. Однако можно показать, что в общем случае, для почти всех начальных условий, условие (ШБП) является и необходимым условием инвариантности пуассоновской меры.

В связи с этим контрпримером заметим, что понятие равновесия макросистемы «не завязано» на условие (ШБП). Так в контрпримере С. А. Пирогова равновесие будет существовать:

$$n_A(\infty) \approx n_{A^*}(\infty) \approx n_B(\infty) \approx n_{B^*}(\infty) \approx n_C(\infty) \approx n_{C^*}(\infty) \approx N/2.$$

Пример (Вильфредо Парето, «Кинетика социального неравенства»).

В городе живет $M \gg 1$ (например, 10 000) пронумерованных жителей. У каждого i -го жителя есть в начальный (нулевой) момент времени целое (неотрицательное) количество рублей $s_i(0)$ (монетками, достоинством в один рубль). Со временем пронумерованные жители (количество которых не изменяется, так же как и суммарное количество рублей) случайно разыгрывают свое имущество. Пусть в момент времени $t \geq 0$ r -й житель имеет k рублей, а l -й житель – m рублей. Тогда $p_{k;m}(t)\Delta t + o(\Delta t)$ есть вероятность того, что жители с номерами r и l ($1 \leq r < l \leq M$) встретятся и попробуют разыграть один рубль по следующему правилу: с вероятностью $1/2$ житель с большим номером отдаёт 1 рубль (если, конечно, он не банкрот) жителю с меньшим номером и с вероятностью $1/2$ наоборот. Будем считать, что $p_{k;m}(t) \equiv \kappa M^{-1}$ ($\kappa > 0$). Пусть $c_s(t)$ – доля жителей города, имеющих ровно s рублей в момент времени t (заметим, что $c_s(t)$ – случайная величина). Пусть $S = \sum_{i=1}^M s_i(0)$, $\bar{s} = S/M$. Тогда:

$$\forall q=0, \dots, S \exists \lambda_q > 0, T_q = O(M): \forall t \geq T_q \rightarrow P\left(\left|\frac{c_s(t)}{C e^{-s/\bar{s}}} - 1\right| \leq \frac{\lambda_q}{\sqrt{M}}, s=0, \dots, q\right) \geq 0.999,$$

где C определяется из условия нормировки: $\sum_{s=0}^S C e^{-s/\bar{s}} = 1$, т.е. $C \approx \bar{s}^{-1}$.

Скорость сходимости оценивается сверху, исходя из оценок в доказательстве эргодической теоремы для однородных марковских цепей с конечным числом состояний. Как показывают численные эксперименты,

оценка $O(M)$ точная. Так, если в городе 10 000 жителей и единица времени – день, то при начальном «социальном равенстве» с вероятностью, близкой к единице, через 20–30 лет (при $\kappa=1$) установится «социальное неравенство». Оказывается (см. также модель Эренфестов), что оценка скорости сходимости $O(\text{poly}(M))$ характерна для большинства макросистем в схеме М.А. Леонтовича (это устанавливается с помощью оценки Добрушина и подсчета кривизны Риччи (неравенств типа Пуанкаре – Чигера) [9]). Отмеченное выше обстоятельство хорошо известно специалистам по имитационному моделированию, как Markov chain Monte Carlo revolution [10].

Заметим, что описанный выше случайный процесс также (как и в модели Эренфестов) обратим во времени. Однако наблюдается необратимая динамика $c_s(t)$ (парадокс Лошмидта).

Этот пример демонстрирует ситуацию, когда число состояний ($\dim \bar{n}$) и число реакций ($|J|$) растут вместе с ростом M . Это обстоятельство, равно как и зависимость $K_{\beta}^{\bar{a}}(\bar{n})$, не позволяют напрямую использовать аппарат, разработанный в [4 – 8], связанный с анализом СОДУ, возникающей при каноническом скейлинге ($M \rightarrow \infty$ так, что $\exists \lim_{M \rightarrow \infty} \bar{n}(0)/M = \bar{c} > \bar{0}$) стохастической марковской динамики. Тем не менее, результаты настоящего примера можно перенести и на общий случай. При этом ключевым местом является существование при термодинамическом предельном переходе $M \rightarrow \infty$, $|J| \rightarrow \infty$ ненулевого финального распределения (см. конец предыдущего пункта и отмеченную там опасность нулевого финального распределения) [11].

Предположим теперь, что множество J не зависит от M , и в начальный момент времени для любого i существует предел $c_i(0) = \lim_{M \rightarrow \infty} n_i(0)/M$, $K_{\beta}^{\bar{a}}(\bar{n}) := K_{\beta}^{\bar{a}}(\bar{n}/M)$. Тогда (Малышев – Пирогов – Рыбко [6]) в произвольный

момент времени $t > 0$ и для любого i существует предел по вероятности (заметим, что $n_i(t)$ – случайные величины, тем не менее $c_i(t)$ – уже не случайные величины) $c_i(t) = \lim_{M \rightarrow \infty}^{\text{п.н.}} n_i(t)/M$. Описанный выше приём называется *каноническим скейлингом*. В результате такого скейлинга приходим к «динамике квазисредних» (терминология В. Вайдлиха [4]):

$$\frac{dc_i}{dt} = \sum_{(\bar{\alpha}, \bar{\beta}) \in J} (\beta_i - \alpha_i) K_{\bar{\beta}}^{\bar{\alpha}}(\bar{c}) \prod_j c_j^{\alpha_j}. \quad (\text{ДК})$$

Эти же уравнения можно получить и по-другому. А именно, как приближенную динамику средних $\bar{c}_i(t) = E[n_i(t)/M]$. Приближенную в том смысле, что при выводе (ДК) используется приближение: $F(\bar{c}_i(t)) \approx E[F(n_i(t)/M)]$ для «достаточно хороших» функций F (например, полиномов). Это верно в случае пикообразного распределения $n_i(t)$.¹

Покажем, во многом следуя Батищевой – Веденяпину [5], что если выполняются условия (ШБП), то траектория (ДК) сходится к неподвижной точке (какой именно, зависит, вообще говоря, от «точки старта»; но можно сказать и точнее: к той единственной неподвижной точке из семейства неподвижных точек, которая принадлежит аффинному многообразию (inv), инвариантному относительно (ДК)).² Для этого, следуя второму методу Ляпунова, введём (минус) *энтропию*: $H = \sum_i c_i \cdot (\ln(c_i/\xi_i) - 1)$ и покажем, что она является функцией Ляпунова для системы (ДК). Посчитаем полную производную H в силу системы (ДК):

¹ Заметим, что этот переход и возможность его использования нуждаются в строгом обосновании (и далеко не всегда правомочны). В качестве примера, укажем популярный в литературе [12, 13] марковский процесс «рождения–гибели» (приводящий к системе уравнений «хищник–жертва»), для которого «флуктуации играют решающую роль, качественно меняя выводы макроскопического анализа».

² Стоит заметить, что аттрактор системы (ДК) даже с постоянными коэффициентами реакции, по-видимому, в общем случае может быть сколь угодно сложным множеством [4].

$$\begin{aligned} \frac{dH}{dt} &= \sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\beta}}^{\bar{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\alpha_j} \cdot \left(\ln \prod_i y_i^{\beta_i - \alpha_i} - \sum_i (\beta_i - \alpha_i) \right) + \\ &+ \sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\beta}}^{\bar{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\alpha_j} \cdot \sum_i (\beta_i - \alpha_i) = \sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\beta}}^{\bar{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\alpha_j} \cdot \ln \prod_i y_i^{\beta_i - \alpha_i}, \end{aligned}$$

где введено обозначение $y_i = c_i / \xi_i$. Заметим, что

$$\sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\beta}}^{\bar{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\alpha_j} = \sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\alpha}}^{\bar{\beta}} \prod_j \xi_j^{\beta_j} y_j^{\alpha_j} = \sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\beta}}^{\bar{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\beta_j}.$$

Таким образом,

$$\frac{dH}{dt} = - \sum_{(\bar{\alpha}, \bar{\beta}) \in J} K_{\bar{\beta}}^{\bar{\alpha}} \prod_j \xi_j^{\alpha_j} y_j^{\beta_j} \cdot \left(\prod_j y_j^{\alpha_j - \beta_j} \cdot \ln \prod_i y_i^{\alpha_i - \beta_i} - \prod_j y_j^{\alpha_j - \beta_j} + 1 \right) \leq 0,$$

поскольку $u \ln u - u + 1 \geq 0$ при $u > 0$, и равенство достигается в одной точке $u = 1$.

Естественно (в виду примера С. А. Пирогова) теперь задаться вопросом: А что будет, если условия (ШБП) не выполняются, однако система (ДК) имеет на внутренности пересечения неотрицательного органа и инвариантного аффинного многообразия (inv) единственную неподвижную точку? Оказывается, имеет место **утверждение**: если эта точка экспоненциально глобально устойчива, то

- 1) все законы сохранения (ДК) определяются (inv);
- 2) около положения равновесия инвариантная мера будет экспоненциально быстро концентрироваться (с ростом M);
- 3) скорость сходимости к равновесию (mixing time [14]) оценивается как $O(\text{poly}(M))$;
- 4) элементы корреляционной матрицы случайного вектора $\vec{n}(t)$ равномерно ограничены по времени;
- 5) предельные переходы $\lim_{M \rightarrow \infty}$ и $\lim_{t \rightarrow \infty}$ перестановочны: $\lim_{M \rightarrow \infty} \lim_{t \rightarrow \infty} * = \lim_{t \rightarrow \infty} \lim_{M \rightarrow \infty} *$.

Обратим внимание, что модель «хищник–жертва», в изложении п. 7.3 книги [13] является хорошим примером того, что может быть, если не выполняется условие устойчивости равновесия.

Результаты [4, 12] и ряда других работ наталкивают на **гипотезу**: аттрактор динамической системы (ДК), который как мы отмечали выше может быть сколь угодно сложным множеством (например, в приложениях типичны случаи предельных циклов, нескольких положений равновесий, и даже хаотических аттракторов), является таким множеством, в малой окрестности которого на больших временах с большой вероятностью будет пребывать рассматриваемая макросистема.

Настоящая статья представляет собой пополненную новыми результатами версию приложения Е. В. Гасниковой в книге [15].

Автор выражает благодарность В. В. Веденяпину, А. В. Гасникову, Е. О. Черноусовой, А. В. Колесникову, В. А. Малышеву, Т. А. Нагапетяну, С. А. Пирогову, А. Н. Рыбко за ряд ценных замечаний.

Литература

1. Кельберт М. Я., Сухов Ю. М. Вероятность и статистика в примерах и задачах. Т. 2. М.: МЦНМО, 2010.
2. Красносельский М. А., Лифшиц Е. А., Соболев А. В. Позитивные линейные системы. Метод положительных операторов. М.: Наука, 1985.
3. Кац М. Вероятность и смежные вопросы в физике. М.: Мир, 1965.
4. Вайдлих В. Социодинамика: системный подход к математическому моделированию в социальных науках. М.: УРСС, 2010.
5. Веденяпин В. В. Кинетическая уравнения Больцмана и Власова. М.: Физматлит, 2001.
6. Malyshev V. A., Pirogov S. A., Rubco A. N. Random walks and chemical networks // Mosc. Math. J. 2004. V. 4. № 2. P. 441-453.

7. Батищева Я. Г., Веденяпин В. В. II-й закон термодинамики для химической кинетики // Мат. мод. 2005. Т. 17. № 8. С. 106-110.
8. Малышев В. А., Пирогов С. А. Обратимость и необратимость в стохастической химической кинетике // УМН. 2008. Т. 63. № 1. С. 3-36.
9. Ollivier Y. A survey on Ricci curvature for metric spaces and Markov chains. 2010.
<http://www.yann-ollivier.org/rech/publs/surveycurvmarkov.pdf>
10. Diaconis P. The Markov chain Monte Carlo revolution // Bulletin (New Series) of the AMS. 2009. V. 49. № 2. P. 179-205.
11. Rybko A., Shlosman S. Poisson hypothesis for information networks I, II // e-prints, 2004.
<http://arxiv.org/abs/math/0406110v1>;
<http://arxiv.org/abs/math-ph/0410053v1>.
12. Гардинер К. В. Стохастические методы в естественных науках. М.: Мир, 1986.
13. Занг В. Б. Синергетическая экономика: время и перемены в нелинейной экономической теории. М.: Мир, 1999.
14. Montenegro R., Tetali P. Mathematical aspects of mixing times in Markov chains. 2006.
<http://people.math.gatech.edu/~tetali/PUBLIS/survey.pdf>
15. Введение в математическое моделирование транспортных потоков: учеб. пособие / Гасников А. В., Кленов С. Л., Нурминский Е. А., Холодов Я. А., Шамрай Н. Б (Приложения: Бланк М. Л., Гасникова Е. В., Замятин А. А. и Малышев В. А., Колесников А. В., Райгородский А. М) / Под ред. А. В. Гасникова – М.: МФТИ, 2010. – 361 с.
<http://zoneos.com/traffic/>, <http://crec.mipt.ru/study/courses/optional/gasnikov/>

Информация об авторах

Data on authors

Гасникова Евгения Владимировна, к.ф.-м.н., НИУ МФТИ, г.
Долгопрудный

Тел. (495) 408-72-90

Электронная почта: egasnikova@yandex.ru

Gasnikova Евгения, Cand. of Physics & Mathematics, MIPT, Dolgoprudny,
Moscow Region

Tel. +7 (495) 408-72-90

E-mail: egasnikova@yandex.ru

ОБОБЩЕННАЯ СТМ-МОДЕЛЬ ТРАНСПОРТНЫХ ПОТОКОВ И ИССЛЕДОВАНИЕ УСТОЙЧИВОСТИ ЕЁ СТАЦИОНАРНЫХ РЕЖИМОВ¹

SOME GENERALIZATION OF CTM-MODEL OF TRAFFIC FLOW AND STABILITY INVESTIGATION

В работе рассматривается обобщение СТМ-модели транспортных потоков (типа клеточных автоматов) К. Даганзо. Исследуются на устойчивость стационарные режимы для предложенной модели.

In the work we consider some generalization of well known model of traffic flow, goes back to the works of K. Daganzo (1994, 1995). At the end of the article we lead some stability properties of proposed generalized model.

Ключевые слова: закон сохранения, схема Годунова.

Key words: conservation law, Godunov scheme.

Во второй половине 40-ых годов и в 50-ые годы XX века в СССР и США интенсивно занимались исследованием процессов, возникающих при взрыве бомбы (см., например, монографии [1], [2]). В частности, большое внимание было уделено изучению начально-краевых задач для уравнения (и систем таких

¹ Работа выполнена при поддержке грантов РФФИ 11-07-00162-а, 11-01-00494-а, 12-01-33007 мол-а-вед; Лаборатории структурных методов анализа данных в предсказательном моделировании, ФУПМ МФТИ, грант правительства РФ дог. 11.G34.31.0073; министерства образования и науки Российской Федерации, соглашение 14.A18.21.0866; гранта Президента РФ № МК-5285.2013.9.

уравнений) типа закона сохранения. В это же время наблюдался и рост приложений, в которых встречаются схожие уравнения. Так в 1955 г. независимо в работах [3] и [4] (см. также [5]) была предложена, по-видимому, первая макроскопическая (гидродинамическая) модель однополосного² транспортного потока, получившая впоследствии название *модели Лайтхилла – Уизема(Уитема) – Ричардса (LWR)*, в которой поток АТС рассматривается как поток одномерной сжимаемой жидкости (часто эту модель называют *моделью Лайтхилла – Уизема*).

В модели LWR предполагается, что

- 1) существует взаимно-однозначная зависимость (уравнение состояния) между скоростью $v(t, x)$ и плотностью (погонной) $\rho(t, x)$ потока;
- 2) выполняется «закон сохранения массы» (количества АТС).

Запись $\rho(t, x)$ – обозначает число АТС на единицу длины в момент времени t в окрестности точки трассы с координатой x .

Аналогично, $v(t, x)$ – скорость *АТС (автотранспортных средств)* в момент времени t в окрестности точки трассы с координатой x . Везде в дальнейшем предполагается, что пространственные масштабы, на которых транспортный поток описывается макроскопическими (гидродинамическими) моделями, значительно превышают характерный размер АТС (т. е. составляют не менее сотни метров). В таком предположении, мы будем интерпретировать $\rho(t, x)$, $v(t, x)$ не как некоторые, должным образом усредненные, величины (см., например, [3], [6]), а как функции, получающиеся при переходе от *микроскопического описания*) к макроскопическому. Иначе говоря, мы считаем что транспортный поток подчиняется некоторой микроскопической модели, в которой детально описывается поведение АТС, в зависимости от обстановки впереди, и эта модель является разностным или дифференциально-разностным аналогом рассматриваемой нами макроскопической модели. Таким образом, корректность, предложенного здесь подхода к определению $\rho(t, x)$, $v(t, x)$

² Полоса бесконечная в обе стороны, движение происходит слева направо (для определенности), нет источников и стоков *автомобильных транспортных (автотранспортных) средств (АТС)*.

основывается на устойчивой аппроксимации макроскопической модели микроскопической. При этом необходимость рассмотрения макроскопических моделей обусловлена, в первую очередь, более простой техникой их исследования и большей наглядностью (в виду гидродинамических параллелей).

Первое предположение выразим условием:

$$v(t, x) = V(\rho(t, x)). \quad (1)$$

Относительно функции $V(\rho)$ делается следующее предположение:

$$V'(\rho) < 0. \quad (2)$$

Обозначим

$$Q(\rho) = \rho V(\rho)$$

– интенсивность потока АТС (количество АТС, проходящих в единицу времени через заданное сечение). Зависимость $Q(\rho)$ часто называют *фундаментальной (или основной) диаграммой*. Для однополосного потока принято считать [5]:

$$Q''(\rho) < 0.$$

Что можно понимать следующим образом: движение по двум одинаковым и независимым полосам с разными плотностями менее «эффективно», чем движение по этим полосам с одинаковой плотностью равной среднему арифметическому первоначальных плотностей.

Второе предположение выразим законом сохранения

$$\int_a^b \rho(t + \Delta, x) dx - \int_a^b \rho(t, x) dx = - \left\{ \int_t^{t+\Delta} Q(\rho(\tau, b)) d\tau - \int_t^{t+\Delta} Q(\rho(\tau, a)) d\tau \right\}.$$

Отсюда следует, что для любого прямоугольного контура Γ в полуплоскости $t \geq 0, x \in \mathbb{R}$ со сторонами параллельными осям (легко показать, что это соотношение справедливо для произвольного кусочно-гладкого контура Γ), выполняется:

$$\oint_{\Gamma} \rho(t, x) dx - Q(\rho(t, x)) dt = 0. \quad (3)$$

В точках гладкости $\rho(t, x)$:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(v\rho)}{\partial x} = \frac{\partial \rho}{\partial t} + \frac{\partial(V(\rho)\rho)}{\partial x} = \frac{\partial \rho}{\partial t} + \frac{\partial Q(\rho)}{\partial x} = 0. \quad (4)$$

Вернемся к соотношению (3). Обратим внимание, что это соотношение может быть выполнено и для разрывной функции плотности $\rho(t, x)$. Причем разрыв функции $\rho(t, x)$ – есть резкое увеличение плотности, что соответствует границе затора. Пусть в момент времени t разрыв находится в точке с координатой x , и

$$\rho(t, x-0) = \rho_-, \quad \rho(t, x+0) = \rho_+.$$

Предположим, что на плоскости $(t; x)$ этому разрыву соответствует кривая L . Возьмем в окрестности точки $(t, x) \in L$ прямоугольный контур (для определенности, зададим ориентацию по часовой стрелке) так, как показано на рис. 1. Будем считать, что ширина контура Γ настолько мала по сравнению с длиной, что интегралом по участкам контура Γ , поперечным к L , можно пренебречь (см. рис. 1).

Тогда из (3) следует, что

$$0 = \oint_{\Gamma} (\rho(t, x) dx - Q(\rho(t, x)) dt) = (\rho_+ c - Q(\rho_+)) \Delta t - (\rho_- c - Q(\rho_-)) \Delta t + o(\Delta t),$$

где $c = dx/dt$ – соответствует наклону касательной к L в точке (t, x) , Δt – длина проекции контура на ось t . При $\Delta t \rightarrow 0$ это равенство переходит в следующее условие (частный случай, условия Стокса (1848)) для скорости движения разрыва c , которое называется (во всяком случае, должно называться, согласно П. Лаксу [9]) **условием Римана – Ранкина – Гюгонио**

$$c = \sigma(\rho_-, \rho_+) = \frac{Q(\rho_+) - Q(\rho_-)}{\rho_+ - \rho_-} \quad (RRH).$$

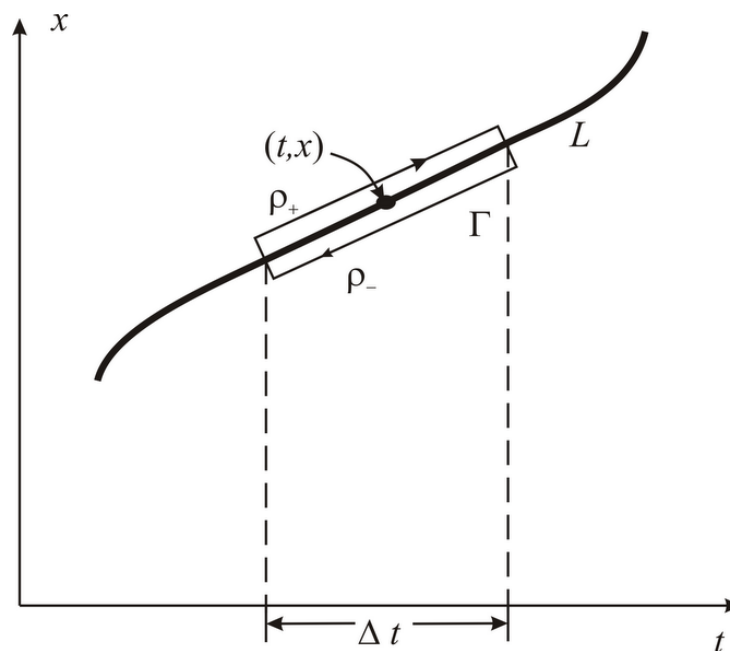


Рис. 1. RRH условие на разрыве

Несмотря на элементарность, модель LWR (а также ее дифференциально-разностные и разностные аналоги) достаточно популярна в прикладных расчетах (в том числе и с многополосными дорогами). Во многом это связано с недостаточным объемом данных, для использования моделей более высокого уровня (поправки, привносимые более тонкими моделями, нивелируются неточностью данных). Ряд современных коллективов исследователей сосредотачивается на решении начально-краевых задач для уравнения (4) на графе транспортной сети. Основные сложности при этом возникают при постановке краевых условий в узлах графа транспортной сети (см., например, [10], [0]).³ Модель LWR (точнее ее разностные и дифференциально-разностные аналоги) хорошо подходит и для управления транспортными потоками. В подтверждение этих слов приведем некоторые идеи, использованные, например, в подходе берклиевской группы (П. Варайя,

³ Стоит отметить, что большое количество исследований сосредотачивается на изучении транспортного потока на отдельном прямолинейном участке транспортной сети с простейшими начально-краевыми условиями. В то время как причиной заторов (согласно К. Даганзо [12]) часто являются «узкие места» (перекрестки, въезды). Поэтому особенно важно (для приложений) создать целостную модель транспортных потоков, адекватную имеющимся данным, включающую в себя описание источников, стоков АТС и поведение АТС в вершинах графа транспортной сети (перекрестки, въезды, съезды и т.п.).

А. А. Куржанский, А. Б. Куржанский, Р. Хоровитц и др. [13]) к управлению дорожным движением.

Из фундаментальной диаграммы (рис. 2) следует, что одному и тому же значению потока АТС соответствуют разные (как правило, две) плотности и, как следствие, разные скорости.⁴ Очевидно, что более выгодным режимом является режим с большей скоростью (см. рис. 2): потоки будут удовлетворены в том же количестве, однако среднее время движения снизится, поскольку движение будет проходить при больших скоростях (и, как следствие, с меньшими плотностями). Задача управления (скажем, светофорами или въездами на основные магистрали) заключается в том, чтобы как можно большую часть времени среднестатистический водитель проводил именно в таких режимах. Исходя из *модели клеточных автоматов* [14], [15] (СТМ – Cell Transmission Model) К. Даганзо = схема Годунова (1959) для (4) + треугольная фундаментальная диаграмма⁵ (см. рис. 3), был предложен способ «оптимального управления» светофорами и въездами на магистрали, а также способ «оптимального разрыхления» однородного потока АТС на магистрали (с помощью светофоров) с целью уменьшения среднего времени в пути [13].

Далее на простейшем примере постараемся (не строго) пояснить, что грамотное разрыхление (например, светофорами) плотного потока на магистрали, может привести к уменьшению среднего времени в пути

⁴ Это обстоятельство является также причиной сложностей, возникающих при постановке краевых условий в узлах графа транспортной сети [10], [0]. Знание характеристик источников и стоков и *матриц перемешивания* в узлах (матриц, характеризующих расщепление потоков в узлах) недостаточно для корректной постановки начально-краевой задачи.

⁵ Заметим, что по зависимости интенсивности потока $Q(\rho) = \rho \cdot V(\rho)$ от плотности ρ в окрестности $\rho_{\max}$ можно определить τ – среднее время реакции водителя, если известна средняя длина АТС L [5] ($L \sim 5,7$ м):

$$\tau = -L/Q'(\rho_{\max}) \text{ (обычно, } \tau \sim 1 \text{ сек.)}.$$

Действительно, путь, пройденный АТС за время τ , $V(\rho)\tau$ не должен превышать расстояния до впереди идущего АТС $1/\rho - L$. Поэтому поведение потока (уравнение состояния) вблизи точки $\rho_{\max} \sim 1/L$ можно описать следующим образом: $V(\rho) = \frac{1/\rho - L}{\tau}$. Откуда имеем в левой окрестности точки $\rho_{\max}$:

$Q(\rho) = -\frac{L}{\tau}(\rho - \rho_{\max})$. Иногда в этих формулах вместо средней длины АТС L фигурирует среднее расстояние между соседними АТС в заторе $d \sim 7,5$ м. Приведенные в этом абзаце формулы активно используются [6], например, Б.С. Кернером при исследовании роста затора [8]. Отметим также, что если известна средняя длина АТС, время реакции водителя и желаемая скорость свободного потока (определяет наклон левой ветви фундаментальной диаграммы), то треугольная фундаментальная диаграмма однозначно строится.

«большинства» АТС. Рассмотрим магистраль, соединяющую два города А и В. Будем считать, что величина потока АТС равна q (см. рис. 2), т. е. наблюдается стабильный (не зависящий ни от чего) поток АТС, выезжающих из А в В. Тогда, возможно два режима: 1) с большей скоростью и 2) с меньшей скоростью (см. рис. 2). Предположим, что сначала имеет место режим 2). «Возмутим», например, с помощью светофоров транспортный поток, т. е. выведем его из стационарного состояния. Тогда, конечно, несколько АТС в результате будут дольше ехать из А в В. Однако это возмущение (в виду «неустойчивости» режима 2) [13], [21], [22]) «при должной его организации»⁶ приведет также и к тому, что «система скатится» в устойчивый («более выгодный») стационарный режим 1) [13], [21], [22]. Следовательно, когда возмущение «рассосется», АТС будут быстрее добираться из А в В, при этом поток, как и оговаривалось, по-прежнему будет q . Выгода разрыхления очевидна! Заметим, что пассажиры в метро, пользующиеся левой («быстрой») стороной эскалатора, неоднократно сталкиваются с похожим явлением. Для дальнейшего, более глубоко, «погружения» в вопросы, связанные с управлением дорожным движением можно рекомендовать [13].

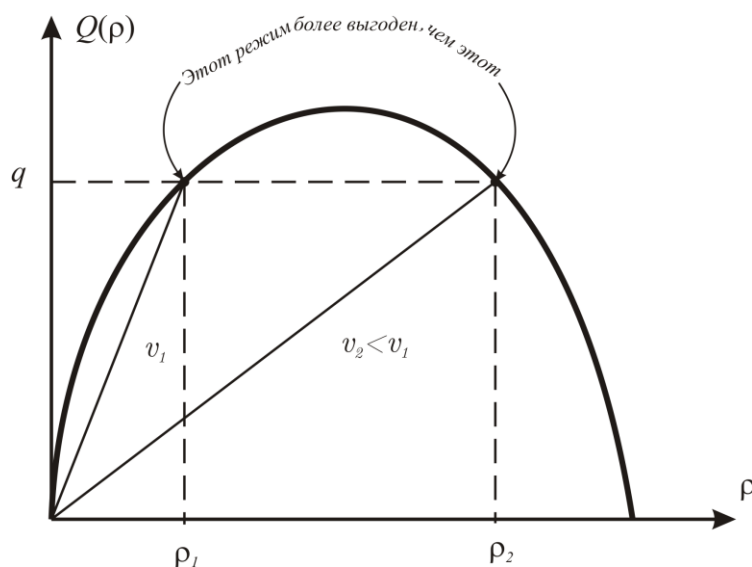


Рис. 2

⁶ Необходимо некоторое время «придержать» АТС на выезде из города А (как следствие, на выезде из города А возникнет очередь - пробка). Придерживать надо до тех пор, пока «плотность АТС на магистрали» не станет меньше ρ_1 (см. рис. 2). После того, как «сдерживание» прекратится, пробка в городе потихоньку «рассосется», и на магистрали установится стационарный режим 1).

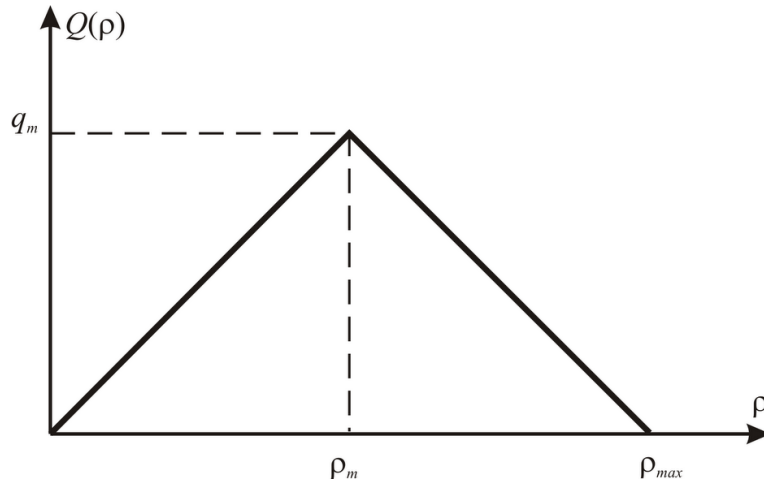


Рис. 3. Треугольная фундаментальная диаграмма

Напомним теперь вкратце, следуя книге П. Лакса [9], в чём заключается упомянутый выше метод численного решения начально-краевой задачи для закона сохранения, предложенный С. К. Годуновым в конце 50-ых годов XX века. Начальное условие $\rho(0, x)$ аппроксимируется кусочно-постоянной функцией

$$\rho^\delta(0, x) = \rho_k, \quad k\delta \leq x < (k+1)\delta,$$

где δ – шаг по пространству, а ρ_k – среднее от $\rho(0, x)$ на промежутке $[k\delta, (k+1)\delta)$, т. е.

$$\rho_k = \frac{1}{\delta} \int_{k\delta}^{(k+1)\delta} \rho(0, x) dx.$$

Задача с начальными данными $\rho^\delta(0, x)$ может быть решена точно. В каждой точке $k\delta$ мы должны решить задачу Римана о распаде разрыва. «Волны», выходящие из двух соседних точек разрыва $k\delta$ и $(k+1)\delta$, не пересекаются, пока $t \cdot c_{\max} \leq \delta/2$, где $c_{\max} = |\max Q'(\rho)|$ – максимальная скорость распространения возмущения. Итак, объединив решения задач Римана, можно получить точное решение. В момент времени $\tau = \delta/(2c_{\max})$ – шаг по времени (такой выбор шага иногда называют *условием Лакса*), мы опять заменим это точное решение приближенной кусочно-постоянной функцией и повторим

процесс. Численные эксперименты говорят о том, что метод Годунова дает хорошее приближение точных решений уравнений LWR. Тем не менее, строгое доказательство устойчивости схемы Годунова в случае систем законов сохранения, имеется, насколько нам известно, лишь для конкретных систем. В заключении заметим, что схема Годунова для LWR модели может быть содержательно проинтерпретирована (см. ниже). Другими словами, можно было ничего не знать про LWR модель и из естественных соображений напрямую прийти к разностной схеме С. К. Годунова (в транспортной литературе принято разностные схемы называть *моделями клеточных автоматов*). Как показывает практика, очень важно (при гидродинамическом описании транспортного потока – по сути, «дискретного» объекта) выбирать разностную схему таким образом, чтобы она могла быть самостоятельно содержательно проинтерпретирована.

Продemonстрируем (следуя М. Л. Бланку (2000), см. [16]-[18]) важность именно треугольной фундаментальной диаграммы одним из способов качественного вывода уравнения состояния $V(\rho)$, исходя из довольно простой СА модели (модели клеточных автоматов, Cellular Automat). Другие доводы в пользу треугольной фундаментальной диаграммы имеются, например, в работах [8], [13]-[15]. Стоит заметить, что наряду с треугольной фундаментальной диаграммой также удобно использовать параболическую фундаментальную диаграмму Гриншилдса (1934) [5], [6].

Рассмотрим кольцевую дорогу, состоящую из n ячеек (клеток). В каждой ячейке может находиться не более одного АТС. Длины всех ячеек одинаковы и равны (условной) единице. Будем также считать, что n достаточно большое. Если брать не кольцевую топологию, а, скажем, бесконечную прямолинейную дорогу (полосу), то по n необходимо будет сделать «*термодинамический предельный переход*» [19] – устремить n к бесконечности, «сохраняя пропорции». Пусть в начальный момент времени в некоторые из ячеек поместили АТС. Обозначим через $0 < \rho \leq 1$ – долю занятых ячеек. Будем считать, что сначала все АТС «смотрят» в следующую по ходу движения

ячейку, а потом те из АТС, для которых эти ячейки оказались свободными независимо от остальных двигаются в свободную ячейку с вероятностью $0 < p \leq 1$. И так происходит на каждом шаге по времени.

Определим *среднюю пространственную* скорость каждого АТС

$$\bar{V}_i^S(m) \stackrel{\text{def}}{=} \frac{1}{\rho n} \sum_{i=1}^{\rho n} V_i(m).$$

Тогда⁷ *средняя временная* скорость каждого АТС

$$\bar{V}_i^T \stackrel{\text{def}}{=} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=1}^N V_i(m) = V = \lim_{m \rightarrow \infty} \bar{V}_i^S(m).$$

Зависимость «средней» скорости V от плотности ρ изображена на рис. 4, из которого становится ясно, что фундаментальная диаграмма, соответствующая рис. 4 будет треугольного типа. В виду простоты модели, несложно качественно объяснить приведенную на рис. 4 зависимость. Для дальнейшего знакомства с СА моделями см. [20].

Перейдем теперь к другому типу клеточных моделей, которые, по сути, являются разностными аналогами LWR модели. Приведем схему, в которую вложены многие модели этого класса. Ограничимся ситуацией магистрали (вообще говоря, многополосной) с въездами и съездами. Разобьем магистраль на клетки (ячейки) – прямолинейные участки дороги длиной не менее сотни метров. Будем считать (не ограничивая общности), что в каждую клетку имеется только один въезд и из каждой клетки имеется только один съезд (см. рис. 5).

⁷ Эргодическая теорема для марковских процессов при $0 < p < 1$ (соответствующие исследования, насколько нам известно, проводились различными коллективами, например, А.П. Буслаевым и А.Г. Таташевым (2010)). При $p = 1$ результат был установлен М.Л. Бланком (2000) [16].

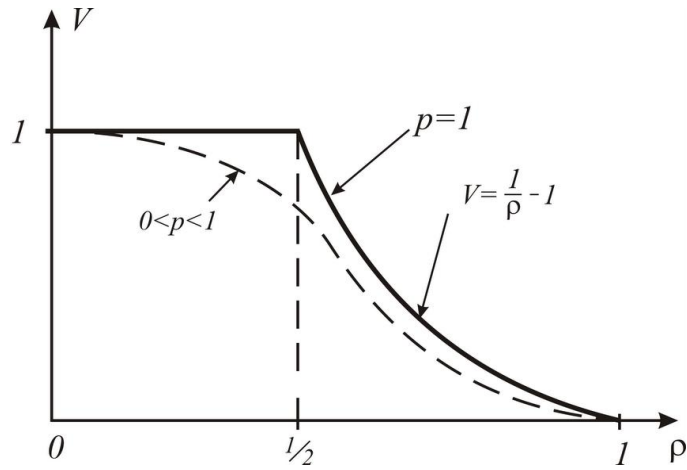


Рис. 4. «Фазовый переход» при $\rho = 1/2$

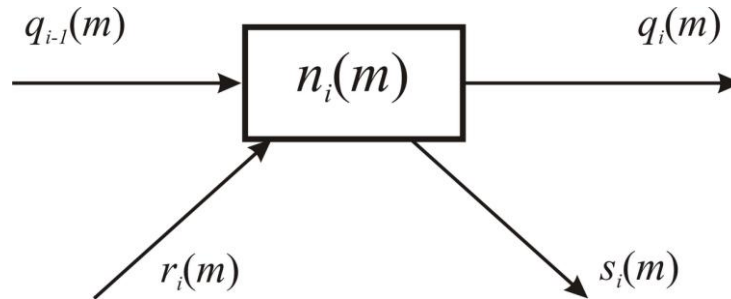


Рис. 5

$$n_i(m+1) = n_i(m) + r_i(m) + q_{i-1}(m) - s_i(m) - q_i(m), \quad s_i(m) = \alpha_i q_i(m),$$

где $n_i(m)$ – число АТС в i -ой клетке в момент времени m .

Так в работах [21], [22] (рассматривалась кольцевая топология транспортной сети без въездов и съездов) полагали $q_i(m) = (1 + \alpha_i)^{-1} Q_i(n_i(m))$ – аналог схемы бегущего счета для модели LWR. Очевидным недостатком этой схемы является возможность следующей «неправдоподобной» ситуации. Предположим, что есть две клетки (без въездов и съездов). Первая клетка полностью загружена (в ней максимальная плотность АТС – «стоячая пробка»), а следующая по ходу движения клетка полностью свободна (в ней нет АТС). Тогда, согласно выбранному способу описания потока АТС, – ничего происходить не будет, т. е. ситуация со временем меняться не будет. В то время как из опыта известно, что АТС начнут «перетекать» в свободную клетку. Тем не менее, исследование «до критических режимов» – «левая (возрастающая)

ветка» фундаментальной диаграммы, с помощью этой модели вполне корректно.

В работах [13]-[15] (рассматривался случай треугольной фундаментальной диаграммы (см. рис. 3), «вершина» которой имеет координаты $(\tilde{n}_i, Q_i^{\max})$) полагали:

$$q_i(m) = \min\{(1 - \alpha_i)v_i n_i(m), Q_i^{\max}, Q_{i+1}^{\max}, (n_{i+1}^{\max} - n_{i+1}(m))w_{i+1}\},$$

– аналог схемы Годунова для модели LWR с треугольной фундаментальной диаграммой (СТМ модель К. Дангазо (1994)). Поясним обозначения:

$$v_i = \frac{Q_i^{\max}}{\tilde{n}_i} \text{ – скорость свободного потока,}$$

$$w_{i+1} = \frac{Q_{i+1}^{\max}}{n_{i+1}^{\max} - \tilde{n}_{i+1}} \text{ – скорость волны от перегрузки,}$$

т. е. скорость роста затора. Другими словами, если перекрыть дорогу (при условии, что движение было достаточно плотным $n_{i+1}(m) \geq \tilde{n}_{i+1}$), то образовавшийся затор будет расти со скоростью w_{i+1} – длина затора через время t после перекрытия будет равна $w_{i+1}t$. Теперь должно становиться ясно, в чем преимущество схемы Годунова, например, над схемой бегущего счета. В заключение заметим, что для длинных участков дороги, включающих довольно много клеток, координаты $(\tilde{n}_i, Q_i^{\max})$ не зависят от индекса i .

Можно обобщить СТМ модель на графы транспортной сети более сложной топологии, чем кольцевая или линейная (магистраль). Правда, для этого потребуется знать матрицу перемешивания в каждом узле графа транспортной сети. Кроме того, нужно дополнительно (по сравнению с магистральной топологией) прописать разрешение «конфликтных ситуаций» (знание матрицы перемешивания может оказаться недостаточно). Например, таких, как следующая: пусть имеется перекресток. Две трети водителей с входящей в перекресток дороги А хотят повернуть на выходящую с перекрестка дорогу С, и две трети водителей с входящей дороги В также хотят повернуть на дорогу С. Будем считать, что двухполосная дорога С может

пропустить 4000 АТС/час (максимальная пропускная способность). Дороги А и В также двухполосные и на обоих из них поток АТС, входящий в перекресток, есть по 4000 АТС/км. Очевидно, что ситуация не доопределена. Для понимания того, что будет происходить, необходимо еще знать, например, режим работы светофора в этом перекрестке (в случае его наличия). Таким образом, реальное расщепление потоков зависит не только от матрицы перемешивания, но и, например, от режима работы светофора. Зная матрицу перемешивания, и «взяв на вооружение» правило работы светофора, можно получить упомянутое обобщение рассмотренных моделей на графы транспортной сети общего вида. Аналогично можно рассмотреть и более сложные развязки.

Как уже отмечалось, несмотря на свою относительную простоту, СТМ модель является одной из наиболее востребованных в приложениях (см., например, [13] и цитированную там литературу).

Что касается изучения аналитических свойств этих моделей (полученных с помощью разностных схем для LWR модели), то, по-видимому [21], [22], проще исследовать все-таки дифференциально-разностные аналоги LWR моделей, т. е. считать, что время течет непрерывно. Разностные схемы (в том числе и упомянутые выше), очевидным образом переделываются в дифференциально-разностные, при этом условие Лакса корректности схемы, аппроксимирующей уравнение LWR (шаг по времени достаточно мал по сравнению с шагом по пространственной переменной), выполняется автоматически.

В связи со всем вышесказанным возникает задача: *исследовать положения равновесия – стационарные режимы (а также бассейны их притяжения, отталкивания) динамической системы на графе транспортной сети общего вида, полученной с помощью дифференциально-разностной схемы Годунова из LWR модели (с треугольной или параболической фундаментальной диаграммой)*. В 2004 г. для схемы бегущего счета исследования в этом направлении были предприняты А. П. Буслаевым и др. [21] для графов специальной структуры (кольцевой,

«цветочной» и т. п.). В 2006 г. А. И. Назаров обобщил результаты статьи [21] на графы общего вида. Однако, как уже отмечалось выше, схема бегущего счета – не самый подходящий вариант для описания транспортного потока во всевозможных состояниях. В диссертации 2007 г. А.А. Куржанского [13] исследовалась асимптотическая устойчивость (глобальная) положений равновесия (образующих многообразие с довольно простой, и полностью описанной, линейной структурой) для СТМ модели (схема Годунова + треугольная фундаментальная диаграмма) транспортных потоков на магистрали (т. е. с простой топологией графа транспортной сети).

Выпишем СОДУ на графе транспортной сети,⁸ используя следующую модификацию правила «пропорциональных приоритетов» [15]:

$$\begin{aligned} \frac{d\rho_i}{dt} = & \min \left\{ \sum_{j:j \rightarrow i} \beta_j^j \min\{\rho_j v_j, Q_j^{\max}\}, \min\{(\rho_i^{\max} - \rho_i) w_i, Q_i^{\max}\} \right\} - \\ & - \sum_{k:i \rightarrow k} \beta_k^i \min\{\rho_i v_i, Q_i^{\max}\} \cdot \min \left\{ 1, \frac{\min\{(\rho_k^{\max} - \rho_k) w_k, Q_k^{\max}\}}{\sum_{l:l \rightarrow k} \beta_k^l \min\{\rho_l v_l, Q_l^{\max}\}} \right\}, \end{aligned}$$

здесь каждое ребро ориентированного графа транспортной сети пронумеровано, ρ_i – плотность потока на i -ом ребре, β_i^j – доля потока АТС на ребре j , ответвляющаяся на ребро i . Обратим внимание, что в общем случае следует считать $\beta_i^j(t, \vec{\rho})$. Причем если учитывать задержки в узлах графа транспортной сети, связанные, например, с наличием светофоров, то, вообще говоря, $\sum_{k:i \rightarrow k} \beta_k^i(t, \vec{\rho}) < 1$.

Утверждение. *Для каждой замкнутой транспортной сети (можно обобщить и на открытые сети) стационарный режим будет устойчивым, если значения стационарных плотностей «лежат» на левых (возрастающих) ветках соответствующих фундаментальных диаграмм $Q_i(\rho_i)$ (зависимость потока от плотности).*

⁸ Для простоты считаем, что нет источников и стоков АТС, в противном случае их следовало бы учитывать, например, подобно тому как это было сделано выше.

Литература

1. Зельдович Я. Б., Райзер Ю. П. Физика ударных волн и высокотемпературных гидродинамических явлений. М.: Физматлит, 2008.
2. Курант Г., Фридрихс К. Сверхзвуковое течение и ударные волны. М.: Издательство иностранной литературы, 1950.
3. Lighthill M. J., Whitham G. B. On kinematic waves: II. Theory of traffic flow on long crowded roads // Proc. R. Soc. London, Ser. A. V. 229. 1955. P. 281-345.
4. Richards P. I. Shock Waves on the Highway // Oper. Res, V. 4. 1956. P. 42-51.
5. Уизем Дж. Линейные и нелинейные волны. М.: Мир, 1977.
6. Traffic flow theory Revised. A state-of-the-art report. TRB, 2001. Revised Transportation Research Board Special Report 165, «Traffic Flow Theory», published in 1975.
7. Луканин В. Н., Буслаев А. П., Трофимов Ю. В., Яшина М. В. Автотранспортные потоки и окружающая среда. М.: ИНФРА-М, Ч. 1, 2. 1998, 2001.
8. Kerner B. S. Introduction to modern traffic flow theory and control. The long road to three - phase traffic theory. Springer, 2009.
9. Лакс П. Д. Гиперболические дифференциальные уравнения в частных производных. Москва–Ижевск: НИЦ «РХД», ИКИ, 2010.
10. Garavello M., Piccoli B. Traffic Flow on Networks. Volume 1 of AIMS Series on Applied Mathematics. AIMS, 2006.
11. Göttlich S., Klar A. Model hierarchies and optimization for dynamic flows on networks. Modeling and optimization of flows on networks. Cetaro (CS), June 15-19, 2009. I.M.E. Courses, 2009.
<http://php.math.unifi.it/users/cime/Courses/2009/01/200914-Notes.pdf>
12. Daganzo C. F. Fundamentals of Transportation and Traffic Operations. New York: Elsevier Science Inc., 1997.

13. Kurzanskiy A. A. Modeling and software tools for freeway operational planning. PhD thesis, Berkeley: University of California, 2007; (see also Xiaotian Sun, PhD thesis, Berkeley: University of California, 2005; Gabriel Clemente Gomes Parisca, PhD thesis, Berkeley: University of California, 2004.)
<http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-148.pdf>;
<http://lihodeev.com/pubs.html>
14. Daganzo C. F. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory // Transp. Res. B., V. 28. № 4. 1994. P. 269-287.
15. Daganzo C. F. The cell transmission model, Part II: Network traffic // Transp. Res. B., V. 29. № 2. 1995. P. 79-93.
16. Бланк М. Л. Точный анализ динамических систем, возникающих в моделях транспортных потоков // УМН, Т. 55(333). № 3. 2000. С. 167-168.
http://www.mathnet.ru/php/archive.phtml?wshow=paper&jrnid=rm&paperid=295&option_lang=rus
17. Blank M. Ergodic properties of a simple deterministic traffic flow model // J. Stat. Phys., V. 111. № 3-4. 2003. P. 903-930. [arXiv:math.DS/0206194](https://arxiv.org/abs/math/0206194)
18. Blank M. Hysteresis phenomenon in deterministic traffic flows // J. Stat. Phys., V. 120. № 3-4. 2005. P. 627-658. [arXiv:math.DS/0408240](https://arxiv.org/abs/math/0408240)
19. Минлос Р. А. Введение в математическую статистическую физику. М.: МЦНМО, 2002.
20. Maerivoet S., De Moor B. Cellular automata models of road traffic // Physics Reports, V. 419. № 1. 2005. P. 1-64. [arXiv:physics/0509082](https://arxiv.org/abs/physics/0509082)
21. Буслаев А. П., Таташев А. Г., Яшина М. В. О свойствах решений одного класса систем нелинейных дифференциальных уравнений на графах // Владикавказский матем. жур., ВНЦ РАН, Т. 6. № 4. 2004. С. 4-18.
22. Назаров А. И. Об устойчивости стационарных режимов в одной системе ОДУ, возникающей при моделировании автотранспортных потоков // Вестник СПбГУ, серия 1. Математика, Механика, Астрономия. № 3. 2006. С. 35-43.

Информация об авторах

Data on authors

Гасников Александр Владимирович, к.ф.-м.н., доцент кафедры математических основ управления НИУ МФТИ, г. Долгопрудный

Тел. (495) 408-72-90

Электронный адрес: gasnikov@yandex.ru

Gasnikov Alexander, Cand. of Physics & Mathematics, Docent at Department Mathematical Annals of Control in NRU MPTI, Dolgoprudny, Moscow Region

Tel. +7 (495) 408-72-90

E-mail: gasnikov@yandex.ru

Холодов Ярослав Александрович, к.ф.-м.н., доцент кафедры вычислительной математики НИУ МФТИ, г. Долгопрудный

Тел. (495) 408-80-88

Электронный адрес: kholodov@cres.mipt.ru

Kholodov Yaroslav, Cand. of Physics & Mathematics, Docent at Department Computing Mathematics in NRU MPTI, Dolgoprudny, Moscow Region

Tel. +7 (495) 408-80-88

E-mail: kholodov@cres.mipt.ru

ИДЕНТИФИКАЦИЯ МОДЕЛИ ДОБЫВАЮЩЕГО СЕКТОРА ЭКОНОМИКИ МОНГОЛИИ*

IDENTIFICATION OF A MODEL OF EXTRACTING SECTOR FOR MONGOLIAN ECONOMY

Модель представляет собой систему дифференциальных уравнений, каждое из которых описывает взаимодействие добывающего сектора экономики Монголии с домашними хозяйствами, внешним рынком, банковской системой и государством.

В ходе проведения численных экспериментов на основе анализа данных экономической статистики Монголии был получен работоспособный вариант модели, качественно верно отражающий положение дел в отрасли.

В данной работе предложен метод оптимизации полученного рабочего варианта модели. Оптимизация производится по набору параметров модели. В качестве критерия близости результатов расчетов и статистических данных выбран индекс Тейла.

Задача оптимизации по набору из двух параметров реализована на языке программирования C++ в виде параллельного пересчета решения задачи Коши при прохождении обоими параметрами области допустимых значений с равномерным шагом.

The model is represented by the system of differential equations. Each of this equations describes interaction between the Extractive sector of Mongolia with Households, Foreign market, Banking system and Government. Calculations with the

* Работа выполнена при поддержке РФФИ (проекты №№ 09-01-90201-Монг_а, 08-01-00377), ПФИ Президиума РАН № 14, ПФИ ОМН РАН № 3.

model were done to find a working variant. Results of calculations for one of these variants were based on analysis of Mongolian economical statistics are presented in our study.

We did experiments with the constructed model. These experiments were based on analysis of Mongolian economical statistics. In the conclusion we offer some results from calculations and analysis of our results. The efficient variant of model qualitatively truly reflecting a state of affairs in branch has been received.

In this work we received method of optimization of the efficient variant of model. Optimization is made on a set of parameters of model. We use Teil index as an affinity criterion of results of calculations and the statistical data.

We realized the optimization problem on a set of two parameters in programming language C ++ in form of parallel recalculation of the model decision when both parameters changes in the area of admissible values with an even stride.

Ключевые слова: модель экономики Монголии, метод оптимизации по набору параметров модели, параллельный алгоритм решения задачи Коши.

Key words: model of economy for Mongolia, method of optimization for parameters of model, parallels algorithm of solution for Cauchy problem.

В современной экономической системе можно выделить ряд государств с активно развивающимися, специфично устроенными экономиками. Представителем подобной страны является Монголия. Экономику Монголии на сегодняшний день представляют два крупных и практически независимых друг от друга сектора экономики:

- сельское хозяйство (в основном пастбищное животноводство) и связанные с ним отрасли обрабатывающей промышленности;
- промышленность (добыча природных ресурсов) и транспорт.

Перестройка экономики Монголии из плановой в рыночную началась относительно недавно. Однако за прошедший период произошли большие изменения в экономике и в социальной жизни страны, как положительные, так

и отрицательные. Годы «перестройки» характеризуются спадом производства, кризисной ситуацией в экономике, увеличением численности безработных, снижением жизненного уровня населения и дифференциацией уровней доходов среди населения.

Перед нами была поставлена задача – построить модель добывающего сектора экономики Монголии, научиться идентифицировать внутренние параметры модели так, чтобы результаты расчетов были наиболее близкими к статистике по заданному критерию оптимизации.

Исследуемая модель развивающейся экономики построена на основе балансового метода. Динамика материальных и финансовых балансов выражается через изменения запасов природных ресурсов, факторов производства и денег. Заданные параметры и переменные модели имеют понятный экономический смысл. Потоки благ в модели описаны посредством линейных дифференциальных уравнений.

В модели учтены особенности экономики Монголии в части налогообложения и внешнеэкономических отношений, особенности связанные с наличием и составом природных ресурсов – в основном это каменный и коксующий уголь и цветные металлы.

При построении модели мы использовали такие экономические агенты как «Добывающий сектор», «Домашние хозяйства», «Банковская система», «Государство», «Внешний рынок». Модель разработана для проведения имитационных экспериментов по построению прогнозов развития и оценки потенциала сектора экономики с целью разработки предложений по активизации деятельности и оценки влияния теневой экономики.

После проведения численных экспериментов на модели нами произведена идентификация ее параметров в первом приближении, таким образом, был получен работоспособный вариант модели. Большая часть параметров модели не может быть оценена напрямую из данных экономической статистики, для их идентификации нужно сравнивать полученные при расчетах на модели временные ряды макропоказателей

области со статистическими временными рядами этих макропоказателей. Технология идентификации параметров предусматривает высокоскоростные параллельные вычисления на кластерной компьютерной системе.

Описание модели

В качестве экономических агентов выделены добывающий сектор X , домашние хозяйства L , банковскую систему B , и Правительство (консолидированный бюджет) G . Производитель поставяет продукцию на внутренний и внешний рынок. Домашние хозяйства предлагают труд и потребляют конечную продукцию. Банковская система выпускает денежные средства, выдает кредиты производителям с целью извлечения банковской прибыли. Правительство региона собирает налоги с производителей и домашних хозяйств. Считаем, что своя цена формируется на каждом рынке, и изменение цены обратно пропорционально изменению запасов соответствующих продуктов.

Для учета в модели теневого оборота, мы предполагаем, что добытый из природной среды продукт производитель делит на легальный и теневой, который не облагается налогами. В результате у производителя оказывается два вида денег – «белые» и «черные». «Черные» деньги могут отмываться, а запас неотмытых денег подвергается штрафным санкциям – «налогообложению» теневого сектора. У потребителя все деньги считаются «белыми», а свой доход потребитель делит по заданным нормам потребления легальных и теневых продуктов всех секторов. Занятое в отрасли население получает «белую» заработную плату и заработную плату «в конверте».

В результате получается вычислимая математическая модель – система равенств и дифференциальных уравнений, описывающая динамику изменения экономических показателей для системы. В пределе в данной системе устанавливается равновесие, поэтому она может быть отнесена к классу CGE моделей.

Для упрощения записи будем использовать стандартные обозначения, которые позволяют легко перейти к обозначениям системы построения и анализа моделей экономики ЭКОМОД. В этих обозначениях показатели и параметры модели снабжены верхними и нижними индексами, причем верхние индексы используются для агентов, а нижние для благ. При этом одни и те же символы могут использоваться для обозначения показателей (или параметров) и верхних и нижних индексов и обозначать разные вещи. В данной работе интенсивные показатели и параметры будем обозначать строчными символами, а экстенсивные показатели и именованные индексы – прописными.

Добывающий сектор.

Объем выпуска Y_X добывающего сектора описан производственной функцией типа Кобба-Дугласа (C-D), зависящей от запаса труда Q_L^X и запаса капитала Q_K^X в отрасли:

$$Y_X = (Q_L^X)^{\alpha_X} \cdot (Q_K^X)^{\beta_X}.$$

Запас собственного продукта Q^X в добывающем секторе растет с добычей, убывает с отгрузкой на внешний и внутренний рынок, часть уходит на инвестиции в производство. Продукция добывающего сектора в модели считается однородной. Большая часть добытых природных ресурсов идет на экспорт в Китай, Россию, США, Великобританию:

$$\frac{dQ^X}{dt} = (1 - q_X)Y_X - (a_X^{XL} + a_X^{XI})Q^X - c_X^X I^X,$$

где $E_X = a_X^{XI}Q^X$ – объем экспорта продукции добывающего сектора,

$I^X = \frac{b_X^I W^X}{p_I^X}$ – объем импорта технологий и основных фондов (промышленная

техника, станки, и прочее), а q_X – доля тени в добывающем секторе X.

Запас капитала Q_K^X (производственные фонды – недвижимость, импортное оборудование) растет с покупкой по цене страны импортера оборудования, уменьшается с амортизацией и с использованием капитала в производстве:

$$\frac{dQ_K^X}{dt} = \frac{b_X^I W^X}{p_I^X} - \mu_K^X Q_K^X.$$

Запас труда Q_L^X в добывающем секторе изменяется с покупкой у агента L открытого труда за официальную ставку заработной платы s_L^X и теневого труда по теневой ставке s_V^X , уменьшается в связи с увольнением, уходом на пенсию, сокращением штата:

$$Q_L^X = \frac{b_L^{XL} W^X}{s_L^X} + \frac{b_V^{XL} B^X}{s_V^X}.$$

Запас открытых («белых») денег W^X в секторе увеличивается при получении кредитов, при продаже товара на внешнем и на внутреннем рынках, за счет получения трансфертов сектору X из консолидированного бюджета (государственные субсидии и прочее), от инвестиций в производство и при поступлении «отмытых» денег $b_B^X B^X$ из теневого оборота; уменьшается от покупки импортного оборудования, от выплаты по з/п и погашения кредиторской задолженности, от выплат по налогам:

$$\frac{dW^X}{dt} = \omega p_X^I E^X + C^{BX} + p_X^L a_X^{XL} Q^X - (b_X^I + b_L^{XL} + b_B^{XB}) W^X - T^{XG} + T^{GX} + b_B^X B^X,$$

где ω – курс тугрика по отношению к доллару.

Считаем, что цена продукции добывающего сектора для внешнего рынка имеет постоянный экспоненциальный рост $p_X^I = p_{X0}^I e^t$.

Считаем, что агент X берет весь предлагаемый агентом B (банковской системой) кредит, однако объем предоставляемого ему кредита C^{BX} ограничен ликвидационной стоимостью производственных фондов, которая считается пропорциональной запасу капитала:

$$C^{BX} = \sigma^X Q_K^X, \text{ где } \sigma^X > 0.$$

Задолженность $Z^X(t)$ агента X банковской системе B прирастает за счет выдачи новых кредитов C^{BX} и начисления текущего процента по кредитам $r(t)$ на имеющуюся задолженность, а уменьшается в силу платежей погашения по задолженности

$$\frac{dZ^X}{dt} = C^{BX} + rZ^X - b_B^{XB} W^X.$$

Налоговая система формализована следующим образом. Предприятия добывающего сектора платят четыре вида налогов, которые рассчитываются по следующим базам, с описанными выше ставками:

– $T_{custom}^{XG} = n_{custom} \omega p_X^I E^X$ – таможенные платежи по ставке $n_{custom} = 10\%$;

– единый социальный налог $T_{social}^{XG} = n_{social} b_L^{XL} W^X$ по ставке $n_{social} = 22,6\%$;

– налог на добавленную стоимость

$$T_{VAT}^{XG} = n_{VAT} (\omega p_X^I E^X + p_X^L a_X^L Q^X - b_B^{XB} W^X - T_{custom}^{XG} - T_{social}^{XG} - T_{axcise}^{XG}) \text{ по ставке } n_{VAT} = 10\%;$$

– и налог на прибыль

$$T_{IT}^{XG} = n_{IT} (\omega p_X^I E^X + p_X^L a_X^{XL} Q^X - (b_B^{XB} + b_L^{XL}) W^X - T_{custom}^{XG} - T_{social}^{XG} - T_{axcise}^{XG} - T_{VAT}^{XG}), \text{ который}$$

взимается в размере 10%, если годовая прибыль составляет менее 3 млрд. тугриков, и 25%, если прибыль составляет более 3 млрд. тугриков.

Общий объем выплат по налогам представляет собой сумму вышеперечисленных отчислений в консолидированный бюджет

$$T^{XG} = T_{custom}^{XG} + T_{social}^{XG} + T_{axcise}^{XG} + T_{VAT}^{XG} + T_{IT}^{XG}.$$

Запас Q_V^X теневого продукта V в секторе X прирастает за счет производства и убывает за счет поставок домашним хозяйствам и нелегальной продажи на внешний рынок:

$$\frac{dQ_V^X}{dt} = q_X Y_X - (a_V^{XL} + a_V^{XI}) Q_V^X.$$

Запас B^X теневых денег в секторе X прирастает при продаже теневого продукта населению L и на внешний рынок; часть b_B^X запаса теневых денег успевают отмыть, часть b_B^{XL} попадает в качестве штрафных санкций в доходную часть консолидированного бюджета правительства страны, а часть b_B^{XL} поступает населению в качестве теневых доходов:

$$\frac{dB^X}{dt} = (a_V^{XL} p_V^L + a_V^{XI} p_V^I) Q_V^X - (b_B^{XL} + b_B^X + b_B^{XG}) B^X.$$

Домашние хозяйства

Безработные в составе экономически активного населения Монголии в модели рассматриваются только в части добывающего сектора X , а в секторе подразделены на безработных в открытой части экономики и безработных теневой части экономики. Безработица увеличивается, если предложения труда превышает спрос на него и уменьшается в противном случае.

Предложение труда соответственно в открытой и теневой частях сектора X определяется долями людей трудоспособного возраста L , желающих работать.

$$Q_L^{LX} = a_L^{LX} L, \quad Q_V^{LX} = a_V^{LX} L.$$

Структура занятости такова: 4% трудоспособного населения (1,488 млн. человек по данным 2003 года) заняты в горно-добывающей промышленности. 6% - в производстве, 6,7% составляет безработица.

Считаем, что рост открытой и теневой ставок заработной платы в секторе X может происходить как при нехватке кадров, так и при росте потребительских цен на продукцию сектора.

$$\frac{ds_L^X}{dt} = \left[\alpha_L^X \left(\frac{b_L^{XL} W^X}{s_L^X} - Q_L^{LX} \right) + \frac{\beta_L^X s_L^X}{p_X^L} \left(\frac{b_X^{LX} W^L}{p_X^L} - a_X^{XL} Q^X \right) \right]_+, \text{ где } \beta_L^X = \delta \alpha_X^L$$

$$\frac{ds_V^X}{dt} = \left[\alpha_V^X \left(\frac{b_V^{XL} B^X}{s_V^X} - Q_V^{LX} \right) + \frac{\beta_V^X s_V^X}{p_V^L} \left(\frac{b_V^{LX} W^L}{p_V^L} - a_V^{XL} Q_V^X \right) \right]_+, \text{ где } \beta_B^X = \delta \alpha_V^L.$$

Здесь и далее используется следующее обозначение: $X_+ = X$, если $X > 0$ и $X_+ = 0$, если $X \leq 0$. Считаем, что доля прироста цен, отражающаяся на росте заработной платы, $\delta \in (0,1)$.

Изменение запаса $Q^X(t)$ конечного продукта X добывающего сектора Монголии, предназначенного агенту L (домашним хозяйствам), определяет изменение индекса потребительских цен p_X^L на продукцию X лесопромышленного комплекса:

$$\frac{dp_X^L}{dt} = \alpha_X^L \left(\frac{b_X^{LX} W^L}{p_X^L} - a_X^{XL} Q^X \right).$$

Изменение запаса $Q_V^X(t)$ теневого продукта X добывающего сектора Монголии, предназначенного агенту L (домашним хозяйствам), определяет изменение индекса потребительских цен p_V^L на продукцию X лесопромышленного комплекса:

$$\frac{dp_V^L}{dt} = \alpha_V^L \left(\frac{b_V^{LX} W^L}{p_V^L} - a_V^{XL} Q_V^X \right).$$

Считаем, что все деньги у домашних хозяйств являются чистыми независимо от источника поступления:

$$\frac{dW^L}{dt} = d^{BL} + b_L^{XL}W^X + b_V^{XL}B^X - (b_X^{LX} + b_V^{LX} + b_L^I)W^L - T^{LG} + T^{GL}.$$

Отчисления в консолидированный бюджет агента L (домашних) T^{LG} складываются из отчислений по таможенным платежам на импорт T_6^{LG} и отчислений по подоходному налогу T_7^{LG} с открытой части дохода.

$$T^{LG} = T_6^{LG} + T_7^{LG}, \quad T_{custom}^{LG} = n_{custom} b_L^I W^L.$$

Банковская система

Описана денежно-кредитная система и система налогообложения для добывающего сектора и домашних хозяйств Монголии с учетом экологических аспектов. Предполагается, что отрасль пользуется кредитными средствами. Максимально возможный объем кредита считается пропорциональным запасу капитала.

Мы предполагаем, что банковские активы складываются из золото-валютного резерва R и задолженности отрасли Z , а пассивы из депозитов собственников предприятий и банков D , которые подчиняются основному финансовому балансу банковской системы

$$\omega R(t) + Z(t) = D(t).$$

Считаем, что золотовалютные резервы меняются в силу платежного баланса

$$\frac{dR(t)}{dt} = \frac{\omega p_X^I E^X - (1 - n_{custom}) b_L^I W^L}{\omega}.$$

Активы банка равны задолженности Z^X отрасли X в оплате по кредиту

$$\frac{dZ^X}{dt} = C^{BX} + rZ^X - b_B^{XB} W^X.$$

Банковская система Монголии не является замкнутой, большую роль в инвестиционных решениях играют филиалы банков с участием иностранного капитала. Поэтому в качестве первого приближения при описании банковской системы будем считать, что часть золотовалютных резервов Монголии обеспечивают резервирование активов Монголии.

Чистая прибыль отрасли:

$$Inc = \omega p_X^I E^X + p_X^L a_X^{XL} Q^X - (b_B^{XB} + b_X^L) W^X - T_{custom}^{XG} - T_{social}^{XG} - T_{axcise}^{XG} - T_{VAT}^{XG},$$

где $d^B = r(t)Z(t)$ – банковская прибыль.

Предполагаем, что резерв банковской системы обеспечивает депозиты при законодательно установленной норме ξ . Таким образом, должно выполняться соотношение

$$\omega R(t) \geq \xi D(t).$$

Государство

Запас денег на счетах консолидированного бюджета прирастает от налоговых поступлений и убывает при трансфертах добывающему сектору экономики и домашним хозяйствам. В модели все налоги остаются на счетах консолидированного бюджета:

$$\frac{dW^G}{dt} = T^{XG} + T^{LG} - T^{GX} - T^{GL},$$

где $T^{GX} = b_X^{GX} W^G$ трансферты добывающему сектору, $T^{GL} = b_L^{GL} W^G$ выплаты по государственным социальным программам.

Текущий дефицит консолидированного бюджета Монголии $D^G(t)$ определяется правой частью $(T^{XG} + T^{LG} - T^{GX} - T^{GL})$ с обратным знаком:

$$D^G = -(T^{XG} + T^{LG} - T^{GX} - T^{GL}).$$

Технология идентификации параметров модели

Для проведения численных экспериментов над моделью проведена идентификация ее параметров, а также оптимизация результатов расчетов при последующей идентификации параметров. При выполнении этой задачи в первую очередь перед нами стояла задача сбора наиболее полных данных статистики, начиная с периода установления в стране рыночной экономики. Используя различные международные инвестиционные и статистические интернет ресурсы, мы осуществили сбор необходимой статистики.

Имея работоспособный вариант и статистические данные, необходимо было провести оптимизацию модели по набору параметров. В качестве критерия близости рассчитанных по модели рядов макропоказателей и статистических данных использован индекс Тейла.

Оптимизация производилась с применением следующего алгоритма:

- выбран набор параметров модели, по которым производится оптимизация (те параметры модели, которые невозможно определить напрямую из статистики);
- на основании наблюдений, полученных в ходе проведенных ранее численных экспериментов, для каждого выбранного параметра задана определенная окрестность допустимых значений (при выходе параметров за пределы множества допустимых значений результаты расчетов теряют интерполяционные свойства);
- реализована программа пересчета критерия близости полученных по модели временных рядов макропоказателей и статистических данных при изменении параметров от нижней до верхней границы окрестности

- допустимых значений с равномерным шагом параллельно по каждому параметру;
- выбраны значения параметров из окрестностей допустимых значений, для которых значение критерия минимально.

На данный момент программа пересчета, основанная на приведенном выше алгоритме, реализована на языке программирования C++ для набора из двух параметров a_X^{XL} , a_X^{XI} . Алгоритм оптимизации решения по параметрам описан следующим образом:

```

param1, param2 - два параметра для прогонки
minP1, maxP1, stepP1 - границы и шаг прогонки первого параметра
minP2, maxP2, stepP2 - границы и шаг прогонки второго параметра
*/
void ProgonkaPoDvumParametram( FILE* file,
    double& param1, double minP1, double maxP1, double stepP1,
    double& param2, double minP2, double maxP2, double stepP2 )
{
    for( double value1 = minP1; value1 <= maxP1; value1 += stepP1 )
    {
        for( double value2 = minP2; value2 <= maxP2; value2 += stepP2 )
        {
            SetStartValues();
            SetStartParams();
            param1 =  $a_X^{XL}$ ;
            param2 =  $a_X^{XI}$ ;
            DoCalculationCycle( nt, file );
            CalculateCoefficients();
            PrintCoefficients();
        }
    }
}

```

Данный метод протестирован на рабочем варианте модели. Результаты расчетов удовлетворительны, наблюдаются положительные тенденции при применении предложенного метода. Например, после проведения оптимизации по двум параметрам рассчитанный по модели временной ряд изменения запаса капитала в добывающем секторе стал значительно ближе к статистическим данным (Рис. 1.).

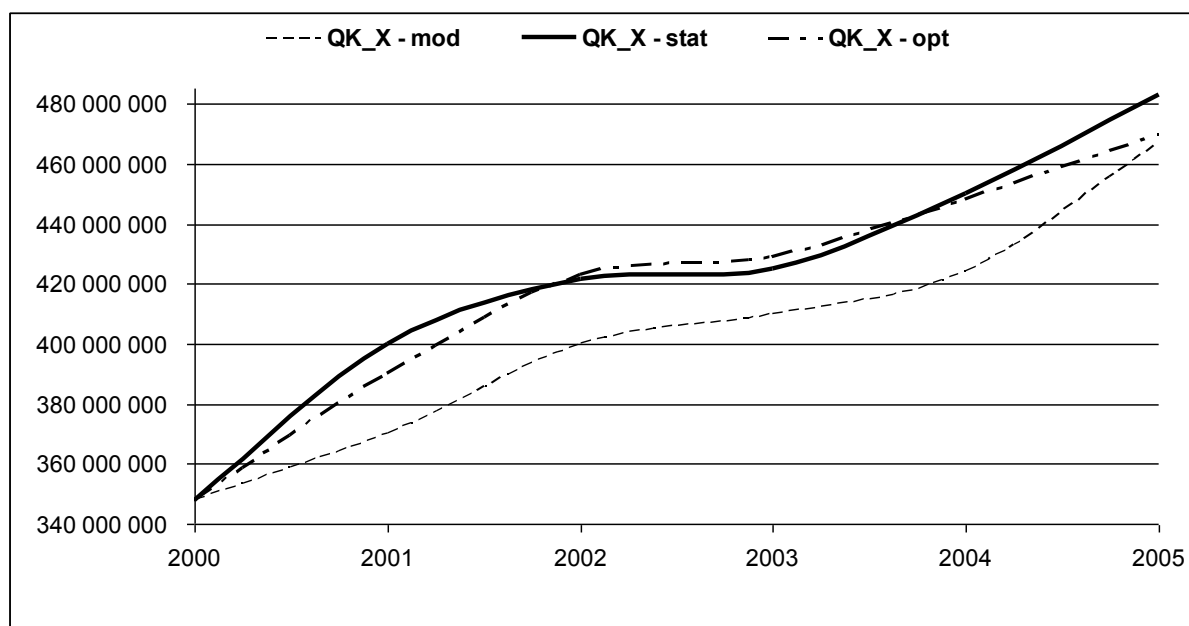


Рис. 1. Сравнение статистического, расчетного и оптимизированного временного ряда запасов капитала в добывающем секторе

Однако оптимизация по неполному перечню параметров не позволяет произвести оптимизацию модели в полной мере. Всего в модели несколько десятков параметров, по каждому из которых необходимо произвести оптимизацию. При этом, выдерживая частоту шага пересчета по каждому параметру на удовлетворительном уровне, мы сталкиваемся с невозможностью проведения численного эксперимента из-за технических ограничений вычислительной техники. Здесь нам на помощь приходят методы параллельных вычислений с применением суперкомпьютеров. Для решения проблемы проведения параллельных

расчетов для оптимизации модели добывающего сектора экономики Монголии планируется применить параллельные вычисления на кластерах ВЦ РАН им. Дородницына.

Литература

1. Горбачев В. А., Оленев Н. Н. Частное решение трехсекторной вычислимой модели экономики Кировской области. – Долгопрудный. Тр. 51 научн. конф. МФТИ. Ч. VII. Управление и прикладная математика. Т. 1. – 2008. – С. 47-50.
2. Дэмбэрэл С., Оленев Н. Н., Поспелов И. Г. Взаимодействие экономических и экологических процессов. – М. : ВЦ РАН, 2003. – 40 с.

Информация об авторах

Data on authors

Оленёв Николай Николаевич, к.ф.-м.н., доцент, старший научный сотрудник ВЦ РАН, доцент кафедры математического моделирования в экономике ФГБОУ ВПО «ВятГУ».

Тел.(499) 783-33-28.

Электронная почта: nolenev@mail.ru

Olenev Nicholas, docent, senior researcher at CCAS, Associate Prof. on Math Modeling in Economics FGEI HPE «VyatSU».

Tel: (499) 783-33-28

E-mail: nolenev@mail.ru

Горбачёв Владимир Александрович, к.ф.-м.н., НИУ РУДН, г. Москва

Тел. (499) 434-53-00

Электронная почта: vovastyi2005@yandex.ru

Gorbachev Vladimir, Cand. Of Physics & Mathematics, NRU RUFP, Moscow

Tel.: (499) 434-53-00

E-mail: vovastyi2005@yandex.ru

ПРОГРАММНЫЙ ПАКЕТ ВЫСОКОТОЧНЫХ МОДУЛЯРНО-ПОЗИЦИОННЫХ ВЫЧИСЛЕНИЙ С ПЛАВАЮЩЕЙ ТОЧКОЙ

HIGH-PRECISION SOFTWARE PACKAGE OF MODULAR-POSITIONAL FLOATING-POINT CALCULATIONS

В работе рассматривается архитектура нового программного пакета для высокоточных вычислений, основанного на модулярно-позиционном представлении чисел с плавающей точкой. Приводится структура и алгоритмы функционирования основных модулей пакета.

Architecture a new software package for high-precision computation which is based a modular-positional floating-point data representation, is considered. The block scheme and algorithms of the main modules of the package are proposed.

Ключевые слова: высокоточные вычисления; программный пакет; модулярно-позиционное представление чисел; параллельные алгоритмы.

Keywords: high-precision computation, software package, modular-positional number representation, parallel algorithms.

ВВЕДЕНИЕ

Практически все современные компьютерные системы, от персональных компьютеров до крупнейших суперкомпьютеров, реализуют 64-битную арифметику с плавающей точкой стандарта IEEE-754, согласно которой под мантиссу числа отводится 53 бита. Для большинства научных и инженерных приложений этого вполне достаточно, тем не менее, быстро увеличивается

число задач, для корректного решения которых необходима гораздо более высокая вычислительная точность [1].

Некоторые задачи требуют для корректного решения двукратного увеличения точности (разрядности мантиссы), другие – четырехкратного, а третьи требуют сотни или более десятичных цифр для получения численно значимых результатов. Такие задачи возникают в самых разнообразных областях знаний: в квантовой физике, при моделировании климата, в электромагнитной теории рассеяния, при изучении кулоновских систем на атомном уровне, в астрофизике, в экспериментальной математике, в вычислительной теории чисел и пр. [1-6].

Производительность (число элементарных операций, выполняемых в единицу времени) современных программных пакетов высокоточных вычислений (GMP, MPFR, MPFR++, HPA, ARPREC, QD, MPFUN90 и пр.) существенно падает при увеличении точности, определяемой разрядностью мантисс операндов.

Это связано с образованием длинных цепочек межзнаковых переносов в ходе выполнения основных арифметических операций над многоразрядными мантиссами чисел. Цепочки переносов требуют сложной алгоритмической обработки, а так же обуславливают принципиальную невозможность распараллеливания арифметических действий на уровне отдельных разрядов операндов.

Ведущие специалисты в области высокоточных вычислений приводят следующие оценки [1]: при вычислениях с четверной точностью (128-битная арифметика, разрядность мантиссы определяется 31 десятичной цифрой) наблюдается пятикратное снижение быстродействия, по сравнению с вычислениями в обычной 64-битной арифметике IEEE. При четырехкратном увеличении точности (восьмерная точность, мантисса состоит из 62 десятичных цифр) время счета увеличивается приблизительно в 25 раз. При использовании вычислений со 100 десятичными цифрами следует ожидать ≈ 50 -кратного увеличения времени решения задачи. Если точность p изменяется в интервале

от 100 до 1000 десятичных цифр (от 332 до 3322 бит), то вычислительные затраты возрастают как ρ^2 . Если $\rho > 1000$, то следует ожидать увеличения вычислительных затрат по закону $\rho \log \rho$ [1]. Проведенные экспериментальные исследования подтверждают данные оценки [7].

Таким образом, актуальной является **задача** разработки эффективных методов, алгоритмов и программных решений высокоточных компьютерных вычислений, обеспечивающих минимизацию зависимости скорости выполнения основных арифметических операций при увеличении точности представления операндов.

В данной работе приводится описание структуры и алгоритмов функционирования *программного пакета высокоточных модулярно-позиционных вычислений*, основанного на перспективной композиции двух различных способов компьютерного представления числовых данных – чисел с плавающей точкой и многомодульной системы остаточных классов, которая называется *модулярно-позиционным представлением с плавающей точкой*.

1. СТРУКТУРА ПРОГРАММНОГО ПАКЕТА ВЫСОКОТОЧНЫХ МОДУЛЯРНО-ПОЗИЦИОННЫХ ВЫЧИСЛЕНИЙ

Структурная схема предлагаемого программного пакета высокоточных модулярно-позиционных вычислений изображена на рисунке 1.

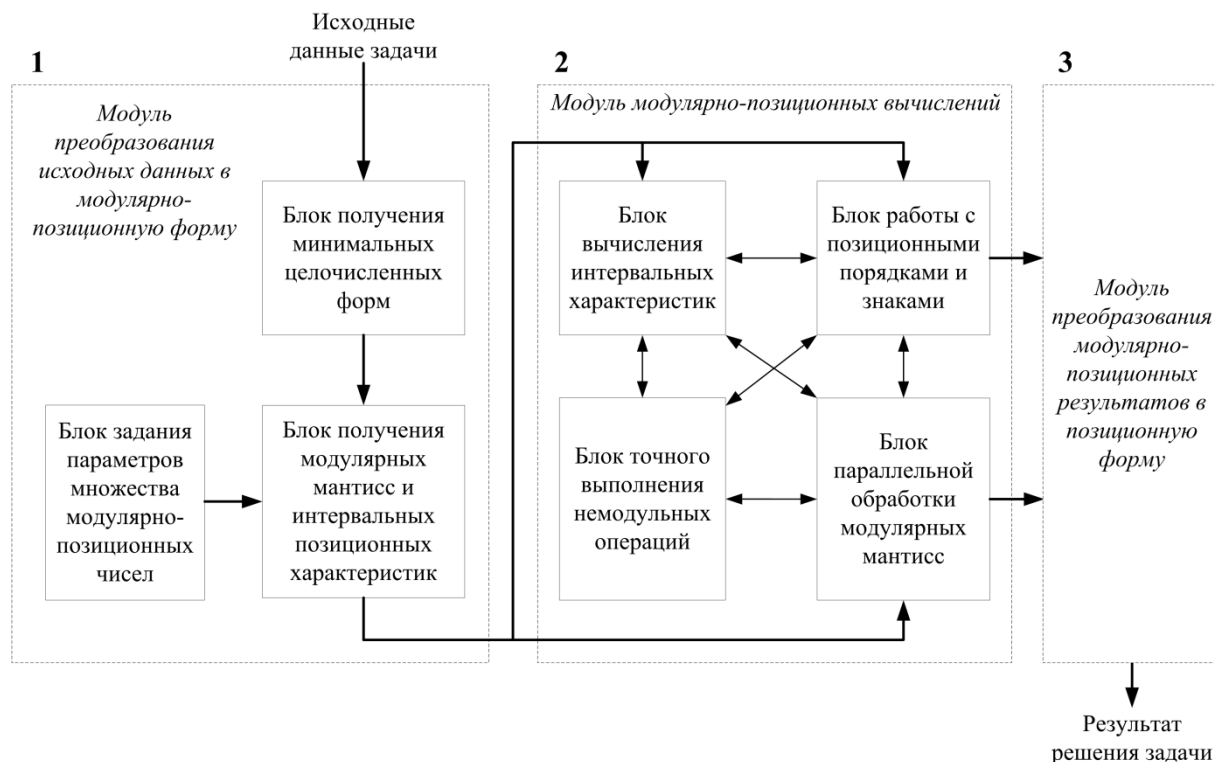


Рис. 1. Схематическая структура программного пакета модулярно-позиционных вычислений

1.1. Первый модуль программного пакета служит для преобразования исходных данных задачи, представленных в форме двоичных чисел с плавающей точкой (стандартных типов double, либо высокоточных представлений MPFR [8]), к модулярно-позиционному представлению.

1.2. Второй модуль (модуль модулярно-позиционных вычислений) является «ядром» предлагаемого решения. Он реализует параллельные алгоритмы выполнения основных арифметических операций над модулярно-позиционными числовыми данными, полученными в первом модуле. В основе его функционирования заложены программные решения поддержки параллельных вычислений для систем с общей памятью.

1.3. Третий модуль программного пакета служит для обратного преобразования полученных модулярно-позиционных результатов решения задачи в позиционную форму представления. Данный модуль содержит набор процедур, обеспечивающих получение удобочитаемого результата и его вывод в файл.

Рассмотрим основные принципы функционирования каждого из модулей более подробно.

2. МОДУЛЬ ПРЕОБРАЗОВАНИЯ ИСХОДНЫХ ДАННЫХ В МОДУЛЯРНО-ПОЗИЦИОННУЮ ФОРМУ

Данный модуль структурно выполнен в виде трех блоков: блока получения минимальных целочисленных форм, блока задания параметров множества модулярно-позиционных чисел и блока получения модулярных мантисс и интервальных позиционных характеристик. Рассмотрим каждый из них более подробно.

2.1. Блок получения минимальных целочисленных форм, который выполняет следующие функции.

2.1.1. Принимает на вход исходные данные решаемой задачи – множество двоичных чисел с плавающей точкой вида

$$x' \triangleq (\mu, e, s) = -1^s \cdot \mu \cdot 2^e, \quad (1)$$

где $\mu = (d_\rho \cdot d_{\rho-1} \dots d_2 d_1)_2$ – рациональная мантисса, определяемая как дробное двоично-рациональное число со значением $\mu = d_\rho + d_{\rho-1}2^{-1} + \dots + d_12^{1-\rho}$; e – порядок (экспонента) – целое число со знаком, изменяющееся в интервале значений $[e_{\min}, e_{\max}]$, а $s \in \{0, 1\}$ – знак числа x' . Запись $x' \triangleq (\mu, e, s)$ следует читать: x' определяется тройкой (μ, e, s) .

2.1.2. Преобразует принятые двоичные числа вида (1) к минимальной целочисленной форме:

$$x'' \triangleq (\mu', \lambda, s) = -1^s \cdot \mu' \cdot 2^\lambda, \quad (2)$$

где μ' – нечетное целое число без знака, называемое минимальной целочисленной мантиссой, такое, что $\mu' = \mu \cdot 2^{\rho - \tau(\mu) - 1}$; ρ – разрядность исходной двоичной рациональной мантиссы μ числа x' вида (1), а $\tau(\mu)$ – числовая избыточность, значение которой равно количеству младших нулевых бит в двоичном представлении $\mu = (d_\rho \cdot d_{\rho-1} \dots d_2 d_1)_2$; $\lambda = e - \rho + \tau(\mu) + 1$ – целое число со знаком, называемое перемещенным позиционным порядком, e – исходный порядок двоичного числа с плавающей точкой x' ; s – знак числа x' .

Представление всякого двоичного числа с плавающей точкой (1) в минимальной целочисленной форме (2) обеспечивает возможность дальнейшего преобразования его мантиссы в систему остаточных классов (СОК). За счет отсечения младших нулевых бит мантиссы обеспечивается, что полученное после преобразования остаточное представление будет минимальным.

2.1.3. Передает полученное множество минимальных целочисленных форм (2), изоморфное множеству исходных данных задачи, на вход блока получения модулярных мантисс и интервальных позиционных характеристик.

2.2. Блок задания параметров множества модулярно-позиционных чисел, основной функцией которого является определение базиса системы остаточных классов [9, 10], в которой будут выполняться все последующие вычисления. В данном блоке выполняются следующие функции.

2.2.1. Задаются основания (модули) СОК, обеспечивающие требуемую вычислительную точность. Для этого выбирается ряд натуральных чисел $p_1, p_2, \dots, p_n$, удовлетворяющих условиям

$$\gcd(p_i, p_j) = 1, \quad \forall i, j \in 1, 2, \dots, n \mid i \neq j,$$

где $\gcd(p_i, p_j)$ – наибольший общий делитель для p_i и p_j . Выбор оснований $p_1, p_2, \dots, p_n$ выполняется исходя из условия:

$$\log_2(p_1 \cdot p_2 \cdot \dots \cdot p_n) \geq 1 - \log_2 \xi,$$

где $\xi < 1$ – заданный предел относительной ошибки представления. Определение попарной взаимной простоты модулей выполняется при помощи алгоритма Евклида.

2.2.2. Для выбранных модулей $p_1, p_2, \dots, p_n$ выполняется вычисление мультипликативных инверсий (констант СОК) $\left| P_i^{-1} \right|_{p_i}, i = 1, 2, \dots, n$ по формуле

$$\left| P_i^{-1} \right|_{p_i} \equiv P_i^{\varphi(p_i)-1} \pmod{p_i},$$

где $P_i = P / p_i$, а $\varphi(p_i)$ – функция Эйлера.

2.2.3. Модули $p_1, p_2, \dots, p_n$ и вычисленные мультипликативные инверсии $\left| P_1^{-1} \right|_{p_1}, \left| P_2^{-1} \right|_{p_2}, \dots, \left| P_n^{-1} \right|_{p_n}$ передаются на вход блока получения модулярных мантисс и интервальных позиционных характеристик.

2.3. Блок получения модулярных мантисс и интервальных позиционных характеристик, который принимает на вход исходные данные задачи – конечное множество $\{x'' \equiv (\mu', \lambda, s)\}$ минимальных целочисленных форм (2) (результат работы первого блока), основания СОК и мультипликативные инверсии (результат работы второго блока). Данный блок выполняет следующие функции.

2.3.1. Преобразует целочисленные мантиссы μ' в систему остаточных классов с основаниями $p_1, p_2, \dots, p_n$:

$$\mu' \xrightarrow{СОК} \tilde{\mu} = \langle m_1, m_2, \dots, m_n \rangle, \quad (3)$$

где кортеж $\tilde{\mu}$ – *модулярная мантисса*, знакопозиции m_i которой, называемые *модулярными разрядами*, представляют собой наименьшие неотрицательные остатки от деления минимальной целочисленной мантиссы μ' на соответствующее основание СОК:

$$m_i \equiv \mu' \pmod{p_i}, \quad i = 1, 2, \dots, n.$$

2.3.2. Для каждой модулярной мантиссы (3), полученной на предыдущем шаге, вычисляется интервальная позиционная характеристика [11] – замкнутый вещественный интервал $I_{\tilde{\mu}/P} = \left[\underline{\tilde{\mu}/P}, \overline{\tilde{\mu}/P} \right]$, представленный парой чисел-

границ, удовлетворяющих условию $\underline{\tilde{\mu}/P} \leq C_{\tilde{\mu}} \leq \overline{\tilde{\mu}/P}$, где $C_{\tilde{\mu}} = \tilde{\mu}/P$ – точное значение отношения модулярной мантиссы $\tilde{\mu}$ к произведению модулей СОК $P = p_1 \cdot p_2 \cdot \dots \cdot p_n$. Границы интервальной характеристики определяются следующим образом [11, 12]:

$$\underline{\tilde{\mu}/P} \approx \left\lfloor \sum_{i=1}^n \downarrow \left(\frac{|m_i \cdot |P_i^{-1}|_{p_i}|_{p_i}}{p_i} \right) \right\rfloor_1, \quad \overline{\tilde{\mu}/P} \approx \left\lceil \sum_{i=1}^n \uparrow \left(\frac{|m_i \cdot |P_i^{-1}|_{p_i}|_{p_i}}{p_i} \right) \right\rceil_1, \quad (4)$$

где $m_1, m_2, \dots, m_n$ – разряды модулярной мантиссы, $|P_i^{-1}|_{p_i}$ – i -ая мультипликативная инверсия, операторы $\downarrow$ и $\uparrow$ означают выполнение округления «вниз» (с недостатком) и «вверх» (с избытком) соответственно, а $| \cdot |_1$ определяет операцию усечения целой части аргумента, т. е. возвращает только его дробную часть.

Таким образом, в результате работы модуля преобразования исходных данных в модулярно-позиционную форму каждое двоичное число с плавающей точкой (1) представляется в виде следующего *модулярно-позиционного числа*:

$$\tilde{x} \triangleq (\tilde{\mu}, I_{\tilde{\mu}/P}, \lambda, s), \quad (5)$$

где $\tilde{\mu} = \langle m_1, m_2, \dots, m_n \rangle$ – модулярная мантисса (3), представленная в системе остаточных классов (СОК) [9, 10] с основаниями $p_1, p_2, \dots, p_n$; $I_{\tilde{\mu}/P}$ – интервальная позиционная характеристика (4) модулярной мантиссы [11], позволяющая повысить эффективность выполнения основных проблемных операций в СОК, таких как сравнение модулярных мантисс по величине, определение знака модулярной мантиссы – результата выполненной операции алгебраического сложения, контроль переполнения модулярных мантисс при сложении; λ – позиционный порядок (экспонента числа), определенный в блоке получения минимальных целочисленных форм, а s – знак модулярно-позиционного числа $\tilde{x}$.

Значение модулярно-позиционного числа (5) формально определяется произведением

$$\tilde{x} = -1^s \cdot \left| \sum_{i=1}^n m_i B_i \right|_p \cdot 2^\lambda,$$

где $B_1, B_2, \dots, B_n$ – ортогональные базисы [9] системы остаточных классов.

Основными достоинствами модулярно-позиционного представления (5) являются следующие положения.

1. Диапазон изменения модулярной мантиссы $\tilde{\mu}$ определяется интервалом $[0, P-1]$, где $P = p_1 \cdot p_2 \cdot \dots \cdot p_n$ – произведение выбранных модулей СОК. Соответственно, количество модулей СОК самым непосредственным образом определяет разрядность позиционного эквивалента $\left| \sum_{i=1}^n m_i B_i \right|_p$ модулярной мантиссы (точность представления), который будет содержать $\rho = \lceil \log_2(p_1 \cdot p_2 \cdot \dots \cdot p_n) \rceil$ значащих разрядов.

2. Модульные операции над модулярными мантиссами, такие как арифметическое сложение, умножение и вычитание для случая, когда уменьшаемое больше вычитаемого, при использовании представления чисел в виде $\tilde{x} \hat{=} (\tilde{\mu}, I_{\tilde{\mu}/P}, \lambda, s)$ могут выполняться *параллельно по всем разрядам* мантиссы, поэтому реализуются эффективно и не требуют больших вычислительных затрат.

3. Для эффективного выполнения немодульных операций, таких как сравнение модулярных мантисс по величине, определение знака мантиссы – результата операции, контроль переполнения при сложении, умножении и пр., в представлении модулярно-позиционного числа (5) включена интервальная характеристика, позволяющая за постоянное время, не зависящее от количества модулей СОК, оценить относительную величину модулярной мантиссы. При этом выполнение перечисленных выше немодульных операций сводится к выполнению аналогичных по смыслу операций в стандартной машинной арифметике с плавающей точкой (каждая граница интервальной характеристики (4) представляется обычным 64-битным числом с плавающей точкой). Операции над интервальными позиционными характеристиками могут быть выполнены *параллельно* с операциями над модулярными мантиссами.

4. В представлении модулярно-позиционного числа (5) в явном виде присутствует порядок (экспонента) λ . Это позволяет расширить аппроксимируемый диапазон вещественных чисел, по сравнению с форматом представления с фиксированной точкой. Благодаря двоичному представлению, операции сравнения и вычитания порядков выполняются тривиальным образом.

3. МОДУЛЬ МОДУЛЯРНО-ПОЗИЦИОННЫХ ВЫЧИСЛЕНИЙ. ОСНОВНЫЕ АЛГОРИТМЫ МОДУЛЯРНО-ПОЗИЦИОННЫХ ВЫЧИСЛЕНИЙ

В модуле модулярно-позиционных вычислений реализуются следующие бинарные операции над модулярно-позиционными числами: выравнивание порядков, сложение, вычитание, умножение, деление, сравнение, округление. Остановимся более подробно на рассмотрении алгоритмов выполнения наиболее важных из них.

3.1. Основные алгоритмы модулярно-позиционных вычислений.

Пусть $\tilde{x} \triangleq (\tilde{\mu}_{\tilde{x}}, I_{\tilde{\mu}_{\tilde{x}}/P}, \lambda_{\tilde{x}}, s_{\tilde{x}})$ и $\tilde{y} \triangleq (\tilde{\mu}_{\tilde{y}}, I_{\tilde{\mu}_{\tilde{y}}/P}, \lambda_{\tilde{y}}, s_{\tilde{y}})$ – два модулярно-позиционных числа вида (5), $\tilde{\mu}_{\tilde{x}} = \langle x_1, x_2, \dots, x_n \rangle$ и $\tilde{\mu}_{\tilde{y}} = \langle y_1, y_2, \dots, y_n \rangle$ – их модулярные мантиссы, представленные в системе остаточных классов с модулями $p_1, p_2, \dots, p_n$, а $I_{\tilde{\mu}_{\tilde{x}}/P} = [\underline{\tilde{\mu}_{\tilde{x}}/P}, \overline{\tilde{\mu}_{\tilde{x}}/P}]$ и $I_{\tilde{\mu}_{\tilde{y}}/P} = [\underline{\tilde{\mu}_{\tilde{y}}/P}, \overline{\tilde{\mu}_{\tilde{y}}/P}]$ – интервальные характеристики модулярных мантисс.

Алгоритм выравнивания порядков чисел $\tilde{x}$ и $\tilde{y}$ состоит из следующей последовательности шагов.

Шаг 1. Определяется разность порядков выравниваемых операндов $\Delta\lambda = \lambda_{\tilde{x}} - \lambda_{\tilde{y}}$.

Шаг 2. Определяется выравнивающая величина $b = 2^{|\Delta\lambda|}$.

Шаг 3.

величину b :

$$\tilde{\mu} = \langle m_1, m_2, \dots, m_n \rangle = \begin{cases} \langle |x_1 b|_{p_1}, |x_2 b|_{p_2}, \dots, |x_n b|_{p_n} \rangle, & \text{если } \Delta\lambda > 0, \\ \langle |y_1 b|_{p_1}, |y_2 b|_{p_2}, \dots, |y_n b|_{p_n} \rangle, & \text{если } \Delta\lambda < 0. \end{cases}$$

Шаг 4. Параллельно с шагом 3 выполняется вычисление относительной величины модулярной мантиис :

$$I_{\tilde{\mu}/P} = \left[\underline{\tilde{\mu}/P}, \overline{\tilde{\mu}/P} \right] = \begin{cases} I_{\tilde{\mu}_{\tilde{x}}/P} \cdot b = \left[\underline{\tilde{\mu}_{\tilde{x}}/P} \cdot b, \overline{\tilde{\mu}_{\tilde{x}}/P} \cdot b \right], & \text{если } \Delta\lambda > 0, \\ I_{\tilde{\mu}_{\tilde{y}}/P} \cdot b = \left[\underline{\tilde{\mu}_{\tilde{y}}/P} \cdot b, \overline{\tilde{\mu}_{\tilde{y}}/P} \cdot b \right], & \text{если } \Delta\lambda < 0. \end{cases}$$

Шаг 5. Контроль переполнения диапазона СОК при умножении осуществляется следующим образом.

5.1) Проверяется выполнение условия $\overline{\tilde{\mu}/P} < 1$. Если условие выполнено, то при умножении на шаге 3 переполнения не произошло. В противном случае осуществляется переход к шагу 5.2.

5.2) Проверяется выполнение условия $\underline{\tilde{\mu}/P} \geq 1$. Если условие выполнено, то при умножении на шаге 3 произошло переполнение. Данная ситуация обрабатывается должным образом. В противном случае осуществляется переход к шагу 5.3.

5.3) Если ни одно из условий, проверяемых на шагах 5.1 и 5.2, не выполнено, то необходимо переопределить интервальные характеристики $I_{\tilde{\mu}_{\tilde{x}}/P}$ и $I_{\tilde{\mu}_{\tilde{y}}/P}$ по выражению (4), повысив точность вычисления их границ, после чего перейти к шагу 4, либо выполнить сравнение полученной на шаге 3 модулярной мантииссы $\tilde{\mu} = \langle m_1, m_2, \dots, m_n \rangle$ с исходной мантииссой $\tilde{\mu}_{\tilde{x}} = \langle x_1, x_2, \dots, x_n \rangle$ или $\tilde{\mu}_{\tilde{y}} = \langle y_1, y_2, \dots, y_n \rangle$ (в зависимости от знака модуля разности порядков $\Delta\lambda$) при помощи точного алгоритма, например потактового перехода к коду ОПСС [10]. При этом, если $\tilde{\mu} \leq \tilde{\mu}_{\tilde{x}}$, то при вычислении $\tilde{\mu}$ на шаге 3 произошло

переполнение и алгоритм завершает свою работу с соответствующим кодом ошибки.

Шаг 6. В качестве общего для операндов порядка λ выбирается меньший из исходных порядков:

$$\lambda = \min\{\lambda_{\tilde{x}}, \lambda_{\tilde{y}}\}.$$

Шаг 7. Выполняется присваивание:

$$\begin{cases} \tilde{\mu}_{\tilde{x}} = \tilde{\mu}, I_{\tilde{\mu}_{\tilde{x}}/P} = I_{\tilde{\mu}/P}, \lambda_{\tilde{x}} = \lambda, & \text{если } \Delta\lambda > 0, \\ \tilde{\mu}_{\tilde{y}} = \tilde{\mu}, I_{\tilde{\mu}_{\tilde{y}}/P} = I_{\tilde{\mu}/P}, \lambda_{\tilde{y}} = \lambda, & \text{если } \Delta\lambda < 0. \end{cases}$$

В результате работы алгоритма оба операнда будут представлены с одинаковым порядком, поэтому могут быть сложены, либо вычтены.

Ярусно-параллельная форма графа рассмотренного алгоритма выравнивания порядков представлена на рисунке 2. Значения рядом с вершиной S_i указывают на трудоемкость $T(S_i)$ соответствующей вершины (трудоемкость – количество элементарных операций, необходимых для перехода на следующую вершину). Трудоемкость всего алгоритма определяется длиной критической цепочки. Дуги, соответствующие критической цепочке, выделены утолщенными линиями.

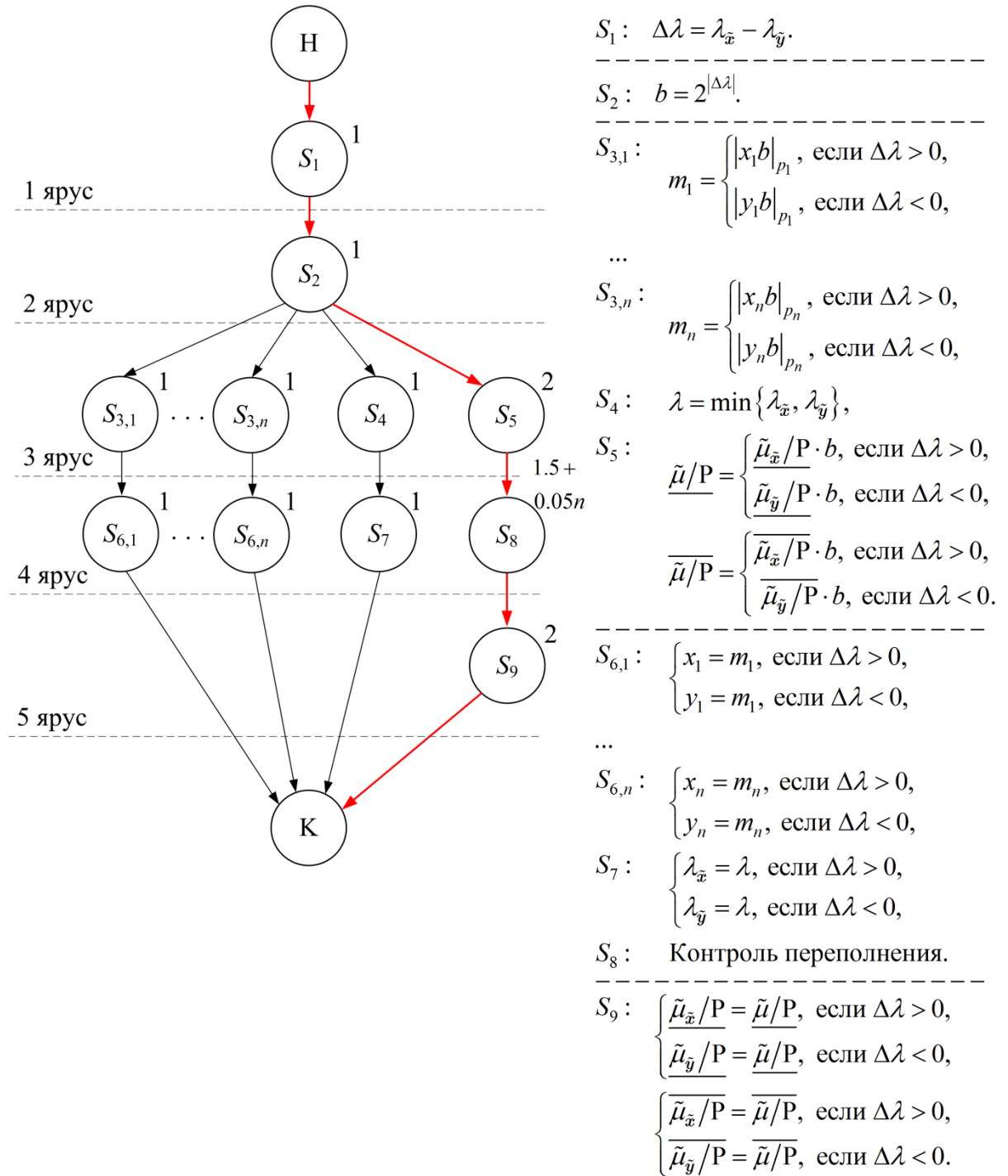


Рис. 2. Ярусно-параллельная форма графа алгоритма выравнивания порядков модулярно-позиционных чисел

Алгоритм арифметического сложения модулярно-позиционных чисел

$\tilde{x} \hat{=} (\tilde{\mu}_{\tilde{x}}, I_{\tilde{\mu}_{\tilde{x}}/P}, \lambda_{\tilde{x}}, s_{\tilde{x}})$ и $\tilde{y} \hat{=} (\tilde{\mu}_{\tilde{y}}, I_{\tilde{\mu}_{\tilde{y}}/P}, \lambda_{\tilde{y}}, s_{\tilde{y}})$ при условии, что $s_{\tilde{x}} = s_{\tilde{y}}$, состоит из следующих шагов.

Шаг 1. Операнды выравниваются до меньшего порядка. В качестве результатного порядка $\lambda_{\tilde{z}}$ выбрать меньший из исходных порядков.

Шаг 2. Значения знакопозиций модулярных мантисс выравненных слагаемых попарно параллельно складываются по соответствующим модулям $p_1, p_2, \dots, p_n$:

$$\tilde{\mu}_{\tilde{z}} = \langle z_1, z_2, \dots, z_n \rangle = \tilde{\mu}_{\tilde{x}} + \tilde{\mu}_{\tilde{y}} = \langle |x_1 + y_1|_{p_1}, |x_2 + y_2|_{p_2}, \dots, |x_n + y_n|_{p_n} \rangle.$$

Шаг 3. Параллельно с шагом 2 определяется интервальная позиционная характеристика результатной модулярной мантиссы по правилам арифметики интервалов [13]:

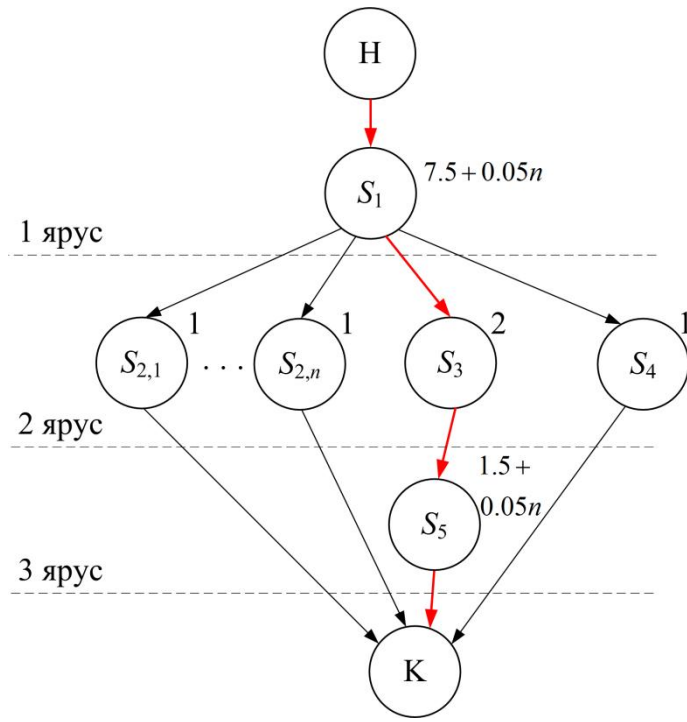
$$I_{\tilde{\mu}_{\tilde{z}}/P} = [\underline{\tilde{\mu}_{\tilde{z}}/P}, \overline{\tilde{\mu}_{\tilde{z}}/P}] = I_{\tilde{\mu}_{\tilde{x}}/P} + I_{\tilde{\mu}_{\tilde{y}}/P} = \left[\underline{\tilde{\mu}_{\tilde{x}}/P} + \underline{\tilde{\mu}_{\tilde{y}}/P}, \overline{\tilde{\mu}_{\tilde{x}}/P} + \overline{\tilde{\mu}_{\tilde{y}}/P} \right].$$

Шаг 4. Контроль переполнения диапазона СОК при сложении модулярных мантисс осуществляется по аналогии с шагом 5 алгоритма выравнивания порядков и состоит в следующем тривиальном сопоставлении границ интервала-суммы $I_{\tilde{\mu}_{\tilde{z}}/P}$ с единицей.

Шаг 5. Определяется знак суммы:

$$s_{\tilde{z}} = s_{\tilde{x}} \wedge s_{\tilde{y}}.$$

В результате выполнения данного алгоритма получается модулярно-позиционная сумма $\tilde{z} = \tilde{x} + \tilde{y} \triangleq (\tilde{\mu}_{\tilde{z}}, I_{\tilde{\mu}_{\tilde{z}}/P}, \lambda_{\tilde{z}}, s_{\tilde{z}})$. Ярусно-параллельный граф алгоритма сложения модулярно-позиционных чисел представлен на рисунке 3.



S_1 : Выравнивание порядков.

$S_{2,1}$: $z_1 = |x_1 + y_1|_{p_1}$,

...

$S_{2,n}$: $z_n = |x_n + y_n|_{p_n}$,

S_3 : $I_{\tilde{\mu}_{\tilde{z}}/P} = \begin{cases} \tilde{\mu}_{\tilde{z}}/P = \tilde{\mu}_{\tilde{x}}/P + \tilde{\mu}_{\tilde{y}}/P, \\ \tilde{\mu}_{\tilde{z}}/P = \tilde{\mu}_{\tilde{x}}/P + \tilde{\mu}_{\tilde{y}}/P, \end{cases}$

S_4 : $s_{\tilde{z}} = s_{\tilde{x}} \wedge s_{\tilde{y}}$.

S_5 : Контроль переполнения.

Рис. 3. Ярусно-параллельный граф алгоритма арифметического сложения модулярно-позиционных чисел

Алгоритм умножения чисел $\tilde{x} \triangleq (\tilde{\mu}_{\tilde{x}}, I_{\tilde{\mu}_{\tilde{x}}/P}, \lambda_{\tilde{x}}, s_{\tilde{x}})$ и

$\tilde{y} \triangleq (\tilde{\mu}_{\tilde{y}}, I_{\tilde{\mu}_{\tilde{y}}/P}, \lambda_{\tilde{y}}, s_{\tilde{y}})$ заключается в следующем.

Шаг 1. Вычисляется произведение модулярных мантисс сомножителей по правилам выполнения умножения в системе остаточных классов [9, 10]:

$$\tilde{\mu}_{\tilde{z}} = \langle z_1, z_2, \dots, z_n \rangle = \tilde{\mu}_{\tilde{x}} \cdot \tilde{\mu}_{\tilde{y}} = \langle |x_1 y_1|_{p_1}, |x_2 y_2|_{p_2}, \dots, |x_n y_n|_{p_n} \rangle.$$

Шаг 2. Порядки операндов складываются:

$$\lambda_{\tilde{z}} = \lambda_{\tilde{x}} + \lambda_{\tilde{y}}.$$

Шаг 3. Определяется интервальная характеристика результатной модулярной мантиссы:

$$I_{\tilde{\mu}_{\tilde{z}}/P} = \frac{I_{\tilde{\mu}_{\tilde{x}}/P} \cdot I_{\tilde{\mu}_{\tilde{y}}/P}}{I_{1/P}} = \left[\frac{\tilde{\mu}_{\tilde{x}}/P \cdot \tilde{\mu}_{\tilde{y}}/P}{1/P}, \frac{\overline{\tilde{\mu}_{\tilde{x}}/P} \cdot \overline{\tilde{\mu}_{\tilde{y}}/P}}{1/P} \right],$$

где $I_{1/P} = \left[\underline{1/P}, \overline{1/P} \right]$ – заранее вычисленный корректирующий интервал,

аппроксимирующий точное значение величины $\frac{1}{P}$.

определяется необходимостью деления модулярных мантисс (а это одна из самых сложных операций в системе остаточных классов).

3.2. Назначение отдельных блоков модуля модулярно-позиционных вычислений.

Реализация приведенных выше алгоритмов реализуется посредством следующих четырех блоков рассматриваемого модуля модулярно-позиционных вычислений: блока параллельной обработки модулярных мантисс, блока работы с позиционными порядками и знаками, блока вычисления интервальных характеристик, блока точного выполнения немодульных операций над модулярными мантиссами.

3.2.1. *Блок параллельной обработки модулярных мантисс* содержит процедуры для выполнения основных модульных операций: сложение, вычитание, умножение, умножение на константу и пр. Программные методы данного блока реализуют SIMD-модель вычислений, которой естественным образом отвечают принципы представления и обработки чисел в системе остаточных классов.

3.2.2 *Блок работы с позиционными порядками и знаками* содержит базовые программные методы сложения и вычитания порядков, анализа знаков модулярно-позиционных чисел и пр. Блок функционирует в стандартной 64-битной целочисленной арифметике.

3.2.3. *Блок вычисления интервальных позиционных характеристик*, основным назначением которого является выполнение основных немодульных операций над модулярными мантиссами: контроль переполнения, определение знака, сравнение. В данном блоке к интервальным характеристикам (5) применяются основные арифметические операции сложения, вычитания, умножения и деления, аналогичные стандартным операциям в арифметике интервалов [13]. При этом корректность результата выполнения немодульной операции определяется однозначным образом, путем проверки необходимого и достаточного условий [11]. Блок функционирует в стандартной 64-битной арифметике с плавающей точкой стандарта IEEE-754 [14].

3.2.4 Блок точного выполнения немодульных операций, реализующий выполнение основных немодульных операций над модулярными мантиссами методом потактового перехода к коду ОПСС, а так же метод итераций Ньютона для деления модулярных мантисс. Обращение к процедурам данного блока осуществляется только в том случае, если интервальные позиционные характеристики, вычисляемые в предыдущем блоке, не позволяют однозначно определить результат выполнения немодульной операции над модулярными мантиссами.

4. МОДУЛЬ ПРЕОБРАЗОВАНИЯ МОДУЛЯРНО-ПОЗИЦИОННЫХ РЕЗУЛЬТАТОВ В ПОЗИЦИОННУЮ ФОРМУ

Данный модуль содержит процедуры, для преобразования полученных с помощью процедур второго модуля результатов решения численной задачи к 64-битному двоичному формату с плавающей точкой стандарта IEEE-754. В основе алгоритмов функционирования данного модуля лежат базовые положения Китайской теоремы об остатках (К.Т.О.) [10], согласно которой позиционное представление модулярной мантиссы $\tilde{\mu} = \langle m_1, m_2, \dots, m_n \rangle$ модулярно-позиционного числа вида (5) определяется следующим образом:

$$\tilde{\mu} = \langle m_1, m_2, \dots, m_n \rangle \xrightarrow{К.Т.О.} \mu' = \sum_{i=1}^n m_i B_i, \quad (6)$$

где $B_1, B_2, \dots, B_n$ – ортогональные базисы [9] СОК.

Для корректного выполнения преобразования (6) используется высокоточная арифметика.

ЗАКЛЮЧЕНИЕ

1. Предложена архитектура нового программного пакета высокоточных вычислений, основная концепция которого заключается в использовании многомодульной системы остаточных классов для увеличения разрядности мантисс чисел с плавающей точкой. Для решения проблем, связанных со сложностью выполнения немодульных процедур в системе остаточных классов,

используются интервальные позиционные характеристики модулярных чисел (мантисс). Все немодульные операции проверяются на корректность, что позволяет избежать получения существенных ошибок в ходе выполнения численного алгоритма.

2. Предлагаемый программный пакет содержит интерфейс языка C++, что позволяет минимизировать сложности при его использовании в готовых приложениях, требующих повышенной вычислительной точности.

3. Прототипом предлагаемого пакета являются разработанные авторами ранее программные решения для выполнения высокоточных матричных вычислений с использованием системы остаточных классов [15, 16]. Опыт экспериментального исследования их быстродействия предопределил следующие особенности реализации предлагаемого в настоящей работе программного пакета:

- необходимость использования эффективных методов и алгоритмов для выполнения немодульных операций в СОК. Предложенные ранее решения [15, 16] имеют ограниченную применимость из-за сложности выполнения таких операций, как сравнение, контроль переполнения, вычитание с возможностью определения знака и пр.

- использование максимума аппаратных возможностей современных аппаратных платформ. Например, для реализации направленного округления при вычислении границ интервальной позиционной характеристики по формулам (4), используются низкоуровневые инструкции управления режимами работы FPU. Это позволяет, во-первых, выполнить округление в соответствии со стандартом IEEE-754, а во-вторых, на порядок ускорить данную процедуру, по сравнению с ее алгоритмической реализацией.

Литература

1. Bailey, D. H. High-Precision Floating-Point Arithmetic in Scientific Computation / D. H. Bailey // Computing in Science and Engineering. – 2005. – Vol. 7, Issue 3. – P. 54-61.

2. Bailey, D. H. High-Precision Computation: Mathematical Physics and Dynamics / D. H. Bailey, R. Barrio, J. M. Borwein // Applied Mathematics and Computation. – 2012. – Vol. 218, Issue 20. – P. 10106-10121.

3. He, Y. Using Accurate Arithmetics to Improve Numerical Reproducibility and Stability in Parallel Applications / Y. He, C. Ding // Journal of Supercomputing. – 2001. - Vol. 18, No. 3. – P. 259-277.

4. Lake, G. From Sir Isaac to the Sloan survey: Calculating the structure and chaos due to gravity in the universe / G. Lake, T. Quinn D. C. Richardson // Proceedings of the 8th ACM-SIAM Symposium on Discrete Algorithms. – Philadelphia : SIAM, 1997. – 10 p.

5. Hauschildt, P. H. The numerical solution of the expanding Stellar atmosphere problem [Electronic resource] / P. H. Hauschildt, E. Baron. – Режим доступа:

<http://hobbes.hs.uni-hamburg.de/PAPERS/jcam-rev2/jcam-rev2.html>

Режим доступа свободный. – Загл. с экрана.

6. Frolov, A. M. Highly Accurate Evaluation of the Few-Body Auxiliary Functions and Four-Body Integrals / A. M. Frolov, D. H. Bailey // Journal of Physics B: Atomic, Molecular and Optical Physics. – 2003. – Vol. 36, No. 9. – P. 1857-1867.

7. Исупов, К. С. Исследование эффективности современных средств поддержки высокоточных вычислений с вещественными числами [Электронный ресурс] / К. С. Исупов, А. Г. Иванов // Общество, наука, инновации (НТК-2012) : Сб. материалов ежегодной всероссийской научно-технической конференции (16-27 апреля 2012 г., г. Киров). – Киров, 2012. – 11 с. – ISBN: 978-5-9822-8061-9. – 1 эл. опт. диск (CD-ROM)

8. The GNU MPFR Library [Electronic resource]. – Mode of access:

<http://www.mpfr.org/> (15.06.2011)

Режим доступа свободный. – Загл. с экрана.

9. Акушский, И. Я. Машинная арифметика в остаточных классах / И. Я. Акушский, Д. И. Юдицкий. – М.: Сов. Радио, 1968. – 440 с.

10. Omondi, A. Residue Number Systems: Theory and Implementation (Advances in Computer Science and Engineering Texts) / A. Omondi, B. Premkumar. – London: Imperial College Press, 2007. – 312 p.
11. Исупов, К. С. Метод выполнения немодульных операций в системе остаточных классов на основе интервальных позиционных характеристик / К. С. Исупов // Фундаментальные исследования. – 2013. – № 4 (3). – С. 566-570.
12. Исупов, К. С. Способ уточненного вычисления приближенной позиционной характеристики для выполнения немодульных операций в системе остаточных классов / К. С. Исупов // Фундаментальные исследования. – 2013. – № 4 (4). – С. 796-800.
13. Кулиш, У. Достоверные вычисления. Базовые численные методы / У. Кулиш, Д. Рац, Р. Хаммер, М. Хокс ; пер. с англ. А. Яковлев. – Москва–Ижевск : НИЦ «Регулярная и хаотическая динамика», 2005. – 496 с.
14. IEEE Standard for Floating-Point Arithmetic. – Introduced 2008-08-29. – New York : Institute of Electrical and Electronics Engineers, 2008. – 70 p.
15. Исупов, К. С. Инструментальный комплекс для проектирования параллельных масштабируемых программ численных расчетов / К. С. Исупов, В. С. Князьков // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики / главный ред. д.т.н., проф. В. О. Никифоров. – СПб., 2010. – № 6(70). – С. 68-72. – ISSN: 2226-1494.
16. Исупов, К.С. Модулярно-позиционный формат и программный пакет для разрядно-параллельных вычислений высокой точности в формате с плавающей точкой / К. С. Исупов // Вестник ЮУрГУ. Серия «Вычислительная математика и информатика». – Челябинск: Издательский центр ЮУрГУ, 2013. – Т. 2, № 4. – С. 65-79. – ISSN 2305-9052.

Информация об авторах

Data on authors

Князьков Владимир Сергеевич, доктор технических наук, профессор, руководитель Научно-образовательного центра «Супервычислительные технологии и системы», профессор кафедры «Электронные вычислительные машины», ФГБОУ ВПО «ВятГУ».

Тел.: 8(912)7343173

Электронная почта: kniazkov@list.ru

Knyazkov Vladimir, Doctor of Technical Science, Prof., Head of Science & Education Centre «High Power Computer Technology & Systems», Prof. of Department «Computer Engineering», FGEI HPE «VyatSU».

Tel.: 8(912)7343173

E-mail: kniazkov@list.ru

Исупов Константин Сергеевич, аспирант, кафедра «Электронные вычислительные машины», ФГБОУ ВПО «ВятГУ».

Тел.: 8(953)674-39-02

Электронная почта: isupov.k@gmail.com

Isupov Konstantin, postgraduate student, department of «Computer Engineering», FGEI HPE «VyatSU»

Tel.: 8(953)674-39-02

E-mail: isupov.k@gmail.com

**ОДНОРОДНЫЙ АРИФМЕТИЧЕСКИЙ ПРОЦЕССОР С
РЕКОНФИГУРИРУЕМОЙ МИКРОАРХИТЕКТУРОЙ ДЛЯ МАССОВЫХ
АРИФМЕТИЧЕСКИХ ОПЕРАЦИЙ**

**HOMOGENOUS ARITHMETICAL PROCESSOR WITH
RECONFIGURABLE MICRO ARCHITECTURE FOR MASSIVE
ARITHMETICAL OPERATIONS**

В статье представлена организация арифметического модулярно-систолического процессора с реконфигурируемой микроархитектурой. Вычислительный модуль процессора организован в виде однородной среды из одноразрядных вычислительных ячеек. Причем, вычислительный модуль может быть секционирован на четыре вычислительных ядра и поддерживает соответственно арифметические операции с 32-, 64-, 128-разрядными числами с фиксированной точкой, как в классических позиционных системах счисления, так и в системах счисления в остаточных классах с различными наборами оснований. Приведена техническая реализация процессора на ПЛИС EP3C120 и результаты экспериментальной апробации.

The paper presents the organization of modular arithmetic-systolic reconfigurable processor microarchitecture. Computer processor module is organized in a uniform medium of single-cell computing. Moreover, the computer module can be partitioned into four cores and supports arithmetic operations, respectively with 32-, 64-, 128-bit fixed-point number, as in the classical positional number systems, and in the number systems in residual classes with different sets of reasons . The technical implementation of the processor to the FPGA EP3C120 and results of experimental testing.

Ключевые слова: Арифметика остаточных классов, систолические вычисления, однородные вычислительные структуры, микроархитектурная реконфигурация, СБИС-процесоры.

Key words: Arithmetic of residuals classes, systolic computing, homogenous computing structures, reconfigurable micro architecture, SBIS-processors.

В статье представляются результаты экспериментальной апробации разработанных авторами динамически реконфигурируемых конвейерных вычислительных структур в базисах модулярной арифметики нового типа: оперативно масштабируемых, поддерживающих высокую скорость вычислений за счет новых алгоритмических и структурных решений, минимизированных по аппаратным ресурсам и обеспечивающих высокую технологичность при их массовом производстве.

Для достижения перечисленных достоинств используются оригинальные способы структурирования потоков для предельного распараллеливания вычислений. В частности, для повышения скорости вычислений используются модулярные системы счисления, которые теоретически обеспечивают независимость образования разрядов числа, в силу чего имеется возможность независимой параллельной обработки разрядов данных. Кроме этого обеспечивается высокая скорость операций контроля вычислений, сокращаются аппаратные затраты при технической реализации исполнительных устройств.

Использование системы остаточных классов позволяет распараллеливать вычисления на уровне разрядов чисел, обеспечивать высокую точность вычислений, повышает надежность вычислений при не высоких аппаратных затратах.

Процессор производит арифметические операции над числами, представленными в модулярной системе счисления по одному из четырех наборов оснований, содержащему 5 остатков (38 бит) и 2 бита, определяющих

набор. На Рис. 1 представлено 32-хбитное число в системе остаточных классов (СОК).

39	32	31	24	23	16	15	9	8	2	1	0
Остаток по модулю №5		Остаток по модулю №4		Остаток по модулю №3		Остаток по модулю №2		Остаток по модулю №1		Группа оснований	

Рис. 1. Модулярное представление 32-х битного числа

Для представления 64-х и 128-ми разрядных чисел модулярное представление выглядит аналогично, отличается лишь количество используемых оснований. Для 64-х битных чисел используются десять восьмиразрядных остатков, а для 128-ми битных – двадцать. Группы оснований при этом должны быть различны.

В системе команд процессора используется универсальный формат всех микроопераций. В 64-битном формате команды (рис. 2) присутствуют следующие поля: смещения операндов, вид адресации, код операции, указатель векторной команды.

63	48	47	32	31	16	15	10	9	3	2	1	0
C3		C2		C1		BA		КОП		B	*	

Рис. 2. Универсальный формат команд

В процессоре используются основные типы адресации (Таблица 1).

Таблица 1

Способы адресации

	Способ адресации	Операнд
	Непосредственная	Операнд в команде
	Регистровая прямая	Операнд в регистре
	Индексная	Смещение в адресном
	Относительная	Смещение в команде

В процессоре задействовано 80 команд, среди которых арифметико-логические команды, передачи управления и другие (Таблица 2).

Общие сведения о командах

Тип команды	Адр	Кол
Команды пересылки данных и ввода/	2	3
Команды арифметико-логические	3	13
	2	17
Команды управления	1	2
	0	11
Команды передачи управления	1	18
	0	3
Привилегированные команды	1	8
Команды работы с аппаратурой	0	5
Всего команд	-	80

В процессоре используется иерархическая структура памяти. Существует как предвыборка команд в очередь команд из оперативной памяти, так и предвыборка операндов в очереди данных из КЭШ памяти. После необходимого преобразования в операционном устройстве результаты поступают в очередь результатов, откуда записываются в КЭШ память по адресу, извлеченному из очереди адресов результатов.

Обмен информацией между ядрами, оперативной памятью и внешними устройствами производится по системе с общей шиной. Структурная схема процессора представлена на рис. 3. КЭШ-память имеет два порта чтения и один порт записи, что обеспечивает поддержку трехадресных команд. Устройство управления процессора децентрализовано, в каждом ядре имеются по три устройства управления, каждое из которых отвечает за свою ступень работы конвейера.

В начале работы первой ступени конвейера происходит анализ готовности остальных ступеней. Если прерываний не произошло, и если очередь данных заполнена не полностью, происходит выборка новой команды из очереди команд (ОК). Параллельно с этим происходит формирование физических адресов операндов и результата в блоке формирования физических адресов (БФФА), после чего адреса выдаются на адресные входа КЭШ памяти данных (КЭШД). Также происходит запись адреса результата в очередь адресов

результатов (ОАР) и запись микрокоманды в очередь микрокоманд (ОМК). Иначе происходит такт ожидания. Если прерывание произошло, то осуществляется работа с аппаратурой прерываний (АП) и формирование адреса дескриптора в IDT, чтение из ОК запрещается, стадии конвейера сбрасываются.

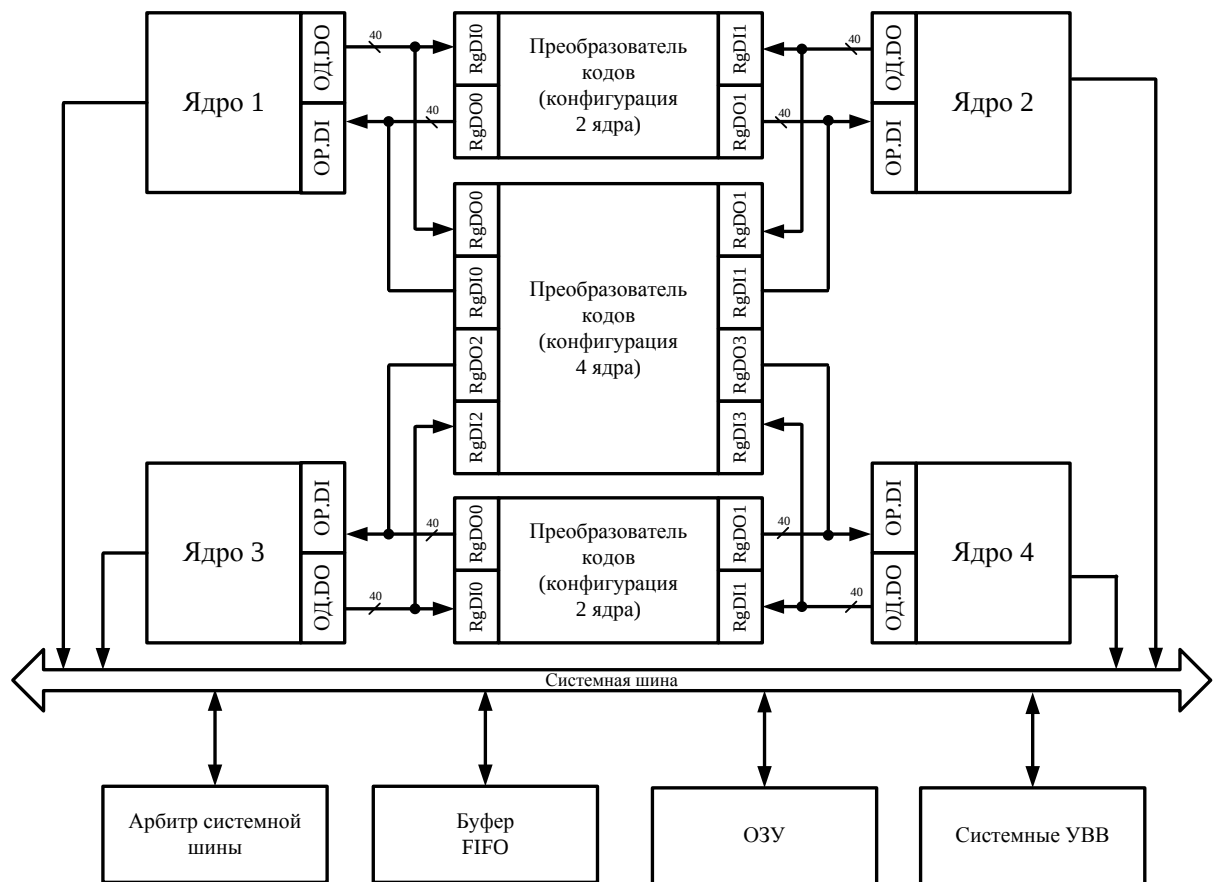


Рис. 3. Структурная схема процессора

Вторая ступень предназначена для записи выбранных из КЭШД операндов в очереди данных ОД1 и ОД2. При этом проверяется, присутствует ли выбираемая строка в КЭШД ядра, при её отсутствии работа ступеней приостанавливается и параллельно проверяются остальные ядра на присутствие строки. Если сигнал a-tag вырабатывается одним из ядер, происходит коммутация с этим ядром и выбранная строка переписывается с КЭШД текущего ядра, в противном случае происходит выборка строки из оперативной памяти (ОП). После этого работа конвейера возобновляется.

Третья ступень производит запись в КЭШД строк, либо отдельных операндов. Если после чтения ОАР адрес принадлежит тэгу в КЭШД, то номер банка сохраняется и происходит чтение ОР, в следующем такте происходит запись в нужный банк памяти КЭШД параллельно новым чтением ОАР. При промахе в КЭШД происходит аналогичная процедура обновления КЭШД, как и при чтении.

Арифметические операции в вычислительном ядре производятся параллельно по нескольким основаниям p_i , их количество n определяется диапазоном P представления чисел: $P = p_1 \cdot p_2 \cdot \dots \cdot p_n$. Представление числа A в СОК обеспечивается наименьшими неотрицательными вычетами α_i по системе взаимно простых модулей

$$p_i (i \in [1, n]): A(\alpha_1, \alpha_2, \dots, \alpha_n)$$

В вычислительном ядре арифметические операции производятся над операндами представленными вычетами α_i и β_i , результат – δ_i .

Каждый вычет представляет собой m -разрядное двоичное число:

$$\alpha(a_m, a_{m-1}, \dots, a_1)$$

Свойства СОК позволяют свести все арифметические операции к операции сложения, которая может выполняться параллельно по всем основаниям:

$$|\alpha_i|_{p_i} + |\beta_i|_{p_i} = |\alpha_i + \beta_i|_{p_i}$$

При этом для невыхода за пределы диапазона представления основания необходимо производить коррекцию результата по правилу:

$$|\alpha_i + \beta_i|_{p_i} = \begin{cases} \alpha_i + \beta_i; 0 \leq \alpha_i + \beta_i < p_i \\ \alpha_i + \beta_i - p_i; p_i \leq \alpha_i + \beta_i < 2p_i \end{cases}$$

Для операции вычитания потребуется ввести представление отрицательных чисел, при этом диапазон $[0, P)$ делится на два интервала: $[0, P/2)$ и $[P/2, P)$ при одном четном основании или $[0, (P+1)/2)$ и $[(P+1)/2, P)$ при отсутствии четного основания. Назовем их соответственно первой и второй половинами заданного диапазона. Числа первой половины будем считать положительными, а числа второй половины – отрицательными. При этом

представление отрицательных чисел задается в дополнительном коде. Так, операция вычитания потребует замены остатков вычитаемого α_i дополнением до модуля p_i :

$$\gamma_i = p_i - \alpha_i$$

Для операции умножения воспользуемся свойствами полей Галуа $GF(p)$, благодаря которым возможно использование изоморфизма между мультипликативной по модулю p_i группой $\alpha_i = \{1, 2, \dots, p_i - 1\}$ и аддитивной модулю $p_i - 1$ группой $\gamma_i = \{0, 1, \dots, p_i - 2\}$. Соответствие задается через порождающий элемент g и определяется выражением:

$$\alpha_i = |g^{\gamma_i}|_{p_i}$$

Тогда умножение над полем $GF(p)$ производится по формуле:

$$|\alpha_i \cdot \beta_i|_{p_i} = g^{|\gamma_i + \phi_i|_{p_i - 1}}$$

Из всех форм деления лишь деление с нулевым остатком является замкнутым относительно операций кольца. Для этой формы деления известно, что делимое представляет собой целое число, кратное делителю, а также известно, что делитель и P являются взаимно простыми. В этом случае:

$$\left| \frac{\alpha}{\beta} \right|_{p_i} = \left| \alpha \cdot \frac{1}{\beta} \right|_{p_i}$$

Тогда можно воспользоваться правилами модулярного умножения, заменив делитель на обратную мультипликативную величину, взятую по модулю:

$$\gamma = \left| \frac{1}{\beta} \right|_{p_i}$$

Для того чтобы производить вышеперечисленные замены, в структуру вычислительного модуля введены таблицы подстановок (ТП).

Общая структура вычислительного ядра представлена на рис. 4.

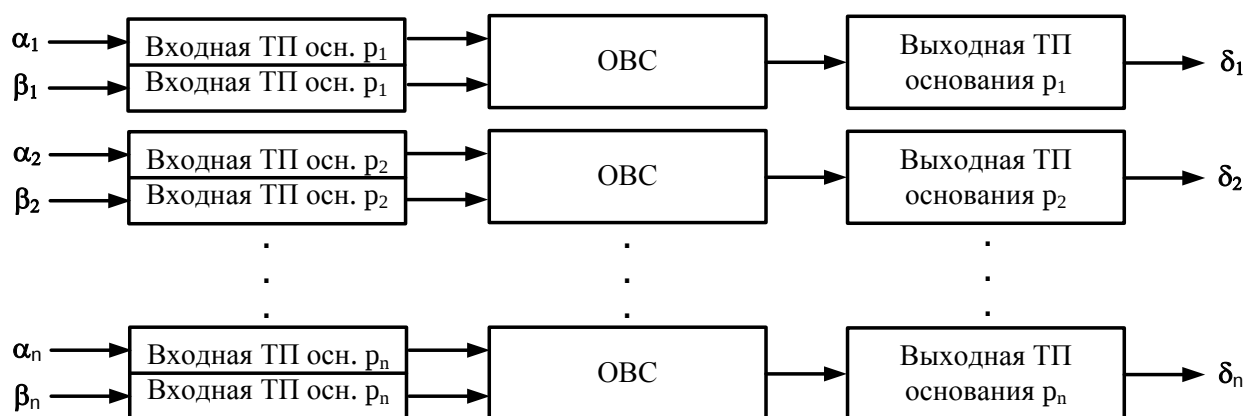


Рис. 4. Структура вычислительного ядра

Вычислительное ядро состоит из вычислительных модулей, число которых равно n . Каждый вычислительный модуль ведет вычисления по заранее определенному основанию и содержит следующие элементы:

а) входная таблица подстановок операнда α_i размерностью $2 \cdot 2^m$, настраиваемая в один из следующих режимов:

- 1) транзит операнда, в случае операции сложения или вычитания (вычисления в аддитивной по модулю p_i группе);
- 2) переход из мультипликативной группы к аддитивной группе по модулю p_i (в случае операции умножения или деления);

б) входная таблица подстановок операнда β_i размерностью $4 \cdot 2^m$, настраиваемая в один из следующих режимов:

- 1) транзит операнда, в случае операции сложения;
- 2) подстановка дополнения операнда до модуля p_i , в случае операции вычитания;
- 3) переход из мультипликативной группы к аддитивной группе по модулю p_i , в случае операции умножения;
- 4) подстановка обратной величины взятой по модулю p_i с переходом из мультипликативной группы к аддитивной, в случае операции деления.

в) ОВС размерностью $m+1$ строк и $m+1$ столбцов базовых элементов (БЭ), предназначенная для сложения операндов и позволяющая осуществлять распараллеливание конвейерного типа вплоть до разряда операнда. Структура ОВС представлена на рис. 5;

г) выходная таблица подстановок результата δ_i размерностью $2 \cdot 2^m$, настраиваемая в один из следующих режимов:

- 1) коррекция результата, в случае операции сложения или вычитания;
- 2) коррекция результата с переходом из аддитивной группы в мультипликативную группу по модулю p_i (в случае операции умножения или деления).

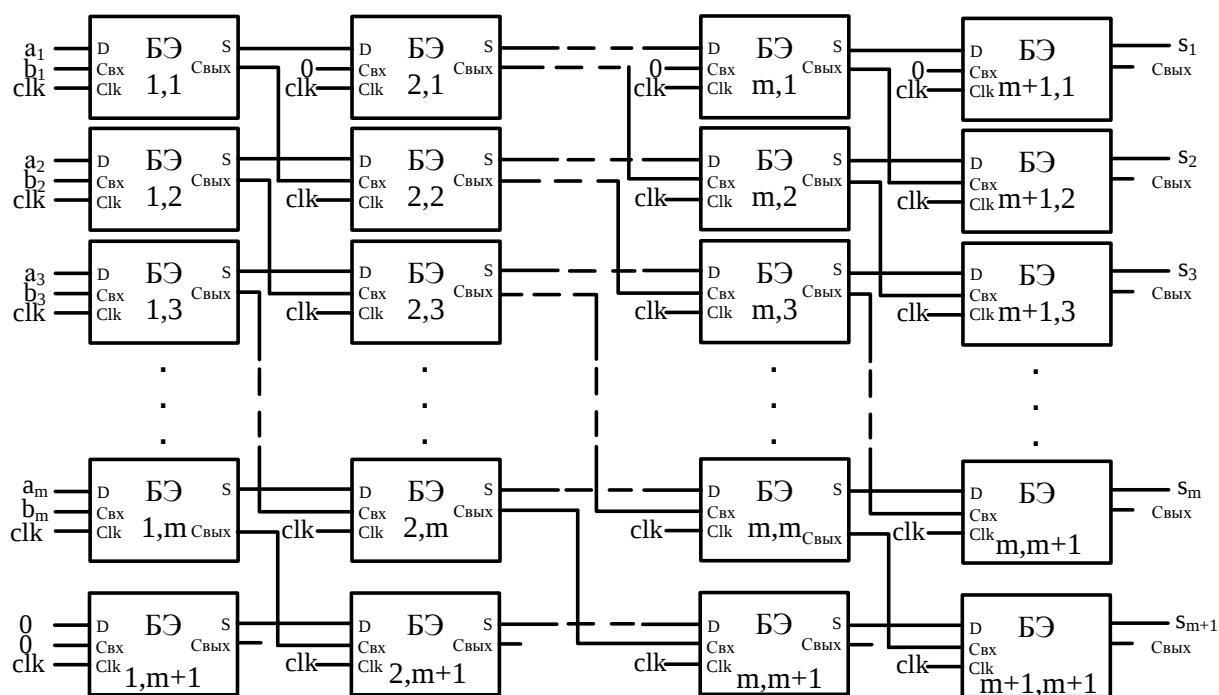


Рис. 5. Структура ОВС

На входы операндов ОВС a_i и b_i подаются соответствующие разряды вычетов операндов, по фронту первого синхросигнала clk данные фиксируются в буферных триггерах. В течение такта происходит формирование промежуточного результата и битов переноса в старший разряд $C_{вых}$. По

следующему фронту синхросигнала сформированные данные фиксируются в следующем слое базовых элементов. Формирование результата продолжится до тех пор, пока не будет единиц переноса. После чего ячейки однородной структуры работают на транзит уже сформированного результата. По пришествию $m+1$ тактов работы результат будет выдан на выходы S_i .

Каждый базовый элемент включает в себя:

- 1) буферную память на D-триггерах T1 и T2 для фиксации текущего состояния разряда D операнда и переноса в старший разряд $C_{вх}$;
- 2) полусумматор на логических элементах XOR и AND осуществляющий сложение бита операнда с битом переноса;

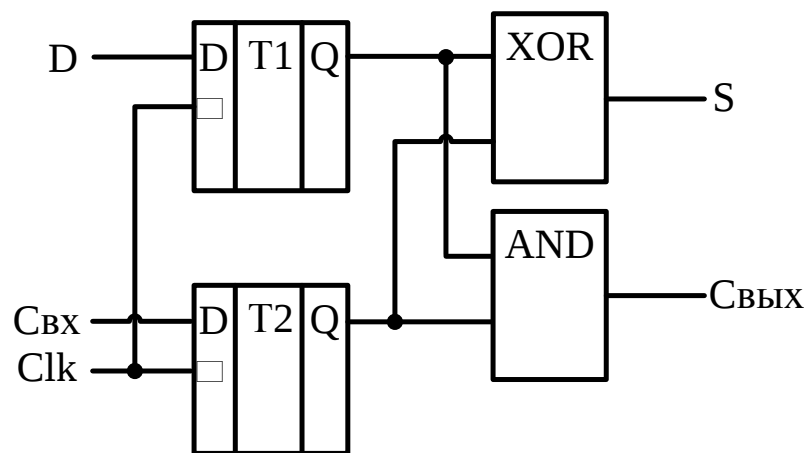


Рис. 6 Структура базового элемента

Таким образом, время одного такта работы ОВС:

$$T_{овс} = t_{триггер} + \max\{t_{AND}; t_{XOR}\}$$

Время заполнения конвейера вычислительного ядра составит:

$$T = t_{вх\ тп} + (m+1) \cdot t_{овс} + t_{вых\ тп}$$

В случае конвейерного режима работы результат операции будет выдаваться каждый такт работы устройства, независимо от типа арифметической операции.

Функциональные схемы процессора спроектированы и промоделированы при помощи САПР Quartus II Web Edition фирмы Altera на ПЛИС EP3C120

семейства Cyclone III, максимальная тактовая частота 500 МГц. Рассмотрим работу процессорного ядра на примере выполнения трех команд, которые уже записаны в сегмент кода оперативной памяти. Прошивка сегмента кода в шестнадцатеричной системе приведена ниже:

- 0:0018000010000008;
- 1:0010000020000008;
- 2:1000200030000002.

На рис. 7, рис. 8, рис. 9 представлены временная диаграмма работы вычислительного ядра. Первая команда производит перевод строки чисел из позиционной системы счисления (ПСС) в СОК, находящейся в ячейке со смещением 0018h сегмента данных, вторая команда также переводит из ПСС в СОК строку чисел, но уже по смещение 0010h. Результат первой команды записывается со смещением 1000h, а второй команды со смещением 2000h. Третья команда производит перемножение строк операндов, полученных в результате работы первых двух команд, результат записывается со смещением 3000h.

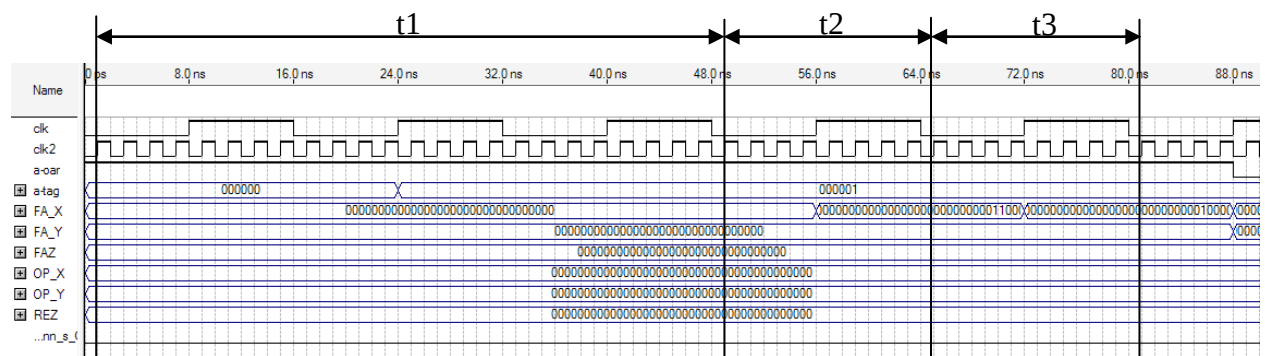


Рис. 7. Временная диаграмма работы ядра процессора (0-88 нс)

Первые три такта t1 работы ядра уходят на инициализацию и запись первой команды из ОП в ОК. В четвертом такте t2 происходит чтение ОК и формирование физических адресов операнда (адрес поступает адресный вход первого порта чтения КЭШД) и результата (адрес записывается в ОАР, а микрокоманда в ОК). В следующем такте t3 вновь происходит то же самое, но уже со второй командой, выбранной из ОК. Кроме того, записывается

выбранная из КЭШД строка операндов первой команды в ОД1, так как присутствует сигнал a-tag.

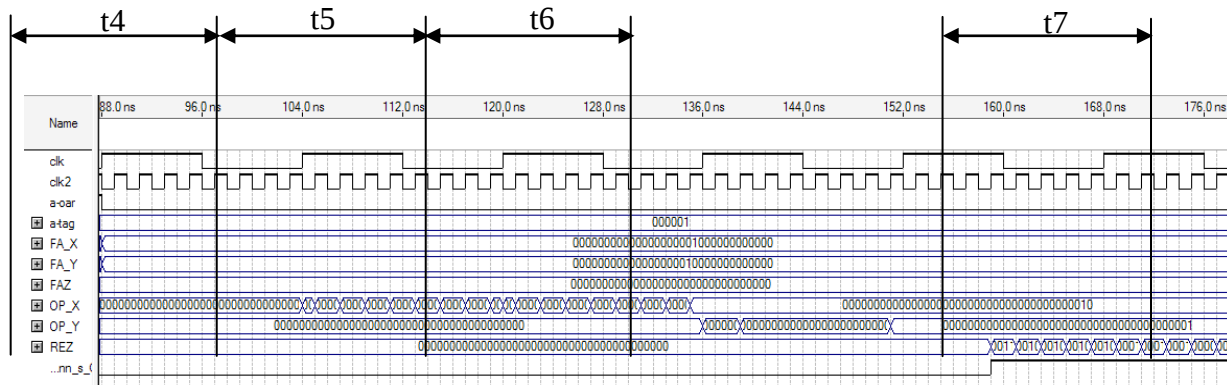


Рис. 8. Временная диаграмма работы ядра процессора (88-176 нс)

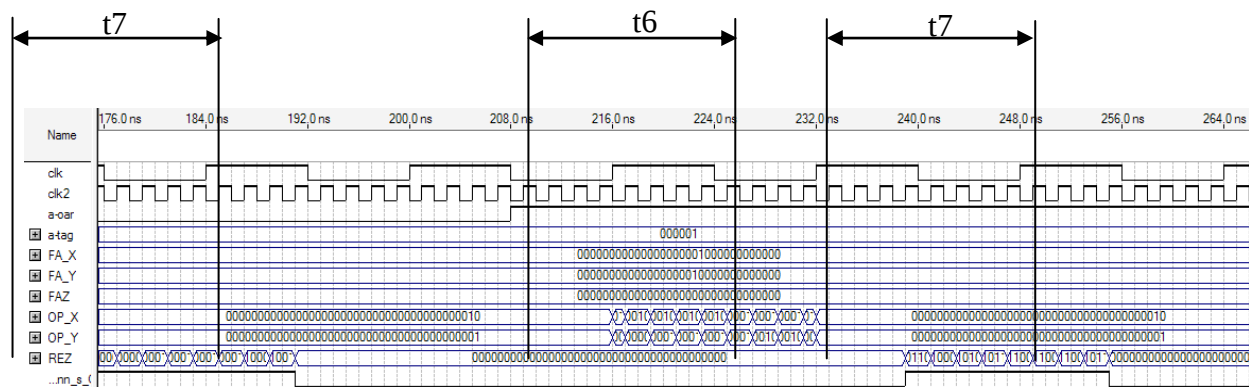


Рис. 9. Временная диаграмма работы ядра процессора (176-264 нс)

В такте t4 выбирается третья команда и при проверке сформированных физических адресов на совпадение в ОАР выявляется конфликт по данным (используются еще не готовые операнды). Выход a-oar принимает значение низкого уровня сигнала, после чего выборка из ОК приостанавливается. ОД1 готова к чтению и в следующем такте начинают последовательно выдаваться операнды из строки с частотой clk2 (500 МГц) на ОУ по шине OP_X[39..0] (t5 строка из первой команды и t6 строка из второй команды).

После заполнения конвейера операционное устройство ОУ результаты преобразования последовательно выдаются на шину REZ[39..0] в интервале t7, сопровождаемые сигналом dann_s_OU, по которому начинается запись результатов в ОР.

Как только результаты первых двух команд будут готовы к чтению, сигнал а-оаг снимается и происходит запись строк, полученных в двух первых командах, в ОД1 и ОД2 соответственно. По шинам ОР_X[39..0] и ОР_Y[39..0] в интервале t_8 операнды поступают в ОУ для перемножения. Произведение первых двух операндов из строк выдается на шину результата REZ[39..0], начиная с 238 нс, а затем каждые 2 нс выдается новое произведение, пока все операнды строки не будут перемножены (интервал t_9).

Затем результирующая строка по выбранному из ОАР физическому адресу будет переписана в КЭШД.

Применение системы остаточных классов позволило распараллелить вычисления между отдельными систолами операционного устройства, что, в свою очередь позволило производить вычисления над операндами следующих разрядностей: 40, 80, 160 бит. Выполнение каждой команды конвейеризировано. В процессоре используются две частоты синхронизации. Операционное устройство работает на максимальной тактовой частоте выбранной ПЛИС EP3C120 семейства Cyclone III фирмы Altera. Эта частота равна 500 МГц. Работа остальных функциональных узлов процессора осуществляется на более медленной частоте - 52,5 МГц. Благодаря трехпортовой КЭШ памяти обеспечена возможность подгружать строки данных каждый такт работы процессора, с тем, чтобы избежать простоев в работе вычислительных ядер. Это позволяет получать результат работы команды каждый такт работы процессора.

Также была проведена экспериментальная апробация функциональных схем процессора в САПР фирмы Altera Quartus II Web Edition. Результаты моделирования подтверждают работоспособность функциональных схем при заданных частотах синхронизации.

В результате экспериментальный вариант модулярно-конвейерного процессора имеют следующие характеристики: быстродействие составляет 36 млн.оп./с; пиковая производительность составляет 500 млн.оп./с; ширина выборки из памяти равняется 64 бита; максимальный объем ОЗУ достигает 4

Гбайт; объем КЭШ памяти ядра равен 80 кбайт; количество прерываний равно 16; форматы данных с ФЗ: 40, 80 и 160 бит; четыре вида адресации в процессоре.

Литература

1.Акушский И. Я. Машинная арифметика в остаточных классах. М.: Советское радио, 1968. – 440 с.

2.Бандман О. Специализированные процессоры для высокопроизводительной обработки данных. Новосибирск: Наука, 1988. – 204 с.

3.Акушский И. Я. Принципы построения высокопроизводительных и надежных процессоров в непозиционных системах счисления. Алма-Ата: Наука, 1980.

4.Алексин Ю. П. Мультиконвейерные вычислительные структуры на однородных средах. Львов: ФМИ АН УССР, 1986. – 74 с.

5.Ахмед Н. Ортогональные преобразования при обработке цифровых сигналов. М.: Связь, 1980. – 271 с.

6.Червяков Н. И. Модулярные параллельные вычислительные структуры нейропроцессорных систем. М.: Физматлит, 2003. – 288 с.

7.Князьков В. С., Федюнин Р. Н. Ячейка однородной структуры. Патент 2300800 С1 от 17.10.2005.

8.Князьков В. С., Федюнин Р. Н. Ячейка однородной структуры. Патент 2295147 С1 от 29.09.2005.

9.Князьков В. С., Федюнин Р. Н. Ячейка однородной структуры. Патент 2285285 С2 от 10.10.2005.

10. Князьков В. С., Волченская Т. В., Дудкин В. С., Пуолуокайнен Д. П. Ячейка однородной структуры. А.С. СССР, № 1335975, Бюл. № 33 от 20.10.1987.

11. Князьков В. С. Однородная вычислительная структура для выполнения операций по заданному модулю. Патент 2005130 614 РФ, G06F

7/00, № 2300800. Заявл. 13.10.05. Оpubл. 10.11. 2007. Бюл. № 31 – 2 с. /Федюнин Р. Н.

12. Князьков В. С. Однородная вычислительная структура для выполнения операций по заданному модулю. Патент 2 310 223 РФ, G06F 7/72. Заявл. 03.10.05. Оpubл. 10.04. 2007. Бюл. № 10 – 2 с. /Федюнин Р. Н.

13. Князьков В. С., Исупов К.С. Система остаточных классов как инструмент для выполнения параллельных высокоточных численных расчетов. Сборник трудов V Всероссийской научной конференции «Математическое моделирование развивающейся экономики и экологии», г. Киров, 5-11 июля 2010, Киров, 2010. С. 86-97.

14. Knyazkov V., Isupov K. Tool complex for performance of precision parallel calculations in basis of matrix algebra. [Электронный ресурс] / «Information technologies in science, education and business – CeBIT 2011, 1-5 march 2011»: Catalog CeBIT 2011 the Ministry of Education and Science of the Russian Federation – Hannover, 2011.- 1 электр. Опт. Диск (CD-ROM) Participants Vyatka State University. //www.cebit.com. – 9 с.

15. Knyazkov V., Osinin I. Modular systolic VLSI processor with a reconfigurable microarchitecture for massive arithmetic operations. [Электронный ресурс] / «Information technologies in science, education and business – CeBIT 2011, 1-5 march 2011»: Catalog CeBIT 2011 the Ministry of Education and Science of the Russian Federation – Hannover, 2011.- 1 электр. Опт. Диск (CD-ROM) Participants Vyatka State University. //www.cebit.com. – 9 с.

Информация об авторах

Data on authors

Князьков Владимир Сергеевич, доктор технических наук, профессор, руководитель Научно-образовательного центра «Супервычислительные технологии и системы», профессор кафедры «Электронные вычислительные машины», ФГБОУ ВПО «ВятГУ».

Тел.: 8(912)7343173

Электронная почта: kniazkov@list.ru

Knyazkov Vladimir, Doctor of Technical Science, Prof., Head of Science & Education Centre «High Power Computer Technology & Systems», Prof. of Department «Computer Engineering», FGEI HPE «VyatSU».

Tel.: 8(912)7343173

E-mail: kniazkov@list.ru

Осинин Илья Петрович аспирант, кафедра «Электронные вычислительные машины», ФГБОУ ВПО «ВятГУ».

Электронная почта: i_osinin@mail.ru

Osinin Il'ya postgraduate student, department of «Computer Engineering», FGEI HPE «VyatSU».

E-mail: i_osinin@mail.ru

ДВЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ВАЛОВОГО РЕГИОНАЛЬНОГО ПРОДУКТА КИРОВСКОЙ ОБЛАСТИ

TWO MODELS OF FORECASTING THE GROSS REGIONAL PRODUCT OF THE KIROV REGION

Рассмотрены два подхода к моделированию динамики валового регионального продукта. По статистическим данным Кировской области построены соответствующие модели. Проведено качественное сравнение моделей и сделаны прогнозы на 3 года.

We considered two approaches to modeling the dynamics of gross regional product. According to the statistics of the Kirov region are made of appropriate models. We made the qualitative comparison of models and forecast for three years.

Ключевые слова: валовый региональный продукт, множественная регрессия, отбор факторов, мультиколлинеарность, системы эконометрических уравнений, временные ряды.

Keywords: gross regional product, multiple regression, factors selection, multicollinearity, systems of econometric equations, time series.

Введение

Традиционные технологии построения регионального экономического прогноза базируются в основном на экспертных методах оценивания. Поиск региональных расчётных схем не является завершённым. Существующие модели часто слишком сложны, что

затрудняет их внедрение.

В первой части настоящей работы рассмотрена структура валового регионального продукта (ВРП) Кировской области, проведён предварительный анализ исходной статистической информации.

Во второй части рассмотрена аддитивная линейная модель [1] для прогнозирования и максимизации ВРП на основе методологии множественной регрессии, которую в данной работе предложено преобразовать в системно-временную модель в соответствии с методологией [2].

В третьей части рассмотрен часто используемый подход построения модели ВРП в виде степенной мультипликативной производственной функции [3] и проведено качественное сравнение моделей.

1. Структура ВРП Кировской области

Анализ отраслевой структуры экономики Кировской области позволяет определить преобладающий промышленный тип развития (с аграрной составляющей) и экстенсивный характер развития основных факторов экономического роста региона. Как видно из рис. 1.1, область специализируется на промышленном производстве, сельском хозяйстве, торговле, транспортных услугах и услугах связи.

На первоначальном этапе моделирования были сгруппированы следующие социально-экономические группы показателей:

1. Отраслевой структуры (объёмы производства, обороты торговли и др.);
2. Инвестиционной активности;
3. Уровня жизни населения;
4. Деятельности предприятий.

Предварительно все показатели были приведены в сопоставимый вид с помощью индекса-дефлятора ВРП и дисконтированы к 2000 году (рис. 1.2)



Рис. 1.1

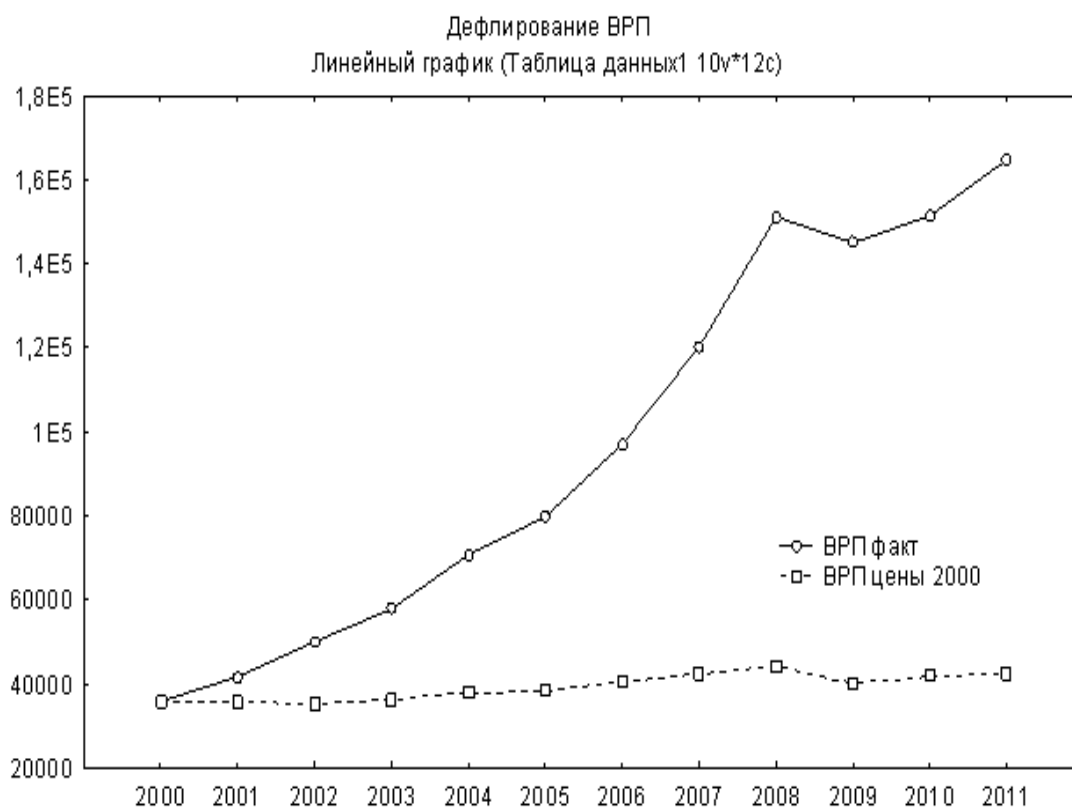


Рис.1.2.

2. Аддитивная модель

За основу методологического подхода к построению модели взята работа [1]. В частности были заимствованы первоначальный состав исходных данных и аддитивный характер взаимодействия переменных модели. Были рассмотрены такие переменные:

- x_1 – ВРП;
- x_2 – доходы населения;
- x_3 – доходы бюджета;
- x_4 – прибыль предприятий;
- x_5 – продукция промышленности;
- x_6 – продукция сельского хозяйства;
- x_7 – объём работ в строительстве;
- x_8 – оборот розничной торговли;
- x_9 – фонд оплаты труда.

Для последующих сценарных проработок апробированы переменные:

- u_1 – инвестиции в основной капитал;
- u_2 – курс доллара;
- u_3 – уровень инфляции (индекс потребительских цен);
- u_4 – расходы бюджета на социальную политику;
- u_5 – субсидии федерального бюджета.

Этот первоначальный состав был подвергнут процедуре отбора факторов с помощью пошагового регрессионного анализа множестве. Кроме того, мы отказались от несистемного построения модели. В результате была получена такая система:

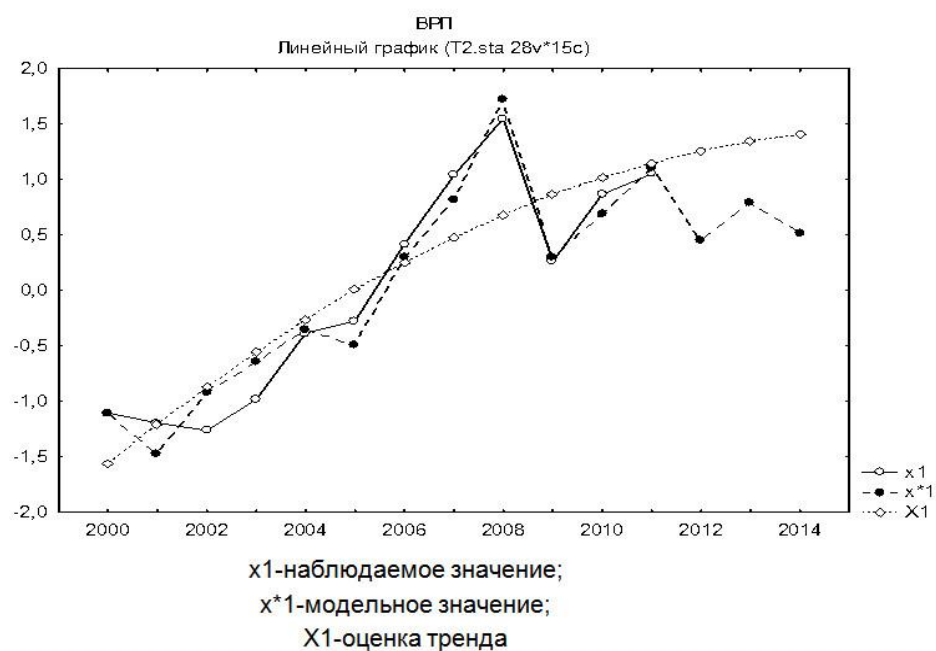
$$\left\{ \begin{array}{l} x_1(t) = 0,41^{***} x_4(t) = 0,57^{***} x_4(t-1) + 0,67^{***} x_7(t-1) - 0,29^{**} u_3(t) + \delta_1 \\ x_2(t) = 0,98^{***} u_1(t) + 0,64^{***} u_2(t) + \delta_2 \\ x_3(t) = 0,96^{***} u_1(t) + 0,56^{***} u_2(t) + \delta_3 \\ x_4(t) = 1,13^{***} x_8(t) + 0,87^{***} u_3(t) + \delta_4 \\ x_5(t) = 0,52^{***} x_6(t) - 1,11^{***} x_8(t) - 0,71^{***} u_2(t) + \delta_5 \\ x_6(t) = -0,81^{***} x_3(t-1) + 0,71^{***} x_4(t) - 0,54^{***} x_7(t) + \delta_6 \\ x_7(t) = 0,47^{***} x_5(t-1) - 0,53^{***} x_6(t-1) - 0,41^{***} u_2(t) + \delta_7 \\ x_8(t) = 0,81^{***} x_5(t) - 0,49^{***} x_6(t) + 0,60^{***} u_2(t) + \delta_8 \\ x_9(t) = 0,48^{***} x_7(t) - 0,81^{***} u_3(t) + \delta_9 \end{array} \right. \quad (2.1)$$

Здесь значения, помеченные тремя звёздочками, соответствуют однопроцентному уровню значимости, двумя – пятипроцентному.

Особенность системного подхода в том, что все переменные модели рассматривались не только как отдельные временные ряды, но и в их взаимодействии, т. е. была использована структурно-временная модель [2]. Системность, в частности, отражает «обратные» взаимодействия переменных: например, не только ВРП является зависимым от совокупности рассматриваемых факторов, но и каждый фактор представляется зависимым от всех остальных (естественно, до этапа отбора статистически значимых факторов), включая ВРП. Результаты прогнозирования ВРП по этой модели представлены на рис.2.1.

Естественно, что прогнозированию подверглись все переменные. По аналогии с [1] были рассмотрены и сценарные варианты развития региона. Для изменения сценариев менялись значения так называемых управляющих переменных. Варианты динамики развития приведены на рис.2.2.

Прогноз величины ВРП Кировской области на 3 года



Сценарные варианты динамики ВРП

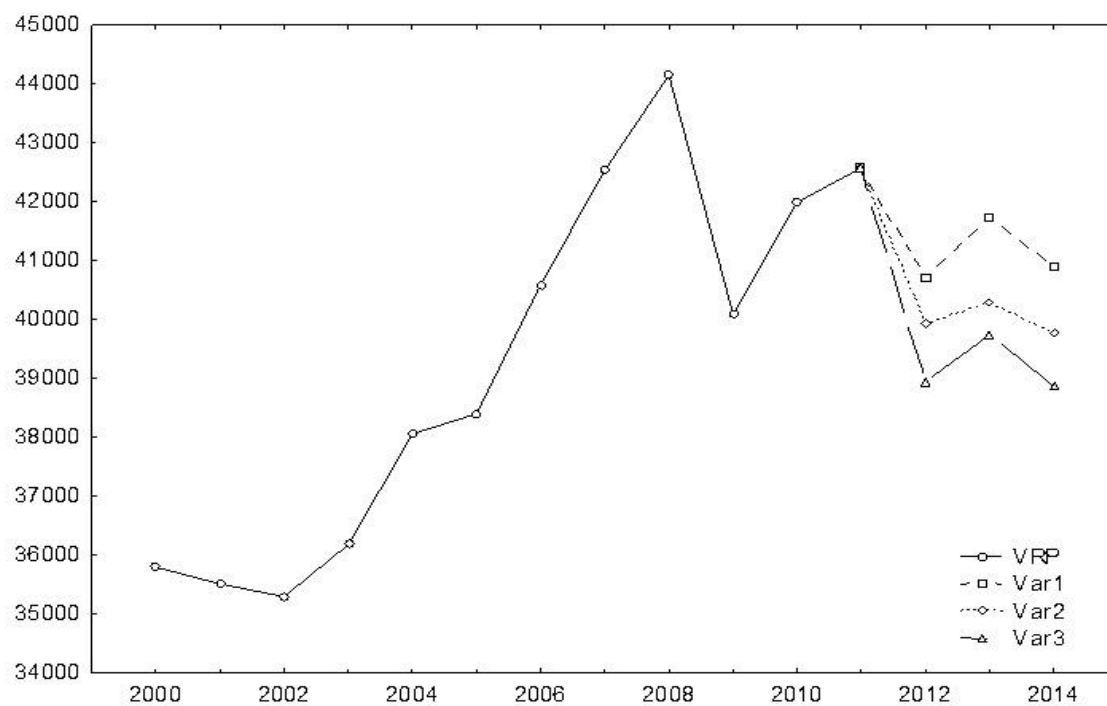


Рис.2.2

3. Мультипликативная модель

Для сравнительного анализа (кроме сценарного) было решено построить модель в виде степенной производственной функции по методологии, предложенной в [3]. Обозначения переменных заимствованы отсюда же. В общем виде модель предложена так:

$$VRP = INCOME^{\alpha^1} FOND^{\alpha^2} S^{\alpha^3} OV^{\alpha^4} OT^{\alpha^5} INV^{\alpha^6} NIOKR^{\alpha^7} RESEARCH^{\alpha^8} ZARPL^{\alpha^9}$$

где VRP – ВРП на душу населения;

INCOME – среднедушевые денежные доходы населения;

FOND – стоимость основных фондов на душу населения;

S – общая площадь жилых помещений на одного жителя;

OV – внешнеторговый оборот на душу населения;

OT – оборот розничной торговли;

INV – инвестиции в основной капитал;

NIOKR – число организаций РИОКР;

RESEARCH – численность персонала НИОКР;

ZARPL – среднемесячная начисленная заработная плата.

В этой модели не рассматривается системное взаимодействие переменных. Вид модели после отбора статистически значимых факторов и результаты прогнозирования представлены на рис.3.1.

Прогноз ВРП по степенной модели

$$VRP = OT^{1,38***} * OV^{0,15***} * INV^{-0,53***}$$

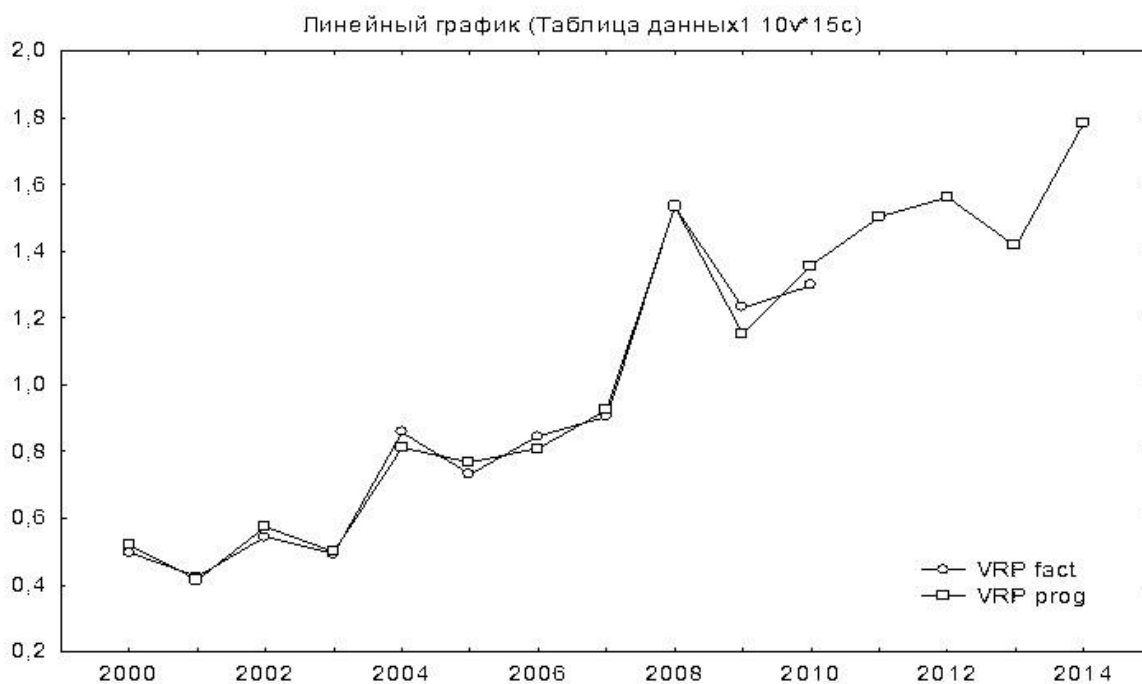


Рис.2.3

Сравнение моделей хоть и показывает, что аддитивная системная модель более информативна, всё же не даёт стопроцентной гарантии её явного преимущества. Чисто качественно (психологически) аддитивная модель вызывает большее доверие хотя бы потому, что она даёт не столь оптимистические прогнозы, чем мультипликативная. А именно оптимизмом и «грешат» многие модели прогнозирования. Кроме того, вызывает сомнение и тот факт, что коэффициент эластичности при инвестициях оказался отрицательным (воруют?).

Литература

1. Андрианов Д. Л. / Целевое управление процессами социально-экономического развития субъектов Российской Федерации:

Монография / Андрианов Д. Л. и др. – Пермь: ЗАО ПРОГНОЗ, 2008. – 239 с.

2. Кошкин Ю. Л. / Структурно-временные модели (СВМ) [Электронный ресурс] / Ю. Л.Кошкин, М. А.Ушакова // Общество, наука, инновации (НТК-2011): ежегод. открыт. всерос. науч.-технич. конф., 18-29 апр. 2011. – электрон. опт. Диск (CD-ROM). – (Социально-экономический факультет. Секция «Математические методы и модели в экономике». Статья № 4).

3. Бакушева Г. В. Модели прогнозирования валового регионального продукта: <http://modeli-prognozirovaniya-valovogo-regionalnogo-producta.htm>

Информация об авторах

Data on authors

Кошкин Юрий Леонидович – к.т.н., доцент ВятГУ

Тел.: (8332)51-17-17-0170

E-mail: Koshyuri@gmail.com

Koshkin Yuri Leonidovich – kand. of Teh. Science, docent VyatSU

Tel.: (8332 research. Tel.: (8332)51-17-17

E-mail: Koshyuri@gmail.com

Сошникова Екатерина Михайловна – преподаватель ВятГУ

Тел.: (8332)71-51-32

E-mail: skate78@rumbler.ru

Soshnikova Ekaterina Michailovna – Ass. VyatSU

Tel.: (8332)71-51-32

E-mail: skate78@rumbler.ru

ИМИТАЦИЯ СТРУКТУРНО-ВРЕМЕННЫХ ДАННЫХ СИСТЕМНОЙ РЕГРЕССИИ

SIMULATION OF THE STRUCTURAL-TEMPORAL DATA SYSTEM REGRESSION

Предложена методика воспроизведения данных для структурно-временной модели регрессии. На первом этапе разыгрываются структурные взаимосвязи между переменными. Затем данные упорядочены во времени.

A procedure is proposed for playing data for structurally-time model regression. At the first stage we will played the structural relationship between variables. Then, the data are ordered in time.

Ключевые слова: регрессия, системы эконометрических уравнений, временные ряды.

Keywords: regression, systems of econometric equations, time series.

Введение

Для оценки эффективности эконометрических моделей и их сравнения в разных комбинациях исходных данных хотелось бы иметь инструмент генерирования управляемых в том или ином смысле данных. В [1] нами были рассмотрены простейшие случаи генезиса данных для парной и множественной регрессии с независимыми факторами, а также простейший случай системной регрессии. При разработке и сравнении методов исследования мультиколлинеарности (метод ридж-регрессии, метод главных компонент, методы отбора в модель только наиболее значимых факторов и др.), нужно управляемо разыграть взаимозависимость факторов, что позволит в

дальнейшем определить в каких условиях какой метод предпочтительнее. Кроме того, навыки разыгрывания позволяют глубже понять суть процесса и представляют самостоятельный интерес.

В первой части настоящей работы показана методика разыгрывания управляемых данных множественной регрессии с заданной мультиколлинеарностью. При этом используется метод главных компонент как средство очистки от случайно возникающих связей независимо разыгрываемых данных.

Во второй части отражено применение изложенной методики с учётом системного взаимодействия переменных.

В третьей и четвёртой частях отражена динамика всех переменных во времени. Сначала показано введение в данные трендовой компоненты (часть 3), а затем и циклической (часть 4).

1. Разыгрывание данных для регрессии с управляемой мультиколлинеарностью

Пусть нужно разыграть данные для линейной регрессии:

$$y = a + b_1x_1 + b_2x_2 + \dots + b_px_p + \varepsilon \quad (1.1)$$

К исходным данным для разыгрывания отнесём параметры $a, b_j (j = \overline{1, p})$, коэффициент детерминации R_y^2 (для разыгрывания остатков ε) и симметричную матрицу R парных коэффициентов корреляции между факторами.

В предлагаемом подходе предварительно разыграем не сами факторы, а как бы их главные компоненты, то есть матрицу F , элементами которой будут значения f_{ij} , где первый индекс – номер разыгрываемого наблюдения, второй – номер переменной.

Главные компоненты по определению независимы. Независимо разыгранные последовательности случайных величин (СВ) с помощью

программных генераторов случайных функций, как правило, оказываются зависимыми. По нашим наблюдениям коэффициент корреляции между такими последовательностями заметно случаен и при разыгрываемой выборке объёмом $n=25$ может колебаться в пределах от 0,05 до 0,6. Разыгрываемые таким образом последовательности (обозначим их матрицу F^* , а её элементы f_{ij}^*) предлагается принять как предварительные. Для получения ортогональных компонент F можно применить упомянутый уже метод главных компонент [2]. Процесс расчёта значений f_{ij} по разыгранным значениям f_{ij}^* предполагает расчёт значений матрицы R^* парных коэффициентов корреляции для предварительных данных, далее – расчёт диагональной матрицы Λ^* собственных значений, элементы которой λ_v^* находятся как корни характеристического уравнения

$$\det(R^* - \Lambda^* E) = 0 \quad (1.2)$$

где E – диагональная единичная матрица.

Далее следует найти собственные векторы V_v как отличные от нуля решения уравнения

$$(R^* - \lambda_v^* E) V_v^* = \underline{0} \quad (1.3)$$

где $\underline{0} = (0 \ 0 \ \dots \ 0)^T$ (символ в степени – признак транспонирования). Далее собственные векторы нормируем в соответствии с выражением

$$U_v^* = \frac{V_v^*}{\sqrt{V_v^{*T} V_v^*}} \quad (1.4)$$

Столбцы A_v матрицы факторных нагрузок A можно вычислить по формуле

$$A_v = U_v \lambda_v^{1/2} \quad (1.5)$$

после чего можно рассчитать матрицу главных компонент

$$F = F^* (A^T)^{-1}.$$

Теперь с помощью независимых главных компонент нужно получить управляемо зависимые между собой данные. Для этого находим собственные числа, собственные векторы и матрицу факторных нагрузок из выражений, подобных (1.2), (1.3), (1.4) и (1.5), но все обозначения, помеченные звёздочкой, заменяем на аналогичные без звёздочки. (Звёздочка была признаком предварительных данных, теперь же у нас есть окончательно матрица главных компонент). Разыгрываемые данные Z в стандартизированном масштабе [3] можно получить по формуле

$$Z = FA^T.$$

Стандартизация предполагает такую связь между элементами матриц Z и X :

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j}, \text{ откуда } x_{ij} = \sigma_j z_{ij} + \overline{x_j},$$

где $\overline{x_j}$, σ_j – среднее значение и среднеквадратичное отклонение j – той переменной, которые также следует отнести к исходным данным.

Для того чтобы разыграть результирующую переменную для (1.1) можно воспользоваться методикой, изложенной в [1].

2. Системные уравнения регрессии

Аналогичные действия можно провести для каждого из уравнений системной регрессии, то есть разыграть данные системы вида:

$$y_i = a_i + \sum_{\substack{k=1 \\ k \neq i}}^m b_{ik} x_k + \varepsilon_i, \quad i = \overline{1, m} \quad (2.1)$$

где m – число уравнений (и смысловых переменных) системы.

Здесь смысловое (например, экономическое в виде издержек и проч., физическое в виде скорости и проч., техническое в виде вязкости и проч. и др.) содержание переменных с одинаковыми индексами совпадает, но роль этих переменных в модели разная. Переменные справа от знака равенства – экзогенные (на этапе идентификации они берутся из выборки, на этапе

прогнозирования они должны быть взяты извне, например, из другой модели [5], из результатов ретроспективного анализа [6], определены методом экспертных оценок [7]), однако на этапе имитации именно они и являются целью разыгрывания. Переменные слева – эндогенные, они, как и остатки, не наблюдаемы и предназначены для оценивания. Разумеется, различаются и числовые значения одинаковых по смыслу переменных (в наблюдениях или же в результате разыгрывания) с их оценками по модели.

Разыгранные данные отражают заданные параметры и заданные парные коэффициенты корреляции между переменными. Если каждое разыгранное наблюдение (совокупность значений всех переменных для него) считать столбцом, то в любом порядке расположения столбцов при идентификации (расчёте значений параметров) метод наименьших квадратов (МНК) даст одни и те же оценки.

Структурно-временная модель [5] (СВМ) предполагает, что данные системы взаимодействуют не только между собой, но и во времени. С этой точки зрения можно считать, что разыгранные данные во времени не упорядочены. Их можно попытаться упорядочить за счёт соответствующих перестановок. Чаще всего во временных взаимодействиях выделяют трендовую и циклическую компоненты. Рассмотрим их разыгрывание пока по отдельности.

3. Разыгрывание тренда

Выберем одну из переменных (любую, например первую) системы (6) в качестве базовой Y_b (например $Y_b = Y_1$). Эта переменная может быть отражена двумя уравнениями:

в виде одного из системных уравнений множественной регрессии:

$$y_b = a_{bs} + \sum_{\substack{k=1 \\ k \neq b}}^m b_{bk} x_k + \varepsilon_b$$

и в виде зависимости от времени $y_b = a_{bt} + f_b(t) + \delta_b$, где функция времени $a_{bt} + f_b(t)$ пусть пока представлена только трендом. Для упрощения будем считать, что тренд T_b линеен (или получен в результате линеаризации), то есть можем записать $T_b = a_{bt} + b_{bt}t$, где t – время.

Возможности упорядочения данных для получения разыгрываемого тренда рассмотрим на примере числовых данных [3, с.54], приведенных в табл.1.

Таблица 1

	0	0	50	00	70	00	50

Здесь l – условный номер наблюдения (для регрессии, отражающей взаимосвязь между переменными, нумерация может быть любой, от этого значения параметров не зависят и в любом случае данные дадут уравнение $y = -5,79 + 36,84x + \varepsilon_y$, на второе уравнение пока отвлекаться не будем). Если сейчас принять $t = l$, то при условии поиска линейного тренда с помощью МНК для переменной y получим: $T_{yt} = T_{yl} = 47,14 + 15,71t$.

Если теперь упорядочить значения y в порядке возрастания, а номера l не менять и принять их за значения t , то получим другое уравнение: $T_y = 20 + 22,5t$. В этом случае коэффициент регрессии максимален, а вариация данных относительно линейного тренда – минимальна. Если n – объём выборки, то число перестановок результирующей переменной равно $n!$, каждая из которых даст своё уравнение. Перебором значений (направленного выбора

найти пока не удалось, но перебор перестановок легко алгоритмизуется, а объём выборки реально не чрезвычайно велик) можно подобрать такую перестановку, которая обеспечит коэффициент регрессии наиболее близкий к желаемому (но не превышающий по модулю максимальное значение для упорядоченных по возрастанию данных). Выход за этот предел всё же возможен, но при условии учёта совместных соотношений трендовой и циклической компонент.

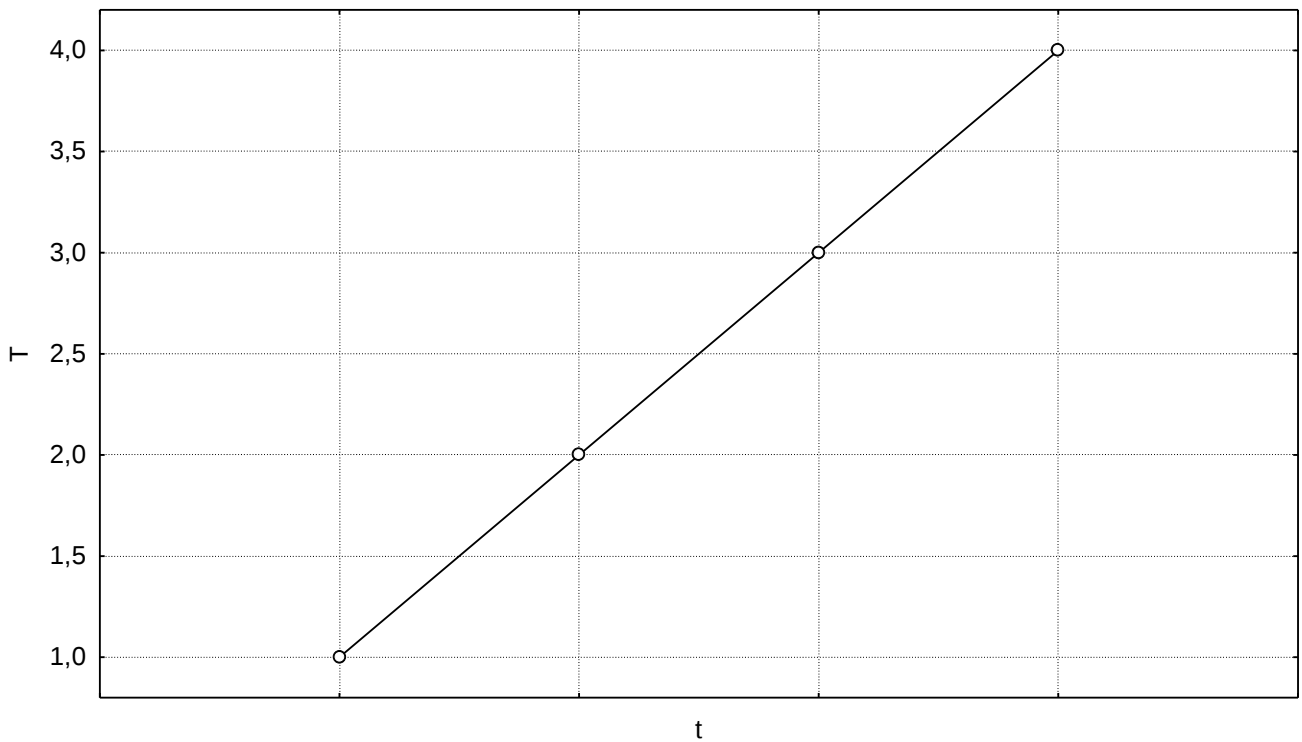
4. Разыгрывание циклической компоненты

Пусть тренд уже разыгран, и нужно «наложить» на него циклическую компоненту. И опять в наших руках только один инструмент: перестановка значений уже разыгранных данных (иначе нарушим уже разыгранные системные взаимосвязи). При таких перестановках (как для разыгрывания тренда, так и для цикличности) системная случайная компонента тоже переставляется, однако системные характеристики надёжности измениться не могут.

Покажем для начала, что если между двумя переменными имеется линейная связь, то путём перестановок значений одной из них (будем иметь в виду зависимую) можно «сконструировать» не только тренд (это уже показано), но и циклическую повторяемость. Как и ранее, не будем «обращать внимание» на случайную компоненту, т. к. с точки зрения системного взаимодействия от перестановок она не изменится.

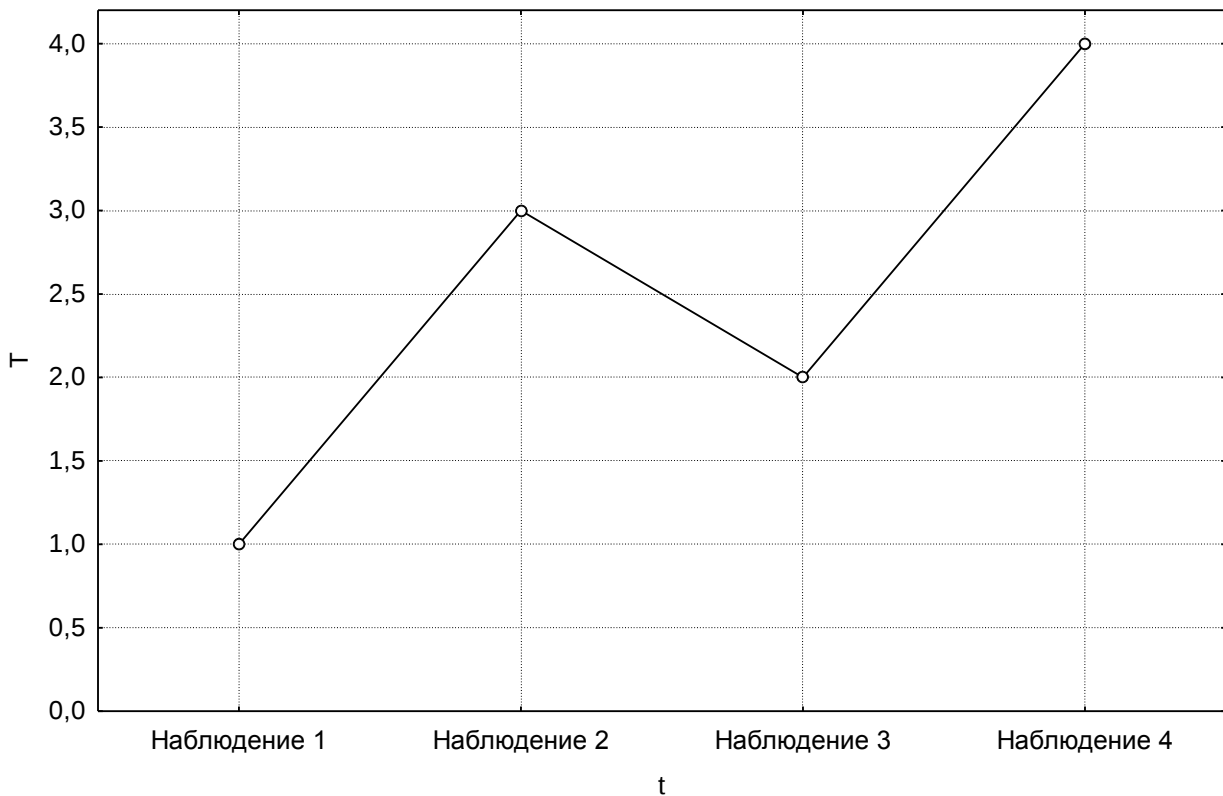
Пусть линейный тренд T сконструирован (рис.1) и нужно получить аддитивно циклическую компоненту S .

Рис.1. Тренд



Каждую дискретную временную точку внутри цикла и повторяющуюся в других циклах принято называть фазой. Пусть для простоты решили разыграть 4 фазы на период (4 квартала в году). Перенесём, например, фазу 2 на место фазы 3, фазу 3 на место фазы 2, а фазы 1 и 4 оставим. В результате получим траекторию для одного цикла (остальные получатся аналогично), причём тренд сохранится (рис. 2).

Рис.2. T+S



В общем случае, если число фаз на период колебания равно q , число возможных перестановок, приводящих к циклам, отличающимся амплитудой, формой и начальной фазой составит $q!$. Правда, некоторые из этих вариантов вырождаются, например, в цикл с нулевой амплитудой или в цикл с большей частотой. Вряд ли при разыгрывании имеет смысл рассматривать все возможные варианты циклической компоненты (при $q=12$ месяцев в году $q!=479001600$), скорее всего, целесообразно для каждого q отобрать характерные базовые варианты (базы данных) из которых затем выбирать подходящий для целей разыгрывания.

Рассмотрим сейчас случаи, когда предварительно разыгранный тренд нельзя использовать для последующего разыгрывания циклов.

Пусть вариант разыгрывания предусматривает отсутствие тренда и такие данные разыграны. Тогда S -компоненту формировать вроде бы не из чего. Чтобы избежать этой ситуации, трендовую компоненту T нужно всё же ввести, но вместе с «антитрендовой» $T_a = -T$. Далее можем сформировать S на базе T_a .

После сложения компонент (в соответствии с аддитивной моделью) тренд и антитренд «аннигилируют» (для мультипликативной модели «аннигиляция» произойдёт если $T_a = 1/T$).

В случае, когда разыгранные системные данные не обеспечивают требуемый размах тренда (даже при упорядочении данных по возрастанию или по убыванию модуль коэффициента регрессии для тренда мал), при разыгрывании S возможно введение двух аннигилирующих трендовых «добавок»: одна из них увеличит размах тренда, другая обеспечит формирование циклической компоненты.

Выше рассмотрен случай разыгрывания временных компонент для выбранной системной базовой переменной. В качестве таковой предпочтительнее доминирующая переменная для последующей идентификации (например, базовая – это валовый продукт, остальные – элементы, его обеспечивающие).

Процедуры введения временных компонент для небазовых переменных проводить по аналогии вряд ли целесообразно. Это может привести к противоречиям в уже разыгранных системных взаимосвязях. Однако в силу линейности этих взаимосвязей временные компоненты «автоматически» введутся во все переменные, если все перестановки данных для базовой переменной провести «синхронно» и для остальных переменных.

В заключение отметим, что управляемые данные как для системной взаимозависимости переменных, так и для временных зависимостей компонент временных рядов, разыграть можно. Это позволит с одной стороны лучше понять механизм генезиса таких данных, а с другой – подбирать и сравнивать методики для идентификации моделей по этим данным, а также использовать эти модели для прогнозирования и имитационного моделирования.

Литература

1. Кошкин Ю. Л. / Генезис данных в эконометрическом моделировании [Электронный ресурс] / Ю. Л. Кошкин, К. А. Мальцев //

Общество, наука, инновации (НТК-2011): ежегод. открыт. всерос. науч.-технич. конф., 18-29 апр. 2011. – /электрон. опт. Диск (CD-ROM). (Социально-экономический факультет. Секция «Математические методы и модели в экономике». Статья № 3)

2. Эконометрика / Под ред. В. С. Мхитаряна. – М.: Проспект, 2009. – 384 с.

3. Эконометрика / Под ред. И. И. Елисеевой. – М.: Финансы и статистика, 2007. – 576 с.

4. Кошкин Ю. Л. / Имитация мультиколлинеарности [Электронный ресурс] / Ю. Л. Кошкин // Общество, наука, инновации (НТК-2012): ежегод. открыт. всерос. науч.-технич. конф., апр. 2012. – /электрон. опт. Диск (CD-ROM). (Социально-экономический факультет. Секция «Математические методы и модели в экономике»)

5. Кошкин Ю. Л. / Структурно-временные модели (СВМ) [Электронный ресурс] / Ю. Л. Кошкин, М. А. Ушакова // Общество, наука, инновации (НТК-2011): ежегод. открыт. всерос. науч.-технич. конф., 18-29 апр. 2011. – /электрон. опт. Диск (CD-ROM). (Социально-экономический факультет. Секция «Математические методы и модели в экономике». Статья № 4)

6. Андрианов Д. Л. / Целевое управление процессами социально-экономического развития субъектов Российской Федерации / Андрианов Д. Л. и др. – Пермь: ЗАО ПРОГНОЗ, 2008. – 239 с.

7. Кошкин Ю. Л. Системность в эконометрическом моделировании. V Всероссийская научная конференция ЭКОМОД-2010 / Сб. трудов. – Киров: ГОУ ВПО «ВятГУ», 2010. – С. 89-97

Информация об авторах

Data on authors

Кошкин Юрий Леонидович – к.т.н., доцент ВятГУ

Тел.: (8332)51-17-17

E-mail: Koshyuri@gmail.com

Koshkin Yuri – kand. of Teh. Science, docent VyatSU

Tel.: (8332)51-17-17

E-mail: Koshyuri@gmail.com

МОДЕЛИРОВАНИЕ ПРОЦЕССА ЛОГИЧЕСКОГО ВЫВОДА С ИСПОЛЬЗОВАНИЕМ ДНК-ЦЕПОЧЕК

MODELLING OF PROCESS OF THE LOGIC CONCLUSION WITH USE DNK-TSEPOCNEK

Описан процесс заполнения базы знаний и представление продукционных правил в виде массива ДНК-цепочек. Рассмотрена реализация механизма прямого логического вывода на биомолекулярном уровне.

The process of filling the knowledge base and representation of production rules in the form of DNA strands massif has been described. Implementation of the mechanism of direct logic conclusion at the biomolecular level was considered.

Ключевые слова: ДНК-вычисления, продукционные правила на базе ДНК, метод прямого логического вывода, ДНК-метод логического вывода.

Keywords: DNA calculation, produktzionny rules on the basis of DNA, a method of a direct logic output, DNA method of a logic conclusion.

Введение

Биокомпьютинг позволяет решать сложные вычислительные задачи, используя методы, принятые в биохимии и молекулярной биологии, организуя вычисления при помощи живых тканей, клеток, вирусов и биомолекул [1,2]. ДНК-вычисления основаны на знаниях о строении и функциях молекулы ДНК и операциях, которые выполняются в живых клетках над молекулами ДНК при помощи ферментов.

Перспективность биомолекулярных компьютеров определяется в первую очередь их способностью работать в биохимической среде (в том числе внутри

живого организма) и взаимодействовать с ней. Вторым преимуществом над полупроводниковыми технологиями является практически неограниченный параллелизм ДНК-вычислений. На уровне отдельных молекул обработка ДНК происходит медленно, со скоростью от 500 до 1000 бит/с, что во много миллионов раз медленнее современных кремниевых процессоров. Но благодаря массовому параллелизму в одной пробирке может одновременно происходить триллион процессов, так что при потребляемой мощности в единицы нановатт может выполняться миллиард операций в секунду.

Развитие парадигмы ДНК-вычислений, кроме развития теории, распадается на следующие важнейшие практические задачи [3]:

- создание эмулятора ДНК-вычислителя;
- разработка языка программирования высокого уровня для ДНК-вычислителя;
- создание банка задач и алгоритмов для ДНК-вычислителя;
- создание среды разработки программ для ДНК-вычислителя.

В докладе рассматривается моделирование ДНК-алгоритма для решения одной из довольно известных проблем: а именно проблемы реализации основных модулей любой интеллектуальной системы – базы продукционных знаний и модуля логического вывода с использованием ДНК-цепочек. Анисимов Д. В. из Санкт-Петербургского государственного университета опубликовал статью «Решение проблемы дедукции с использованием свойств ДНК» [4], в которой детально рассмотрел математическую постановку задачи с использованием молекул ДНК и алгоритмов, основанных на свойствах этих молекул. Однако в ней отсутствуют сведения о принципах построения базы знаний и модуля логического вывода.

Постановка задачи

Проблема реализации основных модулей (базы продукционных знаний и модуля логического вывода с использованием ДНК-цепочек) любой интеллектуальной системы состоит в том, чтобы из множества продукционных

правил базы знаний вывести заключение при наличии множества истинных условий (фактов) и возможного конфликтного множества правил.

Цель компьютерного моделирования ДНК-алгоритма не решение задачи, а создание эмулятора ДНК-вычислителя – программного средства, позволяющего планировать и проводить имитации опытов по ДНК-вычислениям без применения дорогостоящего лабораторного оборудования, а также запускать на выполнение программы для ДНК-вычислителя, скомпилированные в некотором байт-коде. Байт-код эмулятора ДНК-вычислителя представляет собой промежуточный уровень между исходными кодами программы для ДНК-вычислителя и конкретными «операциями низкого уровня». Спецификация байт-кода – вопрос отдельного обсуждения. На первом этапе развития эмулятора это просто средство эффективного хранения программ, а в дальнейшем этот код может развиваться в инструкции для автоматических установок, реализующих интеллектуальную систему на биомолекулярном уровне.

Метод прямого логического вывода

Механизм логического вывода обеспечивает просмотр правил и/или существующих фактов из рабочей памяти и сопоставление этих фактов с базой правил или добавление заключений сработавших правил в рабочую память. При достаточно большом количестве правил для их просмотра требуется достаточно много времени и вычислительных ресурсов.

В продукционной модели основной единицей знаний служит правило (продукция) вида: ЕСЛИ (условие), ТО (действие). Под условием (антецедентом) понимается некоторое предложение, по которому осуществляется поиск в базе знаний, а под действием (консеквентом) – действие, выполняемое при успешном исходе поиска. Действие может быть промежуточным, выступающим далее как условие и целевым (завершающим работу программы).

При помощи правил могут быть выражены пространственно-временные, причинно-следственные, функционально-поведенческие отношения объектов. Правилами могут быть описаны и сами объекты: свойство – объект или объект – набор свойств. Однако чаще описания объектов фигурируют только в качестве переменных внутри правил. В основном продукционная модель предназначена для описания последовательности разных действий и в меньшей степени для структурированного описания объектов. В экспертных системах существует множество алгоритмов, основанных на продукционной модели представления знаний, которая обладает определёнными достоинствами: наглядность и простота модификации; высокий уровень модульности; простой механизм логического вывода.

Отличительная особенность продукционной модели – способность осуществлять выбор правил из конфликтного множества возможных на данный момент времени в зависимости от определённых критериев и характеристик предметной области.

Продукционная модель предполагает гибкую организацию работы механизма вывода. Так, в зависимости от направления вывода возможна как прямая аргументация, управляемая данными (от данных к цели), так и обратная, управляемая целями (от целей к данным). На основе ДНК-вычислений можно реализовать прямой, обратный и комбинированный логический вывод. Прямой вывод применяется в тех случаях, когда выводимые заключения неизвестны. В докладе рассмотрен именно этот метод.

При использовании прямой цепочки рассуждений решается задача: по известным условиям найти последствия.

Прямой вывод используется в продукционных моделях при решении, например, задач интерпретации, когда по исходным данным нужно определить сущность некоторой ситуации или в задачах прогнозирования, когда из описания некоторой ситуации требуется вывести все следствия.

Механизм логического вывода, основанный на прямой цепочке, функционирует следующим образом:

1. Вводится условие;

2. Механизм логического вывода выполняет поиск правил в базе знаний, в условной части которых содержатся истинные факты, введенные в рабочую память;

3. В случае обнаружения всех истинных фактов в условной части правила, данное правило «срабатывает» и заключение этого правила добавляется в рабочую память в виде нового факта;

4. Механизм логического вывода снова выполняет циклический поиск правил в базе знаний. Его работа заканчивается, когда все правила обработаны.

В общем случае прямого вывода на каждом этапе имеется несколько правил, которые можно было бы применить, и возникает проблема выбора наиболее подходящего. Выбор применяемого правила оказывает прямое влияние на эффективность вывода. При большом количестве продукций возникает проблема разрешения конфликта.

Конфликтным набором называется множество правил, которые в одинаковой мере могут быть применены на одном и том же этапе логического вывода. Выбор одного из этих правил называется разрешением конфликта.

Одной из стратегий разрешения конфликта является упорядочивание правил. В качестве критерия выбора используется приоритет или оценка правила.

ДНК-метод решения задачи прямого логического вывода

Данная задача решается с помощью ДНК-цепочек. Этот метод позволяет вначале сгенерировать базу знаний, состоящую из продукционных правил, закодированных ДНК-цепочками. И затем благодаря массовому параллелизму обработки ДНК-цепочек, выполнять необходимые известные биохимические реакции над всеми правилами одновременно.

На данный момент наилучшим из возможных способов решения этой задачи является обычный массовый перебор. При достаточно большом количестве правил в базе знаний для их просмотра требуется достаточно много

времени и вычислительных ресурсов, но, благодаря следующим двум свойствам молекул ДНК, перебор становится, осуществим:

1. Массированный параллелизм цепочек ДНК. Эти цепочки позволяют хранить в них информацию с высокой плотностью и имеют очень простой механизм размножения;

2. Комплементарность Уотсона-Крика. Это свойство позволяет комплементарным парам Аденин (А) – Тимин (Т), Гуанин (Г) – Цитозин (Ц) при образовании двойных цепочек ДНК находить друг друга без стороннего участия. Здесь А, Т, Г, Ц являются основаниями нуклеотидов, из которых состоят полимерные нити, образующие молекулу ДНК.

Введем понятие пробирки [1]. Пробирка – это мультимножество слов (конечных строк) над алфавитом А, Ц, Г, Т. С интуитивной точки зрения пробирка – это вместилище одинарных цепочек ДНК. Цепочки присутствуют в пробирке с некоторыми кратностями, т.е. там может содержаться несколько копий одной и той же цепочки.

Следующие основные операции были первоначально определены для пробирок, т. е. мультимножеств одинарных цепочек ДНК [1]:

– слить – образовать объединение $N1 \cup N2$ (в смысле мультимножеств) двух данных пробирок $N1$ и $N2$;

– размножить – изготовить две копии данной пробирки N . Заметим, что такая операция имеет смысл только для мультимножеств;

– обнаружить – вернуть значение истина, если данная пробирка N содержит, по крайней мере, одну цепочку ДНК, в противном случае вернуть значение ложь;

– разделить (или извлечь) – по данным пробирки N и слову w над алфавитом $\{A, Ц, Г, Т\}$ изготовить две пробирки $+(N,w)$ и $-(N,w)$ такие, что $+(N,w)$ состоит из всех цепочек в N , содержащих w в качестве (последовательной) подстроки, а $-(N,w)$ состоит из всех цепочек в N , которые не содержат w в качестве подстроки;

– разрезать – все нити в N , имеющие подпоследовательность ab , разрезаются между a и b с помощью рестриктаз, двойные нити должны быть разделены, перед тем как их разрезать;

– разделить по префиксу (суффиксу) – по данным пробирки N и слову w , изготовить пробирку $B(N,w)$ (соответственно $E(N,w)$), состоящую из всех цепочек в N , начало (соответственно конец) которых совпадает со словом w .

Для формирования правил из ДНК-цепочек готовится строительный материал: факты в виде специальных ДНК-цепочек, которые кодируются одинаковым числом нуклеотидов. Продукционные правила также кодируются в виде ДНК-цепочек. Операции повторяются, пока появляются в процессе логического вывода новые выведенные из правил новые факты. В результирующей пробирке остаётся хотя бы одна цепочка ДНК или пока не выполнится заданное количество итераций. Для обнаружения ДНК-цепочки используется специальная операция, которая возвращает значение истины, если данная пробирка содержит, по крайней мере, одну цепочку ДНК, в противном случае возвращается значение ложь.

ДНК-алгоритм прямого логического вывода

ДНК-алгоритм прямого логического вывода состоит из следующих этапов:

1. Генерация ДНК-цепочек для кодирования фактов и правил;
2. Определение сработавших ДНК-правил;
3. Разрешение конфликтных ситуаций при срабатывании нескольких ДНК-правил в соответствии с выбранной стратегией;
4. Срабатывание правила при наличии истинных фактов в условной части правила и добавление заключения этого правила в рабочую память. Для этого необходимо выполнить отделение выведенного заключения (нового истинного факта) и его добавление к истинным фактам рабочей памяти;

5. Повторение 2-4 пока появляются новые выведенные факты (в результирующей пробирке имеется хотя бы одна цепочка ДНК) или пока не выполнится заданное пользователем количество итераций.

В начале в растворе получают цепочки ДНК, кодирующие все правила, затем с полученными цепочками производятся необходимые ДНК-операции.

Перед созданием ДНК-цепочек правил необходимо подготовить строительный материал: факты, приоритеты и специальные ДНК-коды «у» и «з». Они кодируются одинаковым числом нуклеотидов. При генерации ДНК-цепочек правил каждый код факта в молекуле правила удлиняется специальным кодом «у», для отличия факта от положения его в условии и заключении (необходимо для операции разделить). Факт, находящийся в заключении, удлиняется с двух сторон специальным ДНК-кодом заключения «з» (необходимо для операции разрезать). На рис.1 представлена структура ДНК-цепочки правила.

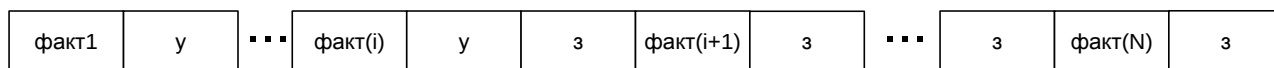


Рис.1 - Структура ДНК-цепочки продукционного правила

К каждому правилу вначале присоединяется ДНК-цепочка кодирующая приоритет. Для кодирования приоритетов используется четверичная система счисления.

Для определения сработавших правил необходимо отделить молекулы правил, в которых не присутствуют известные факты. Это реализуется с применением операции разделить.

Для данной операции используется микроскопический магнитный шарик диаметром порядка одного микрона. К нему притягиваются комплементарные ДНК-коды того или иного не заданного факта удлиненного специальной последовательностью «у», которые выполняют функцию пробы. Затем этот шарик помещается в пробирку с исследуемыми ДНК-цепочками – в результате

к нему (за счет образования водородных связей между комплементарными группами) «притянутся» ДНК-цепочки, в которых присутствует комплементарный код факта. Далее шарик с отсортированными молекулами вынимается и помещается в новый раствор, из которого затем удаляется (при повышении температуры ДНК-молекулы отваливаются от шарика). Данная процедура (сортировка) повторяется последовательно для каждого не заданного факта, и в результате в пробирке останутся только те молекулы, в которых содержатся ДНК-коды всех известных фактов, а значит правила, закодированные данными молекулами, являются сработавшими.

Следующим действием является разрешения конфликта. Для этого в зависимости от первого нуклеотида цепочки разделяются на две пробирки. Если в пробирке с большим приоритетом обнаружено хотя бы одно правило, работа продолжается с ней. Так сортировать до младшего разряда приоритета. Таким образом, на выходе будет получено единственное правило.

После выбора сработавшего правила, необходимо отделить выводимые из него заключения. Для этого необходимо применить две операции разрезать ($N, a \# b$). В первый раз a – специальная последовательность «у», b – специальная последовательность «з». Второй раз a и b – специальной последовательность «з».

За разрезание молекул ДНК внутри цепочки отвечают ферменты – эндонуклеазы, которые могут быть весьма избирательными в отношении того, что они разрезают, где они разрезают и как они разрезают. Сайт-специфичные эндонуклеазы – рестриктазы – разрезают молекулу ДНК в определенном месте, которое закодировано последовательностью нуклеотидов – сайтом узнавания. Разрез может быть прямым, или несимметричным и может проходить по сайту узнавания, или же вне сайта.

Чтобы отделить факты, находящиеся в заключение правил, необходимо применить операцию «разделить по префиксу». В качестве префикса используется специальная маркерная последовательность «з».

За укорачивание молекул ДНК отвечают ферменты – экзонуклеазы, которые могут укорачивать одноцепочечные и двухцепочечные молекулы, с одного или с обоих концов.

Операции повторяются, пока появляются новые выведенные факты (в результирующей пробирке имеется хотя бы одна цепочка ДНК) или пока не выполнится заданное пользователем количество итераций. Для обнаружения ДНК-цепочки используется операция «обнаружить».

Заключение

В данном докладе был описан процесс заполнения базы знаний и представления производционных правил в виде массива ДНК-цепочек. Рассмотрена реализация механизма прямого логического вывода на биомолекулярном уровне. Представлена концепция математического моделирования интеллектуальной системы. Сформированы технические требования к программе моделирования интеллектуальной системы на базе ДНК-цепочек.

После успешных компьютерных экспериментов возможна реализация модулей интеллектуальной системы на биомолекулярном уровне. Данный модуль можно использовать для решения задач интерпретации, когда по исходным данным нужно определить сущность некоторой ситуации или в задачах прогнозирования, когда из описания некоторой ситуации требуется вывести все следствия.

Литература

1. Паун Г., Розенберг Г., Саломаа А. ДНК-компьютер. Новая парадигма вычислений / Под ред. М.В. Волкова. – Москва: Мир, 2003 – 528 с.
2. Борн Д. ДНК-компьютеры обретают логическое мышление // 3D News. – 2009.
http://www.3dnews.ru/news/dnk_komputeri_obretaut_logicheskoe_mishlen

3. Малинецкий, Г. Г., Науменко С. А. Вычисления на ДНК. Эксперименты. Модели. Алгоритмы. Инструментальные средства. – Москва: ИПМ им. М.В.Келдыша РАН, 2005 – 31 с.

4. Анисимов Д. В. Решение проблемы дедукции с использованием свойств ДНК [Электронный ресурс] / Анисимов Д.В. СПГУ – 6с.

5. Ростовцев В. С., Смехова В. М. Принципы разработки модуля базы знаний и модуля логического вывода на базе ДНК-цепочек // Научный прогресс на рубеже тысячелетий. – Польша, 2012. – Т. 24. – 4с.

Информация об авторах

Data on authors

Ростовцев Владимир Сергеевич – к.т.н., доцент кафедры ЭВМ ФГБОУ ВПО «ВятГУ»

Тел.: 89229077847

Электронная почта: rostov_kirov@mail.ru

Rostovcev Vladimir Sergeevich – Cand. Sc. (Engineering), Associate Prof. on Electric Power Systems FGEI HPE «VyatSU»

Tel.: 89229077847

E-mail: rostov_kirov@mail.ru

Смехова Валентина Михайловна – магистрант 2-го года обучения по направлению 230100.68 «Информатика и вычислительная техника» ФГБОУ ВПО «ВятГУ»

Тел.: 89195214511

Электронная почта: ymdpakoiia@mail.ru

Smekhova Valentina – undergraduate on Electric Power Systems FGEI HPE «VyatSU»

Tel.: 89195214511

E-mail: ymdpakoiia@mail.ru

Advanced science

Научный журнал

№ 3 (3)

«Вятский государственный университет» ПРИП ФГБОУ ВПО «ВятГУ» 610000,
Киров, ул. Московская, 36 Тел.: (8332) 64-23-56, <http://vyatsu.ru>